Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

The papers are published under a CC-BY license:
http://creativecommons.org/licenses/by/4.0/

Contents

Editor's note

The Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar (HeadLex) was held at the Polish Academy of Sciences, Warsaw. This conference conjoined the annual Conference on Head-driven Phrase Structure (HPSG) and the annual Conference on Lexical Functional Grammar (LFG) in that a joint program committee and a joint organizing committee were instituted to plan and carry out a joint conference, also resulting in a joint proceedings.

The conference featured 2 invited talks, 32 papers, and 10 posters and were selected by:

Doug Arnold (program committee co-chair) Berthold Crysmann (program committee co-chair), John Lowe (program committee co-chair), Ida Toivonen (program committee co-chair), Anne Abeillé, Alex Alsina, Avery Andrews, I Wayan Arka, Ash Asudeh, Dorothee Beermann, Oleg Belyaev, Emily M. Bender, Adams Bodomo, Tina Bögel, Olivier Bonami, Francis Bond, Kersti Borjars, Bob Borsley, George Aaron Broadwell, Miriam Butt, Rui Chaves, Philippa Cook, Ann Copestake, Mary Dalrymple, Elisabet Engdahl, Daniel Flickinger, Martin Forst, Hyun-Jong Hahm, Dag Haug, Fabiola Henri, Anke Holler, Ron Kaplan, Anna Kibort, Jong-Bok Kim, Tracy Holloway King, Jean-Pierre Koenig, Helge Lodrup, Nurit Melnik, Laura Michaelis, Stefan Müller, Louise Mycock, Tsuneko Nakazawa, Tatiana Nikitina, Rachel Nordlinger, Stephan Oepen, Ryo Otoguro, Agnieszka Patejuk, Gerald Penn, Adam Przepiórkowski, György Rákosi, Frank Richter, Victoria Rosén, Jeffrey Runner, Louisa Sadler, Manfred Sailer, Peter Sells, Liselotte Snijders, Andrew Spencer, Frank Van Eynde, Nigel Vincent, Stephen Wechsler, Shuichi Yatabe, Eun-Jung Yoo, Annie Zaenen.

We want to thank the program committee for putting together a very successful program.

Thanks go to Agnieszka Patejuk (co-organizer) and Adam Przepiórkowski (co-organizer), who were in charge of local arrangements, and their assistants Katarzyna Krasnowska-Kieraś, Jakub Kozakoszczak and Jan Ziółkowski.

Original conference submissions were five page abstracts. The revised versions of the papers underwent an additional round of reviweing. We thank the anonymous reviewers for their help in reviewing.

To ensure easy access and fast publication we have chosen an electronic format. This joint proceedings also functions as the annual proceedings of each of the usually separate conferences.

The proceedings include all the papers except the ones by Liesbeth Augustinus, Tina Bögel, Kersti Börjars and John Payne, Lionel Clément and Sekou Diao, Hyun Jong Hahm, Anna Kibort, Tibor Laczkó, Frank Van Eynde, Nigel Vincent and Kersti Börjars.

Reference patterns in subjunctive complement clauses of Modern Standard Arabic

Tali Arad Greshler
University of Haifa

Livnat Herzig Sheinfux
University of Haifa

Nurit Melnik
The Open University of Israel

Shuly Wintner
University of Haifa

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 4–22

Keywords: control, coreference, corpus study, Modern Standard Arabic

Arad Greshler, Tali, Livnat Herzig Sheinfux, Nurit Melnik & Shuly Wintner. 2016. Reference patterns in subjunctive complement clauses of Modern Standard Arabic. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 4–22. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.1.

Abstract

In this paper we investigate the status of control constructions in Modern Standard Arabic (MSA). MSA has several embedded clause constructions, some of which resemble control in English (and other languages). However, these constructions exhibit some notable differences. Chief among them is the fact that the embedded verb carries agreement features that can indicate both coreference and disjoint reference between a matrix argument and the understood subject of the complement clause. We conducted a thorough corpus-based investigation of such constructions, with a special focus on a search for obligatory control in the language. We show that our findings contradict accepted generalizations (and predictions) proposed by state-of-the-art theories of control, as they indicate that there are no "real" control predicates in MSA. We outline an HPSG analysis that accounts for the MSA data.

1 Introduction

Does Modern Standard Arabic (MSA) have control constructions? MSA has several embedded clause constructions, some of which resemble control in English (and other languages). However, these constructions exhibit some notable differences. Chief among them is the fact that the embedded verb carries agreement features that can indicate both coreference and disjoint reference (dis-reference) between a matrix argument (subject or object) and the understood subject of the complement clause.

The first goal of this paper is to investigate whether all verbs in MSA allow for both coreference and dis-reference, or whether there are predicates which enforce coreference between the understood subject of the embedded clause and a matrix argument. Note that in order to consider the phenomenon from a broad theory-neutral perspective we avoid using the term *control* with all its theoretical implications, unless it is specifically mentioned in the proposals we review. Instead, we distinguish between *co-reference predicates*, which enforce coreference between the subject of their complement clause and one of their arguments, and *free-reference predicates*, which do not pose restrictions on the referent of the embedded subject.

In order to determine whether obligatory co-reference predicates exist in the language we conducted a thorough corpus-based search of such constructions. This empirical investigation was informed by previous insights regarding the distinction between co-reference and free-reference predicates specifically in MSA, in other languages, and from a more general cross-linguistic perspective. We show that our findings contradict accepted generalizations (and predictions) proposed by state-of-the-art theories of control.

[†]This research was supported by THE ISRAEL SCIENCE FOUNDATION (grant No. 505/11). We thank Abdelnaser Jabarin for his helpful advice and judgments regarding the MSA data.

The second goal of the paper is to propose an HPSG analysis of the MSA constructions. To achieve this, we build on existing analyses of the MSA clause structure. We consider whether and how they can be extended to account for the phenomena in the focus of this paper.

The structure of the paper is as follows. We begin Section 2 by briefly reviewing some basic properties of MSA that are relevant to the current study and proceed to discuss *ʔan* clauses, which resemble control constructions, in more depth. In Section 3 we review previous proposals that aim to distinguish between co-reference and free-reference predicates. Our corpus findings are presented in Section 4. In Section 5 we outline an analysis that accounts for the MSA data.

2 Background

2.1 Word order and agreement

Modern Standard Arabic is a *pro*-drop language whose unmarked word order is VSO, yet SVO order is also available. The two word orders differ in their agreement patterns. VSO clauses exhibit partial subject–verb agreement, where the verb agrees with its subject in gender and person, yet its number is invariably singular (1a). SVO clauses, on the other hand, exhibit full subject–verb agreement and therefore the verb bears plural agreement when it has a plural subject (1b). The full/partial agreement distinction is only discernable with plural human subjects. Plural inanimate subjects always trigger singular-feminine agreement.

(1) a. *qaraʔat* *tˤ-tˤaalibaat-u* *l-kitaab-a*
read.3SF *the-students*.PF-NOM *the-book*-ACC
'The female students read the book.'

b. *ʔatˤ-tˤaalibaat-u* *qaraʔna* *l-kitaab-a*
the-students.PF-NOM *read*.3PF *the-book*-ACC
'The female students read the book.'

Finally, *pro*-dropped subjects trigger full agreement on the verb, as demonstrated in (2).[1]

(2) *qaraʔat* *l-kitaab-a*
read.3SF *the-book*-ACC
'She read the book.' (Not: 'They read the book.')

2.2 Complement clauses

MSA has two types of complement clauses, introduced by two principal particles: *ʔan* and *ʔanna*. Example sentences are given in (3a) and in (3b).[2]

[1] See Section 5.1 for an elaboration.

[2] *ʔanna* is a complementizer. However, the syntactic category of *ʔan* is subject to debate and is identified as a functional head, a marker, or a complementizer. Thus, *ʔanna* is glossed as 'that',

(3) a. *qarrara muħammad-un [ʔan yaktuba*
decided.3SM *Muhammad*-NOM(M) *AN write*.3SM-SBJ
r-risaalat-a]
the-letter-ACC
'Muhammad decided to write the letter.'

b. *ʕarafa muħammad-un [ʔanna l-walad-a*
knew.3SM *Muhammad*-NOM(M) *that the-boy*-ACC
sa-yaktubu r-risaalat-a]
will-write.3SM-IND *the-letter*-ACC
'Muhammad knew that the boy would write the letter.'

ʔan clauses and *ʔanna* clauses differ in the following respects. First, the two types of embedded clauses are selected by different predicates. Second, the head of *ʔan* clauses is a verb in the subjunctive mood, while in *ʔanna* clauses it appears in the indicative mood (perfect or imperfect). Third, *ʔan* clauses are verb-initial and when the subject is overt it is marked with nominative case (e.g., (6)), while in *ʔanna* clauses the subject appears clause-initially and bears accusative case. Finally, nothing but negation can intervene between *ʔan* and the subjunctive verb, while in *ʔanna* clauses, as stated, the subject intervenes between *ʔanna* and the verb. In this paper we focus on *ʔan* clauses.

MSA *ʔan* clauses typically appear with no overt subject, yet their unexpressed subject is construed as an argument of the matrix verb. These cases are similar to familiar control constructions in English (and other languages). However, unlike in English, the agreement marking on the subjunctive verb reveals the agreement properties of the intended subject. In (3a) the subjunctive *yaktuba* '*write*' agrees with the matrix subject, Muhammad, which is construed as its understood subject. In (4) the subjunctive *taktuba* '*write*' agrees with the matrix object, Hind, which is construed as its understood subject.

(4) *ʔaqnaʕa muħammad-un hind-an ʔan taktuba*
convinced.3SM *Muhammad*-NOM(M) *Hind*-ACC(F) *AN write*.3SF.SBJ
r-risaalat-a
the-letter-ACC
'Muhammad convinced Hind to write the letter.'

The MSA construction differs from control in English in another respect. The understood subject of the *ʔan* clause and the matrix argument (subject or object) do not necessarily share a reference. Thus, (3a) is actually ambiguous, as the understood subject of the embedded clause can refer to someone other than Muhammad, resulting in the additional reading illustrated in (5).

---

while *ʔan* is glossed as 'AN', without committing to a particular analysis. See Habib (2009) for a discussion of the syntactic category of *ʔan*.

(5) *qarrara* *muħammad-un* *[ʔan yaktuba* *r-risaalat-a]*
decided.3SM *Muhammad*-NOM(M) *AN* *write*.3SM-SBJ *the-letter*-ACC
'Muhammad$_i$ decided that he$_j$ would write the letter.'

In addition, since the understood subject of *ʔan* clauses does not necessarily corefer with the matrix subject (or another argument), the subjunctive verb may exhibit agreement properties which are distinct from those of the matrix verb. As an example, consider (6). The embedded verb bears third-person-singular-feminine (3SF) agreement and can optionally have an overt agreeing subject. This is another property in which MSA *ʔan* clauses differ from control constructions: the embedded clause can have an overt subject.

(6) *qarrara* *muħammad-un* *ʔan taktuba* *(hind-un)*
decided.3SM *Muhammad*-NOM(M) *AN* *write*.3SF-SBJ *(Hind*-NOM(F))
r-risaalat-a
the-letter-ACC
'Muhammad decided that Hind/she would write the letter.'

An additional configuration, which we will refer to here as a 'backward pattern', is one where only the embedded subject is overt (7). In this case, similarly to (3a), its 'forward pattern' counterpart, when the embedded verb and the matrix verb agree the sentence is ambiguous: the unexpressed matrix subject can either be construed as Muhammad, the subject of the embedded verb *yaktuba* '*write*', or as someone else, resulting in the second interpretation presented in (7).

(7) *qarrara*$_{i/j}$ *[ʔan yaktuba*$_i$ *muħammad-un*$_i$ *r-risaalat-a]*
decided.3SM *AN* *write*.3SM.SBJ *Muhammad*-NOM(M) *the-letter*-ACC
'Muhammad$_i$ decided that he$_i$ would write the letter.'
'He$_j$ decided that Muhammad$_i$ would write the letter.'

To conclude, MSA *ʔan* clauses differ from control constructions in other languages in four principle respects: (1) Arabic *ʔan* clauses contain a finite subjunctive verb form; (2) The subjunctive bears agreement features; (3) The subject of the *ʔan* clause does not necessarily corefer with an argument of the matrix predicate; (4) The *ʔan* clause can involve an overt embedded subject. Note that (3) and (4) are independent of each other; there can be an embedded subject in the *ʔan* clause or not, and this subject can corefer with the matrix subject or not.

3 Distinguishing co-reference and free-reference predicates

Examples similar to the introductory examples in (3a) and (4)-(7) are found in reference grammars of MSA (Badawi et al., 2004; Cantarino, 1976; Ryding, 2005). Yet in none of these sources do the authors explicitly distinguish between co-reference and free-reference predicates. Nevertheless, this question is addressed from a functionalist perspective by Persson (2002) and from a generative linguistics perspective in a study by Habib (2009).

Persson (2002), in her corpus-based study of sentential complements in MSA, distinguishes between *ʔan* clauses with an overt embedded subject (and no coreference), and *ʔan* clauses which she describes as clauses in which the embedded subject is deleted under coreference. She argues that the semantic properties of embedding verbs determine their preference for either construction; *manipulative* predicates (*force*, *allow*) prefer coreference, while *cognitive* predicates (desiderative, commentative, fearing, e.g, *want*, *wish*) prefer dis-reference. Persson excludes modality predicates from her study, due to her assumption that they obligatorily require the complement clause subject to be co-referent with the matrix subject. Habib (2009), on the other hand, assumes that there are no "real" control predicates in MSA; all *ʔan* clauses allow for both coreference and dis-reference.

The literature does not seem to have a conclusive answer to the question of whether "real" control exists in MSA. Nevertheless, an interesting parallel is found in Modern Greek (MG), a language which shares a number of syntactic properties with MSA.

Roussou (2009) discusses control and non-control constructions in MG. Like MSA, MG is a *pro*-drop language. MG has two types of complement clauses: *oti* clauses (8), and *na* clauses (9).

(8) *O Yannis pistevi [**oti** to sipiti ine/itan oreo]*
The Yanis.NOM.S believes.S that the house.NOM.3S is/was.3S beautiful
'Yannis believes that the house is/was beautiful.'

(9) *O Kostas matheni [**na** odhiji]*
the Kostas learn.3S PRT drive.3S
'Kostas is learning (how) to drive.'

The distinctions between the two types of complement clauses are reminiscent of those between *ʔan* and *ʔanna* clauses in MSA. The mood of *oti*-complements is always indicative, and their tense is variable. *Na*-complements, on the other hand, have subjunctive mood and invariable present tense. Furthermore, *oti*-complements can be separated from the verb by different elements; *na* must be adjacent to its selecting verb, with only the possibility of a negative element intervening.

Roussou (2009) shows that some MG predicates (e.g., *matheno* 'learn') require the understood subject to be co-referential with the matrix subject, while others (e.g., *thelo* 'want') allow for both a co-referential and a non-coreferential interpretation. As examples consider (9) and (10). In (9), Kostas must be the understood subject of *drive*, while in (10), the understood subject of *drive* can be Kostas or someone else.

(10) *O Kostas theli [na odhiji]*
the Kostas want.3S PRT drive.3S
'Kostas wants (him) to drive.'

With regard to MG, Roussou (2009, p. 1828) suggests that "there seems to be a continuum, which has aspectuals and then modals on the one end and volitionals (and epistemics) on the other". In between, there are predicates which may be closer to either end, and are subject to individual speakers' preferences. Roussou's proposed continuum is shown in Figure 1.

+Control –Control

start --→ *can* ------→ *dare* ------------→ *try* --------→ *want*

Figure 1: The control continuum (Roussou, 2009)

The distinction between obligatory control (OC) predicates and no control (NC) predicates is discussed by Landau (2013) in his comprehensive study of control.[3] Landau proposes a categorical bifurcation between two types of predicates, based on the semantic (in)dependence of the tense of their complement clauses ([T]), as well as their manifestation of overt morphological agreement ([Agr]). The tense specification of complement clauses depends on whether or not their tense is anaphoric to the tense of the matrix clause. Thus, when the complement clause is *tensed* the matrix and embedded events can be temporally mismatched (11a), but when the complement clause is *untensed* they must match (11b).

(11) a. Yesterday, John hoped to solve the problem tomorrow. → infinitive is [+T]
  b. *Yesterday, John managed to solve the problem tomorrow. → infinitive is [-T]

Based on this characterization, Landau categorizes the types of predicates which select tensed or untensed complement clauses.

(12) Predicates which select untensed [-T] complements
  a. Implicatives (dare, manage, remember,...)
  b. Aspectuals (start, stop,...)
  c. Modals (have, need, may,...)
  d. Evaluative adjectives (rude, silly,...)

(13) Predicates which select tensed [+T] complements
  a. Factives (glad, sad, like,...)

[3] Landau distinguishes between OC, NC, and non-obligatory control (NOC). OC and NC occur in complement clauses, while NOC occurs in subject and adjunct clauses. OC and NOC clauses host a PRO subject, and NC clauses host a *pro*/DP subject. PRO in OC is interpreted as a bound variable, which is co-indexed with a co-dependent of the matrix clause. PRO in NOC is logophoric or topic-bound.

b. Propositional (believe, think, claim,...)
c. Desideratives (want, prefer, hope,...)
d. Interrogatives (wonder, ask, find out,...)

The combination of the tense [T] and agreement [Agr] parameters produces four different options, which interact with control. According to Landau's *finiteness rule for Obligatory Control* if a complement clause has slots for both T and Agr, and they are both positive, then no-control (NC) obtains. Otherwise, control is obligatory (OC).

	+T	**-T**
+Agr	NC	OC
-Agr	OC	OC

Table 1: The finiteness rule for Obligatory Control (Landau, 2013)

Landau's (2013) rule implies that if a complement clause is semantically untensed it will enforce obligatory control. Thus, Landau (2013) predicts that "[t]here cannot be a language where modal, aspectual and implicative verbs or evaluative adjectives allow an uncontrolled complement subject" (p. 106).

The picture that emerges from the studies presented so far is that the distinction between co-reference and free-reference predicates is directly linked to their semantic properties. Building on these studies we form predictions regarding the types of predicates associated with each construction. While Habib (2009) does not acknowledge the existence of obligatory coreference in MSA, Persson (2002), Roussou (2009) and Landau (2013) all identify modals as typically coreference-enforcing predicates. Applying Roussou's (2009) continuum to MSA we predict that predicates that are closer to the left end of her continuum would be more likely to enforce coreference. According to Landau's (2013) analysis, given [+Agr], as is the case in MSA *ʔan*-clauses, which are headed by a subjunctive form that shows overt morphological agreement, [+T] implies NC, and [-T] implies OC. Thus, the prediction is that MSA predicates which select an untensed complement clause would be the ones to enforce coreference. These predictions are put to the test in the following section, where we present corpus findings regarding the reference patterns of a representative set of *ʔan*-clause-taking predicates.

4 A corpus study of co-reference and free-reference predicates in MSA

In order to determine whether co-reference predicates exist in MSA we conducted a corpus-based investigation of *ʔan* clauses in contemporary MSA. Our searches focused on representative predicates from Roussou's (2009) continuum and Landau's (2013) classification, in addition to a set of tri-valent manipulative predicates, which were demonstrated by Persson (2002) to prefer coreference and which are typically object control predicates.

The corpus that we used is the 115-million-token sample of the *arTenTen* corpus of Arabic (Arts et al., 2014). This sample has been tokenized, lemmatized and part-of-speech tagged with MADA (Habash & Rambow, 2005; Habash et al., 2009) and installed in the Sketch Engine (Kilgarriff et al., 2004). The morphological tagging of the corpus provides a way of defining queries which target particular person, number and gender features, as well as case and mood. Consequently, we were able to retrieve instances where the matrix predicate and the embedded predicate match in their gender and person agreement, as well as those where there is a mismatch. Furthermore, we could control for the existence or lack of a possible subject (i.e., agreeing nominative noun) following the predicates.[4]

Nevertheless, the search results are not exhaustive. There are numerous instances of erroneous morphological tags, which contributed to false positive results as well as false negatives. Moreover, we decided to favor precision over recall, and limited the distance between the predicates. Consequently, instances with longer NP subjects or intervening adverbials were not retrieved. These limitations notwithstanding, in what follows we provide examples of coreference and dis-reference for a representative set of predicates. Due to the non-exhaustivity of the searches we do not present quantitative data with regard to the distribution of coreference and dis-reference. We do, however, note whether we found dozens of similar examples or whether there were only several examples of disjoint reference.

The corpus search revealed evidence for both coreference and dis-reference with representatives of verbs on Roussou's (2009) continuum, and in Landau's (2013) categories, listed in increasing order by their likelihood to enforce coreference, according to Roussou: the volitional verb *ʔaraada* '*want*', the implicatives *ħaawala* '*try*' and *ʒaruʔa* '*dare*', and the dynamic modal *ʔistatˤaaʕa* '*be able*'. In addition, we found dis-reference examples with the manipulative predicates *ʔaqnaʕa* '*convince*' and *samaħa* '*allow*'. In what follows we present corpus-based examples of coreference and dis-reference with each of the aforementioned predicates.

Volitionals We start at the right end of Roussou's (2009) continuum. Volitionals are predicted by Roussou (2009) and by Landau (2013) to allow free-reference. Consider the volitional *ʔaraada* '*want*' in (14).

[4]For example, to retrieve instances of forward coreference with the verb *ʔaraada* '*want*', we queried for cases where the lemma *ʔaraada* '*want*' is optionally followed by non-verbal material (i.e., the subject, adverbs), then *ʔan* and an adjacent subjunctive verb, which in turn is followed, not necessarily immediately, by a non-nominative noun (i.e., not the subject). Moreover, we restricted the two verbs to share their person and gender properties. The corresponding CQL (Contextual Query Language) query that we constructed was: 1:[tag="verb" & lemma="ʔaraada"] [tag!="verb"] {0,3} [word="ʔan"] 2:[tag="verb" & modus="s"] []{0,2} [tag="noun" & case!="n"] & 1.gender=2.gender & 1.person=2.person.

(14) a. *ʔaraada [ʔan yaʕmala diraasat-an]*
*wanted.*3SM *AN do.*3SM.SBJ *study-*ACC
'He wanted to conduct a study.'

b. *ʔaraada [ʔan yakuuna r-radd-u watˤaniyy-an]*
*wanted.*3SM *AN be.*3SM.SBJ *the-reaction-*NOM(M) *national-*ACC
'He wanted the reaction to be national.'

In (14a), the subject of the embedded predicate corefers with the subject of the matrix predicate; the same person is both the 'wanter' and the 'conductor' of the study. In (14b), on the other hand, the embedded clause involves an overt subject, 'reaction', whose reference is distinct from that of the matrix subject. Our corpus searches revealed dozens of examples of disjoint reference with the predicate *ʔaraada* '*want*'.

Implicatives Moving left on Roussou's (2009) continuum, we found dozens of examples of disjoint reference with the predicate *ħaawala* '*try*', indicating that it is indeed a free-reference predicate. While the matrix and embedded verbs share a subject in (15a), in (15b) the matrix verb bears 1P agreement while the embedded verb bears 3SM agreement and has an overt subject. Clearly, the two subjects do not share a reference.[5]

(15) a. *ħaawala r-raʒul-u [ʔan yatakallama maʕa-na]*
*tried.*3SM *the-man-*NOM *AN speak.*3SM.SBJ *with-us*
'The man tried to speak with us.'

b. *ʔinna-na nuħaawilu [ʔan yataħaddaθa sˤamt-u-na]*
*indeed-we try.*1P *AN speak.*3SM.SBJ *silence-*NOM*-our(*M*)*
'We are trying to make our silence speak.'

The implicative 'dare' is closer to the left end of Roussou's (2009) continuum and is classified in Landau's (2013) categorization as an untensed predicate. Thus, the prediction is that it will enforce coreference, or in other words, be an OC predicate. However, as (16b) shows, this is not the case. MSA *ʒaruʔa* '*dare*' allows free-reference between the embedded subject and its subject; the verb 'be' in (16b) has its own overt subject, 'her opinion', and does not match in agreement with the matrix verb, 'dare'. Admittedly, the disjoint reference example presented here is the only one we were able to find with this predicate. Note, however, that *ʒaruʔa* '*dare*' in itself is an infrequent verb (12.93 per million instances), with substantially fewer attestations of it followed by an *ʔan* clause (1.36 per million).

[5]Interestingly, the disjoint reference examples (15b) and (16b) involve a possessive pronominal clitic on the embedded subject, which refers back to the matrix subject. Such "indirect" coreference with a matrix argument is frequent in dis-reference examples, and can also be expressed as an object clitic on the embedded verb, but it is not obligatory. This coreference creates cohesion between the two events denoted by the two clauses.

(16) a. *laa yaʒruʔu raʒul-un [ʔan yaquula l-ħaqiiqat-a fi*
*not dare.*3SM *man*-NOM *AN say.*3SM.SBJ *the-truth*-ACC *in*
l-zawaaʒ-i]
the-marriage-GEN

'No man dares to say the truth in the marriage.'

b. *lan taʒruʔa [ʔan yakuuna raʔy-u-haa ɣayr-a*
*never dare.*3SF *AN be.*3SM.SBJ *opinion*-NOM-*her(*M*) not*-ACC
musaanid-in li-lmaɣrib-i]
supportive-GEN *to-Morocco*-GEN

'She will never dare that her opinion would be non-supportive of Morocco.'

Manipulatives Tri-valent manipulatives do not appear in Roussou's (2009) continuum, yet Persson (2002) identifies them as the ones which generally impose a coreference restriction. Obtaining exhaustive results with predicates from this class was even more complex than obtaining them with 'subject-control' predicates. However, here too we find evidence of both types of reference relations, with several instances of disjoint reference. (17b) is a disjoint reference example of the predicate *ʔaqnaʕa* '*convince*', and (18b) is a similar example of the predicate *samaħa* '*allow*'.

(17) **Convince**

a. *wa-fi l-masaaʔ-i kaanat malaak qad ʔaqnaʕat*
and-in the-evening-GEN *was.*3SF *Malak(*F*) already convinced.*3SF
waalid-a-haa [ʔan yaʔmura saaʔiq-a-hu
father-ACC-*her AN order.*3SM.SBJ *driver*-ACC-*his(*M*)*
l-xaas^{ʕ}s^ʕ-a bi-ʔiis^{ʕ}aal-i buuʒaa ʔila qaryat-i-hi]
the-private-ACC *in-delivering Buja to village*-GEN-*his*

'And in the evening, Malak had already convinced her father to order his private driver to deliver Buja to his village.'

b. *ʔaqnaʕnaa-hum [ʔan yuʕayyina huwa*
*convinced.*1P-*them AN appoint.*3SM.SBJ *he.*NOM
l-ħukuumat-a]
the-government-ACC

'We convinced them that he would appoint the government.'

(18) **Allow**

a. *iðaa lam nasmaħu li-l-ʔameriikaan-i [ʔan yamurruu min*
*if not allow.*1P *to-the-Americans*-GEN *AN pass.*3PM.SBJ *from*
ʔaraad^{ʕ}ii t-turkiyya]
territory the-Turkish

'If we don't allow the Americans to pass from Turkish territory'

b. *fa-mawqiʕ-u-hu l-ʔiʒtimaaʕiyy-u laa yasmaħu lahu [ʔan*
and-status-NOM-*his the-social*-NOM *not allow.*3SM *to.him AN*
yakuuna bnu-hu fii haaða l-makaan-i]
*be.*3SM.SBJ *son*-NOM-*his in this the-place*-GEN
'And his social status does not allow him that his son will be in this place.'

Modals Modals like 'can' are close to the left (OC) end of Roussou's (2009) continuum and are classified as untensed by Landau (2013). The prediction is therefore that they would enforce coreference. This prediction, however, does not hold. We found dozens of instances of the predicate *ʔistatˤaaʕa* '*be able*' in which the embedded subject does not corefer with the matrix subject. One such case is (19b), in which the predicate is used as a dynamic modal expressing intention or willingness. The matrix subject is a *pro*-dropped first-person-plural subject while the embedded subject is the third-person-singular-feminine 'government'.

(19) a. *lam ʔastatˤiʕ [ʔan ʔasmaʕa sˤawt-a-hu ʔaw ʔaraa-hu]*
*not be.able.*1S *AN hear.*1S.SBJ *voice*-ACC-*his or see.*1S.SBJ-*him*
'I couldn't hear his voice or see him.'

b. *lan nastatˤiiʕa [ʔan tataħammala l-ħukuumat-u*
*never be.able.*1P.SBJ *AN carry.*3SF.SBJ *the-government*-NOM(F)
kaamil-a n-nafaqaat-i]
all-ACC *the-expenses*-GEN
'We will never be able (to accept the fact that) the government will carry all the expenses.'

All the *ʔan*-clause selecting predicates that were investigated in our corpus study turned out to be free-reference predicates, as instances of disjoint reference with them were attested. Importantly, we found disjoint reference examples of modals, which were predicted to enforce coreference. Consequently, we tentatively conclude that MSA does not have predicates which enforce coreference.

5 Towards an analysis

5.1 An HPSG analysis of VSO and SVO clauses in MSA (Alotaibi & Borsley, 2013)

The syntactic structure of VSO and SVO Arabic clauses has been thoroughly discussed in the literature (Fassi Fehri, 1993; Mohammad, 2000; Aoun et al., 2010; Alotaibi & Borsley, 2013, among others). The main challenge is the subject–verb agreement asymmetries between SVO and VSO clauses described in Section 2.1. The analysis put forth by Aoun et al. (2010) and elaborated and cast in HPSG by

Alotaibi & Borsley (2013) proposes that clause structure in MSA is invariantly VSO, where number agreement is suppressed (see Figure 2).

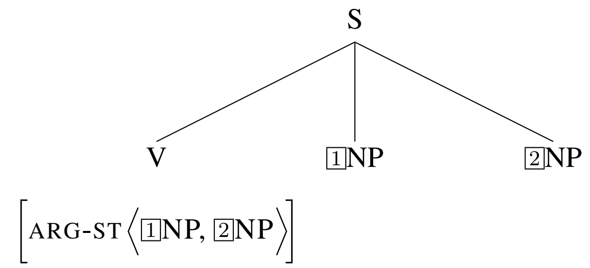


Figure 2: VSO

In the two constructions where full agreement on the verb is found, namely SVO structures and *pro*-drop, the manifestation of full person–number–gender agreement is triggered by the existence of a clitic, referred to as *pro*, which realizes an unexpressed subject. This account, proposed by Alotaibi & Borsley (2013), echoes the analysis proposed by traditional Arab grammarians. In SVO structures what looks like a preverbal subject is in fact a topic which is associated with *pro* subject resumptive pronoun (see Figure 3). This analysis is supported by the fact that subject arguments in SVO clauses are required to be definite.

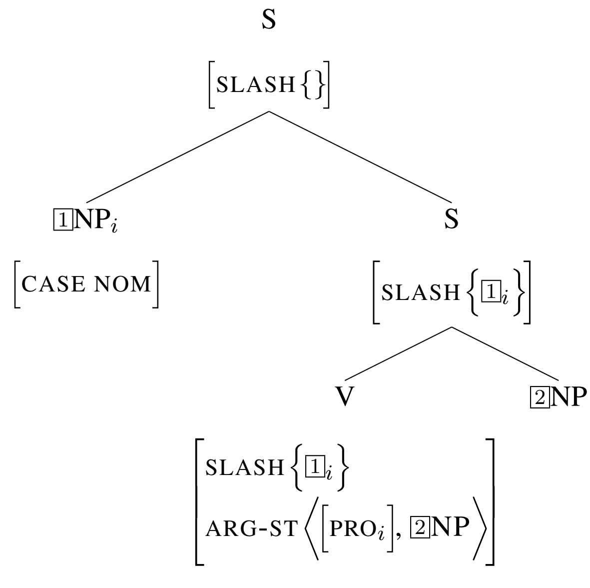


Figure 3: SVO

What is crucial for the current discussion is the idea, which originated in traditional grammar and was adopted and formalized by Alotaibi & Borsley (2013), that *pro*-dropped subjects always trigger full agreement. As we show in the next section, extending this notion to the analysis of MSA coreference constructions can account for the agreement patterns observed in forward constructions but not for the ones observed in backward constructions.

5.2 A *pro*-drop analysis of co-reference and free-reference in MSA

For languages like Modern Greek (and other languages), which have co-reference predicates and free-reference predicates, it is natural to assume that each is associated with a distinct syntactic structure. With regard to MSA, however, if it is indeed the case that it does not have predicates which enforce coreference (i.e., OC predicates), a straightforward analysis would be to assume one structure for both coreference and dis-reference, namely, a free-reference structure.

Consider the schematic representation in Figure 4. Constructions with *ʔan* complement clauses are simply structures with two independent subjects. The matrix verb combines with its subject (lexical NP or *pro*) and with its *ʔan*-clause complement.[6] This complement clause is preceded by a complementizer/marker *ʔan*. The clause itself is in a VSO configuration and is headed by a subjunctive verb. Its subject is either a lexical NP or *pro*.[7]

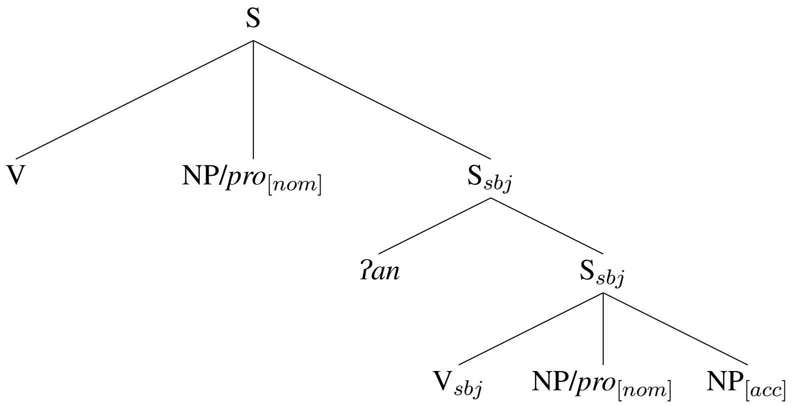


Figure 4: *ʔan* clause complement – coreference/dis-reference

This kind of analysis is based on the *pro*-drop property of MSA; each of the clauses, the matrix clause and the embedded clause, can independently either have an overt subject or a *pro*-dropped subject. There are no constraints on the agreement relations between the two predicates, and therefore they do not need to match. Consequently, what can be construed as subject control is in actuality just a case where the two subjects have identical agreement features, and one of them, either the matrix subject in the backward pattern, or the embedded subject in the forward pattern (or both) is *pro*-dropped. As such, the proposed *pro*-drop analysis does not

[6] Note that the NP/$pro_{[nom]}$ node is an abbreviated notation to indicate the possibility of either using a lexical NP or *pro*-dropped subject and does not imply the existence of empty categories in syntax.

[7] When both the matrix and the embedded subjects are overt NPs, coreference is impossible.

assume a special structure for control and builds on the grammar of MSA to predict the possible constructions of *ʔan*-clause-taking predicates. It is similar in spirit to the analysis proposed by Habib (2009) for all *ʔan* clauses in MSA, and by Roussou (2009) for no-control in Modern Greek, which is also a *pro*-drop language.

Let us first focus on the proposed structure for the forward pattern. In what follows are examples of *ʔan* clauses with plural human subjects, for which the full/partial agreement distinction is observable.

(20) a. *qarrarat tˤ-tˤaalibaat-u [ʔan taktubna r-risaalat-a]*
decided.3SF the-students.PF-NOM AN write.3PF-SBJ the-letter-ACC

b. *ʔatˤ-tˤaalibaat-u qarrarna [ʔan taktubna r-risaalat-a]*
the-students.PF-NOM decided.3PF AN write.3PF-SBJ the-letter-ACC

'The female students decided to write the letter.'

In (20a), the plural human subject follows the matrix verb, thus triggering partial agreement, as exhibited by the 3SF agreement on the verb. The embedded verb, on the other hand, exhibits plural agreement, since it involves a *pro*-dropped subject, which according to the grammar of MSA triggers full agreement (see (2)). In (20b), on the other hand, the plural human subject precedes the matrix verb, resulting in full (3PF) agreement on the matrix verb. The embedded verb exhibits full agreement due to its *pro*-dropped subject, just like it does in (20a).

Figures 5 and 6 illustrate the syntactic structure of the forward pattern, with its two variations. The proposed *pro*-drop analysis along with the assumption that *pro*-dropped subjects trigger full agreement (Alotaibi & Borsley, 2013) correctly predict the agreement variations observed in the forward pattern.

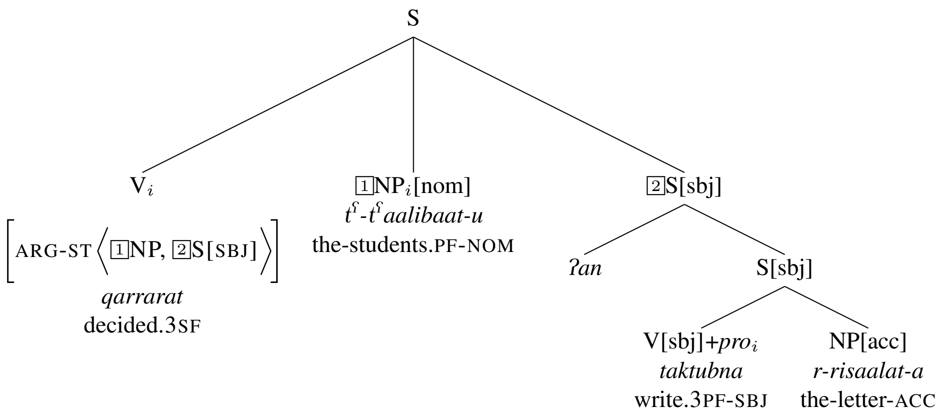


Figure 5: VSO forward pattern - coreference

Moving on to the backward pattern, consider the structure given in Figure 7. According to the *pro*-drop analysis, the matrix verb in this case is *pro*-dropped and the embedded subject is overt. Assuming that *pro*-dropped subjects always trigger

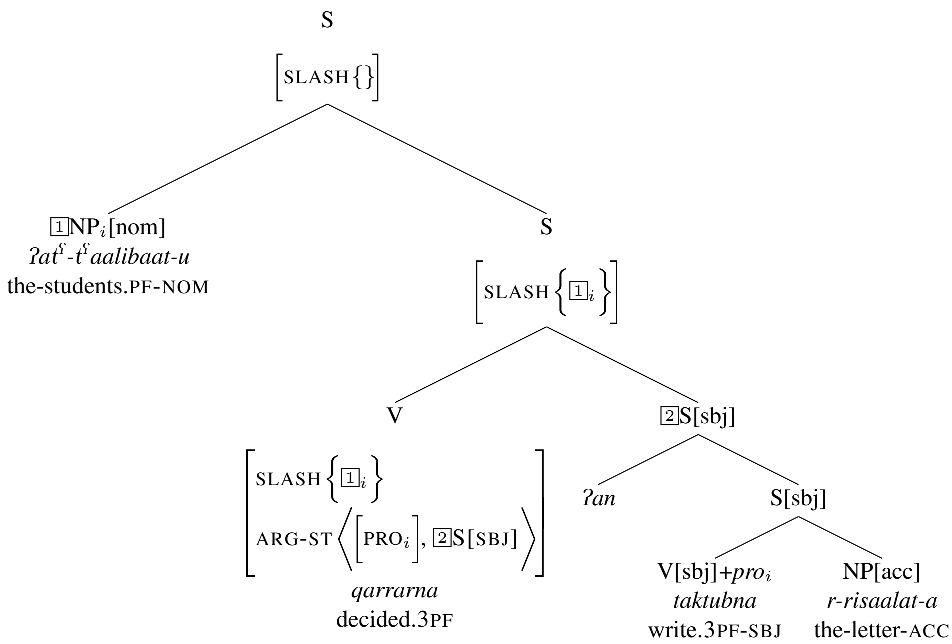


Figure 6: SVO forward pattern - coreference

full agreement, we expect the matrix verb in the backward coreference pattern to exhibit full agreement with the embedded subject.

This, however, is not what is revealed by corpus data. Searches for backward patterns with overt embedded subjects which are both human and plural retrieved only instances in which the matrix verb exhibits partial agreement with the subject. Consider (21), a corpus example of a backward pattern involving a human plural embedded subject.

(21) *takaaliif-u l-ʕilaaʒ-i l-baahiðˤat-u llati [laa*
costs-NOM the-treatment-GEN the-exaggerated-NOM that not
yastatˤiiʕu [ʔan yataħammalu-ha l-fuqaraaʔ-u]]
be.able.3SM AN bear.3SM.SBJ-it the-poor.PM
'The exaggerated costs of the treatment that the poor are not able to bear'

The embedded subject *l-fuqaraaʔ-u* 'the poor' follows the embedded verb *yataħammalu-ha* 'bear', triggering partial agreement on it, as predicted by the grammar of MSA. However, the singular agreement on the matrix verb *yastatˤiiʕu* 'be able' is not predicted by the *pro*-drop analysis and the assumption that *pro* subjects trigger full agreement. Thus, although the *pro*-drop analysis predicts the agreement variations attested in the forward pattern, it makes the wrong prediction with regard to the backward pattern.

Wurmbrand & Haddad (2016) explore backward raising patterns among *ʔafʕaal ʔalmuqaaraba* 'verbs of appropinquation' in Standard Arabic (SA). This verb class

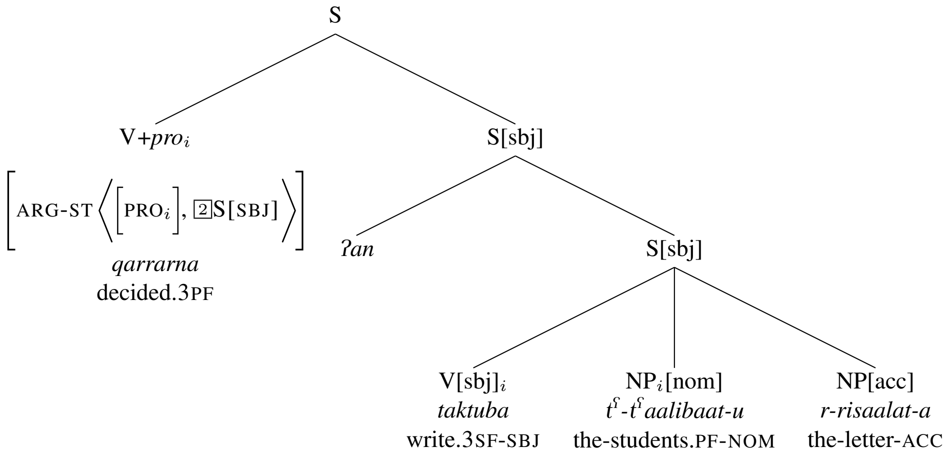


Figure 7: Backward pattern - coreference

encompasses three semantic types: verbs of proximity, verbs of hope, and verbs of inception (Wright, 2007). Wurmbrand & Haddad argue that these verbs can appear in one of two backward raising patterns: one where the matrix verb exhibits full agreement with the overt subject in the embedded clause, and one where agreement is only partial (excluding number).[8] The two patterns are demonstrated in (22).

(22) a. *ʔawʃakna [(ʔan) tanʒaħa t^ʕ-t^{ʕ}aalibaat-u]*
*were.about.to.*3PF *(AN) succeed.*3SF.SBJ *the-students.*PF-NOM

b. *ʔawʃakat [(ʔan) tanʒaħa t^ʕ-t^{ʕ}aalibaat-u]*
*were.about.to.*3SF *(AN) succeed.*3SF.SBJ *the-students.*PF-NOM

'The female students were about to succeed.'

Wurmbrand & Haddad propose that this pattern is unique only to Standard Arabic verbs of appropinquation, and is not found with other raising predicates or control predicates. They attribute the agreement alternation to the different positions of the unpronounced raised copy (pre-verbal and post-verbal).[9] However, the authors acknowledge that although the full agreement case is the one that conforms with prescriptive grammar, they were not able to find naturalistic instances of this structure in contemporary newspapers. They did find instances of the second pattern, where the matrix verb exhibits partial agreement with the embedded subject. Nevertheless, they assume that both orders are available in Standard Arabic. Our corpus study reveals, contrary to Wurmbrand & Haddad, that the backward pattern

[8] Wurmbrand & Haddad also discuss an impersonal backward pattern, where the matrix verb exhibits default 3SM agreement. This pattern is not relevant to the current discussion.

[9] The alternating agreement is a crucial factor in their analysis since it provides evidence for the structural effects of the deleted higher copy of the subject. This, according to Polinsky & Potsdam (2006), is a necessary condition for "real" backward raising, as opposed to cases of long-distance agreement between the matrix and the embedded predicates.

also exists with "control" predicates. Like them, we were not able to find instances of the full agreement pattern in the corpus.

Corpus-based usage data contradict prescriptive grammar and the internal logic of the grammar for "control" predicates, just like it does for verbs of appropinquation. We believe that this discrepancy can be ascribed to the special circumstances of MSA, which is a language that is not spoken natively by any of its speakers.[10] Thus, we propose that the use of partial agreement in the backward pattern is motivated by analogy to the partial subject–verb agreement found in simple VSO clauses. This type of reasoning may explain why although the partial agreement pattern conflicts with the internal logic of MSA grammar, it is the pattern which speakers choose to use. Certainly, more work is needed for a complete analysis that accounts for the full range of patterns found with *ʔan* complement clauses. This remains an open issue for future work.

6 Conclusion

We showed that there is no evidence for the existence of obligatory coreference (or control) with *ʔan*-clause-taking predicates in MSA. A one-structure *pro*-drop analysis with no specific assumptions regarding control accounts for most of the data, but does not align with the agreement pattern attested in the backward construction. The integration of these data into the theory requires some additional assumptions, which seems to involve extra-grammatical factors, related to the non-native status of MSA.

References

Alotaibi, Mansour & Robert D. Borsley. 2013. Gaps and resumptive pronouns in Modern Standard Arabic. In Stefan Müller (ed.), *Proceedings of the 20th international conference on HPSG*, 6–26. Stanford: CSLI Publications.

Aoun, Joseph E., Elabbas Benmamoun & Lina Choueiri. 2010. *The syntax of Arabic*. Cambridge University Press.

Arts, Tressy, Yonatan Belinkov, Nizar Habash, Adam Kilgarriff & Vit Suchomel. 2014. arTenTen: Arabic corpus and word sketches. *Journal of King Saud University - Computer and Information Sciences* 26(4). 357 – 371. doi: http://dx.doi.org/10.1016/j.jksuci.2014.06.009. http://www.sciencedirect.com/science/article/pii/S1319157814000330.

Badawi, El Said, Michael Carter & Adrian Gully. 2004. *Modern written Arabic: A comprehensive grammar*. Routledge.

[10]MSA is the literary standard of the Arab world, but it is acquired in school. The mother tongue of its speakers is some regional dialect of Arabic.

Cantarino, Vicente. 1976. *Syntax of modern Arabic prose*, vol. 3. Bloomington: Indiana University Press.

Fassi Fehri, Abdelkader. 1993. *Issues in the structure of Arabic clauses and words*. Dordrecht: Kluwer.

Habash, Nizar & Owen Rambow. 2005. Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop. In *Proceedings of the 43rd annual meeting of the association for computational linguistics (acl'05)*, 573–580. Ann Arbor, Michigan: Association for Computational Linguistics. http://www.aclweb.org/anthology/P/P05/P05-1071.

Habash, Nizar, Owen Rambow & Ryan Roth. 2009. MADA+TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization. In Khalid Choukri & Bente Maegaard (eds.), *Proceedings of the second international conference on Arabic language resources and tools*, The MEDAR Consortium.

Habib, Rania. 2009. The syntax of the Standard Arabic *ʔan* and *ʔanna*. In Kleanthes K. Grohmann & Phoevos Panagiotidis (eds.), *Selected papers from the 2006 Cyprus syntaxfest*, 159–194.

Kilgarriff, Adam, Pavel Rychly, Pavel Smrz & David Tugwell. 2004. The Sketch Engine. In *Proceedings of EURALEX*, 105–116.

Landau, Idan. 2013. *Control in generative grammar: A research companion*. Cambridge University Press.

Mohammad, Mohammad A. 2000. *Word order, agreement, and pronominalization in Standard and Palestinian Arabic*. Amsterdam: John Benjamins.

Persson, Maria. 2002. *Sentential object complements in Modern Standard Arabic*, vol. 2. Lund University.

Polinsky, Maria & Eric Potsdam. 2006. Expanding the scope of control and raising. *Syntax* 9(2). 171–192.

Roussou, Anna. 2009. In the mood for control. *Lingua* 119(12). 1811–1836.

Ryding, Karin C. 2005. *A reference grammar of Modern Standard Arabic*. Cambridge University Press.

Wright, William. 2007. *A grammar of the Arabic language*. New York: Simon Wallenberg.

Wurmbrand, Susi & Youssef A Haddad. 2016. Cyclic spell-out derived agreement in Arabic raising constructions. In Youssef A. Haddad & Eric Potsdam (eds.), *Perspectives on Arabic linguistics XXVIII: Papers from the annual symposium on Arabic linguistics, Gainesville, Florida, 2014*, 193–228. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Externally and internally headed relative clauses in Marori

I Wayan Arka

Australian National University / Udayana University

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 23–42

Keywords: relative clauses, information structure, definiteness constraints, Papuan languages, discontinuous expressions

Arka, I Wayan. 2016. Externally and internally headed relative clauses in Marori. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 23–42. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.2.

Abstract

This paper discusses relative clauses (RCs) in Marori, showing that this language unusually has almost all of relative clause types, from headed/headless, externally/internally headed, single-/double-headed, to pre-/post-head, to attached/detached RCs. Special attention is given to internally headed relative clauses (IHRC). It is argued that Marori IHRCs are of the restrictive or non-maximalising type, which accounts for certain intriguing properties, such as their indefiniteness constraints and the possibility for RC stacking.

1 Introduction

Marori (ISO 639-3: mok; a subgroup-level isolate, TNG/Papuan, highly endangered, around a dozen fluent speakers) is, like many other Papuan languages, predominantly verb final with free word order. It is perhaps unusual as far as its relative clause typology (RCs) is concerned.[1] It has almost all of relative clause types: headed and headless RCs, externally and internally headed RCs, pre- and post- head RCs, as well as detached RCs or co-relatives. In addition, all grammatical relations (subject, objects, obliques and adjuncts) are relativisable.

Internally headed relative clauses (IHRCs), while constrained, may give rise to ambiguity, if out of context. This is discussed in detail in section 4.4. Thus, either the patient ‘bench’ or the instrument ‘club’ can be understood as the relativised noun in the following example of IHRC in Marori:[2]

(1) [Keme na **njaj**=i **samagau** ngge terme-ben]$_{IHRC}$
REL 1SG bench=U club with 3SG.U.M.hit-1SGNrPST

tamba keiwei nggu-f
PERF damaged 3SG.M.U.AUX-NrPST

a) ‘The **bench** that I hit with the club was damaged.’
b) ‘The **club** with which I hit the bench was damaged.’

However, there is an intriguing definiteness constraint, which can disambiguate them. For example, if one were a proper name (further

1 I gratefully acknowledge the support of ARC Discovery Grant (DP10100307). For helpful discussion and feedback, I thank the anonymous reviewers and audience at the HeadLex 2016 Conference and the Departmental Seminar at Sydney University (23/09/2016) in particular Ron Kaplan, Tracy H. King, Doug Arnold, Bill Palmer, Jim Martin, and Sebastian Fedden. I also thank my research assistant, Agus Mahuze, and Marori language consultants in particular Pak Paskalis and Pak Mikael (both deceased), Pak Lukas and Pak Willem, for their help with data and my fieldwork in Merauke.

2 Abbreviations, alphabetically ordered: 1,2,3 (first, second and third person), A (Actor), Acc (accusative), ARG (argument), AUX (auxiliary), F (feminine), FUT (future), Gen (genitive), GF (grammatical function), HAB (habitual), LOC (locative), IRR (irrealis), MID (middle), NML (nominaliser), NOM (nominative), NPL (nonplural), M (masculine), MP (macro present), NrPST (near past), O (object), P (Patient), PL (plural), POSS (possessive), PRES (present), Q (question marker), REAL (realis), REL (relativiser), RmPST (Remote Past), S (intransitive subject), STAT (stative), SG (singular), TOP (topic), U (undergoer).

discussed later in section 4), such ambiguity would not arise. This definiteness constraint of IHRCs in Marori can arguably be accounted for in terms of Grosu's (2012) semantic typology of RCs; that is, IHRCs in Marori are essentially of the restrictive type, having non-specific indefinite intersective force. Proper names are nominals with unique and inherently definite referents, incompatible with the intersective force of the restrictive RC type. The findings on RCs reported in this paper provide a good empirical basis for the typological and theoretical study of RCs.

The paper is organised as follows. After an overview of the clausal structure and RC marking in Marori (section 2), the basic facts and the salient properties of RCs are outlined in section 3. The discussions given in section 4 provide an explicit analysis of Marori RCs in terms of their c-structure properties (4.1), lexical entries and functional annotations involved (section 4.2), and the demonstration of how the analysis works (section 4.3). That IHRCs in Marori belong to the non-maximalising type is discussed in section 4.4. Finally, the conclusion is given in section 5.

2 Clausal structure and marking RCs in Marori

The morphosyntax of the finite clausal structure in Marori is depicted in (2). It captures the following salient properties of Marori morphosyntax: (i) it is non-configurational (i.e. no VP); (ii) clausal word order is typically verb-final, but it allows postverbal arguments; (iii) argument NPs are freely ordered subject to certain information structure constraints (Arka 2016); (iv) the grammar shows a clear verb-noun distinction with the verb being the clausal head and inflected, (v) the predicate can be complex with the NP preceding the inflected (light/auxiliary) functioning as a lexical predicate; (vi) argument marking is semantically transparent, with core NPs flagged for the U role with the clitic *=i;* and (vii), verb inflections show TAM information and agreement with the prefix and suffix indexing U and A arguments respectively. Illustrative examples are given in (3)-(4).

(2)	DP*(*=i*)	,	[PREF:U-V.Root- SUFF:A]$_V$
	ARG(=U)/PRED		(inflected)

(3)	a.	*mbe=na*	*kundo-ru*	b.	*pa=na*	*ter=∅-me-ru*
		PART=1SG	run-1SG.FUT		soon=1SG	hit=3-AUX-1SG.FUT
		'I will run.'			'I will hit him/her.'	

(4)	a.	pa=ka=i	kara	ku-nggo.
		soon=2SG=U	sick	2SG-AUX.3NPL.FUT
		'You will be sick.'		

b.	tat,	tamba	kwon	k-imb-ra-f.	
	grandfather	already	misquito	2SG-bite-PL-3.NrPST	
	("you")			OBJ	SUBJ
	'Granpa, you're bitten by misquitos.'				

c.	Thomas	fis	mara=i	nde-∅-f	nggambe.
	Thomas	yesterday	stone=U	3SG.bring-3A-NrPST	there
	'Thomas brought the stone there yesterday.'				

A relative clause (RC) is an adjunct within NP, which itself is part of the nominal structure (DP) whose structure is shown in (5). The D (Determiner) comes before or after the NP. The adjunct RC and other elements are freely ordered within the NP. Sentence (1) exemplifies a postnominal RC in Marori. Note that I assume all nodes in the c-structure are optional.

(5) DP → NP, D. b. NP → PossP, NUM, XP:ADJUNCT, N.

(1) efi moipur ki=kwundo-f tamba soron
that child SG.REL=run-NrPST PERF fall.NrPST
'The child who just ran away already fell off.'

The RC marker in (1) is *ki=*, the shortened form of *kei/kefi,* which is also used as a spatial proximal (PROX) deictic in Marori.[3] The full set of RC markers in Marori is given in (2). These markers consist of the PROX formative *k-*, with the stems *efi/em(nd)e*, which are actually the third person pronouns in Marori. The same forms are also used as demonstratives. The form *keme* has general number, usable for any number.

(2) RC markers in Marori:

SG	NSG	GENERAL NUMBER
kefi/ kei/ki	*kemnde*	*keme*

3 Defining RCs and the salient RC properties in Marori

3.1 Definition and challenges

There are three important related aspects in the definition of RCs: syntactic, semantic and pragmatic. Syntactically, a RC is a subordinate clause, functionally an adjunct within a nominal structure. Semantically, it delimits the reference of the nominal by specifying the role of the referent of that NP in the situation described by the RC (Andrews 2007). This defining semantic function of the RC is, however, true only for the RC of the restrictive type. A close scrutiny of the semantics of RCs, however, reveals a complex constellation when other types of RCs are taken into account (further discussed in 4.4). Closely linked to its semantics is the information structure aspect of the RC: a RC introduces a contrast set into the discourse and the referent of the relativised nominal is being focused, in contrast to other referents in the set.[4] For example, in a sentence like (3), the relative clause (within the square brackets) singles out one soldier (implying that there were other soldiers in the discourse). This is done by referring the agentive-subject role of the solder in the event (i.e. firing the shot). The

3 Marori has a complex deictic system, showing a four-way spatial opposition (speaker-proximal, hearer-proximal, semi-distal, and distal) which cross-cuts a three-way number distinction (SG vs. NSG vs. neutral).

4 In this paper, I adopt a traditional LFG analysis, where information structure units (TOPIC and FOCUS) are integrated into f-structure. A precise analysis capturing the semantic/information structure properties of RCs, that allows for (contrastive) set interpretation of FOCUS, would perhaps involve independent information/semantic structures as proposed in Krifka (2008) and Butt (2014), in which the notions of Common Ground (CG) Content and CG Management are important.

relativised nominal is therefore analysed as bearing the discourse function of FOCUS in the proposed LFG-based analysis (further discussed in sections 4.2-4.3).

(3) The soldier [who fired the shot]$_{RC}$ was suspended until completion of the Military Police investigation.

Properties of relative clauses have been of typological and theoretical interest. Typologically, RCs provide a good window to how languages differ in the typological space of complex clause formation. The topics of intense research include, among others, (i) complexity in structure and marking, investigating questions such as headedness in RC (headed vs. headless, externally vs. internally headed); (ii) strategies to encode RC dependency (gapping vs. pronominal copy) and the related restriction showing which grammatical roles are possibly relativised (cf. the accessibility hierarchy (Keenan and Comrie 1977, Comrie and Keenan 1979, Keenan and Comrie 1979), and (iii) RC semantics, investigating the different meanings associated with different RC types (restrictive, vs. non-restrictive/appositive vs. maximalising), which may account the different behaviours of RCs (Arnold 2007, Grosu 2012).

RCs pose a challenge to any theory of grammar, as their complexities involve constraints across all components of the grammar, from semantics to morphosyntax and information structure. This is particularly true for a language that has complex morphology, which also shows different types of RC within same grammatical system. In this paper, it is demonstrated that LFG (Bresnan 1982, Dalrymple 2001, Bresnan et al. 2015, among others) is well equipped to handle the complexities of RCs in such a language, namely Marori. The different types of RC in Marori are outlined in the next subsection. The LFG-based analysis, given in section 4, can be straightforwardly captured in other lexically-based frameworks such as HPSG (Sag, Wasow, and Bender 2003).

3.2 Basic facts: different faces of RCs in Marori

Marori appears to be unusual in that it shows different types of RC. Almost all types are attested in this language: headed/headless, externally/internally headed, pre-/post-head, attached/detached. Clear cases of these types are presented in this section. There is an issue of the identification of double-headed RC, to be addressed later in section 5.

The externally headed relative clause (EHRC) exemplified in (1) is straightforward and needs no further comment. Of particular interest is the internally headed relative clause (IHRC) given in (1). Recall that this IHRC is ambiguous: the patient object and the instrument can be equally relativised. This ambiguity effect is an important issue to be discussed in considerable depth in this paper.

When the referent is clear from the context, the relativised noun is often unexpressed. This gives rise to a headless relative clause, exemplified in (4). The relativiser itself, e.g. *keme* in (5)a can be also dropped, giving rise to a structure shown in (5)b. The syntactic status of this structure is unclear, and ambiguous between IHRC without a relativiser (reading i) or simply two juxtaposed free clauses (reading ii). Further investigation is needed for this.

(4) kefi kwara keswemi tentara=te
REL drum 3SG.M.hit.3SG.PRES REL soldier=BE.3NPL.PRES
'the one who is playing the drum is a soldier'

(5) a. Koro [keme na ife-ben]$_{EHRC}$ tamba kundo-f
dog REL 1SG see-1.NrPST PERF run-NrPST
'The dog that I saw ran off.'

b. [Koro — na ife-ben] tamba kundo-f
dog 1SG see-1.NrPST PERF run-NrPST
i) 'The dog that I saw ran off.'
ii) 'I saw the/a dog (and it) ran off.'

RCs in Marori can appear within or outside their nominal structures. Within the nominal structure, they can be pre- or post-head RCs. Examples given so far are of the post-head RC type. Example (6)a shows a pre-head RC inside its nominal structure whereas examples (6)b shows a detached postverbal RC. This is not unique to RCs in Marori; an ordinary adjunct in Marori can float away from its NP too, as seen in (6)c.

(6) a. [keme na fis ndon-du] ujif]
REL 1SG yesterday 3SG.F.bring.here-1SG.PRES bird

tamba yaba nggwo-f
already dead AUX3SG.F-NrPST

'The female bird that I brought here yesterday already died.'

b. Efi purfam Johni naw te
DET person John named BE.3NPL.PRES

kefi nam bosik yaba ngguf.
RELPOSS pig die AUX-NrPST
'The person whose pig already died is (called) John.'

c. Na mara=i pemje-ben kwebu-wen.
1SG stone=U step.on-1.NrPST sink-SG.NML
'I stepped on the submerged stone.'

Another salient property of Marori grammar is that all grammatical relations, arguments and adjuncts, are relativisable. The previous examples show the relativisation of subject in (1) and (4), patient object in (1), stimulus object in (5), theme object in (6)a, possessor in (6)b and instrumental adjunct i n (1). The following examples illustrate the relativisation of other roles: recipient object (7) and a locative oblique (8).

(7) Maria=i keme njomo-bon bosik, tamba kurye-f.
maria=U REL give.FUT -1PST pig PERF return-NrPST
'Maria who I gave a pig has gone back (home) taking it with her.'

(8) Efi njaj [fis **keme**=na kufamon]
DET bench yesterday REL=1SG sleep-1SG.DUR.NrPST

tamba rafonngin.
PERF broken
'The bench on which I slept yesterday is already broken.'

An important empirical point with a theoretical implication worth mentioning here is the marking and relativisation of obliques and adjuncts. Obliques and adjuncts in Marori must be flagged by their relevant postpositions when they are not relativised. For example, the postpositional clitic =*ku* is obligatory in (9)a, marking the locative-goal. When relativised as in (9)b, =*ku* is not present, either inside or outside the RC. Therefore, this dependency in the RC cannot be accounted for in terms of a filler-gap analysis, as the category of the filler (NP) and that of the gap (PP) are distinct. It must be accounted for in terms of referential identity at the level of functional structure (f-structure), further discussed in the next section.

(9) a. John mara=i sour=ku monjo-f.
John stone=U house=LOC throw-3NPL.NrPST
'John threw a stone to the house.'

b. Efi sour=e [keme John __ mara=i monjo-f].
that house=part REL John stone =U throw-3NPL.NrPST

tamba kewei nggu-f.
PERF damaged AUX-NrPST
'The house at which John threw a stone has been damaged.'

4 Discussion and LFG analysis

In this section, I propose an LFG analyis to account for the RCs in Marori. The analysis consists of the c-structure analysis (4.1) and the functional constraints captured by the lexical entries of the relativisers and the annotations in the c-structure (4.2). The demonstration of how the analysis works is given in 4.3. Finally, intriguing facts about IHRCs with their possible ambiguity and disambiguation are discussed with reference to the semantic constraints of IHRCs 4.4.

4.1 Structural properties

As mentioned in section 2, the RC is an adjunct, part of an NP within DP. Internally there is good evidence to support that the RC is a CP, with the relativiser in C position, and [Spec, CP] filled in by an XP, e.g. adjunct modifying the RC. This is exemplified in (10) below. Thus, the adjunct *fis* 'yesterday', while showing up before the relativiser *keme*, modifies the RC.

The structure in (10) is for EHRC. Likewise, the IHRC has a CP structure but the mother nodes (NP and DP) are not branching (i.e., without their respective heads) as seen in structure (11).[5] .

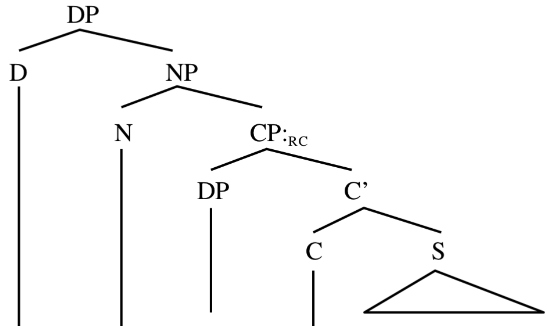
Our c-structure rule can correctly capture the empirical point of the IHRC where the determiner (*efi*) modifying the noun head *ujif* 'bird' shows up outside the relative clause CP, as shown in (12).

5 Note that, since we adopt the conception that all nodes in c-str are optional, the c-structure in (16) can be simplified by not showing the non-branching NP node.

4.2 Lexical entries and CP annotations

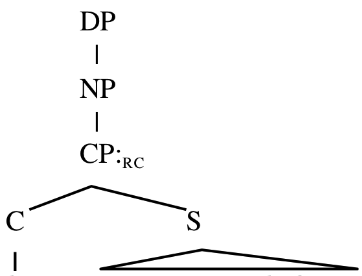
The second part of the analysis deals with the information specified in the lexical entry of the relativiser and the associated annotations on the c-structure of the RC to ensure that both EHRC and IHRC are correctly parsed or generated.

(10)



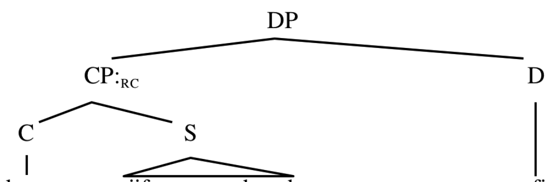
Efi *njaj* ***fis*** *keme* *na* *kufamon*
DET bench yesterday REL 1SG sleep-1SG.DUR.NrPST
'The bench on which I slept yesterday'

(11)



keme *na* *njaj* *kufamon*
REL 1SG bench sleep-1SG.DUR.NrPST
'The bench on which I slept.'

(12)



keme ujif na ndon-du efi
REL bird 1SG 3SG.F.bring-1SG.MPRES the.SG
'the bird which I brought here'

The entries of the general relativiser *keme* and the singular relativiser *kefi* are given in (13). Each carries a set of equations by which the information associated with the selected relativised NP in the RC, which bears FOCUS and a particular GF (grammatical function), is shared by the matrix DP. *Keme* and *kefi* differ only in the INDEX specification that regulates the agreement: unspecified for *keme* vs. specified for singular for *kefi*, indicated by [ ]α and [SG]α respectively. The alpha subscript means that the exact value of the index is yet to be determined, depending on the referential features of the nominal in a given RC. This index equation, as part of the agreement constraint, allows for the possibility that more than one noun could be possibly relativised, e.g. as in example (1).

(13) a.	*keme*	b. *kefi*
	C	C
	(↑FOCUS INDEX)= [ ]α	(↑FOCUS INDEX)= [SG]α
	((↑FOCUS PRED)='pro')	((↑FOCUS PRED)='pro')
	(↑TYPE)=relative	(↑TYPE)=relative
	(↑GF^{+})= (↑FOCUS),	(↑GF^{+})= (↑FOCUS),

The joint specifications ((↑FOCUS PRED)='pro') and (↑TYPE)=relative mean that the relativiser is a relative pronoun and that its pronominal function is optional. This is to capture the fact that in the absence of any overt noun, the relativiser itself is referential; see the case of headless RCs, example (4). When an overt relativised head noun is present (either inside or outside the RC), however, this pronominal meaning is not used. Instead, the PRED value of the head noun is used.

The last line in (13), (↑GF^{+})= (↑FOCUS), indicates that the relativised GF also bears the discourse function of FOCUS. The notation + means that you can have a path here; e.g. OBJ POSS (i.e. the relativisation of the possessor of OBJ). This equation and the index equation (e.g., (↑FOCUS INDEX)= [ ]α) in the entry (which imposes agreement) determine that the specific GF of a particular relativised NP is identified within the embedded RC. In example (1), for instance, there are three GFs (SUBJ, OBJ and instrumental ADJUNCT) eligible to be picked up by the equations of *keme*. However, the matrix auxiliary (*ngguf*) requires '3SG.M' agreement. This excludes the possibility of picking up the embedded SUBJ '1SG' as the relativised nominal. (Being a pronoun also excludes it from the relativisation; further discussed in the next subsection.) Given their compatible referential features, OBJ and ADJUNCT are then possibly relativisable, giving rise to ambiguity, as expected.

In addition to the functional equation in the entry of the relativiser, we need a set of equations on the CP (relative clause) node, shown in (14). The equations regulate the function of the RC and the flow of information, in particular to ensure that the mother node DP picks up its referential information available from the embedded IHRC. The equation ↓∈(↑ADJUNCT) says that the RC is a member of the ADJUNCT set, associated with the mother's GF. The optional ((↑PRED)=(↑ADJUNCT ∈ FOCUS PRED)) means that the functional head of the DP (i.e. the relativised nominal PRED) is supplied by the RC when the mother DP does not have its head PRED. In this way, we capture the essence of an IHRC, where the relativised PRED comes from an NP internal to the embedded clause. If no NP supplies the PRED internally from the embedded RC, given the entry of the relativiser specified in ((13), then the relativiser itself will supply the functional head of the nominal, namely (↑FOCUS PRED)='pro'. In this way, we capture how the headless RC gains the pronominal interpretation translatable as ‘the one (who/which)...’ in spite of the missing head. Finally, the selection of the relativised NP is highly constrained by the agreement system. This is imposed by the equation ($\uparrow$INDEX)=($\uparrow$ADJUNCT $\in$ FOCUS INDEX). The notation $\in$ represents a path into an adjunct set.

(14)

DP
($\uparrow$GF)= $\downarrow$
|
CP
$\downarrow\in$($\uparrow$ADJUNCT)
(($\uparrow$PRED)= ($\uparrow$ADJUNCT $\in$ FOCUS PRED))
($\uparrow$INDEX)= ($\uparrow$ADJUNCT $\in$ FOCUS INDEX)

4.3 Accounting for RC properties in Marori

Having outlined the lexical entries and c-structure annotations, we are now ready to account for RC intricacies in Marori. We start in this section with the straightfoward case of an IHRC, exemplified by (15). The partial c-structure of this sentence is shown in (16)a. Its f-structure is shown in (16). Note that the matrix verb *nggwof* requires [3.SG.F] agreement (not shown in the c-structure but shown in the f-structure as index *j*).

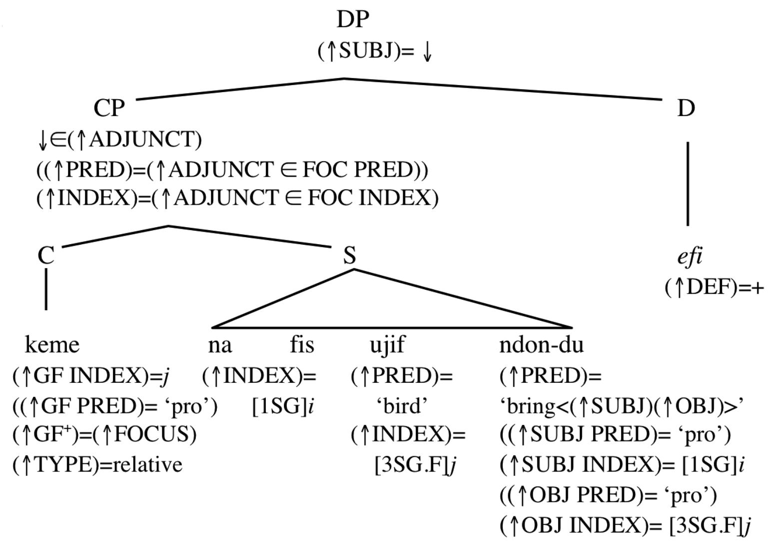
What is special about the IHRC here is that the definite determiner *efi* ‘that.SG’ modifying the noun inside the RC is external to the RC. The proposed analysis, as seen in the f-structure, correctly captures that the [DEF +] information ends up specifying the nominal head PRED ‘bird’, thanks to the equations on the CP node. These equations in effect spread the PRED and INDEX values (tag [1], j) from inside the embedded relative clause (S) to higher nodes in the DP, allowing them to be specified by the determiner carrying [DEF +]. The same mechanism in our analysis accounts for the Marori fact that the quantifier *usindu* ‘all’ or numeral *yanadu* ‘two’ can also quantify a relativised NP inside the RC from a position external to the RC. This is shown in (17).

(15)	[[keme	na	fis	**ujif**	ndon-du]$_{RC}$]$_{NP}$	**efi**
	REL	1SG	yesterday	bird	3SG.F bring.here-1SG.PRES	that.SG

tamba	yaba	nggwo-f
already	dead	AUX3SG.F-NrPST

‘The female bird that I brought here yesterday already died.’

(16) a.



b.

```
[ ...
  SUBJ  [ PRED      [1]:'bird'
          INDEX     j: 3SG.F
          DEF       +
          ADJUNCT { [ TYPE        relative
                      FOCUS   [2] [ PRED    [1]
                                    INDEX   j
                                    N-TYPE  common ]
                      PRED    'bring<(SUBJ)(OBJ)>'
                      SUBJ        [ PRED    'pro'
                                    INDEX   i: 1SG
                                    N-TYPE  pronoun ]
                      OBJ     [2]
                      ADJUNCT     'yesterday'          ] } ] ]
```

(17)

[keme	na	fis	**ujif**	kein-du]$_{RC}$	**usindu**
REL	1SG	yesterday	bird	3NSG.bring.here-1SG.PRES	all

tamba	yaba	nggorforo-f
already	dead	AUX.3PL-NrPST

'All of the birds that I brought here yesterday already died.'

IHRCs lead to ambiguity, a phenomenon also known in IHRCs in other languages. The ambiguity can be now straightforwardly accounted for in our analysis. The relevant example, repeated as (18), is the case where two or more NPs inside the RC whose index features are not in clash with those imposed by the agreement in the matrix predicate. They are therefore eligible to be picked up as the functional head of the nominal. These are 'bench' and 'club' (readings (a)-(b)), but not the pronoun *na* '1SG' (reading (c), which is excluded due to a feature clash).

(18)	[Keme	na	**njaj**=i	**samagau**	ngge	terme-ben]$_{IHRC}$
	REL	1SG	bench=U	club	with	3SG.U.M.hit-1SGNrPST

tamba	keiwei	nggu-f
PERF	damaged	3SG.M.U.AUX –NrPST

a) 'The **bench** that I hit with the club was damaged.'
b) 'The **club** with which I hit the bench was damaged.'
c) *'I, who hit the bench with a club, was damaged.'

The f-structure for reading (b) is given in (24). As seen, the nominal PRED 'club' (tag [1], INDEX *k*) is inside the RC as an adjunct. The postpositional *ngge* 'with' is treated like a prepositional case marker flagging the ADJUNCT TYPE of instrument (i.e. carrying no PRED attribute). The whole ADJUNCT value is shared by FOCUS, due to the relativisation enforced by the specification in the entry of *keme* (cf. (24), tag [2]).

(19)

```
[ PRED   'damaged<(SUBJ)>'
  TNS    NrPST
  SUBJ   [ PRED      [1]
           INDEX     k: 3SG.M
           ADJUNCT { [ TYPE      relative
                       FOCUS     [2]
                       PRED  'hit<(SUBJ)(OBJ)>'
                       SUBJ        [ PRED    'pro'
                                     INDEX   i: 1SG
                                 [3] N-TYPE  pronoun ]
                       OBJ         [ PRED    'bench'
                                     INDEX   j:3SG.M
                                 [4] N-TYPE  common ]
                       ADJUNCT  { [ TYPE   instrument
                                    PRED   [1]: 'club'
                                    INDEX  k: 3SG.M
                               [2]  N-TYPE common ] } ] } ] ]
```

Given the [3SG.M] SUBJ agreement of the matrix structure required by the verbal auxiliary *ngguf*, the value specifications of OBJ 'the bench' (tag

[4], i.e. ‘bench’, index *j*) also satisfy the requirement. Hence, OBJ [tag [4] is predicted to be possibly selected as FOCUS (i.e. relativised) and matrix SUBJ. The embedded SUBJ, carrying an index of [1SG] is expected to cause a clash in agreement and cannot be selected as the matrix SUBJ, as predicted.

A brief note is needed for detached RCs, as in example (6)b. Detached RCs can be straightforwardly handled in the proposed analysis as floating headless RCs. Thus, the relativiser supplies the PRED (i.e. ‘pro’) value, taking the detached noun head as its antecedent anaphorically. Sentence (6)b literally means ‘The person_i is named John, the one_i whose pig already died’, where the relativiser *kefi* ‘the one’ refers to ‘the person’ (indicated by index *i*). For reasons of space, no c-structure/f-structure representations are given here, but such representations could be determined with special care given to the annotation of c-structure nodes, to ensure that the right information is passed to the adjunct set of the relative clause.

4.4 Semantic constraints and IHRC typology

This section addresses the definiteness constraint of IHRC in Marori, a phenomenon also observed in other languages such as Lakhota (Williamson 1987, Culy 1990), Dàgáárè (Bodomo and Hiraiwa 2010), Kobon and Wappo (Dixon 2009:331). In what follows, the nature of the constraints is outlined, but no formalisation is proposed. Such formalisation should be handled as part of the semantic structure representation.

We have seen that two NPs are equally relativisable, giving rise to ambiguity. However, when one of NPs in the IHRC is a proper name, no ambiguity arises, as seen in example (20). The proper name cannot be relativised in IHRCs in Marori.

(20) [Keme Markus bosik =i ife -f]$_{\text{IHRC}}$
REL Markus pig=U 3SG.M.see-3NPL.NrPST

tamba kundo -f
PERF run-3SG.NrPST

i) ‘The pig that Markus saw ran off.’
ii) * ‘Markus who saw the pig ran off.’

The effect of the definiteness constraint in Marori is also observed with quantification. Recall that the NP relativised in the IHRC can have its quantifier *usindu* ‘all’ outside the RC; example (17). The universal quantifier *usindu* presupposes definite referents, and when it quantifies a noun internally within the IHRC it resists relativisation. Thus we have no ambiguity in the following example:

(21) [Kemde usindu meninggon=i purfam paar]$_{\text{IHRC}}$
REL all child.PL=U person money

njemba-b tamba sra-f
3.give-3PL.NrPST already go.PL-NrPST
i)‘The people who gave money to all the children already went away.’
ii) *‘All the children who were given money by the people already went away.’

I argue that the definiteness constraint, as observed in the preceding examples, is a logical consequence of the typological property of the Marori

IHRC, which is essentially a restrictive type. The restrictive RC is known to have the salient features of non-specific indefinite and intersective force at the level of the RC (Grosu 2012). A highly definite or unique referent like a proper name does not allow the intersective interpretation and is therefore inconsistent with the restrictive RC. Evidence for Marori IHRCs as restrictive is given below, but I will first outline Grosu's (2012) typology of (IH)RC briefly.

Grosu (2012) distinguishes three semantic types of RCs: restrictive, appositive and maximalising. The first two are exemplified from English RCs shown in (22)a-b. In (22)a, there were more than three boys at the party, and only three of them had beards; thus, the RC imposes a further restriction to the denotation of 'boys'. The RC information is essential for the identification of the referent of the head noun. In (22)b, there were only three boys in the party; all of them had beards. The RC imposes no referential restriction, and the RC can therefore elide without affecting the identification of the referent of the relativised noun.

(22) a. At the party, I saw only [[three boys [*who had beards*]$_{RC}$]$_{NP}$ (restr.)

b. At the party, I saw only [[three boys, [*who had beards*]$_{RC}$]$_{NP}$ (apps.)
(Grosu 2012:452, ex. (6))

The maximalising RC is like the appositive RC, in that the denotation is already specific/definite. They are different in the locus of the specificity/definiteness. In the appositive RC, it is fully defined in the matrix NP, as seen in the meaning of (22)b above. In the maximalising RC, the definiteness is fully defined within the RC itself, exemplified by (23). The maximalising RC is, in a way, like the restrictive RC, in that its information is highly essential. It is 'super restrictive', making the referent of the relativised noun maximally definite; e.g. when the noun is referentially plural it gives rise to the totality of plural meaning, as seen in the example from Japanese in (24)b below. In short, maximalising relatives have the characterisation of having strong definite import, presupposing the relativised noun to be maximally definite; hence, dropping *the* in (23) would degrade its acceptability.

(23) I suddenly noticed [the three books [that there were on your desk]$_{RC}$]
(i.e. 'there were exactly three books on your desk and I suddenly noticed them')
(Grosu 2012: 453)

Japanese provides a good illustration involving a maximalising IHRC. It should be noted that Japanese has both EHRC and IHRC, but the IHRC is of the maximalising RC type. The EHRC is exemplified in (24)a and its IHRC counterpart is in (24)b (Grosu's (18) and (20)). Note the difference in meaning as seen in the free translation. The maximalising relative encodes the definite totality of the cookies to be brought by Taro to the party, which is not the case in (24)a.

(24) a.

Taro-wa	[[Yoko-ga	reezoko-ni	__	irete-oita]
Taro-TOP	Yoko-NOM	fridge-Loc		put-AUX

hotondo-no	**kukkii-o]**	paatii-ni	motte	itta
almost-all-Gen	cookie-Acc	party-to	brought	

'Taro brought to the party almost all the cookies that Yoko had put in the fridge.'

b.

Taro-wa	[[Yoko-ga	reezoko-ni	**hotondo-no**	**kukkii-o**
Taro-TOP	Yoko-NOM	fridge-Loc	almost-all-Gen	cookie-Acc

irete-oita]-no-o	paatii-ni	motte	itta
put-AUX-NML-Acc	party-to	brought	

'Yuko put almost all the cookies in the fridge and Taro brought {them, *some} to the party.'

Turning to Marori, I argue that IHRCs in this language are restrictive, not maximalising. That is, they do not presuppose definite/specific denotation of the relativised nominal. First evidence for this comes from the fact that the IHRC in Marori can have the full range of quantificational forces, including an existential force, a salient property of the restrictive relative. This is shown in example (25), where the referent of the relativised noun 'boy' is indefinite. It should be noted that, while not presupposing definiteness, the IHRC does provide specifications which make the referent of the relativised noun definite, with the possibility of the (strong) determiner overtly present at the matrix level; see example (15).

(25)

Ka	einda	kefi	meipur	di	nie=fi	kuye-den
2SG	3.search.2PL	REL	child.SG	FUT	1NSG=with	stay-1DU.PRES

'You look for a child who wants to stay together with me.'

Other evidence that IHRCs are restrictive comes from the fact that they have intersective import, providing a restriction to the denotation of the associated noun in the same way as an ordinary adjunct. To understand this, first consider the restrictive relative clause in English in (26). The denotation of the object bought in (26) is the one in the intersection of sets of 'books', 'cheapest things', and 'things which are not paperbacks'.

(26) I bought the cheapest book which was not a paperback.
(Arnold 2007, ex. (1b))

(27) a. I've never spoken to Kim, who plays poker.

b. *I've never spoken to Kim who plays poker.
(Arnold 2007, ex. (1b))

In contrast to (26), applying a restrictive RC to the proper name Kim in (27) results in downgraded acceptability as seen in (27)b. This is attributed to the referential uniqueness of a proper name which is inconsistent with the intersective interpretation of restrictive RC, which requires a set of referents. Thus, it is not suprising to see why a restrictive RC is not possible with a proper name.

Arguably, the same semantic constraint accounts for why proper names are not relativisable in the IHRC type in Marori, as seen in (20), giving rise to no ambiguity. The same is true for NPs by the universal quantifier *usindu* 'all', which requires definite referents; see example (21).

Another related effect of this intersective import is the introduction of contrast set into the discourse by the restrictive RC. Then, this contrast set can be accessed anaphorically. In English, this is done by an expression like *others*, as in (28)a. Non-restrictive RCs do not introduce such a set; hence the downgraded acceptability of sentence b. The equivalent example in Marori is (29), where the contrast set introduced by the IHRC in the first clause is accessed by *now* in the second IHRC. The possibility of clauses like in (29) provides supports that Marori IHRC is of the restrictive type.

(28) (a) I like puzzles which require imagination and creativity, and others that just depend on knowledge. [Restrictive]

(b) #I like puzzles, which require imagination and creativity, and others that just depend on knowledge. [Non-restrictive]
(Arnold 2007, examples (4a-b))

(29)

Keme=na	ujif	kein-du,	keme	Jhon	**now**	kein
REL=1SG	bird	bring-1SG.MPRES	REL	John	other	bring.NPST

tamba	yaba	nggorforof
PERF	dead	AUX.PLURAL-NrPST

'The bird which I brought, (and) the other ones which John brought were dead.'

Given the intersective interpretation, it is therefore possible, and indeed natural, to stack restrictive RCs to provide further intersective specifications to make the denotation more specific. Crucially, both restrictive and appositive RCs can have the stacking, exemplified by (30) and (31). This is not, however, possible with the maximalising RCs (Grosu 2012), example (32). In (30), the depicted (three) referents are on the intersection of sets of individuals who are 'boys', 'bearded', and 'wearing no shoes'. The boldfaced RC further restricts the individuals picked up by the italicised RC, leaving open the possibility that the speaker may have seen additional bearded boys who wore clothes. In the appositive RC (31), the boldfaced RCs provide additional information but the referent of the relativised noun ('the Honourable member') remains unique (without any possibility of others). In (32), the boldfaced RC is unacceptable in its full version. That is, according to Grosu (2012:454), the denotation of the construction is already fully determined by the italicised relative, leaving no interpretation for it.

(30) At the party, I saw only [three boys *who had beards* **who wore no clothes**]. (Grosu 2012: example 8a)

(31) I fear the Honourable Member, *who nobody trusts*, **who nobody believes, who not even his own supporters listen to,** has finally run out of time. (Arnold 2007: example 56; bold and italics added)

(32) I suddenly noticed [the three books *that there were on your desk* **that (*there) had earlier been on my desk**].

In Marori, stacking of IHRCs is possible, as seen in (33). This supports the analysis that IHRCs in Marori are restrictive, not maximalising in nature. Lakhota also has a restrictive IHRC, and is expected to allow RC stacking, as seen in example (34):

(33)

Na	fis	purfam=i	eyew=nda-mon
1SG	yesterday	person=U	eye=3.AUX-1SG.NrPST

[*kefi*	*koro*	*imbirif*	**kefi**	**kundo-f**]
REL	dog	bite-NrPST	REL	run-3NrPST

'Yesterday I was looking for (the/a) person [that dog bit that ran away].'

(34)

[Ogle	eya	šapšapa	cha]	agli	pi	wachi	ki	lena	e
shirt	some	dirty	indef	take-home	Pl	1-want	the	these	be

'These are the shirts that are dirty that I want them to take home.' (Grosu 2012: 455)

5 Final remarks and further research

In this final section, I will provide a brief summary of the facts and analysis, and then point out further research needed.

This paper has demonstrated the different types of relative clauses in Marori. The paper hopefully contributes to the empirical basis for the typological and theoretical studies of RCs. Marori is unusual in that it has almost all of the relative clause types: headed and headless; if headed, both externally and internally headed; if externally headed, either attached in the nominal structure or detached from it; if attached, either pre- or post-nominal. Of particular interest is the morphosyntax and semantics of the IHRC. It has been demonstrated that LFG formalism is well equipped to capture the structural intricacies and intriguing properties of IHRCs, such as the possibility for ambiguity/non-ambiguity, which relates to nominal types and definiteness. Nominals with unique and definite referents, such as proper names and pronouns, cannot be relativised in the IHRC. It has been argued that this is due to the fact that IHRCs in Marori are of the restrictive type, not the maximalising one, in Grosu's typology. The salient semantic feature of the restrictive RC is that it has non-specific indefinite and intersective import. As such, it provides some specification to a referent of a set, introducing a contrast to the set. This intersective import is therefore semantically incompatible with a unique definite referent where no contrast (set) is possible. As outlined in this paper, salient properties of IHRCs in Marori, such as the possiblility of stacking IHRCs or the inability of proper names and pronouns to be internally relativised, have a good semantic basis.

Most of the data presented in this paper was elicited. Future studies on RCs in Marori must therefore include an investigation of the distribution of different types of RCs in natural texts. The corpus-based study is expected to further reveal and illuminate the semantic-discourse constraint of specificity/definiteness in RCs.[6] It has been mentioned in this paper that a

6 A Marori corpus is being developed as part of the Sothern New Guinea (SNG) and the ELDP projects. The corpus consists of audio-audio recordings with

pronoun and a proper name cannot be relativised in single-headed RCs in Marori (and also other languages such as Lakota (Culy 1990:168)). However, there is a surprising twist, which shows that such a pronoun can be relativised, provided that it is expressed in a double-headed RC. That is, the pronoun is expressed twice, internally within the RC and also externally in the matrix position as shown in (35). This kind of nominal doubling, giving rise to a double-headed RC, is attested in other Papuan languages with IHRCs for disambiguation, e.g. as seen in the contrast in examples in (36) from Kobon (Dixon 2009:331). It is not completely clear at this stage of our research whether structure (35) in Marori and (36)b in Kobon are in fact two juxtaposed free clauses, rather than complex structures with embedded RCs. This could be the case for Kobon, for which the alternative translation would be 'the boy hit the girl (and) I know the girl.' A similar two-free-clause analysis in Marori, however, would have a problem in accounting for the presence of the RC marker *keme*. Dropping *keme*, which is possible in Marori, would indeed make the structure analysable as two juxtaposed free clauses. In the presence of *keme*, the structure in (35) can perhaps be analysable as belonging to a double-headed RC. Further investigation is needed what happens if proper names, and even common nouns, are forced to appear in double-headed RCs in Marori.

(35) keme **na** bosik=i ife-ben tamba=**na**
REL 1SG pig=U 3SG.M.see-1SG.NrPST PERF=1SG

kundo-bon
run-1SG.NrPST

'I, who saw the pig, ran off.'

(36) a. [ña pai pak-öp]$_{RC}$ yad nöŋ-bin
boy girl hit-PERF.3SG 1SG perceive-PERF.1SG
i. 'I know the girl who the boy hit.'
ii. 'I know the boy who hit the girl.'

b. [ña **pai** pak-öp]$_{RC}$ **pai** yad nöŋ-bin
boy girl hit-PERF.3SG girl 1SG perceive-PERF.1SG
i. 'I know the girl who the boy hit.'
ii. *'I know the boy who hit the girl.'

Finally, a RC-related issue which is of particular interest and needs further investigation is the connection between (clausal) nominalisation and the nature of structural embedding with its constraints in head-final (OV) languages like Marori. There is an intriguing behaviour of IHRCs in Marori with a preference for discontinuous post-verbal RCs. This is perhaps what is expected for an OV language, in line with the finding reported in the literature that the reduction of preverbal arguments in SOV languages is a compensatory strategy to reduce the heavy cost in production and comprehension (Hawkins 2004, Ueno and Polinsky 2009, and the references therein). Further research in this area will include an in-depth corpus

transcriptions in ELAN, covering a range of topics. This corpus will be made available to the public in due course through PARADISEC (http://www.paradisec.org.au/) and ELAR (http://www.elar-archive.org/).

investigation of Marori, preferably including comparison with (OV/VO) languages with IHRCs, to gain further empirical evidence for any analysis proposed.

References

Andrews, Avery. 2007. "Relative clauses." In *Language typology and syntactic description*, edited by Timothy Shopen, 206-236. Cambridge: Cambridge University Press.

Arka, I Wayan. 2016. Information structure and its morphosyntactic resources in Marori. Paper presented at The Workshop on Information Structure at TUFS (18-20 February 2016).

Arnold, Doug. 2007. "Non-Restrictive relatives are not orphans." *Journal of Linguistics,* 43 (2):271-309.

Bodomo, Adams, and Ken Hiraiwa. 2010. "Relativization in Dàgáárè and its typological implications: left-headed but internally-headed." *Lingua,* 120:953-983.

Bresnan, Joan. 1982. *The mental representation of grammatical relations.* Cambridge, Massachusetts: the MIT Press

Bresnan, Joan, Ash Asudeh, Ida Toivonen, and Stephen Wechler. 2015. *Lexical-Functional Syntax (second edition).* Oxford: Wiley-Blackwell.

Butt, Miriam. 2014. "Question and Information Structure in Urdu/Hindi " In *Proceedings of LFG14*, edited by M Butt and T. H King, 158-178. CSLI, http://web.stanford.edu/group/cslipublications/cslipublications/LFG/19/lfg14.html.

Comrie, Bernard, and Edward L. Keenan. 1979. "Noun phrase accessibility revisited." *Language*, 55 (3):649-664.

Culy, Christopher Douglas. 1990. *The syntax and semantics of internally headed relative clauses*, Linguistics Department, Stanford University, Stanford.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*, *Syntax and semantics*. San Diego: Academic Press.

Dixon, R.M.W. 2009. *Basic Linguistic Theory*. Vol. 2. Oxford: Oxford University Press.

Grosu, Alexander. 2012. "Towards a more articulated typology of internally headed relative constructions: the semantic connection." *Language and Linguistics Compass*, 6 (7):447-476.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Keenan, E, and B. Comrie. 1979. "Data on the Noun Phrase Accessibility Hierarchy." *Language*, 55:333-51.

Keenan, Edward L., and Bernard Comrie. 1977. "Noun phrase accessibility and universal grammar." *Linguistic Inquiry*, 8:63-99.

Krifka, Manfred. 2008. "Basic Notions of Information Structure." *Acta Linguistica Hungarica* 55 (3–4):243–276.

Sag, Ivan A, Thomas Wasow, and Emily M Bender. 2003. *Syntactic theory: a formal introduction (second edition)*. Stanford: CSLI Publications.

Ueno, Mieko, and Maria Polinsky. 2009. "Does headedness affect processing? A new look at the VO-OV contrast." *Journal of Linguistics*, 45:675-710.

Williamson, J. 1987. "An Indefiniteness Restriction for Relative Clauses in Lakhota." In *The Representation of (In)definiteness*, edited by E. Reuland and A. ter Meulen. Cambridge, MA.: The MIT Press.

One of those constructions that really needs a proper analysis

Doug Arnold
University of Essex

Christopher Lucas
SOAS, University of London

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 43–63

Keywords: HPSG, English, syntax, semantics, agreement mismatch

Arnold, Doug & Christopher Lucas. 2016. One of those constructions that really needs a proper analysis. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 43–63. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.3.

Abstract

We describe, and provide an HPSG account of, a hitherto little studied English construction (of which the title of the paper is an instance) involving an agreement mismatch: a partitive construction in which a plural nominal is apparently modified by a singular relative clause.

1 Introduction

Example (1) seems impeccable to almost all English speakers:

(1) This is one of those problems that really bothers me.

This should be surprising, since it involves an 'agreement mismatch', a failure of agreement between a plural nominal head (*those problems*) and a singular modifier – the relative clause *that really bothers me* contains a third person singular verb, indicating that the subject is singular, as made explicit in (2). Normal agreement would produce (3).

(2) This is one of those problems$_{pl}$ [which Δ_{sg} really bother**s** me].
(3) This is one of those problems$_{pl}$ [which Δ_{pl} really bother me].

We think any remotely plausible grammatical theory that provides an account of partitive constructions and relative clauses will predict that (2) is ungrammatical, and that speakers should reject it in favour of (3), but they do not. On the contrary, the 'mismatch' construction exemplified in in (2) is widely attested, and normally goes un-noticed (some slightly edited corpus examples can be seen in (4), the first three are from the British National Corpus; as a reading aid, here and below we underline the plural noun and item in the relative clause that signals that there is a mismatch). It is in fact far more frequent than the alternative construction, even in relatively careful writing, at least with typical nouns such as *things* – searching Google Books for *one of the things that bothers me*, which is an instance of the mismatch construction, gives about 5860 results; searching for the corresponding form with regular agreement (i.e. with *bother* in place of *bothers*) gives 698 hits.[1]

[†] As well as HeadLex 2016 in Warsaw, versions of this paper have been presented at the Third European Workshop on HPSG ('Auf nach Frankfurt', in Frankfurt, November 2015), and the 2016 meeting of the Linguistics Association of Great Britain. We are grateful to participants at those meetings, to our colleagues at Essex and SOAS, and anonymous referees for HeadLex 2016 for discussion, comments, and support. Remaining deficiencies are entirely our responsibility.

[1] We assume here that (2) and (3) are equivalent, except that the mismatch construction is more frequent. This may be an over-simplification: for example, while the present authors think there is not much to choose between (2) and (3), we find it much more natural to talk about *one of the things that bugs me about you*, where there is an agreement mismatch, than to use the plural verb form *bug* in the relative clause.

(4) a. Dr Hemingway and colleagues [. . .] have also found one of the genes that makes malaria-transmitting mosquitoes resistant to pesticides such as DDT. [AKD/871]

b. This generation of vipers has again bitten one of the hands that was stretched out in blessing it. [B1J/1984]

c. The Cullen report is widely recognised as one of the most excellent reports that has ever been produced on matters that affect industrial safety. [HHX/19354]

d. It was, and remains, one of the best goals that has ever been scored at Carrow Road. . . [Edward Couzens-Lake, *Norwich City in the Eighties*, Amberley Pub., Stroud, 2015]

While the mismatch construction in (2) has been noted before (e.g. Huddleston and Pullum (2002, p506), Pinker (2014, p250) and in some prescriptive grammars, e.g. Burchfield (2004, p30,550)) it has not received much attention in the formal literature: de Hoop et al. (n.d.) is the only discussion we are aware of.[2] The goal of this paper is to provide a relatively detailed discussion of the construction, by exploring a number of potential, but flawed, approaches (Section 2). On the basis of this, in Section 3 we will outline a rather straightforward, and descriptively plausible, formal account, using very standard HPSG apparatus. Section 4 notes some problems and open questions.[3]

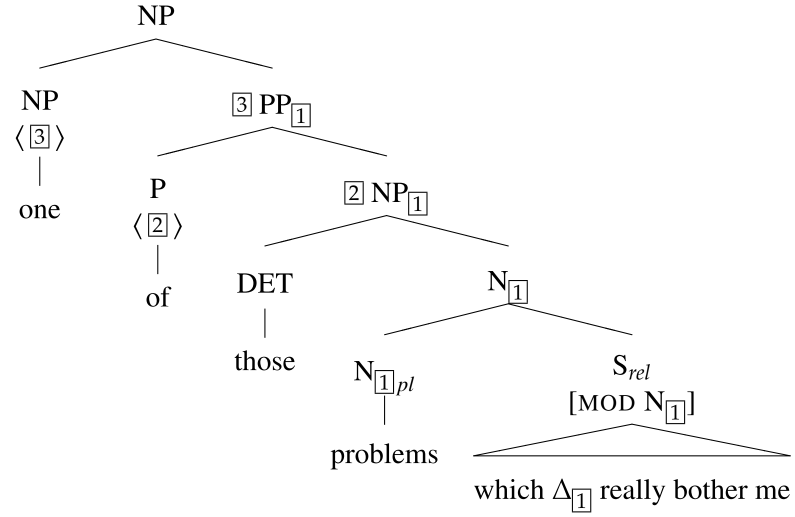
2 The Problem, and Some Non-solutions

It is useful to begin with discussion of cases involving 'normal' agreement, such as (3), for which (5) is a plausible representation. Overall, what we have in (5) is a partitive construction where the 'quantity word' is *one*. Following existing analyses of partitive constructions in HPSG (e.g. Kim (2002), Flickinger (2008), Kim and Sells (2008)), we assume that *one* is a nominal here, which selects a PP complement headed by *of*, which must itself contain a plural NP. We assume that *of* is a non-predicative preposition here, so that it has the same content (CONT) as its complement NP. As a consequence it has the same INDEX value, which we have indicated as a boxed subscript (as is standard in HPSG, we assume indices are feature bundles containing PERSON, NUMBER, and GENDER attributes). As one would expect, this index is also shared by the

[2] A good collection of links and prescriptive grammar sources that mention the construction can be found at http://english.stackexchange.com/questions/232255/is-this-correct-one-of-the-things-that-makes-him-great-is.

[3] Our discussion here is restricted to English. We have some indications that a similar construction exists in several other languages, including Maltese and Spanish. However, we will not pursue this here. See de Hoop et al. (n.d.) for discussion of a similar construction in Dutch.

(5)



head noun *problems*, which is plural, so $\boxed{1}$ in (5) stands for [NUM *pl*].[4]

Our assumptions about relative clauses reflect the analysis of Sag (1997), except that we take (restrictive) relative clauses to be modifiers of N (or perhaps Nom), rather than NP, and that subject relatives involve a gap (both are for presentational reasons, neither has significant implications for the analysis). The fact that relative clauses are nominal modifiers is indicated by their MOD value.[5] General principles require that in head-adjunct structures the SYNSEM value of the head daughter must be identical to the MOD|SYNSEM value of the adjunct daughter, with the consequence that the index of the head (SYNSEM|LOCAL|CONTENT|INDEX) must be identical to the index in the MOD value of the adjunct: since *problems* is $N_{\boxed{1}}$, the relative clause *that really bother me* must be [MOD $N_{\boxed{1}}$], as it is in (5). The grammar of relative clauses involves a set-valued feature REL, which in English is constrained to contain at most one element. This element is identical to the index of the modified noun, and (in the case of *that*-relatives as in (5)) the index of the relativized NP. Hence in (5) the relative clause is [MOD $N_{\boxed{1}}$], and the index of the subject is also $\boxed{1}$, i.e. [NUM *pl*]. Normal agreement processes require the third person plural form *bother*, which is what we have. Similar constraints apply in *wh*-relatives and non-finites.

[4]Alternative analyses might treat *one* as a determiner associated with a phonologically empty head noun which takes the PP as a complement, or perhaps involve some kind of a ellipsis (cf. *one ~~problem~~ of those problems*). Technical details aside, nothing in what follows depends on this.

[5]Much recent work uses the SELECT feature in place of MOD: changing this would have no effect on the analysis here.

As regards the semantics, Sag's analysis of relative clauses involves a constraint like (6). In words: if a relative clause modifies a nominal whose index is $\boxed{1}$ and whose restrictions are $\boxed{2}$, it produces a nominal whose index is $\boxed{1}$, and whose restrictions are $\boxed{2}$ plus the propositional content of the relative clause.

(6)

$$\begin{array}{cc}
\multicolumn{2}{c}{\text{N}} \\
\multicolumn{2}{c}{\left[\text{CONT}\ \left[\begin{array}{ll}\text{INDEX} & \boxed{1} \\ \text{RESTR} & \boxed{2} \cup \{\boxed{4}\}\end{array}\right]\right]} \\
\text{N} & \text{S}_{rel} \\
\left[\text{CONT}\ \boxed{3}\left[\begin{array}{ll}\text{INDEX} & \boxed{1} \\ \text{RESTR} & \boxed{2}\end{array}\right]\right] &
\left[\begin{array}{ll}\text{MOD} & \left[\text{CONT}\ \boxed{3}\right] \\ \text{CONT} & \boxed{4}\ \textit{proposition}\end{array}\right]
\end{array}$$

In the case of a normal plural relative modifying a plural nominal as we have in (5), this will amount to (7), where the propositional content of the relative clause has been added to the restrictions of the noun. Intuitively, the value of CONT on the mother node here describes a plurality X such that X is a plurality of problems, and where this plurality bothers the speaker.

(7)

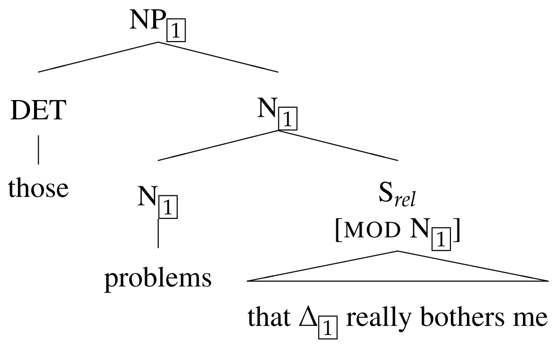
$$\begin{array}{cc}
\multicolumn{2}{c}{\text{N}} \\
\multicolumn{2}{c}{\left[\text{CONT}\ \left[\begin{array}{ll}\text{INDEX} & \boxed{1} \\ \text{RESTR} & \{\, \textit{problems}(\boxed{1}), \textit{bother}(\boxed{1}, \textit{me})\,\}\end{array}\right]\right]} \\
\text{N} & \text{S}_{rel} \\
\left[\text{CONT}\ \boxed{3}\left[\begin{array}{ll}\text{INDEX} & \boxed{1} \\ \text{RESTR} & \{\, \textit{problems}(\boxed{1})\,\}\end{array}\right]\right] &
\left[\begin{array}{ll}\text{MOD} & \left[\text{CONT}\ \boxed{3}\right] \\ \text{CONT} & \textit{bother}(\boxed{1}, \textit{me})\end{array}\right] \\
| & \triangle \\
\text{problems} & \text{that bother me}
\end{array}$$

We can now see quite sharply what is problematic about the mismatch construction from (2), repeated here:

(2) This is one of those problems$_{pl}$ [that Δ_{sg} really bothers me].

The obvious analysis will give a representation as in (8) for the downstairs NP.

(8)



Just like the case of normal agreement in (5), the mismatch construction in (8) contains an N and a relative clause, and there is an index $\boxed{1}$, which must be the same everywhere indicated. The difference is that in (8) it clearly *cannot* be the same everywhere, because the subject of *bothers* must be [INDEX | NUM *sg*], and *problems* is clearly [INDEX | NUM *pl*], because of its morphology and the plural determiner *those*. And of course, (8) is completely ungrammatical if it appears anywhere apart from the construction we are discussing here:

(9) *those problems which really bothers me

On the face of it, this construction poses a serious challenge for standard accounts of agreement in HPSG e.g. Pollard and Sag (1994), Kathol (1999), Wechsler and Zlatić (2003).

There is clearly something wrong: such a construction should be impossible, but not only is it possible, it seems to be generally used, and even preferred, by native speakers. But it is far from obvious *what* is wrong. In the remainder of this section we will consider a number of more or less plausible suggestions.

Perhaps the best starting point for an analysis is a consideration of other agreement mismatches, of which there are several in English. It seems to us very unlikely that the mismatch construction can be analysed as one of them.

Examples like (10) show a plural expression (*ten days in Florida*) being treated as singular (cf the singular determiner *that*, and singular agreement on the verb). Huddleston and Pullum (2002, p354) talk about the plural nominal here being 'respecified' as singular (cf also Maekawa (2015), other cases of 'respecification' are discussed in Pollard and Sag (1994, Ch2)):

(10) [That ten days we spent in Florida] was fantastic.

But this is rather unlike our construction. What we have in (10) is a plural NP respecified as singular (denoting a single entity – a group or collection), our construction is rather the reverse in that it involves a singular predicate being understood as plural.

Measure phrases (pseudo-partitives) also show some odd (variable) agreement behaviour, witness (11):

(11) a. That pile of problems that has puzzled philosophers down the ages...
b. That pile of problems that have puzzled philosophers down the ages...

But of course our construction does not involve a measure phrase, and what we see in (11) seems to be a straightforward matter of high vs. low attachment – the relative clause can either be attached high, and interpreted as modifying *pile* (in which case it is singular) or attached low, modifying *problems*, in which case it is plural. As will become clear below, this line of analysis is of no help with our construction.

Some measure phrases seem to be 'transparent' to number, e.g. *a lot of problems* seems to be internally singular (because of the singular article), but it is externally plural. For these the only option is plural agreement:

(12) [A lot of problems] have/*has been solved today.

However, notice that with measure phrases like this the mismatch construction is never allowed:

(13) a. One of those problems$_{pl}$ that Δ_{sg} annoys you ...
b. *A lot of those problems$_{pl}$ that Δ_{sg} annoys you ...

Conceivably, an alternative analysis might be that there is something special about the head noun, e.g. that in (8) *problems* is not plural. But this seems utterly implausible: not only does *problems* have plural morphology, it has a plural determiner (*those*), and in partitives involving countable nouns the downstairs NP is *always* plural (and only countables will be possible with a numeral like *one*) – a fact which makes intuitive sense given the meaning of the partitive, which involves selecting from a set or collection.

A potentially more plausible approach might involve the idea that there is something odd about the relative clause. While one cannot rule this out completely, the difficulty is that there are no obvious restrictions on the kind of relative clause that is possible in the mismatch construction. The example we have discussed so far is a subject relative with *that*. But relatives involving *which* and other relative pronouns are possible:

(14) a. This is one of those problems which really bothers me.
b. She is one of those people who really annoys me.

As (15) shows, the relativized NP need not be a 'top-level' subject:

(15) This is one of those problems that$_i$ [we think [Δ_i deserves urgent attention]].

Examples involving relativised subjects of finite verbs are the most obvious, because of agreement phenomena, but examples involving non-subject relatives can also be constructed. In (16) the relativised NP is the object of *add*, so there is no indication of its number on the verb, but it must be singular, because it is co-indexed with the singular *itself*. Similarly, in (17) the relative pronoun *who* is interpreted as the object of *leave*, and is co-indexed with singular *himself*. (18) makes the same point: *who* is the object of *understand*, it must be singular because it is co-indexed with singular *his*.

(16) This is one of those numbers (that) you can add Δ_i to itself$_i$ to get an interesting result.

(17) He is one of those people (who$_i$) you should leave Δ_i strictly to himself$_i$.

(18) He is one of those patients (who$_i$) you can't understand Δ_i until you have met his$_i$ mother.

Notice that in each of these *that* or the relative pronoun are optional, i.e. bare relatives are possible. The following exemplify some other kinds of relative clause. The examples in (19) involve pied-piping, and (20) shows examples with non-finite relative clauses:

(19) a. He's one of those people [about whom$_i$] even his$_i$ best friends have reservations.
b. He's one of those candidates [about whose$_i$ electoral prospects] not even his$_i$ strongest supporters could be certain.
c. He may turn out to be one of those musicians [whose$_i$ appeal] is only clear when you actually see him$_i$ live.

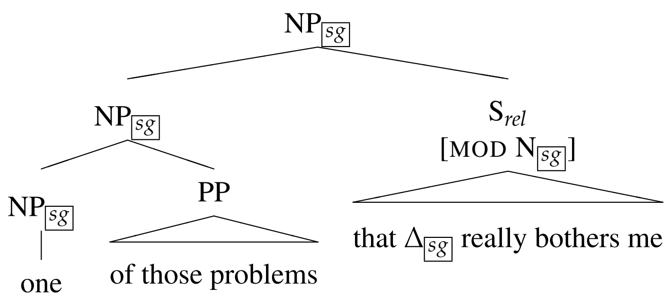
(20) a. His sister had married one of the first merchants [to establish himself as a plantation owner in Virginia].
b. The Weisswurst is one of those sausages [intended to be eaten without its skin].

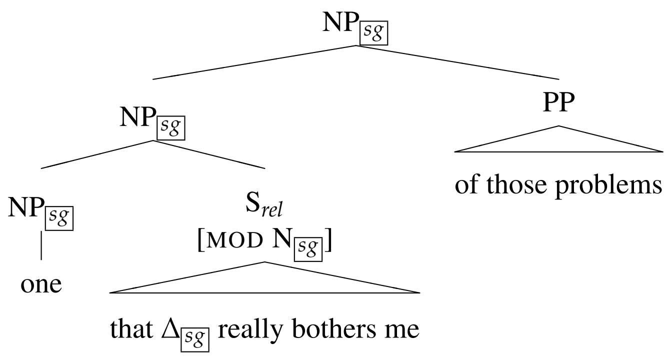
What this suggests is that there is nothing inherently odd about the kind of relative clause that appears in the mismatch construction, and consequently no motivation for introducing a special kind of 'mismatch relative' which only appears in this construction. This means that we are stuck with the idea that the construction involves a singular relative (i.e. one specified as [MOD $N_{\boxed{sg}}$]).

The following is a more initially appealing, but still fatally flawed, approach. Externally, partitive NPs with the quantity word *one* trigger singular agreement, and in general behave like normal singular indefinites, regardless of whether they show normal or 'mismatch' agreement. For example, (21) shows mismatch NPs in several syntactic environments (complement of existential *be*, normal subject, object of a preposition, and coordinated with a normal singular indefinite). This is not at all surprising, of course: one would expect an NP whose quantifier is *one* to be singular and behave in this way, and in general a partitive like *one of those problems* and a singular indefinite like *a problem* are very similar in meaning.

(21) a. There's [one of those letters that always annoys me] in the post.
b. [One of those letters that annoys me] has found its way into the post.
c. I have just torn up [one of those letters that always annoys me].
d. It's either [[a circular] or [one of those letters that always annoys me]].

This might lead one to try to analyse the singular relative clause in the mismatch construction as a modifier of the singular NP *one of those problems*, or the quantity word *one*, rather than the plural noun *problems*, assigning a structure along the lines of (22) or (23) (of course, (23) cannot be a representation of the surface syntax, but the surface syntax could presumably be derived by assuming some form of extraposition).[6]

(22)


(23)


One problem with these analyses can be seen from the interpretation of the pronoun *them* in (24).

[6]Taking the relative clause as a modifier of *one* in this way seems to be the basis of de Hoop et al. (n.d.)'s approach.

(24) This is one of those problems that really bothers me. I wish I could ignore *them*.

The interpretation of *them* in (24) involves the intersection of 'problems' and 'things that bother me' – what the speaker wishes she could ignore is not just a set of problems, it is a set of 'problems that bother her'. To get this reading, the singular relative clause *that really bothers me* must be interpreted as a restrictive modifier of *problems*, exactly like a normal plural relative. It cannot be interpreted as a modifier of *one* or *one of those problems*.[7]. Strong reinforcement of this point can be seen by considering examples involving negative polarity items (NPIs) and superlative adjectives. As is well known, superlative adjectives create contexts which permit NPIs like *ever* in their scope. Witness the contrast in (25): (25a) is grammatical, but without the superlative *most impressive* the NPI *ever* is disallowed, making (25b) ungrammatical.

(25) a. the most impressive goals that have ever been seen in this stadium
b. *the goals that have ever been seen in this stadium

Unsurprisingly, this contrast can be seen in partitives involving normal agreement:

(26) a. one of the most impressive goals that have ever been seen in this stadium
b. *one of the goals that have ever been seen in this stadium

We assume that what is happening here is that the superlative operator scopes over the adjective, the nominal that the adjective modifies, and all other intersective modifiers of that nominal.

The important point here that exactly the same pattern can be seen with 'mismatch' relatives:

(27) a. one of the most impressive goals that has ever been seen in this stadium
b. *one of the goals that has ever been seen in this stadium

In this example, and in attested examples like (4c) and (4d), we see NPIs in relative clauses in the mismatch construction. This is natural if the relative clause is in the semantic scope of the superlative (as it is in the 'normal' cases),

[7]Treating the relative clause as a modifier of *one of those problems* as in (22) amounts to treating it as a non-restrictive relative, which requires *which* in place of *that* (in general, *that* relatives cannot be interpreted as non-restrictive):

(i) This is one of those problems, which really bothers me.

This is interpreted like *This is one of those problems, and it really bothers me*, and as one would expect, it provides '(those) problems', but not 'problems that bother me' as an antecedent for *them*.

e.g. if mismatch examples involve the same basic structure as 'normal' ones. But it is inexplicable if the input to compositional interpretation is a structure like (22), where the relative clause is only in the scope of *one*.

A natural reaction to agreement mismatches in general is to wonder whether it may be possible to exploit the distinction between INDEX and CONCORD agreement: INDEX agreement is 'semantic' – to do with denotation and how discourse entities are individuated and tracked across discourse; CONCORD agreement is simply 'formal' or morphosyntactic (see, e.g. Kathol (1999), Wechsler and Zlatić (2003)).[8] This approach is also not promising. It is not at all clear how this could work in practice, but the general idea might be that combining a relative clause with its head (sometimes) involves CONCORD, where the CONCORD and INDEX values can differ. One problem is that it is normally assumed that for number values (as opposed to gender values), CONCORD and INDEX are identical, so this would be something of an innovation.[9] A second problem would be over-generation: there seem to be no other cases of singular relatives modifying plural nouns – as noted above (see (9)) examples like (28) are normally completely ungrammatical – so one would have to find some way of restricting the domain of application of the relevant principles to just the partitives we are discussing here (e.g. those with *one* as the quantity word).

(28) *I want to talk about [those problems which has been bothering me].

More seriously, this sort of approach would provide at most a partial solution, because, though it might deal with some issues on the syntactic side of things, it does not address the semantic issue at all. It is very hard to avoid the assumption that the antecedent N in the mismatch construction is semantically plural, hence [INDEX | NUM *pl*] (e.g. it clearly denotes a plurality). Moreover, relevant aspects of agreement inside the relative clause involve INDEX values: subject-verb agreement, and pronoun antecedent agreement are generally assumed to involve INDEX values, so we have clear evidence that the relative pronoun in the mismatch construction is [INDEX | NUM *sg*]. What this means is that whatever we do about CONCORD values, we will still be faced with combining a semantically plural predicate (corresponding to the head N) with a seman-

[8] We are grateful to an anonymous referee for mentioning another suggestion, which we think is on the same lines, and subject to the same objections. Müller (1999) introduced a feature REF which is used for reference tracking, and which is at least partially independent of morphosyntactic properties, leaving the INDEX feature free to reflect morphosyntactic properties.

[9] Though perhaps it cannot be ruled out entirely. Wechsler and Zlatić (2003, p50ff) discuss a class of collective nouns in Serbo-Croatian (the *deca*-type) which trigger singular agreement inside NP, but antecede plural pronouns, and Corbett (2006, Ch5) lists a number of other mismatches which might invite this kind of analysis, so this alignment of number values can be only a default generally. However, dissociating INDEX | NUM and CONCORD | NUM would still be a significant innovation for English, where it is otherwise not attested.

tically singular predicate (corresponding to the relative clause), and this will not be straightforward. Standard views on the semantics of plurality (see, e.g. Winter and Scha, 2015) regularly draw a sharp distinction between pluralities and ‘ordinary’ individuals: a plurality is a collection of ordinary individuals. The difference is that whereas, for example, the ultimate component parts of a plurality of people are individual persons, the component parts of an individual person are not. Sometimes this is analysed as a difference of logical type (pluralities are treated as sets, of type $\langle e, t \rangle$, whereas ordinary individuals are of type *e*), or as a difference among sorts of individual of the same type (e.g. pluralities satisfy an ‘atomicity’ requirement which ordinary individuals fail to satisfy). Whatever analysis is chosen, compositional semantics will deliver a singular predicate for *that bothers me*, and a plural predicate for *problems*, and combining them will not be a straightforward matter. An analysis which deals only with the syntax of the mismatch construction does not provide a solution.[10]

For those familiar with the LFG literature, a natural response to feature mismatches is to try to employ the ‘restriction’ operator introduced into LFG by Kaplan and Wedekind (1993). Intuitively, where *P* is a collection of attributes and values, the notation $\boxed{1} \backslash P$ should be read as: the value $\boxed{1}$ with the attributes and values in *P* ‘restricted out’. If $\boxed{1}$ is as in (29a), then $\boxed{1} \backslash [\text{INDEX} \,|\, \text{NUM } sg]$ will be as in (29b).

(29) a. $\begin{bmatrix} \text{INDEX} & \begin{bmatrix} \text{PER} & \mathit{3} \\ \text{GEN} & \mathit{neut} \\ \text{NUM} & \mathit{sg} \end{bmatrix} \end{bmatrix}$ b. $\begin{bmatrix} \text{INDEX} & \begin{bmatrix} \text{PER} & \mathit{3} \\ \text{GEN} & \mathit{neut} \end{bmatrix} \end{bmatrix}$

The leading intuition here would be that the constraint on phrases consisting of a noun and a relative clause given in (6) is too strict in requiring *identity* of index values between head noun and relative clause, what we should require is identity modulo the number value. (6) could be replaced by (30), with the addition highlighted.

[10] This is one reason why we do not consider other purely formal approaches, such as the following. The reason there is a mismatch between INDEX *sg* and INDEX *pl* is that *sg* and *pl* are maximal types. We could thus avoid a mismatch if we introduced a common sub-type (*sg-pl*, say), to which SG $\wedge$ *pl* would resolve. This might provide a solution to the morphosyntactic problem (assuming one could avoid the massive overgeneration that would result in this value being available everywhere), but it would leave the conceptual/semantic issues unresolved.

(30)

$$\begin{array}{c}\text{N}\\ \left[\text{CONT}\ \left[\begin{array}{l}\text{INDEX}\ \boxed{1}\\ \text{RESTR}\ \boxed{2}\cup\{\boxed{4}\}\end{array}\right]\right]\end{array}$$

$$\begin{array}{c}\text{N}\\ \left[\text{CONT}\ \boxed{3}\left[\begin{array}{l}\text{INDEX}\ \boxed{1}\\ \text{RESTR}\ \boxed{2}\end{array}\right]\right]\end{array} \qquad \begin{array}{c}\text{S}_{rel}\\ \left[\begin{array}{l}\text{MOD}\ \left[\text{CONT}\ \boxed{3}\ \backslash\left[\text{INDEX}\,|\,\text{NUM}\ \mathit{sg}\right]\right]\\ \text{CONT}\ \boxed{4}\ \mathit{proposition}\end{array}\right]\end{array}$$

The intention here would be to remove the singular specification in the MOD value of the relative clause: though the internal makeup of a particular relative clause may result in a MOD value which requires an index value like (29a) (i.e. a relative clause that can only modify a third person singular neuter nominal), the specification in (30) would replace (29a) with (29b), and allow a singular relative clause to combine with a nominal regardless of its number, which seems to be what we want in the mismatch construction.

Unfortunately, this is not a solution. Leaving aside the issue of overgeneration (how would we ensure that this method of combining nominals and relative clauses only applies in the construction under discussion?), the problem is that though the restriction in (30) may allow us to ignore the mismatch between the singular specification in the MOD value of the relative and plural on the modified noun, it does not avoid the fact that we will be trying to combine incompatible values, as will be seen elsewhere in the structure. For example, in a singular relative clause like *that* $\Delta_{\boxed{1}}$ *bothers me when I think about it*$_{\boxed{1}}$, the singular index $\boxed{1}$ appears on the subject, in the specification of the properties of the subject of the verb *bothers*, in the semantics of the first argument of *bothers* (i.e. *bothers(*$\boxed{1}$*,me)*), on the co-indexed pronoun *it*, and in the REL value of the relative clause, and in all these places it must be [NUM *sg*]. But the same index also occurs in the content of the head noun *problems*, where it is [NUM *pl*]. The restriction operator in (30) seems to allow us to ignore one feature mismatch, but it does not allow us to ignore all the others. To get the effect we want, we would have to restrict out the value of INDEX | NUM *everywhere* in the relative clause structure, which is of course impossible, because we cannot know where it may occur, ahead of time.[11]

[11] A reader familiar with the use of the restriction operator in LFG may wonder why this is not a problem for its use in LFG. The difference is this: in LFG, the operator is used on f-structures, which are 'models', that is solutions to constraints, and re-entrance in f-structures means that all paths to a particular value are equivalent. Of course, the same is true in the models of HPSG constraints. However, the structures that one sees in presentations of HPSG – such as (30) – do not represent models or solutions to constraints, but the constraints themselves, and so changing

However, one might try to develop a different intuition about (30), along the following lines. While it is true that (30) combines a nominal with a plural index and a relative clause with a singular index in several places, (30) is the only place where a link is made. One could think of the restriction operator as severing this link: then the noun can remain plural, and the relative clause can remain singular. The problem with this is that if we sever the link between the index on the nominal and the indices in the relative then we cannot get the intersective semantics we require. To get the right semantics we need something like the normal plural agreement case, as in (31a), which will produce restrictions like (31b).

(31) a. problems$_{\boxed{1}}$ that $\Delta_{\boxed{1}}$ bother me
b. $problems(X) \wedge bother(X, me)$

If we sever the link between the index on the nominal and the indices in the relative, we will get something like (32a), and restrictions like (32b), where y is an unbound variable.

(32) a. problems$_{\boxed{1}}$ that $\Delta_{\boxed{2}}$ bothers me
b. $problems(X) \wedge bothers(y, me)$

It is not immediately obvious how one should interpret unbound variables, but one idea is that they are existentially bound by default. This would give something like (33b) as the interpretation of (33a). This will be true if there is some unique plurality X of problems, and Kim solved X, and some entity y (potentially unrelated to X) that bothers me. This does not capture the meaning of (33a).[12]

(33) a. Kim has solved one of the problems that bothers me.
b. $\exists X.unique(X) \wedge problem(X) \wedge \exists y.bothers(y, me) \wedge solved(Kim, X)$

To summarise: we have in the mismatch construction a genuine agreement mismatch, involving a conflict of INDEX values, in which a normal, syntactically singular, relative clause must be interpreted as a restrictive modifier of a plural nominal. We have looked at a number of possible approaches, all of which more or less implausible or inadequate. In the following section we will remedy this.

one path to a value does not necessarily change all. It is easy to be misled by the fact that both LFG and HPSG use the same structures (attribute-value matrices) in such different ways.

[12] Alternative ways of dealing with unbound variables include assigning them universal force (which would wrongly make (33a) entail that everything bothers me), or treating them as a indicators of ungrammaticality (which would wrongly make (33a) ungrammatical).

3 An HPSG Analysis

Though we saw above that the problem of the mismatch construction cannot be solved by having the relative clause modify *one*, it is clear that *one* plays a crucial role in the construction, since it is not a general property of quantity words that they allow the mismatch construction:[13]

(34) *Two/*Some/*Many*/*All/*None of the problems that bothers you have been solved.

The obvious (and we think correct) way to capture this is to treat the relative clause as a dependent (specifically, an optional complement) of *one*. In what follows we will develop this proposal.

The following is a plausible starting point for normal partitive *one* (as it appears in normal partitives e.g. *one of the problems*).

(35)
$$\begin{bmatrix} & PP_{of} \\ \text{ARG-ST} & \left\langle \begin{bmatrix} \text{LOC}\,|\,\text{CONT} & \begin{bmatrix} \textit{npro} \\ \text{INDEX} \ \boxed{X} \\ \text{RESTR} \ \boxed{R} \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{SS}\,|\,\text{LOC} & \begin{bmatrix} \text{CONT} & \begin{bmatrix} \textit{one-part-rel} \\ \text{INDEX} \ \boxed{x}\begin{bmatrix}\text{NUM} & \textit{sg}\end{bmatrix} \\ \text{RESTR} \ \left\{\boxed{x} \in \boxed{X}\right\} \cup \boxed{R} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

According to this, partitive *one* takes a PP *of* complement, which has a plural index $\boxed{X}$ and some restrictions $\boxed{R}$ (for clarity and readability we will from now on indicate singular indices with lowercase variables like $\boxed{x}$, and plural ones with uppercase variables like $\boxed{X}$). It takes this index and these restrictions and produces content whose relation is the '*partitive-one-rel*', and whose index is singular $\boxed{x}$, which is a part of the plurality. For example, given a PP like *of those problems*, whose content is as in (36), the content of *one of those problems* will be as in (37).[14]

---

[13] In fact, not all forms of *one* license the construction, in particular determiner *one* (i.e. *one* that appears with an explicit noun) does not license it – examples like the following require a plural relative:

(i) *one token/example/instance of the problems that bother/*bothers me

In what follows, all lexical entries for one should be understood as relating to nominal *one* (i.e. specified as SYNSEM | LOCAL | CAT | HEAD *noun*).

[14] Since *of* is a non-predicative preposition, its content is the same as that of its NP complement, i.e. a *nominal-object*, specifically an *npro* (non-pronominal).

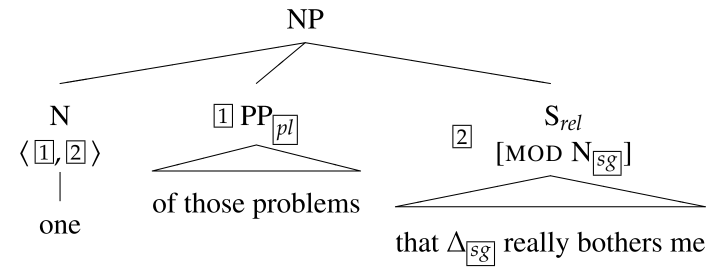
(36) $\begin{bmatrix} \textit{npro} & \\ \text{INDEX} & \boxed{X} \\ \text{RESTR} & \{\, \textit{those\_problems}(\boxed{X}) \,\} \end{bmatrix}$

(37) $\begin{bmatrix} \textit{partitive-one-rel} & \\ \text{INDEX} & \boxed{x} \\ \text{RESTR} & \{\, \boxed{x} \in \boxed{X}, \textit{those\_problems}(\boxed{X}) \,\} \end{bmatrix}$

Treating the relative clause as a complement of *one* involves providing it with an additional lexical entry, for which (38) is a first approximation, and where the additions with respect to (35) have been highlighted ((35) and (38) can easily be collapsed into a single entry). This will give structures like (39).

(38) $$\begin{bmatrix} \text{ARG-ST} & \left\langle \overset{\textit{PP}_{\textit{of}}}{\begin{bmatrix} \text{LOC} \mid \text{CONT} \begin{bmatrix} \textit{npro} \\ \text{INDEX } \boxed{X} \\ \text{RESTR } \boxed{R} \end{bmatrix} \end{bmatrix}}, \overset{\textit{Relc}}{\begin{bmatrix} \text{REL } \{\boxed{y}\} \\ \text{LOC } \begin{bmatrix} \text{CONT } \boxed{p} \end{bmatrix} \end{bmatrix}} \right\rangle \\ \text{SS} \mid \text{LOC} & \begin{bmatrix} \text{CONT} \begin{bmatrix} \textit{partitive-one-rel} \\ \text{INDEX } \boxed{x} \begin{bmatrix} \text{NUM } \textit{sg} \end{bmatrix} \\ \text{RESTR } \left\{ \boxed{x} \in \boxed{X} \right\} \cup \boxed{R} \cup \boxed{R'} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(39)



The relative clause specified in (38) has a REL value containing the (singular) index $\boxed{y}$ with a proposition $\boxed{p}$ as its content. Since in relative clauses like this the index that appears in the REL value is the index of the relativised NP, for an example like *that* Δ *bothers me*, this proposition will be something like *bothers*($\boxed{y}$, *me*).

What (38) does not address is how the restrictions of the relative clause should be added to those of the PP (i.e. we have not specified the relationship between $\boxed{R'}$ and the content of the relative clause). Notice that we cannot identify $\boxed{y}$ with either of the other indices in (38): we cannot identify it with $\boxed{X}$ because $\boxed{y}$ is singular, whereas $\boxed{X}$ is plural; and if we identify it with $\boxed{x}$ we will get the wrong reading, where it is only a single problem that bothers the speaker (cf. the ‘one (which bothers me) of those problems’ reading that we would get from representations like (22) and (23), above). However, we can capture the intended reading if we ‘distribute’ the interpretation of the relative clause across the parts of the plurality $\boxed{X}$. This can be implemented if $\boxed{R'}$ in (38) is specified as in (40), giving Figure 1 as the full entry.

(40)
$$\boxed{R'} = \left\{ \begin{bmatrix} \text{QUANTS} & \begin{bmatrix} \textit{foreach} \\ \text{INDEX} \ \boxed{y} \\ \text{RESTR} \ \left\{ \textit{partof}(\boxed{y},\boxed{X}),\ \textit{atomic}(\boxed{y}) \right\} \end{bmatrix} \\ \text{NUCLEUS} & \boxed{p} \end{bmatrix} \right\}$$

Intuitively, the restriction this adds is that every atomic part of the plurality X satisfies p, which is the content of the relative clause. In the case of *one of those problems that bothers me*, this additional restriction is that for every y which is an atomic part of the plurality of problems X – that is, every individual problem – y bothers the speaker. This is spelled out in (41a), which we can abbreviate as (41b), giving (42) as the overall content. In words, this amounts to ‘one of those problems each of which bothers me’.

(41) a.
$$\left\{ \begin{bmatrix} \text{QUANTS} & \begin{bmatrix} \textit{foreach} \\ \text{INDEX} \ \boxed{y} \\ \text{RESTR} \ \left\{ \textit{partof}(\boxed{y},\boxed{X}),\ \textit{atomic}(\boxed{y}) \right\} \end{bmatrix} \\ \text{NUCLEUS} & \textit{bother}(\boxed{y},\textit{me}) \end{bmatrix} \right\}$$

b. $\{\ \textit{foreach}(\boxed{y},\ \boxed{y} \in \boxed{X},\ \textit{bother}(\boxed{y},\textit{me}))\ \}$

(42)
$$\begin{bmatrix} \textit{partitive-one-rel} \\ \text{INDEX} \ \boxed{x} \\ \text{RESTR} \ \{\ \boxed{x} \in \boxed{X},\ \textit{those\_problems}(\boxed{X}), \textit{foreach}(\boxed{y},\ \boxed{y} \in \boxed{X},\ \textit{bother}(\boxed{y},\ \textit{me}))\ \} \end{bmatrix}$$

This is the right interpretation, and it correctly reflects the idea that the relative clause is interpreted as a singular throughout, in the sense that the relative pronoun and everything that is co-indexed with it is singular, but is nevertheless interpreted as describing a property of every element of the plurality introduced by the head noun. Notice in particular that it correctly predicts that the antecedent of *them* in (24), repeated here, is a collection of problems that bother the speaker (because it is a collection of problems *each of which* bother the speaker). In this way, it avoids the problems associated with the sort of analysis represented in (22) and (23).[15]

[15] In the previous section, we noted cases where a mismatch relative clause contains a negative

$$\left[\begin{array}{ll}
\text{ARG-ST} & \left\langle \overset{PP_{of}}{\left[\text{LOC}\,|\,\text{CONT}\begin{bmatrix}\textit{npro}\\ \text{INDEX}\ \boxed{X}\\ \text{RESTR}\ \boxed{R}\end{bmatrix}\right]},\ \overset{Relc}{\begin{bmatrix}\text{REL}\ \left\{\boxed{y}\right\}\\ \text{LOC}\ \left[\text{CONT}\ \boxed{p}\right]\end{bmatrix}} \right\rangle \\
\text{SS}\,|\,\text{LOC} & \left[\text{CONT}\begin{bmatrix}\textit{partitive-one-rel}\\ \text{INDEX}\ \boxed{x}\left[\text{NUM}\ sg\right]\\ \text{RESTR}\ \left\{\boxed{x}\in\boxed{X}\right\}\cup\boxed{R}\cup\left\{\begin{bmatrix}\text{QUANTS} & \begin{bmatrix}\textit{foreach}\\ \text{INDEX}\ \boxed{y}\\ \text{RESTR}\ \left\{partof(\boxed{y},\boxed{X}),\ atomic(\boxed{y})\right\}\end{bmatrix}\\ \text{NUCLEUS} & \boxed{p}\end{bmatrix}\right\}\end{bmatrix}\right]
\end{array}\right]$$

Figure 1: Lexical Entry for *one* in the Mismatch Construction

(43) This is one of those problems that really bothers me. I wish I could ignore *them*.

In this section we have proposed a rather straightforward solution to the problems posed by the mismatch construction, which captures the key distributional facts (dependence on partitive *one*), and gets the semantics right, by showing how a singular relative clause can be interpreted as holding of a plurality. There remain, however, a number of complexities and some open questions which we have not addressed. These are outlined in the final section.

4 Open Questions

One obvious objection to the analysis we have presented is that it involves treating a relative clause as a complement (of *one*), whereas relative clauses are normally adjuncts. It is not clear to us how important this objection is. First, cases where heads select as complements expressions that are normally adjuncts are well-known. For example, the verbs *word* and *treat* both select adverbs, as can be seen in (44a) and (44b), and (44c) suggests that adverbs are a necessary part of resultative constructions:[16]

polarity item. We assume that adding the restrictions given in (40) to the restrictions of the PP (and hence the lower NP, and the Nominal it contains) is sufficient to put them in the scope of the superlative, and thus account for the possibility of negative polarity items. This is different from (and much simpler than) the account that we gave in the abstract for the conference, and at the conference itself, which involved the use of apparatus for extraposition. We no longer think this is necessary.

[16]Of course, the analysis of at least some adjuncts as complements has been a well-established in the HPSG literature since Bouma et al. (2001). As regards relative clauses specifically, it seems to us arguable that relative clauses associated with superlative adjectives are complements of the

(44) a. I worded the letter *(carefully).
b. The management has treated Sandy *(contemptuously).
c. This book reads *(easily).

The following is a more subtle objection. Our analysis treats the relative clause in the mismatch construction as a complement of *one*, and intuitively, one would expect this to explain why the relative clause is singular. But our analysis does not really capture this (or captures it only rather indirectly). According to our analysis, the reason the relative clause is singular is because the index $\boxed{y}$ is singular, and the reason this is singular is because of the atomicity condition (*atomic*($\boxed{y}$)) – the condition that it be anchored to an atomic part of the plurality denoted by the partitive noun (it is reasonable to assume that atomic parts are always associated with singular variables). Notice, in particular, that there is no real 'agreement' between *one* and the relative clause, e.g. $\boxed{x}$, the index of *one*, figures nowhere in the part highlighted part of Figure 1, which is the part of the lexical entry for *one* that is associated with the relative clause. Again, it is not clear to us how compelling this objection is.

We have described the mismatch construction as crucially depending on the presence of partitive *one*, and given a lexical account. But this is a simplification. First, a number of expressions that include partitive *one* seem to permit the construction.

(45) a. At least one of the problems that bothers me has been solved.
b. More than one of the problems that bothers me has been solved.
c. Every one of the problems that bothers me has been solved.
d. Not one of the problems that bothers me has been solved.
e. Not a single one of the problems that bothers me has been solved.

Moreover, it seems that some other words allow it, at least for some speakers (the following are from the BNC):

(46) a. *another* of those volcanoes which was thought to be extinct until something nasty happened. [ASR/837]
b. *another* of those chores which is easier to carry out during post-production editing... [CBP/901]
c. an individual programme fitted for *each* of those who is going on. [ASY/1463]

It is not obvious how to extend our basic account to deal with these data – it would of course be straightforward to deal with the examples in (46) by providing a lexical entry for *another* parallel to that for *one*, but the examples in (45) seem more challenging. Notice in particular that the impossibility of (47a)

superlative, e.g. in cases like *This is the best [(that) we could hope for at present].*

and (47b) indicates that it is not simply a matter of the mismatch construction being allowed whenever the quantity expression contains *one*. All of this raises the possibility that there may be something more general and more interesting going on.

(47) a. *Less than one of the problems that bothers me has been solved.
  b. *Fewer than one of the problems that bothers me has been solved.

A further, and in some ways more worrying, issue with our analysis is the following. The intuition expressed by our analysis is that the quantity word *one* allows a singular adjunct to be interpreted distributively over the elements of the plural in the partitive. This obviously suggests that the construction should be impossible with relative clauses involving non-distributive predicates – that is, predicates that cannot be applied to ordinary singular individuals, but only to pluralities. There is something in this, as the following data suggest. The predicate *sleep in separate beds* cannot be predicated of singular individuals, as (48a) shows, and as (48b) suggests, it is problematic in a singular relative clause in the mismatch construction.

(48) a. *He sleeps in separate beds. (vs. They sleep in separate beds.)
  b. ???one of those people who sleeps in separate beds

The problem is that, while (48b) seems to us to be bad, it is not nearly as bad as it should be (in particular, not as bad as (48a), which is awful). Other examples are even better: the present authors find (49b) and (50b) acceptable, though they involve predicates which should not permit singulars. There is clearly more to investigate here.

(49) a. *She is numerous. (vs. They are numerous.)
  b. one of those crazy people who is so numerous on demonstrations these days

(50) a. *He meets every week. (vs. They meet every week.)
  b. one of the people who meets every week to discuss semantics

References

Bouma, Gosse, Malouf, Rob and Sag, Ivan A. 2001. Satisfying Constraints on Extraction and Adjunction. *Natural Language and Linguistic Theory* 19(1), 1–65.

Burchfield, R. W. 2004. *Fowler's Modern English Usage*. Oxford: Oxford University Press, revised 3rd edition edition.

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

de Hoop, Helen, Hogeweg, Lotte and Ramacher, Stefanie. n.d. Singular Agreement in ‘Special’ Partitive Constructions, unpublished ms.

Flickinger, Dan. 2008. Transparent Heads. In Stefan Müller (ed.), *The Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar*, pages 87–94, Stanford, CA: CSLI Publications.

Huddleston, Rodney and Pullum, Geoffrey K (eds.). 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Kaplan, Ronald M. and Wedekind, Jürgen. 1993. Restriction and correspondence-based translation. In *Proceedings of the 6th Conference of the Association for Computational Linguistics European Chapter*, pages 193–202, Utrecht University.

Kathol, Andreas. 1999. Agreement and the Syntax-Morphology Interface in HPSG. In Robert D. Levine and Georgia Green (eds.), *Studies in Contemporary Phrase Structure Grammar*, pages 209–260, Cambridge: Cambridge University Press.

Kim, Jong-Bok. 2002. On the structure of English partitive NPs and agreement. *Studies in Generative Grammar* 12, 309–338.

Kim, Jong-Bok and Sells, Peter. 2008. *English Syntax: An Introduction*. Stanford, CA: CSLI Publications.

Maekawa, Takafumi. 2015. ‘Agreement mismatch’ between sort/kind/type and the determiner. In Stefan Müller (ed.), *Proceedings of the 22nd International Conference on Head-Driven Phrase Structure Grammar, Nanyang Technological University (NTU), Singapore*, pages 136–156, Stanford, CA: CSLI Publications.

Müller, Stefan. 1999. *Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche*. Linguistische Arbeiten, No. 394, Tübingen: Max Niemeyer Verlag.

Pinker, Steven. 2014. *The Sense of Style: the Thinking Person’s Guide to Writing in the 21st Century*. London: Allen Lane.

Pollard, Carl J. and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Sag, Ivan A. 1997. English Relative Clause Constructions. *Journal of Linguistics* 33(2), 431–484.

Wechsler, Stephen and Zlatić, Larisa. 2003. *The Many Faces of Agreement*. Stanford, CA: CSLI Publications.

Winter, Yoad and Scha, Remko. 2015. Plurals. In Shalom Lappin and Chris Fox (eds.), *The Handbook of Contemporary Semantic Theory*, pages 77–113, Oxford: Blackwell, second edition.

Switched control and other “uncontrolled” cases of obligatory control

Dorothee Beermann

Norwegian University of Science and Technology, Trondheim, Norway

Lars Hellan

Norwegian University of Science and Technology, Trondheim, Norway

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 64–82

Keywords: deontic modal, deontic infinitives, suasive verbs

Beermann, Dorothee & Lars Hellan. 2016. Switched control and other “uncontrolled” cases of obligatory control. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 64–82. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.4.

Abstract

The paper presents an analysis of *control switch* in German and Norwegian, as exemplified in the German pair *Ich verspreche ihm zu kommen* 'I promise him to come' vs. *Ich verspreche ihm kommen zu dürfen* 'I promise him to be allowed to come'. The phenomenon is induced by deontic modals in the context of suasive verbs of communication. The analysis is cast both in LFG and HPSG framework, in both cases deploying a pronounced feature-based semantic component. Our core assumption is that a *normative agent* is computed on top of control relations.

1 Introduction[1]

1.1 Background and objective

Both LFG and HPSG assume that obligatory control is lexically specified. While the controller is realised syntactically, the controllee corresponds to the unexpressed subject of the complement clause. In LFG control is seen either as functional control, that is the sharing of an f-structure (Bresnan 1982), or as obligatory anaphoric control, that is as a semantic relation only (Dalrymple 2001). In HPSG control resides in sharing of indices, but also here the scope of these indices can range from referential pointers (not unlike what is found in LFG for obligatory anaphoric control) to feature structures. In the latter case both theories assume control as unification.

In this work we will discuss some of the relevant mechanisms of the two frameworks in connection to the analysis of a regular pattern of ‘control switch’ in German and Norwegian. The pattern arises with verbs of communication which express wishes, desires, commitments or judgements, such as:

German:
anflehen, überreden,versprechen, bitten, beschuldigen
Norwegian:
bønnfalle, overtale, love, be, anklage/beskylde
(English, respectively: ‘beseech’, ‘persuade’, ‘promise’, ‘beg’, ‘accuse')

We will call verbs in this group ***suasive verbs of communication*** [2](Mair 1990. Quirk et al. 1985). When suasive verbs select a modal infinitival complement

[1] We are grateful for comments from the participants at the HeadLex 2016 conference, and for insightful advice from two anonymous reviewers.

with the modal verb *dürfen* or deontic *können* (German), or *få* in its modal use (Norwegian), a complex verbal chain is formed, and an apparent switch of control can be triggered.

Switch control has not only been noticed in configurational studies of CONTROL (e.g., Ruzicka 1999), but has also been discussed from a pragmatic/conceptual perspective combined with corpus studies (Mair 1990), or as a feature-based approach using experimental techniques (Köpcke and Panther 1991). Our approach is restricted to what one might call 'deontic switch' following Rusicka (op. cit.). Our aim is to show how aspects of the *dürfen*-related phenomena follow from specific semantic factors of these constructions. The analytic designs of LFG and HPSG allow us to model the effect of these factors on top of the general mechanisms underlying control.

1.2 Switch from object control to subject control in German and Norwegian

In German, the transitive verb *anflehen* generally induces object control, cf. (1a):

(1a) Er fleht mich an zu kommen
He beseech.PRES me PRTCL to come.INF
'He beseeches me to come'

In combination with the modal verb *dürfen* and deontic *können,* object control switches to subject control:

(1b) Er fleht mich an kommen zu dürfen
He beseech.PRES me PRTCL come.INF to may.INF
'He beseeches me to be allowed to come'

For Norwegian, a similar pattern is observed. In (2a) the logical subject of *komme* is *meg* 'me'. In (2b), when combined with *få* in its modal use as part of the infinitive, object control changes to subject control:

(2a) Han ba meg om å komme
He ask.PST me PREP to come.INF
'He asked me to come'

[2] Suasive verbs expressing orders such as German *befehlen,* Norwegian *befale, beordre* ('order'), do not group with the other suasive verbs relative to the phenomenon discussed here.

(2b) Han ba meg om å få komme
He ask.PST me PREP to get.INF come.INF
'He asked me to be allowed to come'

Få also has aspectual uses, as described, e.g., in Lødrup (1996); here we are focusing on its 'deontic' use.

1.3 Switch from subject control to object control in German and Norwegian

The German verb *versprechen* is a subject control verb, but in combination with *dürfen* and deontic *können* the construction receives an object control interpretation:

(3a) Ich verspreche ihm zu kommen
I promise.PRES him to come.INF
'I promise him to come.'

(3b) Ich verspreche ihm kommen zu dürfen
I promise.PRES him come.INF to may.INF
'I promise him to be allowed to come.'

In Norwegian, the verb *love* 'promise' shows a similar pattern: In (4a) the logical subject of *komme* is *'jeg'*, in (4b) it is *'han'*:

(4a) Jeg lovet ham å komme '
I promise.PST him to come.INF
I promised him to come'

(4b) Jeg lovet ham å få komme
I promise.PRES him to get.INF come.INF
'I promised him to be allowed to come'

In contrast, *wollen* as well as its Norwegian counterpart *ville*, which have a volitional modal base, do not affect lexically determined control relations.

1.4 The phenomenon in English

For English, Radford (1985:381) discusses an example with an object-control verb which receives a subject-control interpretation. While *John pleaded with* *me to go* means that I should go, *John pleaded with me to be allowed to go* states that I should allow John to go. Also in the case of an unlikely interpretation, a default object-control pattern may be overridden by a subject-control interpretation such as in the case of *He asked his boss to have an afternoon off* (Mair 1990). These cases of switched control seem marginal for English, but not so for the cases we discuss in German and Norwegian, where modal verbs are used widely and systematically in embedded infinitives (see also Stiebels (2015)).[3]

2 'Control switch' – diagnosis

Control switch constructions are composed of two verbal predicates: the modal non-finite predicate and the matrix predicate. We would like to treat modals with a deontic modal base such as *få, dürfen* and *können* as three-place relations with a normative agent as first argument, an addressee who gets a permission as second argument, and an action as third argument.

Suasive verbs describe communications about what we may call ***negotiable situations***. Thematically the situation is instantiated by a promiser/persuader and an addressee, and the lexicalised control pattern encodes whether the promiser/persuader subject or the addressee object is under negotiaton as the agent of the situation under discussion. *Promise* type verbs feature the promiser as this prospective agent, while for the *beseech* type the adressee is construed as this agent. In a deontic context a normative agent is introduced under whose regime the prospective agent of the embedded infinitive will have to act. Under obligatory control the normative agent is always bound to one of the expressed arguments, and in this way the switched control pattern described here is borne. What thus makes suasive verbs in construction with deontic modals unique is that they, with obligatory control, provide a formal device to express, next to understood logical subjects, also the understood normative agent. This is interesting in itself since in constructions headed by modal verbs, the normative agent normally remains unexpressed. In this respect switch control structures are the exception rather than the rule. What we try to show here is how constraint-based linguistic formalisms can be adapted to give a theoretically grounded representation for which of the realised arguments in an obligatory control construction needs to be interpreted as the normative agent, and which one as the logical subject of the embedded event.

Illustrating the configuration, the suasive verbs of the type *promise* have three semantic arguments (x y P) overtly realised, so that the sentence *He promises*

[3] Hypothetically speaking, an English counterpart could have been like:
**He promised me to may go*.

him to come has roughly the semantic structure in (5), with x as the promiser, y as the addressee, and P as the action to be conducted by x:

(5) PROMISE(x y P(x)).

In a sentence like (3b), repeated,

(3b) Ich verspreche ihm kommen zu dürfen
I promise.PRES him come.INF to may.INF
'I promise him to be allowed to come.'

the normative agent introduced by d*ürfen* is bound to the first argument of *versprechen,* the promiser subject instantiated in (5) as x. Thus we get the semantic pattern in (6b) for the switched pattern, as opposed to the 'normal' pattern in (6a). In both schemata identical letters indicate referential identity, underlined letters indicate the bearer of the deontic control relation.

(6) a. PROMISE [***x*** y [***x*** come]] *subject control*
b. PROMISE (<u>x</u> ***y*** [PERMIT (<u>x</u> ***y*** [***y*** come]] *object control*

The opposite pattern obtains for the *beseech* type of verbs (e.g., (1) and (2)):

(7) a. BESEECH [x ***y*** [***y*** come]] *object-control*
b. BESEECH (***x*** <u>y</u> [PERMIT (<u>y</u> ***x*** [***x*** come]] *subject-control*

From a formal linguistic point of view, a question is how we can construct a semantics which allows us to express the patterns discussed. Our representation of a deontic agent in (6b) and (7b) as a semantic argument on a par with standard arguments leaves some room for consideration, partly because the 'visibility' of this argument is far less apparent than for other types of implicit arguments. Thus, even when the deontic verb appears as matrix finite verb, the overt subject is semantically that of the addressee, whereas the deontic agent is still only understood; cf. (8):[4]

(8) a. Er darf kommen (German)
He may.PRES come.INF
'He may come'

[4] Following standard practice, 'evidence' for an implicit agent in the case of passives is seen as residing in the existence of an active counterpart, where the overt subject instantiates the agent.

b. Han får komme (Norwegian)
He get.PRES come.INF
'He gets the possibility to come'

The only 'visibility' relative to *få* resides in the circumstance that in Norwegian, this implicit agent can be accessed through an adjunct as in (9a), with much the same form and meaning as the adjunct for demoted agents in passive (9b), and for understood causers (9c):

(9) a. Hun får gå på fotballkamper av søstrene sine
She get.PRES go.INF on football-matches by sister.PL her.PL
'She is-permitted-to watch football matches by her sisters'

b. Hun ble kritisert av søstrene sine.
She be.PST criticize.PASS by sister.PL her.PL
'She was criticized by her sisters'

c. Treet blåste ned av vinden
tree.DEF blow.PST down by wind.DEF
'The tree blew down by the wind'

The German counterparts of *få* - *dürfen* and *können* - do not offer similar adjunct possibilities.[5]

The visibility signs just mentioned, and the circumstance that deontic modals do indeed induce a control switch, in our view warrants representing the modal agent as playing a part in the semantic argument structures as indicated. However, distinct from the standard array of arguments representing situational participants, this modal agent represents a social index already exposed by a suasive verb.

A point to be noted is that the patterns now described arise only when both of the arguments of the suasive verb are overtly expressed, as in all of our examples so far. We discuss this point in subsection 5.2 below.

A remark is in order also concerning the assumed addressee/permissee argument in (6b) and (7b). In the constructions at hand with a suasive matrix verb, the relation between the deontic agent and the controlled agent is always direct. This is also reflected in our analyses where we assume for HPSG and LFG that an embedded PERMIT relation is a 3-place relation between a normative agent, a permissee and a situation. However, using the predicate

[5] *Få* can also be used as a plain transitive verb, like in (i), and with a similar 'donor' agent explicitly stated as in (ii), presumably counting as implicit also in (i):
(i) Han fikk boken *'He received the book'*
(ii) Han fikk boken av presten *'He received the book by the priest'*
German modal verbs are different in this respect.

name ‘PERMIT’ does not quite capture the relation intended, namely (quoting from above) ‘*normative agent under whose regime the prospective agent of the embedded infinitive will have to act*’: the relation could well be less direct than permission in the normal sense.[6] A predicate name like ‘FACILITATOR’ might have been better, but we leave it at the shorter name.

3 HPSG-style representation

Using an HPSG format,[7] the switch from subject to object control illustrated in (3b) and (4b) can be represented as in Figure 1, with coindexation for referential identity. The semantics corresponding to the schematic display in (6b) is found under SEM (with ARG0 representing a situational index; for expository convenience we use English predicate names in the semantics).

$$\begin{bmatrix} \text{ARG-ST} & \left\langle \text{NP}\left[\text{INDX } \boxed{1}\right], \text{NP[INDX } \boxed{2}\text{]}, \text{VP}\begin{bmatrix}\text{SPR} & \left\langle \text{NP}\left[\text{INDX } \boxed{2}\right]\right\rangle \\ \text{INDX} & \boxed{3}\end{bmatrix}\right\rangle \\ \text{SEM} & \begin{bmatrix} \text{PRED} & \text{promise-rel} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \\ \text{ARG3} & \begin{bmatrix} \text{PRED} & \text{permit-rel} \\ \text{ARG0} & \boxed{3} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \\ \text{ARG3} & \begin{bmatrix}\text{PRED} & \text{come-rel} \\ \text{ARG1} & \boxed{2}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

Figure 1 HPSG representation of (3b) *Ich verspreche ihm kommen zu dürfen*

[6] Like in a possible case described by a reviewer: “*ich versprach ihm zu Peters Party kommen zu dürfen.* may involve a permitter distinct from me (I may have some influence on Peter or some organiser to be able to make that commitment).”

[7] We stay essentially within the frame of works such as Pollard and Sag (1994), Sag and Wasow (1999), Sag et al. (2003), and the ‘Matrix’ architecture underlying some of the HPSG computational grammars (cf. Bender et al. 2010). Thus, the ‘ARG’ attributes are as in Pollard and Sag (op.cit.) and Bender et al. (op cit.), while the leanness of the feature structure approximates that of the other two; some further simplifications are made for the purpose of exposition. An algorithmically tractable implementation of the analysis can be attained using the ‘Matrix’ architecture as indicated in footnote 8, and most likely other architectures as well.

The contribution of *dürfen/få* per se is indicated in Figure 2 (the referential index of ARG1, indicated by the boxed reentrancy number ‘1’, deliberately left free):

$$\begin{bmatrix} \text{ARG-ST} & \left\langle \text{NP}\begin{bmatrix}\text{INDX } \boxed{2}\end{bmatrix}, \text{VP}\begin{bmatrix}\text{SPR } \left\langle \text{NP}\begin{bmatrix}\text{INDX } \boxed{2}\end{bmatrix}\right\rangle \\ \text{INDX } \boxed{3}\end{bmatrix}\right\rangle \\ \text{SEM} & \begin{bmatrix} \text{PRED permit-rel} \\ \text{ARG1 } \boxed{1} \\ \text{ARG2 } \boxed{2} \\ \text{ARG3 } \begin{bmatrix}\text{ARG0 } \boxed{3}\\ \text{ARG1 } \boxed{2}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

Figure 2 Representation of *dürfen/få* as in (8a) *Er darf kommen*

In *Er darf kommen* the subject of *kommen* will bind the index indicated with the boxed number ‘2’in Figure 2, whereas the index indicated with the boxed number ‘1’ remains un-instantiated syntactically. It gets instantiated only when *versprechen* combines with *kommen zu dürfen*, imposing its subject control pattern, resulting in the constellation shown in Figure 1. Notably, the ARG1 of *dürfen* which now gets bound is not the index associated with the *subject* of *dürfen*, but the index of the *permitter*.

Thus, what here has to act as the lexical specification of *versprechen* is the structure in Figure 3, requiring identity between the two ARG1’s (on the paths ‘SEM|ARG1’ and ‘SEM|ARG3|ARG1’), and between the ARG2 of *promise* and the referent of the syntactic subject of the VP (indicated by the boxed number ‘2’):

$$\begin{bmatrix} \text{ARG-ST} & \left\langle \text{NP}\begin{bmatrix}\text{INDX } \boxed{1}\end{bmatrix}, \text{NP[INDX } \boxed{2}\text{]}, \text{VP}\begin{bmatrix}\text{SPR } \left\langle \text{NP}\begin{bmatrix}\text{INDX } \boxed{2}\end{bmatrix}\right\rangle \\ \text{INDX } \boxed{3}\end{bmatrix}\right\rangle \\ \text{SEM} & \begin{bmatrix} \text{PRED promise-rel} \\ \text{ARG1 } \boxed{1} \\ \text{ARG2 } \boxed{2} \\ \text{ARG3 } \begin{bmatrix}\text{ARG0 } \boxed{3}\\ \text{ARG1 } \boxed{1}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

Figure 3 Representation of *versprechen* targeted for its use in control switch

This contrasts with the specification of the downstairs syntax that the lexical specification of *versprechen /love/promise* would normally be assumed to carry, exhibited in Figure 4, which in this case would wrongly equate the one who gets permission with the one who promises:

$$\begin{bmatrix} \text{ARG-ST} & \left\langle \text{NP}\begin{bmatrix}\text{INDX } \boxed{1}\end{bmatrix}, \text{NP[INDX } \boxed{2}\text{]}, \text{VP}\begin{bmatrix}\text{SPR } \left\langle \text{NP}\begin{bmatrix}\text{INDX } \boxed{1}\end{bmatrix}\right\rangle \\ \text{INDX } \boxed{3}\end{bmatrix} \right\rangle \\ \text{SEM} & \begin{bmatrix} \text{PRED promise-rel} & \\ \text{ARG1 } \boxed{1} & \\ \text{ARG2 } \boxed{2} & \\ \text{ARG3} & \begin{bmatrix}\text{ARG0 } \boxed{3} \\ \text{ARG1 } \boxed{1}\end{bmatrix}\end{bmatrix} \end{bmatrix}$$

Figure 4 Representation of *versprechen/love* for 'standard' cases

How would the grammar ensure that the right version of *versprechen /love/promise* is used in each case? Nothing in the syntax or semantics would be formally ill-formed in case the wrong combinations are made (control switch using the lexical specification in Figure 4, or standard case using the lexical specification in Figure 3); hence it will seem that the lexical specification in Figure 3 must be expanded with some reference down into the VP ensuring that it is headed by a *suasive* verb, and that the lexical specification in Figure 4 must be expanded with reference down into the VP ensuring that it is *not* headed by a suasive verb. While the elegance of such a scenario can be left partly up to its implementation in an explicit grammar, it at least is an asset to the analysis that the representation of *dürfen/få* as such is kept constant across cases with control switch and cases where they occur as a matrix verb by themselves, as in (8) above.[8] Thus, although control switch involves a tight dependency between the

[8] In the online HPSG grammar *NorSource* of Norwegian (http://regdili.hf.ntnu.no:8081/linguisticAce/parse), the syntactic and MRS-semantic analyses of the following sentences display the approach here argued for, in the 'Matrix' type formalism (cf. Bender et al. 2010). It may be noticed that the same item ***få*** occurs in all analyses, whereas special 'switch' lexical items are used for both matrix verbs.

Jeg får komme 'I get-to come', showing abstract 'permitter'
Hun lovet meg å komme 'she promised me to come', *normal subject control*
Hun lovet meg å få komme 'she promised me to get-to come', *switched control*
Hun ber meg om å komme 'she asks me to come', *normal object control*
Hun ber meg om å få komme 'she asks me to get-to come', *switched control*

In this implementation crucial use is made of the attribute 'XARG', which can be set distinct from a verb's ARG1 (which corresponds to its 'logical subject'), so as to, for instance, represent a verb's subject in cases where the verb's syntactic subject is distinct from its 'logical' subject. The grammar can be found at https://github.com/Regdili-matrix verb and *dürfen/få*, the analysis maintains the compositionality of the construction, in that each part has its specific semantic contribution to the overall reading.

4 Formal analysis in LFG-style representation

In LFG obligatory control is captured by means of lexically induced functional control equations. *Versprechen* has, next to the meaning we are interested in here, an epistemic reading with an upstairs non-thematic-subject - the corresponding f-structure is Figure 5a:

$$\begin{bmatrix} PRED & 'versprechen\ < XCOMP > \ SUBJ' \\ & \\ SUBJ & \boxed{1}\begin{bmatrix} PRED & 'pro' \\ NUM & SG \\ PERS & 3 \end{bmatrix} \\ XCOMP & \begin{bmatrix} PRED\ 'regnen < pro > & \\ SUBJ\ \boxed{1} & \end{bmatrix} \end{bmatrix}$$

Figure 5a Functional control: Representation of *'Es verspricht zu regnen* [9]

We are here interested in the *equi construction* for which we assume obligatory anaphoric control. Coindexation indicates referential identity, as shown in Figure 5b.[10]

NTNU/NorSource/tree/master, and screenshots from analyses of some of the above sentences at http://headlex16.ipipan.waw.pl/programme.

[9] The English Iness XLE web grammar (http://clarino.uib.no/iness/xle-web, accessed 08.02.16) suggests functional control for the verb ‘promise’.

[10] In an architecture where one also makes use of an s-structure component, referential dependencies do not need to be shown in f-structure. In our representations we however stay with the conventions.

$$
\begin{bmatrix}
PRED & \text{'}versprechen \ < SUBJ, OBJ, COMP > \text{'} \\
SUBJ & \begin{bmatrix} PRED & \boxed{1}\text{'}pro\text{'} \\ NUM & SG \\ PERS & 3 \end{bmatrix} \\
OBJ & \begin{bmatrix} PRED & \text{'}pro\text{'} \\ NUM & SG \\ PERS & 3 \\ CASE & DAT \end{bmatrix} \\
COMP & \begin{bmatrix} PRED \ \text{'}kommen < SUBJ > \text{'} \\ SUBJ \ \boxed{1}\text{'}pro\text{'} \end{bmatrix}
\end{bmatrix}
$$

Figure 5b Obligatory anaphoric control: *'Ich verspreche ihm zu kommen.'*

In the LFG literature, the discussion of control is in many cases confined to the discussion of f-structure, but semantic approaches using linear logic, also covering control, are available (Dalrymple 1999, Asudeh 2005, next to others).
In order to formalise switched control, we need an explicit semantic representation. Working within feature semantics (Fenstad et al. 1985, Halvorsen and Kaplan (1995)), we use a formally grounded flexible format to relate co-reference and shared-argument configurations. Halvorsen and Kaplan formalise their approach by the composition of mappings, with an attribute-value type s-structure σ and a reversed f-function Φ^{-1}. This is what we will use to describe switch control as outlined in section 2.

As outlined in section 2, control constructions containing a suasive matrix verb and an embedded modal infinitive require that one of the arguments of the matrix verb is the understood normative agent. thus giving rise to the construction's deontic controller.[11] Although we use the same formal device to state normative and semantic dependencies, we have in mind that semantic participants and deontic controllers do not have the same conceptual status, and that thematic indices most likely should therefore be distinguished from normative ones. In this study, though, we compute deonticity as piggybacking on control relations and thus as semantic assignments that lead to the observed switched control patterns. As for now we suggest to introduce a conditioned functional control equation which allows us to introduce normative constraints on top of the already existing lexical semantic constraints when required by a deontic infinitival complement. This can be done in the lexicon using semantic equations as shown for *verspechen*/promise in (10). Note that if the if-then constraint is not met, standard subject control (Figure 5b) will result.[12]

[11] We discuss deontic control in the case of implicit arguments in section 5.2.

[12] We use σ'^{-} to denote an inverse function, as suggested by Halvorsen for phi (Halvorsen 1995: 283). Here we work with an inverse sigma function instead, that is a mapping from s-str to f-str.

(10) **versprechen**

(M* PRED) = ‘versprechen <SUBJ, OBJ, COMP >’
(M* SUBJ) = (M* COMP SUBJ)
if (M* COMP MOD) = deontic **then**
σ(M* ARG1) = σ (M* ARG3 ARG1) and
σ(M* ARG2) = σ (M* ARG3 AGR3 ARG1) and
σ(M* ARG3 ARG1) = $\sigma^{'-}$ (M* SUBJ PRED) and
σ(M* ARG3 ARG3 ARG1) = $\sigma^{'-}$ (M* COMP COMP SUBJ PRED)
= $\sigma^{'-}$ (M* COMP SUBJ PRED)
= $\sigma^{'-}$ (M* OBJ PRED)

If the conditional constraint is met, we, as a procedure, first tie up referential identity in two purely semantic constraints, and then add two inter-modular constraints that define the relation between the deontic semantics and its functional realisation. In f-structure the deontic controller is only realised once as the matrix subject while the matrix object is instantiated as the understood subject of the modal infintive (as in (3b) *Ich verspreche ihm kommen zu dürfen).* The combined f- and s-structure correspondence is shown in Figure 6:

$$\left[\begin{array}{ll} PRED & \text{'}versprechen < SUBJ, OBJ, COMP > \text{'} \\ SUBJ & \left[\begin{array}{ll} PRED & \boxed{1}\ \text{'}pro\text{'} \end{array}\right] \\ OBJ & \left[\begin{array}{ll} PRED & \boxed{2}\ \text{'}pro\text{'} \\ CASE & DAT \end{array}\right] \\ COMP & \left[\begin{array}{ll} PRED & \text{'}dürfen < SUBJ, COMP > \text{'} \\ SUBJ & \boxed{2}\ \text{'}pro\text{'} \\ MOD & DEONTIC \\ COMP & \left[\begin{array}{ll} PRED & \text{'}kommen < SUBJ > \text{'} \\ SUBJ & \boxed{2}\ \text{'}pro\text{'} \end{array}\right] \end{array}\right] \end{array}\right]$$

f-structure

σ^{-1}

$$\left[\begin{array}{ll} \text{REL} & \textit{versprechen} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \\ \text{ARG3} & \left[\begin{array}{ll} \text{REL} & \textit{PERMIT} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \\ \text{ARG3} & \left[\begin{array}{ll} \text{REL} & \textit{kommen} \\ \text{ARG1} & \boxed{2} \end{array}\right] \end{array}\right] \end{array}\right]$$

s-structure

Figure 6 Control resolution in deontic non-finite verb chains headed by suasive verbs of communication

5 Discussion

5.1 Passive inside the infinitival complement

Changes in control relations can be induced also through passive voice inside the infinitival complement. Examples from both languages are given in (11):

(11a) Ich versprach ihm abgeholt zu werden (German)
I promise.PST him pick-up.PASS to be.INF
'I promised him to be picked up'

(11b) Jeg lovet ham å bli sitert (Norwegian)
I promise.PST him to be.INF cite.PASS
'I promised him to be cited'

In both cases the reading that perhaps comes first to mind is one of object control, but given a suitable context they could also have subject control, and in both cases either reading necessitates a certain understood constellation of power or influence in the actual situation. Thus, in object control interpretation of (11), the 'I' can in both cases be easily understood as having enough influence in the situation to bring about the event expressed by the infinitival clause.

If a deontic modal is added, however, as in (12), the object control interpretation in both cases seems to be the only one available (i.e., *he* will be picked up, not I, and *he* will be cited, not I):

(12a) Ich versprach ihm abgeholt werden zu dürfen (German)
I promise.PST him pick.up.PASS be.INF to may.INF
'I promised him to be allowed to be picked up'

(12b) Jeg lovet ham å få bli sitert (Norwegian)
I promise.PST him to get.INF be.INF cite.PASS
'I promised him to be allowed to be cited'

The corresponding – converse - judgments seem valid for object control verbs; the sentences in (13) both seem ambiguous (although perhaps with preference for subject control), while the sentences in (14) can only have one reading:

(13a) Ich bat ihn abgeholt zu werden (German)
I ask.PST him pick.up.PASS to be.INF
'I asked him to be picked up'

(13b) Jeg ba ham om å bli sitert (Norwegian)
I ask.PST him about to be.INF cite.PASSP
'I asked him to be cited'

(14a) Ich bat ihn abgeholt werden zu dürfen (German)
I ask.PST him pick.up.PASS be.INF to may.INF
'I asked him to be allowed to be picked up'
('I' as the one to be picked up)

(14b) Jeg ba ham om å få bli sitert (Norwegian)
I ask.PST him about to get.INF be.INF cite.PASS
'I asked him to be allowed to be cited'
('I' as the one to be cited)

With regard to their role in serving as source of control switch for the verbs in question, there thus seems to be a contrast between downstairs passives and deontic modals. While the former tend to induce ambiguity in control interpretation, but can be disambiguated through use of the deontic modal, the deontic modals do not induce ambiguity. The mutual independence of these sources is also shown by the circumstance that either type can appear without the other, and that they can occur together.

We will not venture into any analysis of passives whereby the optional effect of control switch could follow,[13] since this ambiguity does not reside in anything like the deontic predicate we have assumed for *dürfen/få*. As just shown, the introduction of the deontic modal *resets* the ambiguity back to the pattern we have described.

5.2 Implicit objects as controllers

As noted at the end of section 3, the control switch patterns we describe obtain only when both the subject and the object are overtly expressed. If there is no overt object, constructions without deontic modals tend to display much variation in their control possibilities; thus, in German (15a) is possible with the

[13] We therefore will also not make any concrete suggestions as to how the switch effect connected to passives ought to be implemented in a grammar like the one mentioned in footnote 9, except for the following If an approach involving 'XARG' is used for passives analogously to what was indicated in footnote 8 for *få*, then the lexical representation of the modal *få* has to be split into one variant taking an active infinitive complement, in essence like the one considered in the text, and one variant taking a passive complement, whose 'XARG' will then have to be equated with the XARG of *få*.

interpretation indicated, but not subject control, whereas the Norwegian counterpart (15b) can have neither construal:

(15) a. Ich bat das Zimmer zu verlassen
I ask.PST the room to leave.INF
'I asked everyone to leave the room'

b. * Jeg ba om å forlate rommet
I ask.PST about to leave.INF room.DEF
'I asked to leave the room'

However, if a deontic modal is indeed used, both constructions turn into exclusively subject control patterns:

(16) a. Ich bat das Zimmer verlassen zu dürfen
I ask.PST the room leave.INF to may.INF
'I asked to be allowed to leave the room'
('I' leaving the room)

b. Jeg ba om å få forlate rommet
I ask.PST about to get.INF leave.INF room.DEF
'I asked to be allowed to leave the room'
('I' leaving the room)

What is observed here is thus the same *resetting* effect of using the deontic modal as was noted at the end of the previous subsection.[14]

This confirms the picture of an item which deterministically seeks out an overtly expressed controller in the matrix clause.

6 Concluding remarks

We have argued that the control patterns found in 'control switch' constructions containing the deontic modals *dürfen* (German) and *få* (Norwegian) are a case of deontic indexation. When a deontically headed infinitive is embedded under a suasive verb of communication, a deontic controller is introduced. This deontic controller is identified with the referent of one of the arguments of the matrix clause.

With their concise designs of syntactic-semantic representation, both of the frameworks LFG and HPSG allow for the articulation of these constellations.

[14] A similar observation is made in Doliana and Sundaresan (2016, p. 9), in a discussion of related phenomena.

Conditionals formulated for LFG (cf. (10) correspond to the double lexical specification of suasive verbs mentioned in the HPSG analysis.

This phenomenon of control switch is ‘deterministic’ – one relation of obligatory control is switched into the opposite relation of obligatory control. It is thereby different from the situation of ambiguous control relations as sometimes arise when the embedded infinitive is in passive form. Interesting in this context is that embedded deontic passives, as opposed to simple embedded infinitival passives, are unambigous in their control resolution.

A possible weakness of our approach is that we formally treat deontic indices on a par with thematic ones, which invites the question why the deontic controller introduced as the ARG1 of the deonic predicate is never openly realized, as opposed to the subject (or ARG1) of a standard predicate. As we see the facts, in these constructions deonticity is computed together with obligatory control relations, and so the deontic dimension must be allowed to – so to say – piggyback on the thematic dimension. In representing this circumstance, our approach nevertheless seems justified.

References

Asudeh, Ash. 2005. Control and semantic resource sensitivity. *Journal of Linguistics 41, 465–511.*

Bender, Emily M. and Scott Drellishak and Antske Fokkens and Laura Poulson and S. Saleem. 2010. Grammar Customization. In *Research on Language & Computation*, Volume 8, Number 1, 23-72.

Beermann, Dorothee. To appear. Infinitives: a comparative German-Norwegian study. In Hellan, Lars and Andrej Malchukov and Michela Cennam. (eds) *Contrastive Studies in Verbal Valency.* Amsterdam: J.Benjamins Publ. Co.

Bresnan, Joan. 1982. Control and Complementation. In: Bresnan, Joan (ed) *The Mental Representation of Grammatical Relations.* Cambridge, Mass.: The MIT Press.

Dalrymple, Mary. 1999. *Semantics and Syntax in Lexical Functional Grammar.* Cambridge, Mass.: The MIT Press.

Dalrymple, Mary. 2001. *Lexical Functional Grammar.* Syntax and Semantics, volume 34. New York: Academic Press

Dalrymple, Mary, Kaplan, Ronald M., Maxwell III, John T. & Zaenen, Annie (eds). 1995. *Formal Issues in Lexical-Functional Grammar*. Stanford CA: CSLI.

Doliana, Aaren & Sandhya Sundaresan. 2016. Proxy Control: A new species of obligatory control under modality. *In: Linguistische Arbeitsberichte* LAB. Band 93. Katja Barnickel, Matías Guzmán Naranjo, Johannes Hein, Sampson Korsah, Andrew Murphy, Ludger Paschen, Zorica Puškar &

Joanna Zaleska (Hrsg.): *Replicative Processes in Grammar*. Institut für Linguistik der Universität Leipzig.

Fenstad, Jens Erik.and Per-Kristian Halvorsen and Tore Langholm and Johan van Benthem. 1985. *Equations, schemata and situations: a framework for linguistics semantics. Technical Report 29. Stanford University: CSLI.*

Halvorsen, Per-Kristian. 1995. Situations and semantic descriptions in lexical-functional grammar. In Dalrymple et al. (eds).

Halvorsen, Per-Kristian and Ronald M. Kaplan. 1995. Projections and Semantic Descriptions in Lexical-Functional Grammar. *Formal Issues in Lexical-Functional Grammar*. Mary Dalrymple and Ronald M. Kaplan and John T. Maxwell III and Annie Zaenen (eds). Standford: CSLI.

Kiss, Tibor. 1995. *Infinitive Komplementation. Neue Studien zum deutschen Verbum infinitum.* Tübingen: Niemeyer.

Köpcke, Klaus-Michael, and Klaus-Uwe Panther. 1991. Kontrolle und Kontrollwechsel im Deutschen. In: *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 44, 143–166.

Lødrup, Helge. 1996. The Theory of Complex Predicates and the Norwegian Verb få 'get'. *Working Papers in Scandinavian Syntax 57, 1996, 76-91.*

Mair, Christian. 1990. A Contrastive Analysis of Object Control in English and German. *Papers and Studies in Contrastive Linguistics* 25 (1990): 85-101.

Polinsky, Maria. 2013. Raising and control. *In: The Cambridge Handbook of Generative Syntax*, ed. by Marcel den Dikken. Cambridge: Cambridge University Press; 20

Pollard, Carl J. and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: Chicago University Press.

Quirk, Randolph and Greenbaum, Sidney and Leech, Geoffrey and Svartvik, Jan. 1985. A Comprehensive Grammar of the English Language. London: Longman.

Radford, Andrew. 1985. *Transformational Syntax: A student's guide to Chomsky's Extended Standard Theory*. Cambridge: Cambridge University Press.

Ruzicka, Rudolf. 1999. *Control in Grammar and Pragmatics: A cross-linguistic study.* Amsterdam: John Benjamins Publ. Co.

Sag, Ivan A. and Thomas Wasow. 1999. *Syntactic Theory. A Formal Introduction.* Stanford: CSLI Publications.

Sag, Ivan A. and Thomas Wasow and Emily Bender. 2003. *Syntactic Theory.* Stanford: CSLI Publications.

Stiebels, Barbara. 2015. Control. In Tibor Kiss & Artemis Alexiadou (eds.) *Syntax – theory and analysis. An international handbook* (HSK 42), 412-446. Berlin: de Gruyter.

Resources
http://typecraft.org

Online grammars for Norwegian:
for LFG http://clarino.uib.no/iness/xle-web,
for HPSG http://regdili.hf.ntnu.no:8081/linguisticAce/

Ergative gender agreement in Dargwa “backward control” or feature sharing?

Oleg Belyaev
Lomonosov Moscow State University and Institute of Linguistics RAS

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 83–103

Keywords: agreement, person hierarchy, gender, obviation, Dargwa

Belyaev, Oleg. 2016. Ergative gender agreement in Dargwa “backward control” or feature sharing? In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 83–103. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.5.

Abstract

Dargwa languages have two types of agreement at clause level: gender and person agreement. In the general case, person agreement is hierarchical (speech act participants prefered to 3rd persons), while gender agreement is with the absolutive (S/P) argument. Two exceptions to this pattern have been observed in some dialects: first, some auxiliary verbs have a gender agreement slot which can be controlled by both ergative and absolutive arguments; second, adverbials agreeing in gender can agree with either ergative or absolutive if they are located at clause edges. A proposed explanation of this behaviour is through effectively splitting each clause into two layers, with the top layer having its own zero absolutive position, coreferential with either the subject or the direct object of the lower layer. In this way, the general rule that gender agreement is with the absolutive can be preserved. In this paper, I argue that the data of Ashti Dargwa do not support the Backward Control theory. Peripheral adverb agreement and auxiliary gender agreement are independent phenomena, while auxiliary agreement can be explained by splitting the 3rd person based on topicality, as in proximate-obviative systems. This allows us to preserve the conventional account of clause structure while framing the data of Dargwa in a wider typological context.

1 Introduction

Dargwa[1] is a group of East Caucasian languages spoken in central Daghestan (Russia). Like most East Caucasian languages, their key features include (non-rigid) SOV word order, morphological (dependent-marking) consistent ergativity, and a rather high degree of morphological complexity. Another feature of Dargwa that is shared with the majority of languages of this family is a system of grammatical gender agreement. Unlike most other branches of East Caucasian, Dargwa also has person agreement on the verb in finite clauses.

Person and gender agreement in Dargwa are largely morphologically distinct and governed by separate sets of rules. Gender agreement is, at clause level, straightforwardly controlled by the Absolutive argument (S/P), regardless of its position or grammatical function. The controller of person agreement is, in contrast, chosen between subject and object (A and P) by a complex set of rules that is mainly governed by the person hierarchy ($1, 2 > 3$ or $2 > 1 > 3$, depending on the language).

In some varieties of Dargwa, however, gender agreement on the auxiliary verb and on adverbs in clause-peripheral positions can optionally be controlled by the Agent-like argument. Sumbatova (2014) has proposed to explain this behaviour by splitting all clauses into two tiers, the lower tier headed by the lexical verb and the higher tier, by the auxiliary. The thematic arguments of the verb are located at the “lexical” tier, while the auxiliary has its own (absolutive) subject position. This position is always filled by a PRO which is anaphorically backward controlled by either the subject or the object of the lower tier. It is this zero subject that the

1. This research has been supported by the Russian Science Foundation, project no. 14-18-02429.

auxiliary and all peripheral adverbs agree with, thus allowing us to maintain a uniform absolutive-control rule for gender agreement.

In this paper, I will use the data of Ashti Dargwa to show that, at least for this dialect, the Backward Control hypothesis is redundant compared to a simpler solution that dispenses with a strict distinction between "gender" and "person" agreement, instead tying these phenomena to their syntactic positions. All agreement on auxiliaries is thus, in effect, conditioned by the same rules that govern "person" agreement, regardless of the features that they happen to display. The seemingly exceptional pattern of "gender agreement with A" is a purely morphological fact that follows from certain auxiliaries having a gender agreement slot. This solution allows us to maintain a traditional, single-tier f-structure, while capturing all the complexities of Dargwa agreement and making more generalizations than the Backward Control analysis.

2 Agreement in Ashti Dargwa

In this section, I will describe the core agreement system of Ashti, the main variety discussed in this paper. This system also serves as a representative example of agreement in Dargwa as a whole.

2.1 Gender

Like all other Dargwa languages, Ashti has a system of three genders, masculine, feminine and nonhuman. All of these are semantically transparent. In the plural, the distinction is only between human (masculine + feminine) and nonhuman. The gender markers have the same form across all morphological positions:

	SG	PL
M	*w*	*b*
F	*j*	
N	*b*	*d*

Gender agreement regularly occurs in several contexts, of which the most frequent are:

- prefix on most verb stems;
- suffix on attributive forms;
- suffix on essive nouns and certain adverbs.

Attributive forms generally agree with the head that they modify. Items appearing at clause level, i.e. verbs and adverbials, agree with the P/S (absolutive) argument.[2] For example, in (2), both the verb *j-us.aj* and the adverb *wac'ac:i-j* 'in the forest' agree with *pat'imat* 'Patimat' in feminine singular; the verbal agreement slot is prefixal, the adverbial agreement slot is suffixal.

2. The abbreviations in glossing follow the Leipzig Rules, apart from the following labels: ATTR: attributive, ESS: essive (static location), HPL: human plural, NPL: neuter plural, PTCL: particle. A list of abbreviations is found at the end of this paper. Morpheme boundaries are simplified to the extent that it does not affect the translation or the phenomena under discussion. Whenever existing morpheme boundaries in the first line have been suppressed for convenience, the dot (".") is used instead of the hyphen. The frame denotes the agreement controller: this helps to separate its annotation from the annotation of the agreement morphemes and other highlighted material, which uses boldface.

(1) Ashti

$\boxed{\textit{pat'imat}}$	***j**-id.až.i*
P.(F)	F-went.out

'**Patimat** went out.'

(2)

murad-li	*wac'a.c:i-**j***	$\boxed{\textit{pat'imat}}$	***j**-us.aj*
M.(M)-ERG	in.forest-F	P.(F)	F-caught

'Murad caught **Patimat** in the forest.'

2.2 Person

Person agreement morphology is found in most independent sentences and on certain dependent forms. There are several sets of synthetic (morphologically bound) person markers, and one clitic set that is used together with non-finite verbs to form periphrastic paradigms and in nonverbal predication. The clitic set only distinguishes number in the 2nd person; the 2nd person plural marker is homonymous with the 1st person, and the 2nd person singular marker is distinct. The 3rd person is marked by the so-called "copula" *sa-b* or by zero, depending on the paradigm. Importantly, the copula contains the gender marker as a suffix (in the table, neuter *-b* is used for illustration).

	SG	PL
1	*=da*	
2	*=di*	*=da*
3	(*=sa-b*)	

The morphological structure of most synthetic paradigms is largely the same:[3] there are separate exponents for person and number in the 1st and 2nd person and no number distinction in the 3rd person. In Ashti, the most widely used set of this type is the preterite set:

	SG	PL
1	*-d*	*-d-a*
2	*-t:i*	*-t:-a*
3	*-aj*, *-in*, *-i*	

Agreement in Ashti is hierarchical and fits into the general framework provided in Sumbatova (2011) and Belyaev (2013). Descriptively, the choice of the controller obeys the following rules:

- In intransitive clauses, person agreement is with S.
- In transitive clauses, the controller is chosen between A and P:
  - **A=3,P=3 → 3** if both arguments are not speech act participants (SAPs), i.e. 3rd person, the verb is in the 3rd person;
  - **A=1/2, P=3 → A** if A is a SAP while P is 3rd person, the verb agrees with A;
  - **A=3, P=1/2 → P** if A is 3rd person while P is a SAP, the verb agrees with P;

3. The only exception is the so-called optative set, which follows the pattern of the clitic set: 1p., 2pl. *-a*, 2sg. *-i*, zero in third person.

A=1/2, P=1/2 ⟶ P if both arguments are SAPs, agreement is controlled by P (the absolutive argument).

This is illustrated by the following examples:

(3) $\boxed{\textit{di-l}}$ *murad us.a-**d***
me-ERG M.(M) [M]caught-1[SG]
'**I** caught Murad.' (A = 1, P = 3 ⟶ agreement with 1)

(4) *murad-li* $\boxed{\textit{du}}$ *us.a-**d***
M.(M)-ERG I(M) [M]caught-1[SG]
'Murad caught **me**.' (A = 3, P = 1 ⟶ agreement with 1)

(5) *di-l* $\boxed{\textit{u}}$ *us.a-**t:i***
I-ERG thou(M) [M]caught-2[SG]
'I caught **you**.' (A = 1, P = 2 ⟶ agreement with 2)

(6) *u-dil* $\boxed{\textit{du}}$ *usa-**d***
thou-ERG I(M) [M]caught-1[SG]
'You caught **me**.' (A = 2, P = 1 ⟶ agreement with 1)

(7) $\boxed{\textit{murad-li}}$ $\boxed{\textit{rasul}}$ *us-aj*
M.(M)-ERG R.(M) [M]caught-3
'**Murad** caught **Rasul**.' (A = 3, P = 3 ⟶ agreement with 3)

As a generalization, one can say that agreement in Ashti is controlled by two hierarchies: the person hierarchy ($1, 2 \succ 3$, or SAP ≻ non-SAP) and the grammatical function hierarchy (OBJ ≻ SUBJ), with the former being dominant. Hence, as a general rule, the highest-ranking argument on the person hierarchy controls agreement; if both subject and object have the same rank, the controller is the object. Naturally, since there is no number distinction in the 3rd person, there is no way to distinguish between the controllers, thus the last part of the rule is only observed when both arguments are SAPs.

It is worth mentioning that, for the purposes of agreement, Ergative and Dative-marked transitive subjects behave in the same way:

(8) $\boxed{\textit{dam}}$ *murad u̯lħ.i-**d***
me.DAT M.(M) [M]saw-1
'**I** saw Murad.'

(9) *murad.li-j* $\boxed{\textit{du}}$ *u̯lħ.i-**d***
M.(M)-DAT I(M) [M]saw-1
'Murad saw **me**.'

Thus, for the purposes of this paper, "ergative subjects", "ergative agreement", "A agreement" and similar terms should be understood as referring to both Ergative and Dative-marked arguments.[4]

4. Unlike some other languages with hierarchical agreement, only core arguments, i.e. S, A and P, can serve as controllers.

3 The Backward Control hypothesis

3.1 Tanti data

Most Dargwa dialects, with few variations, follow a pattern similar to the above. However, in certain varieties, this clear picture faces problems if one considers the behaviour of the gender agreement marker that is found on the copula. In particular, Sumbatova (2014) has shown that in Tanti Dargwa, which otherwise has exactly the same agreement syntax as Ashti, the copula[5] can agree alternatively with the absolutive or the ergative:

(10) Tanti

murad-li	*t'ant'i-b*	*qali*	*b-irq'.u.le*	*sa-**j***
M.(M)-ERG	in.T.-N	house(N)	N-building	COP-M

(11)

murad-li	*t'ant'i-b*	*qali*	*b-irq'.u-le*	*sa-**b***
M.(M)-ERG	in.T.-N	house(N)	N-building	COP-N

'**Murad** is building **a house** in Tanti.'

Examples (10)–(11) represent periphrastic constructions that are typical for Dargwa. A non-finite lexical verb (participle or converb) is accompanied by an auxiliary, in this case the 3rd person "copula" *sa-b*. This form of the auxiliary is remarkable in that it incorporates a gender marker.[6] Unlike most other elements agreeing in gender at clause level, this auxiliary can alternatively agree with the ergative or the absolutive argument of the clause.

Based on the data provided in Sumbatova (2014) and Sumbatova and Lander (2014), the choice of agreement controller in 3rd person contexts seems to be based mostly on topicality.[7] More specifically, the "default" option seems to be subject agreement, with the object only "overtaking" agreement control only in case it possess a "higher degree of topicality" (there is no precise formulation of this notion given in the paper):

(12) a.

se.li.ž	*ʕeˁli*	*ʕeˁla*	*χ:ʷe*	*it.u.se=de?* —	*hi.ti-li*	*dila*	*uc:i*
why	thou:ERG	thy	dog(N)	hitting=2	it(N)-ERG	my	brother(M)

uc.ib.le	*=sa-**b***	/	*sa-**j***
bitten	COP-N	COP-M	

'Why are you hitting your dog? — It hit my brother.' (Sumbatova and Lander 2014, 453)

5. "Copula" is a traditional term for what is essentially a 3rd person auxiliary in languages like Ashti, and an auxiliary stem (not limited to the 3rd person) in languages like Tanti. It does act as a copula in nonverbal predication, but so do the person markers *=da* and *=di*, for which the term is not usually employed.

6. The diachronic origin of the gender distinction in the copula is not clear. It is homonymous with the absolutive form of the 3rd person personal-"reflexive" pronoun *ca‹b›i* (*sa-b* in Tanti). Such pronouns in East Caucasian are closer to personal pronouns (pronominals) than actual reflexives; hence, this may be an example of the transition from a pronoun to a copula (Li and Thompson 1977). However, the situation is far from clear, as the oblique forms of the "reflexive" come from a different source (sg. *cin-*, pl. *ču-*), which means that the absolutive form may itself be an innovation.

7. It is emphasized that this is not a strict rule; in particular, even arguments explicitly marked by focus can control gender agreement. The authors provide no explanation for this behaviour.

b. *ʕeˤla uc:i.li-ž se b-it.arg.ur.se? —* $\boxed{\textit{hi.t}}$ *ca*
thy brother(M)-DAT what(N) N-happened that(M) one
*χ:ʷe-li uc.ib =s:a-**j** / *=s:a-b*
dog(N)-ERG bite COP-M COP-N

'What happened to your brother? A dog bit him.' (ibid.)

According to Sumbatova, in the answer in (12a), the topic is the subject 'dog', yet the direct object 'brother' can also control agreement because it nevertheless possesses a high degree of "topicality" due to its human reference. In contrast, in (12b), the topic is the direct object 'brother', and the subject 'dog' cannot control agreement because it is neither topical nor higher than the subject on the animacy hierarchy.

Overall, while the discussion of the conditions on agreement in Sumbatova and Lander (2014) is rather vague and ultimately unconvincing (as the authors themselves admit), it seems rather clear that gender agreement of the auxiliary in languages like Tanti is a syntactic phenomenon that is conditioned by information structure, in particular by topicality and relative prominence on the animacy hierarchy. The specific way these factors interact needs further elaboration, but the core of the phenomenon seems reasonably clear. It must also be observed that a similar conclusion is reached for Akusha gender agreement in van den Berg (2001).

It is important that in Tanti both options seem to be available even if one of the arguments is a SAP. The autors admit that absolutive control in the case of an SAP subject and a non-SAP object is only marginally possible if there is "emphasis" on the absolutive:

(13) $\boxed{\textit{ʕaˤli}}$ *rurs:i quli-r r-alt.un.ne sa-**j**=de*
thou:ERG girl(F) in.house-F F-keeping COP-M=2SG

'You are keeping the girl at home.'

(14) *ʕaˤli* $\boxed{\textit{rurs:i}}$ *quli-r r-alt.un.ne sa-**r**=de*
thou:ERG girl(F) in.house-F F-keeping COP-F=2SG

'You are keeping the girl home alone.' (Sumbatova 2014)

Unfortunately, what is meant by "emphasis" in this case is not clear. It may be the adverb 'alone' in the translation of (14) that is responsible for the "emphatic" reading, but in this case the claimed association with gender agreement and topicality seems dubious at best: if anything, 'alone' is closer to a focus marker. Regardless of the interpretation, the grammaticality of this example shows that the controller of gender agreement on the auxiliary is independent from the controller of person agreement.

For Tanti there is no data on what happens when both arguments are SAPs, or when a SAP is in the direct object position.

3.2 Proposed solution

The solution proposed in Sumbatova (2014) is to divide the clause into two layers, one headed by the auxiliary (roughly corresponding to IP) and the other headed by the lexical verb (roughly corresponding to VP) and stipulate that the upper layer has its own subject position. This position is always filled by a null PRO which is backward controlled by either the subject or the object of the lower layer:

(15) ergative agreement (= ex. 10)

[$\Delta_{i(ABS)}$ [***murad-li**$_i$* *t'ant'i-b* *qali* *b-irq'ule*] *=sa-**j***]
M.(M)-ERG T.-N[ESS] house(N) N-doing=COP-M

(16) absolutive agreement (= ex. 11)

[$\Delta_{i(ABS)}$ [*murad-li* *t'ant'i-b* ***qali**$_i$* *b-irq'ule*] *=sa-**b***]
M.(M)-ERG T.-N[ESS] house(N) N-doing=COP-N

In this case, the seemingly exceptional pattern of ergative gender agreement on the auxiliary is fully regular: the copula agrees not with the ergative argument of the lower tier, but with its own absolutive subject that is coreferential with that argument.

Sumbatova's analysis mainly rests on two pieces of independent evidence: the behaviour of adverbs and the behaviour of non-finite forms.

3.2.1 Adverbs

In Tanti, an additional piece of evidence to confirm this theory is the behaviour of adverbial elements. While as a general rule adverbs agree with the absolutive argument, they can agree with the transitive subject NP if they are located at the left or right edge of the clause:

(17) [*maˁħaˁmmad.li.š:u-**w*** / ***-b*** [$\boxed{\textit{rasul-li}}$ *dig* *b-uk:-un-ne*] *=sa-**j***]
chez.M.-M -N R.(M)-ERG meat(N) N-eating COP-M

'At Muhammad's place **Rasul** is eating **meat**.' (Sumbatova 2014)

This behaviour is easily explained if we assume that such adverbs are actually adjoined at the higher (IP) layer of the clause. Like the auxiliary, they agree with the absolutive argument within the domain, which happens to be the zero absolutive. Again, the advantage of this solution is that no additional agreement patterns have to be introduced.

However, the data of Ashti put the relevance of these data for the analysis of auxiliary gender agreement into doubt. Just like Tanti, Ashti allows "peripheral" adverbs to agree with the ergative argument:

(18) Ashti

*wac'a.c:i-**w*** / *wac'a-c:i-**j*** $\boxed{\textit{rasul-li}}$ *pat'imat* *j-u:s-u* *li-**w***
in.forest-M in.forest-F R.(M)-ERG P.(F) F-catching be-M

'In the forest Rasul is catching Patimat.'

However, it seems that there is no direct connection between this phenomenon and auxiliary agreement. The Backward Control theory predicts that adverbs can only agree with the ergative argument if the auxiliary also agrees with this argument. Indeed, since every adverb that appears at the edge of the clause can be potentially identified as adjoined at either IP or VP level, absolutive agreement is predicted to always be available. Ergative agreement, in contrast, is only predicted to occur if the "zero absolutive" at the upper level is coreferent with the ergative argument, which should trigger ergative agreement on the copula as well. But this prediction is not borne out: ergative agreement on peripheral adverbs is available even if the auxiliary agrees with the absolutive, as seen in (19).

(19) *wac'a.c:i-**w*** / *wac'a.c:i-**j*** *rasul-li* $\boxed{\textit{pat'imat}}$ *j-u:s.u* *li-**j***
in.forest-M in.forest-F R.(M)-ERG P.(M) F-catching be-F

In this example, the auxiliary agrees in feminine with the absolutive argument, which, in the Backward Control analysis, means that the zero absolutive in the higher clause is coreferent with the direct object. This should make ergative agreement of peripheral adverbs with the ergative argument impossible, yet it seems to be no less grammatical than in examples like (18).

Thus, while the phenomenon is interesting, it does not seem to have any relation to the issue of auxiliary agreement — or at least, this is not true for all varieties. A possible explanation is that such adverbs head secondary predications with their own internal subjects. An alternative explanation, suggested by an anonymous referee, might be an analysis along the lines of the Constructive Case approach used in Nordlinger (1998), on the assumption that the adverb in (19) adjoins not at clause level, but to one of the arguments. Whatever conclusion is eventually reached, we may say that, whatever the merits of the Backward Control analysis in explaining auxiliary agreement, adverb agreement has no direct connection with this phenomenon.

3.2.2 Relative clauses

One of the arguments that seems to favours the Backward Control analysis is the fact that in Tanti participial relative clauses with nonverbal predicates, the copula can agree either with the ergative of the embedded clause or the head noun of the NP:

(20) Tanti

a. [*rasul-li* *waw-ne* *d-ič:ible* *sa-**r**-se*] $\boxed{\textit{rursi}}$
R.(M)-ERG flower(N)-PL NPL-gave COP-F-ATTR girl(F)

b. [$\boxed{\textit{rasul-li}}$ *waw-ne* *d-ič:ible* *sa-**w**-se*] *rursi*
R.(M)-ERG flower(N)-PL NPL-gave-COP-M-ATTR girl(F)

c. * [*rasul-li* $\boxed{\textit{waw-ne}}$ *d-ič:ible* *sa-**d**-se*] *rursi*
R.(M)-ERG flower(N)-PL NPL-gave-COP-NPL-ATTR girl(F)

'the girl to whom Rasul gave flowers' (Sumbatova and Lander 2014, 469)

It is claimed that the Backward Control hypothesis explains this better than possible alternatives. In a "conventional" view, one would have to assume a separate set of agreement rules for relative clauses, whereas the Backward Control hypothesis allows to assume anaphoric control uniformly.

However, without an explicit analysis, it is not clear how exactly the Backward Control hypothesis makes analyzing such examples simpler. The control pattern in (20b) has to be different from the control patterns used in finite clauses in any case. Both probably have to be construction-specific. Furthermore, the ungrammaticality of (20c) can be due to the fact that the NP 'flowers' is neither animate nor a topic, and thus it is outranked by ‘girl’: the same explanation is provided by the authors for (12b), and it is not clear why it would not work in this case. Thus the examples do not provide enough data for reaching any valid conclusions.

Perhaps more importantly, the agreement of the verb in a participial relative clause with the head seems to be independent from agreement of the copula in Dargwa in general. The relevant evidence comes from another Dargwa variety, Shiri (field data). In this language, under no circumstances can the copula agree with the ergative:

(21) Shiri

rasul-li	*pat'imat.li-ž*	$\boxed{\textit{wa̰w-ni}}$	*d-ik:.ib.li*	*ca‹d›i* /	* *ca‹w›i*
R.(M)-ERG	P.(F)-DAT	flower(N)-PL	NPL-gave	COP‹NPL›	COP‹M›

‘Rasul has given flowers to Patimat.’

Thus, apparently, the Backward Control analysis is not applicable to this Dargwa variety. However, the attributive marker *-zi-b* on the relative clause can alternatively agree with the absolutive argument of the relative clause or the head of the relative clause:

(22) a.

[*rasul-li*	$\boxed{\textit{wa̰w-ni}}$	*d-ik:.ib-žu-**d***]	*rursi*
R.(M)-ERG	flower(N)-PL	NPL-given-ATTR-NPL	girl(F)

b.

[*rasul-li*	*wa̰w-ni*	*d-ik:.ib-zi-**r***]	$\boxed{\textit{rursi}}$
R.(M)-ERG	flower(N)-PL	NPL-given-ATTR-F	girl(F)

‘The girl to whom Rasul gave flowers.’

In (22a), the gender marker on the attributive suffix agrees with the absolutive argument of the relative clause, while in (22b), it agrees with the head of the relative clause. The distribution of controllers is different — ergative vs. NP head in Tanti and absolutive vs. NP head in Shiri — but the phenomena seem to be of the same nature. They require further exploration, but, given the absence of ergative gender agreement in finite clauses in Shiri, probably have no direct bearing on the question discussed in this paper.

4 Ergative gender agreement in Ashti

In the previous section, I have attempted to demonstrate that neither the behaviour of non-finite forms nor the agreement of adverbs — the hallmarks of the Backward Control analysis — seem to be valid arguments in favour of the Backward Control hypothesis. The latter, therefore, only serves to explain the auxiliary gender agreement and some of its properties. It is therefore preferable to integrate these data into one of the prior analyses rather than propose a new one. In this section I will test the Backward Control hypothesis on the data of Ashti Dargwa, which overall seems to have the same system of both gender and person agreement as Tanti does.

4.1 Preliminary remarks

For all the similarities, there is an important difference between Ashti and Tanti verb systems that makes the study of gender agreement somewhat more complicated in Ashti. The difference consists in the fact that, unlike Tanti which allows the gender-marked copula in all persons, most paradigms in Ashti only use it in the 3rd person under negation:

(23) Ashti

du *w-ax.ul* ***=da*** / ****sa-w=da***
I(M) M-going 1 COP-M=1
'I **am** going.'

(24) *u* *w-ax.ul* ***=di*** / ****sa-w=di***
thou(M) M-going 2 COP-M=2
'You **are** going.'

(25) *murad* *w-ax.ul* *(****sa-w****)*
M.(M) M-going[3] COP-M
'Murad **is** going.'

(26) *murad* *w-ax-ul* ***a-sa-w***
M.(M) M-going NEG-COP-M
'Murad **is not** going.'

This means that the number of forms where one can test for gender agreement on the auxiliary is limited to 3rd person negative contexts, which is obviously not enough for establishing a definitive analysis.

Thankfully, Ashti also possesses a series of periphrastic forms utilizing the so-called existential verbs in the position of auxiliaries. There are altogether four existential verbs in Ashti, which consist of a stem marking the location of the object or action relative to the speaker and a gender marker (*le-b* 'near speaker or hearer', which is also the neutral existential form; *te-b* 'away from speaker and hearer, on the same level'; *k'e-b* 'away from speaker and hearer, above'; *χe-b* 'away from speaker and hearer, below'). Existential verbs additionally agree in person and number by attaching auxiliaries from the clitic set, the 3rd person being zero-marked.

When used without an additional lexical verbs, existentials are used in a number of sentence types with nonverbal predications, in particular, in expressions of existence, location, and predicative possession. Existentials can also serve as auxiliaries in place of the more widespread person clitics. Periphrastic forms using such auxiliaries are structured exactly like their corresponding "unmarked" versions, for example:

(27) *murad* ***ʔu̯q'ˤ.u̯n*** ***li-w***
M.(M) [M]going be-M[3]
'Murad is going.'

(28) *du* ***ʔu̯q'ˤ.u̯n*** ***li-w=da***
I(M) [M]going be-M=1
'I am going.'

Using existential periphrastic forms provides the extra advantage of having the gender marker on the auxiliary in each person, not just in the 3rd person. In Sumbatova (2014), it has been demonstrated that existential forms have exactly the same agreement behaviour as ordinary periphrastic verb forms. To the extent that it can be tested, the same seems to apply in Ashti, so using these forms appears justified.

4.2 Auxiliary agreement

4.2.1 3rd person

Just like in Tanti, the auxiliary in Ashti can agree in gender with A in the 3rd person:

(29) *rasul-li* *pat'imat* *j-u:s.u* *li-j* / *li-w*
R.(M)-ERG P.(F) F-catching be-F be-M
'Rasul is catching Patimat.'

Again, as in Tanti, this seems to correlate with topicality, although more research is needed in order to establish the specific factors that are responsible for the choice of agreement controller. In Ashti, a further complication is that, as mentioned above, the number of forms which exhibit auxiliary gender agreement is rather low, and they are rarely found in natural texts, hence there is not enough data to test the topicality hypothesis.

However, Kubachi Dargwa, a much larger variety that is very closely related to Ashti, has generally the same verbal system and agreement rules, but, unlike Ashti, does use the copula in the 3rd person in all periphrastic verb forms. There is also a large number of texts available in Kubachi. From the collection of stories about Mullah Nasruddian (Šamov 1994), the relation between 3rd person auxiliary gender agreement and topicality is readily seen, such as in the following examples.

(30) a. *na* *q:ala.l* *saʁ.ib,* *wagzal.li-b* *čumadan* *sa*
now to.Mamedkala when.he.reached at.station-N[ESS] bag(N) one
hambal.li.c:e *b-ič:.ib=sa-**w***
to.porter N-gave=COP-M
'When he [Mullah Nasruddin] reached Mamedkala, at the station he gave his bag to a porter.'

b. *jiš.te* *χulžin* *d-ač:.ib* *kʷi‹d›ič.ib.li=sa-**d***
these(HPL) bag(NPL) NPL-having.found return‹NPL›=COP-NPL
malla.c:e
to.Mullah
'Having found the bag, they returned it to the Mullah.'

The sentences in (30), although they are from different texts and thus not a "minimal pair", illustrate the difference between the two agreement patterns rather clearly. The first sentence is in the very beginning of the text, which, like all texts in the collection, about Mullah Nasruddin; the bag, in contrast, has not been mentioned in the preceding context. Hence it is rather clear that here Mullah is the topic (the sentence is about his actions), while the bag is part of the focus.

Conversely, the second sentence is among the final sentences of the text, which detail the fate of a bag that had earlier been stolen from the Mullah. The subject, "they", refers to the people of Amuzgi, and it is inconsequential to the narrative, which in this case concerns the fate of the bag. Thus the bag is the topic, accordingly triggering gender agreement.

(31) *du-dil ha.ʔ.ila-žu-d si.k'al.dix ʔa̱:ʔa̱-dil duč:i.al haʔ.ib-žu-**d**=sa-**d***
I-ERG said-ATTR-NPL the.matter hen-ERG at.night said-ATTR-NPL=COP-NPL
(Mullah, why has the judge acquitted you without you even saying anything?) 'At night the hen has already explained the matter for me.'

The agreement pattern in (31) is easily explained by information structure. The nighttime event that the Mullah refers to is his giving the chicken to the judge as a bribe. In the context of the question and his answer, it is clearly part of the focus, not the topic, which is the Mullah or the event of his acquittal.

(32) a. *wah, malla, si uk'.u.t.nu, allah-le duna e:k*
oh mullah(M) what(N) art.thou.saying Allah(M)-ERG world(N) six
*bac.le a-sa-**b**=q'al, e:k:-il sa-**b** b-a:q'.ib-zi-**b***
in.month NEG-COP-N=PTCL six-day COP-N N-done-ATTR-N
'Oh, Mullah, what are you saying, God created the world in six days, not months!'

b. *e:k:-il b-a:q'.ib-zi-**w**=sa-**w** b-uk'.ne dammi=ja=q'el*
six-ATTR N-done-ATTR-M=COP-M N-that.is.said to.me=also=PTCL
b-ak'u.q'a.nnu
N-is.known
'I do also know that it is said that He has created the world in six days.'
(... but would you believe me if I told you that?)

The sentences in (32) are from the same text. The one in (a) is a statement made by a pious worshipper in response to the Mullah's sermon on God creating the world in six months. Here, the topic under discussion is the world and the timeline of its creation. Consequently, the copula and the predicative attributive form all agree with the absolutive argument 'world'.

The sentence in (b) is the Mullah's reply; in it, he quotes the traditional Biblical/Quranic statement 'God created the world in six days', as clear from the embedding of this clause under *b-uk'-ne*, the masdar (verbal noun) form of the verb 'to say'. This traditional narrative concerns God and his actions, and not the world in particular; thus, the copula agrees in the (overtly unexpressed) ergative subject 'God'.

4.2.2 1st and 2nd persons

When only 3rd person participants are considered, nothing in Ashti or Kubachi seems to contradict the Backward Control hypothesis. However, when one of the arguments is 1st or 2nd person, and the other is 3rd person, gender agreement can only be with the SAP argument:

(33) a. $\boxed{\textit{di-l}}$ *pat'imat j-u:s.u li-**w=da** / *li-j=da / *li-w / *li-j*
me(M)-ERG P.(F) F-catching be-M=1 be-F=1 be-M be-F
'**I** (m.) am catching Patimat (f.).' (1 > 3)

b. *pat'imat-li* $\boxed{\textit{du}}$ *u:s.u li-**w=da** / *li-j=da / *li-w / *li-j*
P.(F)-ERG I(M) [M]catching be-M=1 be-F=1 be-M be-F
'Patimat (f.) is catching **me** (m.).'

Similarly, when both arguments are SAPs, gender agreement can only be with the absolutive:

(34) a. *di-l* $\boxed{\textit{u}}$ *j-u:s.u li-**j=di** / *li-w=di / *li-w=da / *li-j=da*
me(M)-ERG thou(F) F-catching be-F=2 be-M=2 be-M=1 be-F=1
'I (m.) am catching **you** (f.).'

b. *u-dil* $\boxed{\textit{du}}$ *u:s.u li-**w=da** / *li-j=da / *li-j=di / *li-w=di*
thee(F)-ERG I(M) [M]catching be-M=1 be-F=1 be-F=2 be-M=2
'You (f.) are catching **me** (m.).'

This means that, when one of the arguments is a SAP, the controller of gender agreement on the auxiliary must be the same as the controller of person agreement: in (33), the non-3rd-person argument; in (34), where both arguments are SAPs, the direct object.

This behaviour is not predicted by the Backward Control hypothesis. If gender agreement on the auxiliary were triggered by a zero absolutive argument that is backward controlled by an NP in the lower layer of the clause, there would be no direct connection between person agreement and gender agreement. While (33) could be explained by SAPs being more likely topics than 3rd person participants, no such explanation is available for (34): there is no reason why, among two SAP participants, the absolutive is somehow more inherently topical than the ergative. In fact, the pattern here directly contradicts what is observed in the 3rd person, where the default option is subject, not object agreement.

What we see from the examples above is that, whenever the person agreement controller is clearly visible (i.e. when one of the arguments is an SAP), gender agreement on the auxiliary must be controlled by the same argument. This strongly suggests that in the 3rd person as well, the competition between A and P agreement does not involve a separate syntactic mechanism, but is based on the same pattern as person agreement in other instances.

4.3 Generalization

To conclude, there seems to be no positive evidence in favour of the "Backward Control hypothesis" in Ashti. Rather, as a general rule, the controller of gender agreement on the auxiliary is identical to the controller of person agreement.[8]

8. A similar line of reasoning can be found as early as Magometov (1963, 155), with the exception of outdated terminology. Following traditional grammar, Magometov considers subject person agreement as the unmarked case, and object agreement as a passive construction. Person agreement with the absolutive, and gender agreement of the copula in the absolutive, are thus treated in the same way.

This is true for all examples where SAP controllers compete with non-SAP controllers, or when two SAP controllers compete with each other. Extrapolated to the topicality split in the 3rd person, it means that we also have to split 3rd-person controllers into two classes, giving the following general rule:

SAP vs. non-SAP SAP wins

- A = 1, P = 3 : 1
- A = 2, P = 3 : 2
- A = 3, P = 1 : 1
- A = 3, P = 2 : 2

SAP vs. SAP P argument wins

- A = 1, P = 2 : 2
- A = 2, P = 1 : 1

non-SAP vs. non-SAP "topic" wins

- A = 3_{TOP}, B = 3 : 3_{TOP}
- A = 3, B = 3_{TOP} : 3_{TOP}

In terms of the person hierarchy, this can be captured by splitting the 3rd person into two "persons": the topical and the non-topical 3rd person. This gives us the following two hiearchies, with the same agreement rule as described above:

Person $1{,}2 \succ 3_{\mathrm{TOP}} \succ 3$

Grammatical relations $\mathrm{P} \succ \mathrm{A}$

If we call the "topical" third person *proximate* and label it as "3", and rename the "non-topical" third person to *obviative*, labeling it as "3′", the Ashti system looks like a typical proximate-obviative system, typologically well-known, especially in North America, cf. e.g. Aissen (1997). The only difference from a typical proximate-obviative system is that in Ashti, the distinction is not marked on NPs, but is only relevant for selecting the agreement controller. But the latter is also encountered in proximate-obviative systems, and it is typical to find topicality play a role for 3rd persons, but not for SAPs.

5 Analysis

5.1 "Person" vs. "gender"

The generalization provided in the preceding section does not resolve a key problem for the traditional view of agreement in Dargwa languages. Under this account, "person" (i.e. hierarchical) agreement reflects not only person and number, but also gender features. This means that we can no longer provide simple separate rules for person and gender, but have to explicitly state the pattern that is used for each individual agreement morpheme.

However, it is well-known that at least the term "gender agreement" is somewhat misleading in East Caucasian anyway. In particular, "gender" agreement may involve person, as has been convincingly argued for Archi in Corbett (2013). Exactly the same phenomenon that is described by Corbett occurs in Dargwa, including Ashti. This phenomenon consists in the fact that the neuter pl. marker *-d-* is used for 1/2PL arguments instead of the expected human pl. marker *-b-*:[9]

9. Even though the morpheme is *-d-* both for 1st/2nd person plural and for neuter plural, I gloss

(35) *rasul.li-j* [*du*] *j-u̯lħ.i̯-d*
R.(M)-DAT I(M) F-saw-1
'Rasul saw me (f.).'

(36) *rasul.li-j* [*nus:a*] ***d**-u̯lħ.i̯-d-a* / *****b**-u̯lħ.i̯-d-a*
R.(M)-DAT we 1PL-saw-1-PL HPL-saw-1-PL
'Rasul saw us.'

(37) *rasul.li-j* [*du=ba murad*] ***d**-u̯lħ.i̯-d-a* / *****b**-u̯lħ.i̯-d-a*
R.(M)-DAT I=and M.(M) 1PL-saw-1-PL HPL-saw-1-PL
'Rasul saw me and Murad.'

While there have been attempts to describe this phenomenon by stipulating a special "fourth gender" specifically for the pronouns 'we' and 'you (pl.)', (37) shows that this solution does not work: the marker *-d-* surfaces even if 1st person plural is resolved syntactically and not provided in the lexicon. Therefore, we have to conclude that "gender" markers in Ashti do genuinely mark person in addition to gender and number.

This conclusion having been made, the claim that "person" agreement markers encode gender features does not seem as radical. In fact, it seems that a view that connects agreement rules with agreement features is not adequate for Dargwa. It seem more productive to speak of two kinds of agreement *patterns*: the "absolutive" pattern and the "hierarchical" patterns. While the former typically involves gender and number features and the latter, person and number, this is merely a tendency that can be overridden by individual morphemes being specified for additional features. In this way, while singular absolutive-controlled ("gender") morphemes in Dargwa are only marked for gender and number, plural ones are also marked for person. Similarly, most hierarchically controlled ("person") morphemes only mark person and number, but certain auxiliaries also have a slot for gender.

Having two different patterns of agreement might be a problem in some approaches, but LFG allows us to use two different mechanisms: feature sharing and co-specification of agreement features.

5.2 Feature sharing

Feature sharing is a syntactic mechanism that has been most recently elaborated in Haug and Nikitina (2015), specifically in an LFG framework. This approach assumes that certain kinds of agreement are better described not in the traditional LFG co-specification approach, but by the controller and target having separate AGR features that are equated using the LFG mechanism of structure sharing, as in the following example:[10]

---

it as 1PL or 2PL in the former case, because in this function it does not in any way encode the neuter feature.

10. To simplify the discussion, I am using labels in double quotation marks, such as "be", as a shorthand for complete f-structures, or those parts of f-structures that are not shown in the example (including PRED features).

(38) *¡Qué desgraciad-as somos las mujer-es!* 'How unfortunate we women are!' (Ackema and Neeleman 2013)

$$\left[\begin{array}{ll} \text{“BE”} & \\ \text{TENSE} & pres \\ \text{AGR} & \left[\ \right] \\ \text{SUBJ} & \left[\begin{array}{ll} \text{“WOMAN”} & \\ \text{AGR} & \left[\begin{array}{ll} \text{PERSON} & 1 \\ \text{NUMBER} & pl \\ \text{GENDER} & f \end{array}\right] \end{array}\right] \end{array}\right]$$

The lexical entry for the verb does not specify the features of the SUBJ directly, as in traditional LFG accounts, but rather specifies the clause-level AGR feature, which is syntactically equated (shared) with the AGR feature of the subject:

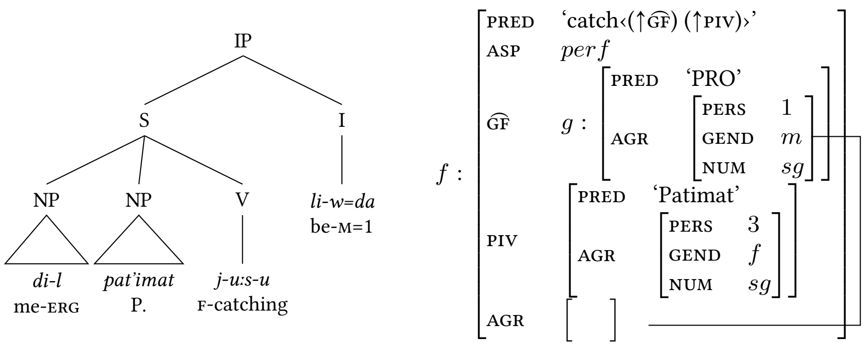
(39) *somos* $(\uparrow \text{PRED}) = $ 'be‹(↑SUBJ) (↑PREDLINK)›'
$(\uparrow \text{AGR PERS}) = 1$
$(\uparrow \text{AGR NUM}) = pl$
$(\uparrow \text{SUBJ AGR}) = (\uparrow \text{AGR})$

This allows us to separate the (lexical) specification of agreement features themselves from the (possibly syntactic or postsyntactic) identification of the agreement controller.

Using this advantage of feature sharing, Alsina and Vigo (2014) have used feature sharing to describe systems where the agreement controller is selected based on a competition of several candidates. Indeed, such systems pose a problem for the codescription approach, as the controller is not tied to a particular grammatical function. Using feature sharing allows us to relegate the job of choosing the controller from the LFG grammar itself to some external filter; in this case, Optimality Theory.

Within this system, absolutive ("gender") agreement in Dargwa would be treated in the conventional, "co-descriptive" way, because it is always tied to a particular grammatical function and thus there is no need for a separate mechanism. For simplicity, I will assume that Dargwa is syntactically ergative in the sense of Falk (2006), such that A/S is identified with $\widehat{\text{GF}}$, and P/S, with PIV; in a transitive clause, the latter is always structure shared with OBJ, which is not shown here for simplicity. Then gender agreement merely specifies the features of PIV. This decision does not have any bearing on the phenomena under discussion.

The hierarchical ("person") agreement, in contrast, involves the sharing of the clause-level AGR feature bundle with the AGR feature of either of the two core participants, as in the following f-structure:



The choice of the particular agreement controller is not part of the LFG grammar, but is relegated to Optimality Theory in a way similar to the earlier analysis proposed in Belyaev (2013). The motivation for OT as discussed in that earlier paper is that an alternative analysis would require complex disjunctions on the agreement markers in order to capture the patterns. This would make the description of cross-dialectal variation more difficult, while also lacking the intuitive appeal of the OT approach.

The use of feature sharing allows us to simplify the earlier version of the OT approach, which relied on an m-structure feature TH to carry the features of the person agreement controller because there is no appropriate position at f-structure. This role is now assigned to the clause-level AGR feature, eliminating the need for a special position at m-structure or any other level.

The following lexical entries can achieve the needed behaviour:[11]

(40) *b-i:q'-ul* V $(\uparrow \textsc{pred}) = \text{`do}\langle\widehat{\textsc{gf}}\ \textsc{piv}\rangle\text{'}$
$(\uparrow \textsc{piv agr gend}) =_c n$
$(\uparrow \textsc{piv agr num}) =_c sg$

(41) *li-w=da* I $(\uparrow \textsc{pred}) = \text{`be}\langle\textsc{comp}\rangle\text{'}$
$\{(\uparrow \textsc{agr}) = (\uparrow \widehat{\textsc{gf}}\ \textsc{agr}) \mid$
$(\uparrow \textsc{agr}) = (\uparrow \textsc{piv agr})\}$
$(\uparrow \textsc{agr pers}) =_c 1$
$(\uparrow \textsc{agr gend}) =_c m$
$(\uparrow \textsc{agr num}) =_c sg$

Gender agreement is thus done in the traditional way, through codescription, while person agreement is handled by AGR feature sharing. The choice between two alternative options for identifying the controller of hierarchical agreement is then relegated to Optimality Theory (Bresnan 2000; Lee 2004). The input is an incomplete f-structure without the clause-level AGR features. The constraints of Belyaev (2013) have to be redefined to refer to this AGR instead of the m-structure feature TH. A further constraint for preferring 3rd person topics to non-topics,[12]

11. The rules for S and IP are trivial: I assume that S and I are co-heads of IP; I contains the auxiliary or the finite verb, while the last constituent of S is either a non-finite verb form or a nonverbal predicate.

12. Another possibility is to literally introduce a "fourth person", as is done in traditional Algonquian linguistics, cf. e.g. Akmajian and Anderson (1970). This would make the OT constraints

AGR-3$_{\text{TOP}}$, should also be added:

AGR-2 $(f\ \text{AGR PERS}) =_c 2$

AGR-1 $(f\ \text{AGR PERS}) =_c 1$

AGR-3$_{\text{TOP}}$ $(f\ \text{AGR PERS}) =_c 3$

$((\text{AGR}(f\ \text{AGR}))_\sigma\ \text{DF}) =_c \text{TOPIC}$

AGR-$\widehat{\text{GF}}$ $(\widehat{\text{GF}}\ \text{AGR}(f\ \text{AGR}))$

AGR-PIV $(\text{PIV AGR}(f\ \text{AGR}))$

These constraints can be illustrated by the f-structure at the previous page. Since they have to be evaluated for clausal f-structures, f in the constraints is identified with f in the sample f-structure. The first two constraints, AGR-1 and AGR-2, are trivial in that they simply specify the person feature of the agreement controller.

The third constraint, AGR-3$_{\text{TOP}}$, is more complicated. It consists of two statements. The first statement defines the controller as 3rd person. The second statement is meant to constrian the information structure function of the agreement controller. This is done by using the inside-out Functional Uncertainty equation $(\text{AGR}(f\ \text{AGR}))$. Since $(f\ \text{AGR})$ is structure shared with the AGR feature of the agreement controller, this equation can in principle lead to two different f-structures: f (trivially returning back to the starting point) and g, i.e. the subject f-structure. The feature DF of the resulting f-structure's projected s-structure is then constraint. On the assumption that only arguments and adjuncts, but not finite clauses, have discourse functions, this means that the f-structure defined by the equation $(\text{AGR}(f\ \text{AGR}))$ can only be g.Effectively, what this equation does can be restated in the following way: "the f-structure whose AGR feature is shared with the clausal AGR feature must be a topic".

The use of inside-out Functional Uncertainty in the fourth and fifth constraints is similar, but these are existential equations: they check whether the argument with which agreement is shared is a $\widehat{\text{GF}}$ or a PIV.

The rankings for individual varieties mostly stay the same, with the exception of AGR-3$_{\text{TOP}}$. This constraint dominates AGR-PIV and AGR-$\widehat{\text{GF}}$ in languages like Ashti and Tanti, and at the end of the hierarchy for those languages where there is no competition between 3rd person controllers. The ranking for Ashti is thus AGR-1 $\vee$ AGR-2 $\succ$ AGR-3$_{\text{TOP}}$ $\succ$ AGR-PIV $\succ$ AGR-$\widehat{\text{GF}}$ (on constraint disjunction see Crowhurst and Hewitt 1997).

The following tableaux illustrate how these constraints work to ensure correct agreement patterns (I am using sentences as shorthand for underspecified f-structures in the input field):

- A = 1sg.m, P = 3sg$_{\text{TOP}}$.f

di-l pat'imat j-us-u …	AGR-1 $\vee$ AGR-2	AGR-3$_{\text{TOP}}$	AGR-PIV	AGR-$\widehat{\text{GF}}$
☞ li-w=da (A)		*	*	
li-j (P)	*!			*

simpler by relegating the choice between "third" and "fourth" person to the lexicon.

- A = 1sg.m, P = 2sg.f

di-l u j-us-u …	Agr-1 $\vee$ Agr-2	Agr-3$_{\text{TOP}}$	Agr-piv	Agr-$\widehat{\text{GF}}$
li-w=da (A)		*	*!	
☞ li-j=di (P)		*		*

- A = 3sg$_{\text{TOP}}$.m, P = 3sg.m

rasul-li pat'imat j-us-u …	Agr-1 $\vee$ Agr-2	Agr-3$_{\text{TOP}}$	Agr-piv	Agr-$\widehat{\text{GF}}$
☞ li-w (A)	*		*	
li-j (P)	*	*!		*

6 Conclusions

In this paper, I have analyzed exceptional patterns of gender agreement in Ashti Dargwa. While it has been proposed to analyze similar facts in other varieties through a complex two-tiered clause structure with anaphoric backward control, my data do not support this hypothesis, and a simpler solution seems preferable. While exceptional gender agreement on adverbs seems to be a completely independent phenomenon, gender agreement on the auxiliary in Ashti merely reflects the gender feature of the person agreement controller, and thus requires only a modification of the hierarchical account by splitting the 3rd person based on topicality, such that the person hierarchy in Ashti is $1, 2 > 3 > 3'$. The OT approach of Belyaev (2013) can then be applied with only minor modifications. It can also be simplified by using agreement sharing, dispensing with a separate position at m-structure proposed in that earlier paper.

Apart from being conceptually simpler, this analysis also has the advantage of being typologically more well-motivated. A split of the 3rd person based on topicality is well-known as obviation, and having the obviative vs. proximate distinction is typical for languages having hierarchical agreement or alignment systems.

However, the analysis is still preliminary to the extent that we do not have enough information on the information structure conditions on agreeement, and not enough data on different Dargwa varieties is available.

References

Ackema, Peter, and Ad Neeleman. 2013. "Subset controllers in agreement relations." *Morphology* 23 (2): 291–323.

Aissen, Judith. 1997. "On the syntax of obviation." *Language* 73 (4): 705–750.

Akmajian, Adrian, and Steven Anderson. 1970. "On the rise of the fourth person in Navajo, or Navajo made harder." *International Journal of American Linguistics* 36:1–8.

Alsina, Alex, and Eugenio M. Vigo. 2014. "Copular inversion and non-subject agreement." In *Proceedings of the LFG14 Conference,* ed. by Miriam Butt and Tracy Holloway King, 5–25. Stanford, CA: CSLI Publications.

Belyaev, Oleg. 2013. "Optimal agreement at m-structure." In *Proceedings of the LFG13 Conference,* ed. by Miriam Butt and Tracy Holloway King. Stanford: CSLI Publications.

Bresnan, Joan. 2000. "Optimal syntax." In *Optimality Theory: phonology, syntax and acquisition,* ed. by Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer, 334–385. Oxford: Oxford University Press.

Corbett, Greville G. 2013. "The unique challenge of the Archi paradigm." *Proceedings of BLS 37:* 52–67.

Crowhurst, Megan J., and Mark Hewitt. 1997. "Boolean operations and constraint interaction in Optimality Theory." ROA 229.

Falk, Yehuda N. 2006. *Subjects and Universal Grammar: an explanatory theory.* Cambridge: Cambridge University Press.

Haug, Dag, and Tatiana Nikitina. 2015. "Feature sharing in agreement." *Natural Language and Linguistic Theory.* Forthcoming.

Lee, Hanjung. 2004. "Minimality in a lexicalist Optimality Theory." In *Minimality effects in syntax,* ed. by Arthur Stepanov, Gisbert Fanselow, and Ralf Vogel, 241–288. Berlin: Mouton de Gruyter.

Li, Charles N., and Sandra A. Thompson. 1977. "A mechanism for the development of copula morphemes." In *Mechanisms of syntactic change,* ed. by Charles N. Li, 419–444. Austin, TX: University of Texas Press.

Magometov, Aleksandr A. 1963. *Kubačinskij jazyk. (Issledovanie i teksty).* [Kubachi: Grammar and texts]. Tbilisi: Mecniereba.

Nordlinger, Rachel. 1998. *Constructive case: evidence from Australian languages.* Dissertations in Linguistics. Stanford: CSLI Publications.

Šamov, Ibragim A. 1994. *Pritči o mulle Nasreddine* [Stories about Mulla Nasruddin]. Makhachkala: Izdatel'stvo "Jupiter".

Sumbatova, Nina. 2011. "Person hierarchies and the problem of person marker origin in Dargwa: facts and diachronic problems." In *Tense, aspect, modality and finiteness in East Caucasian languages,* ed. by Gilles Authier and Timur Maisak, 131–160. Diversitas Linguarum 30. Bochum: Universitätsverlag Dr. N. Brockmeyer.

Sumbatova, Nina R. 2014. "V poiskax podležaščego: kontrol' klassnogo soglasovanija i priznaki grammatičeskogo prioriteta v darginskom jazyke" [Looking for the subject: Gender agreement and grammatical priority in Dargwa]. In *Jazyk. Konstanty. Peremennye. Pamjati Aleksandra Evgen'eviča Kibrika,* [Language. Constants. Variables. In memoriam Alexander E. Kibrik]. Saint Petersburg: Aletheia.

Sumbatova, Nina R., and Yury A. Lander. 2014. *Darginskij govor selenija Tanty: grammatičeskij očerk, voprosy sintaksisa* [The Dargwa dialect of Tanti: A grammatical sketch, syntactic issues]. Moscow: Jazyki slavjanskix kul'tur.

van den Berg, Helma. 2001. *Dargi folktales: oral stories from the Caucasus with an introduction to Dargi grammar.* Leiden: Research School of Asian, African / Amerindian Studies.

On pseudo-non-finite clauses in Welsh

Olivier Bonami
Université Paris Diderot

Robert D. Borsley
University of Essex

Maggie Tallerman
Newcastle University

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 104–124

Keywords: Welsh, finiteness, syntax-morphology relation, periphrasis.

Bonami, Olivier, Robert D. Borsley & Maggie Tallerman. 2016. On pseudo-non-finite clauses in Welsh. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 104–124. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.6.

Abstract

Languages differ in how they employ finite and non-finite clauses. Welsh finite and non-finite clauses have a similar distribution to their counterparts in English. However, it doesn't look like this because Welsh has certain finite clauses which look rather like non-finite clauses. We examine two types of pseudo-non-finite clauses: finite *bod* clauses and finite *i* clauses. We argue that both cases are instances of a mismatch between syntax and morphology, while the latter only involves periphrasis. We provide an HPSG analysis capturing similarities and differences between these two constructions and canonical finite and nonfinite clauses.

1 Introduction

Languages[1] differ in how they employ finite and non-finite clauses. Welsh finite and non-finite clauses have a similar distribution to their counterparts in English. However, it doesn't look like this because Welsh has certain finite clauses which look rather like non-finite clauses (Tallerman 1998, Borsley et al. 2007, chap. 3). One type just involves the verb *bod* 'be', while the other involves all verbs. The two types of clause differ in important ways and pose rather different challenges for grammatical theory.

2 Data

We will generally focus on spoken varieties of Welsh, commenting on distinctions between such varieties and Literary Welsh, which is not a spoken variety, where necessary. Although there are significant differences between northern and southern dialects, we will generally abstract away from these in what follows.

2.1 Finite verbs in Welsh

Apart from *bod* 'be', Welsh verbs have three synthetic subparadigms: future, past, and conditional. *Bod* has two additional subparadigms, present and imperfect. In Table 1 we provide a very partial illustration.

In various situations these forms may undergo one or other of the mutation processes that affect initial consonants in Welsh. Thus, we may have *gerddith* or *fydd* as a result of soft mutation or *cherddith* as a result of aspirate mutation.[2]

[1]We are grateful to David Willis, Ian Roberts and Nigel Vincent, various anonymous reviewers, and the audience at the HeadLex 2016 conference, for a number of helpful comments. The first author's work was partially supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-LABX-0083).

[2]Aspirate mutation only affects forms beginning with a voiceless plosive. Hence *bod* is not affected. There are many examples of mutation below. We only comment on them when they are relevant to our analyses.

	Cerdded 'walk'	*Bod* 'be'
Future	*cerddith*	*bydd*
Past	*cerddodd*	*buodd*
Conditional	*cerddai*	*byddai*
Present	—	*mae*
Imperfect	—	*roedd*

Table 1: Third person forms of *cerdded* 'walk' and *bod* 'be'.

Welsh also expresses various aspectual meanings with a form of *bod*, an aspectual particle, and a non-finite verb, but this is not particularly important in the present context.

It also seems necessary in Welsh to distinguish between positive and negative forms of finite verbs and between main clause and complement forms. Positive main clause verbs may be preceded by the particle *mi* or *fe* and and when they are, they show soft mutation.[3] (1) is a typical example:

(1) Mi/Fe gerddith Emrys i 'r dre.
PRT walk.FUT.3SG Emrys to the town
'Emrys will walk to the town.'

Negative main clause verbs are not preceded by *mi* or *fe*, but they generally show soft mutation or in the case of some verbs aspirate mutation, and they generally co-occur with the negative post-subject adverb *ddim*:[4]

(2) Gerddith/Cherddith Emrys ddim i 'r dre.
NEG.walk.FUT.3SG Emrys NEG to the town
'Emrys will not walk to the town.'

Positive complement clause verbs are not preceded by the particles *mi* and *fe* and do not show soft mutation:

(3) Dywedodd Megan [cerddith Emrys i 'r dre].
say.PAST.3SG Megan walk.FUT.3SG Emrys to the town
'Megan said Emrys will walk to the town.'

In negative complement clauses verbs generally co-occur with *ddim* and show essentially the same mutation as in main clauses. They may also be preceded by the particle *na*.

(4) Dywedodd Megan [na cherddith Emrys ddim i 'r dre].
say.PAST.3SG Megan NEG walk.FUT.3SG Emrys NEG to the town
'Megan said Emrys will not walk to the town.'

[3] If the particle is not present, the verbs may or may not show mutation.
[4] For more on the form of negative verbs, see (Borsley & Jones, 2005, chap. 3).

The present tense of *bod* is generally not preceded by *mi* or *fe* in positive main clauses, and negative forms show suppletion.

(5) Mae Emrys yn yr ardd.
be.PRES.3SG Emrys in the garden
'Emrys is in the garden.'

(6) Dydy Emrys ddim yn yr ardd.
NEG.be.PRES.3SG Emrys NEG in the garden
'Emrys is not in the garden.'

We will see in the next section that *bod* also has a distinctive form in complement clauses.

2.2 *Bod*-clauses

The present and imperfect forms of *bod* are fine in main clauses and in relative clauses or other unbounded dependency clauses, e.g. *wh*-interrogatives, but the present forms and for some speakers the imperfect forms too are ungrammatical in complement clauses:

(7) Mae Elen yn darllen y llyfr.
be.PRES.3SG Elen PROG read.INF the book
'Elen is reading the book.'

(8) Roedd Elen yn darllen y llyfr.
be.IMPF.3SG Elen PROG read.INF the book
'Elen was reading the book.'

(9) y llyfr [mae / roedd Elen yn ei ddarllen]
the book be.PRES.3SG be.IMPF.3SG Elen PROG 3SGM read.INF
'the book that Elen is/was reading'

(10) Pa lyfr [mae / roedd Elen yn ei ddarllen]?
which book be.PRES.3SG be.IMPF.3SG Elen PROG 3SGM read.INF
'Which book is/was Elen reading?'

(11) * Mae Aled yn credu [mae Elen yn
be.PRES.3SG Aled PROG believe.INF be.PRES.3SG Elen PROG
darllen y llyfr].
read.INF the book
'Aled believes that Elen is reading the book.'

(12) % Mae Aled yn credu [roedd Elen yn
be.PRES.3SG Aled PROG believe.INF be.IMPF.3SG Elen PROG
darllen y llyfr].
read.INF the book
'Aled believes that Elen was reading the book.'

Instead of present forms of *bod* and for some speakers imperfect forms as well, what looks like the non-finite form *bod* appears:

(13) Mae Aled yn credu [bod Elen yn darllen y llyfr].
be.PRES.3SG Aled PROG believe.INF be.INF Elen PROG read.INF the book
'Aled believes that Elen is/was reading the book.'

We will call the complement clause in such examples a *bod*-clause.

If the subject of a *bod*-clause is pronominal, *bod* shows agreement in the form of a preceding clitic:

(14) Mae Aled yn credu [ei bod hi 'n darllen y llyfr].
be.PRES.3SG Aled PROG believe.INF 3SGF be.INF she PROG read.INF the book
'Aled believes that she is/was reading the book.'

Ordinary non-finite verbs also show agreement in the form of a clitic with a following pronoun, but the pronoun can only be an object because they never have a following subject. Here are some typical examples:

(15) a. Dylai Aled weld Elen.
ought Aled see.INF Elen
'Aled ought to see Elen.'

b. Dylai Aled ei gweld hi.
ought Aled 3SGF see.INF she
'Aled ought to see her.'

We will return to this contrast in section 2.4.

There is one situation in which present and imperfect forms of *bod* may appear in complement clauses. This is in complement clauses affected by an unbounded dependency such as the following (Willis, 2000, 2011; Borsley, 2013):[5]

(16) Beth mae Aled yn credu [mae Elen yn ei ddarllen]?
what be.PRES.3SG Aled PROG believe.INF be.PRES.3SG Elen PROG 3SGM read.INF
'What does Aled believe that Elen is reading?'

(17) Beth mae Aled yn credu [roedd Elen yn ei ddarllen]?
what be.PRES.3SG Aled PROG believe.INF be.IMPF.3SG Elen PROG 3SGM read.INF
'What does Aled believe that Elen was reading?'

[5]Some speakers have *bod* in such sentences, but others prefer present and imperfect forms.

We will show below that *bod*-clauses are a type of finite clause in spite of the form of the verb. We will call them *pseudo-non-finite clauses*.

2.3 Finite *i*-clauses

A second type of pseudo-non-finite clause involves all Welsh verbs. Past tense forms of Welsh verbs appear in main clauses and in the negative complement clauses to epistemic and declarative verbs. However, in some varieties, and especially Literary Welsh, they are not used in positive complement clauses to such verbs.

(18) Aeth Mair adre'.
go.PAST.3SG Mair home
'Mair went home.'

(19) Meddyliodd Aled [nad aeth Mair ddim adre'].
think.PAST.3SG Aled NEG go.PAST.3SG Mair NEG home
'Aled thought that Mair had not gone home.'

(20) % Meddyliodd Aled [aeth Mair adre'].
think.PAST.3SG Aled go.PAST.3SG Mair home
'Aled thought that Mair had gone home.'

Instead of a positive complement clause with a past tense verb, what looks rather like an English *for-to* clause appears:[6]

(21) Meddyliodd Aled [i Mair fynd adre'].
think.PAST.3SG Aled to Mair go.INF home
'Aled thought that Mair had gone home.'

Here we have a clause introduced by what looks like the preposition *i* 'to', for' (hence the gloss). As we will see shortly, there is evidence of various kinds that these clauses are finite, and we will call them finite *i*-clauses. In the complements to other classes of matrix verb we find i-clauses that are clearly non-finite, as in the following example.

(22) Disgwyliodd Aled [i Elen ddarllen y llyfr].
expect.PAST.3SG Aled to Elen read.INF the book
'Aled expected Elen to read the book.'

[6] In Colloquial Welsh, some varieties allow past tense forms in affirmative complement clauses. The *i*-clause construction is often considered rather formal, and may be avoided on these grounds. Some speakers may use a perfect clause, as in (i).

(i) Meddyliodd Aled [bod Mair wedi mynd adre'].
think.PAST.3SG Aled be.INF Mair PERF go.INF home
'Aled thought that Mair had gone home.'

The word *i* occuring in finite and nonfinite *i*-clauses has the same irregular inflectional paradigm as the preposition *i*, indicated in Table 2. Just like the preposition *i*, it agrees with a following pronoun in gender, person and number, but the morphology unexpectedly neutralizes number distinctions in the first and second person.

ABSOLUTE		*dan*	
		SG	PL
AGREEING	1	*dana*	*danon*
	2	*danat*	*danoch*
	3.M	*dano*	*danyn*
	3.F	*dani*	

dan 'under'

ABSOLUTE		*i*	
		SG	PL
AGREEING	1	*i*	*i*
	2	*i*	*i*
	3.M	*iddo*	*iddyn*
	3.F	*iddi*	

i 'to'

Table 2: Inflectional paradigm of two prepositions.

Non-finite *i*-clauses are negated by the negative verb *peidio* (which only has non-finite and imperative forms and is mutated here):

(23) Disgwyliodd Aled [i Elen beidio â darllen y llyfr].
expect.PAST.3SG Aled to Elen NEG with read.INF the book
'Aled expected Elen not to read the book.'

This is not possible in a finite *i*-clause; finite i-clauses are always positive.

(24) * Meddyliodd Aled [i Mair beidio â mynd adre'].
think.PAST.3SG Aled to Mair NEG with go.INF home
'Aled thought that Mair had not gone home.'

Apart from this finite and non-finite *i*-clauses seem to have the same internal structure.

2.4 Evidence that *bod*-clauses and finite *i*-clauses are really finite.

Bod-clauses and finite *i*-clauses are used in contexts where a finite clause is expected; the corresponding finite clauses are unexpectedly ungrammatical. In that sense, they fill a gap in a paradigm of finite constructions. They can also coordinate with ordinary finite clauses:

(25) Dywedodd Aled [fod Mair wedi mynd yn barod] a
say.PAST.3SG Aled be.INF Mair PERF go.INF PRED ready and
[byddai Gwen yn mynd yn fuan].
be.COND.3SG Gwen PROG go.INF PRED soon
'Aled said that Mair had gone already and that Gwen would be going soon.'

(26) Meddyliodd Aled [i Alys fynd adre'] a [byddai Mair
think.PAST.3SG Aled to Alys go.INF home and be.COND.3SG Mair
yn mynd hefyd].
PROG go.INF too
'Aled thought that Alys had gone home and that Mair would be going too.'

Bod-clauses and finite *i*-clauses do not allow a reflexive subject with an antecedent in the main clause:

(27) * Dywedodd Aled [(ei) fod ei hun wedi gadael].
say.PAST.3SG Aled 3SGM be.INF 3SGM REFL PERF leave.INF
*'Aled said that himself had left.'

(28) * Dywedodd Aled [iddo 'i hun fynd].
say.PAST.3SG Aled to.3SGM 3SGM REFL go.INF
*'Aled said that himself had gone.'

In this, they are like finite clauses and unlike uncontroversial non-finite causes:

(29) * Dywedodd Aled [eith ei hun].
say.PAST.3SG Aled go.FUT.3SG 3SGM REFL
*'Aled said that himself will go.'

(30) Dymunai Aled [iddo 'i hun ddarllen y llyfr].
wish.COND.3SG Aled to.3SGM 3SGM REFL read.INF the book
'Aled would want himself to read the book.'

Bod-clauses also resemble finite clauses in their internal structure. As noted above, they show verb-subject order, which is like finite clauses and unlike non-finite clauses. They also show negation with the post-subject negative adverb *ddim*:

(31) Mae Aled yn dweud [bod Mair ddim yn barod].
be.PRES.3SG Aled PROG say.INF be.INF Mair NEG PRED ready
'Aled says that Mair isn't ready.'

In addition, they allow expletive *yna* 'there', which appears in finite clauses but not in non-finite clauses:

(32) Mae Gwyn yn meddwl [bod yna ddafad yn yr
be.PRES.3SG Gwyn PROG think.INF be.INF there sheep in the
ardd].
garden
'Gwyn thinks that there is a sheep in the garden.'

(33) Mae yna ddafad yn yr ardd.
be.PRES.3SG there sheep in the garden
'There is a sheep in the garden.'

(34) * Disgwyliodd Gwyn [i yna fod dafad yn yr ardd].
expect.PAST.3SG Gwyn to there be.INF sheep in the garden
'Gwyn expected there to be a sheep in the garden.'

In contrast, as noted above, finite *i*-clauses seem to have essentially the same internal structure as non-finite *i*-clauses, differing from non-finite *i*-clauses only in that they cannot be negated.

3 Analyses

3.1 Preliminaries

Although both contain what looks like a non-finite verb, both *bod*-clauses and finite *i*-clauses seem to be finite and to fill a gap in a paradigm of otherwise finite constructions. In the following pages we will develop analyses which capture this fact but also capture the difference between the two clause types, the fact that *bod*-clauses have a finite internal structure, while finite *i*-clauses have essentially the same structure as non-finite *i*-clauses.

Our analyses rely on three mechanisms familiar from recent work in realisational morphology. First, we assume that the morphology-syntax interface involves a distinction between two sets of morphosyntactic features (Sadler & Spencer, 2001; Stump, 2006; Bonami, 2015) corresponding to the syntactic and the morphological view of the inflectional paradigm. Following Bonami (2015), we implement this distinction by distinguishing between the value of HEAD, which provides a syntactic view of the paradigm, and the value of INFL, a feature carried only by words, which serves as the input to inflectional morphology.

Second, the morphology-syntax interface is governed by a principle of morphosyntactic blocking (Andrews, 1990; Koenig, 1999). Specifically, we assume with Stump (2006, 2015) that a paradigm is licensed by a set of conditional statements such that if the antecedents of two statements stand in a subsumption relation, only the most specific statement may apply. In HPSG terms, we assume a distinguished set of conditional interface statements whose antecedent restricts attention to a particular SYNSEM value and whose consequent specifies the relationship between the HEAD and INFL features; in this paper we highlight the special status of these statements by typesetting them with a grey background. We then define a closure operation of *Pāṇinian strengthening* over these statements that makes them mutually incompatible.[7] This is a variant of the implementation of Pāṇini's Principle by Bonami & Crysmann (2013) and Crysmann & Bonami (2016) generalized to the syntax-morphology interface.

[7] Pāṇinian strengthening may be defined as follows. Let S be the set of interface statements, and $\sigma \rightarrow \tau$ be one particular such statement. Let $\{\sigma_1 \rightarrow \tau_1, \ldots, \sigma_n \rightarrow \tau_n\} \subset S$ be the set of all statements whose antecedent is strictly more specific than σ, i.e., for each i, $\sigma_i \models \sigma$ and $\sigma \not\models \sigma_i$. Replace $\sigma \rightarrow \tau$ by $(\sigma \wedge \neg\sigma_1 \wedge \cdots \wedge \neg\sigma_1) \rightarrow \tau$.

Third, we follow Spencer (2013) in assuming that the forms of a lexeme need not all belong to the same part of speech; this allows us to state that in some circumstances an otherwise verbal lexeme is realised by a complementizer.[8]

3.2 *Bod*-clauses

Bod-clauses seem fairly straightforward. They are essentially syntactically finite but morphologically non-finite. Present forms of *bod* and for some speakers imperfect forms too have non-finite morphology under certain circumstances, namely in a complement clause not affected by an unbounded dependency. Given a distinction between syntactic and morphological finiteness it is not difficult to accommodate these clauses.

We assume that verbs which are syntactically finite are normally morphologically finite as a result of the following constraint which applies to all words and captures the idea that in the canonical situation, HEAD and INFL information match:

(35) $$[\ ] \rightarrow \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{INFL} & \boxed{1} \end{bmatrix}$$

In the case of *bod*-clauses this will be overridden by a more specific constraint. The other machinery that we need here is fairly standard HPSG machinery. Following much earlier work, we assume a feature LID, whose value is unique to each distinct lexeme: a lexeme, the words that realise it, and the phrases headed by such words all have the same unique LID value. To handle tense we assume a TMA (TENSE-MOOD-ASPECT) feature with the system of values in Figure 1.

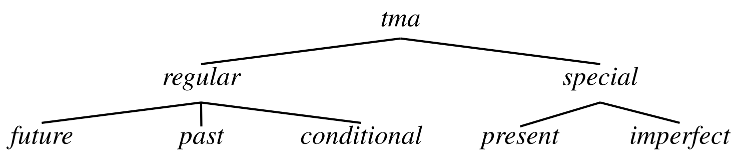


Figure 1: Values for the feature *tma*.

This allows us to say that ordinary verbs only have [TMA *regular*] finite forms, and will also allow us to accommodate speakers who have *bod* instead of both present and imperfect forms. We also need to distinguish complement clauses on the one hand from main clauses and unbounded dependency clauses on the other. We will do this with a feature STATUS with values *main*, *udc*, *subord*(*inate*); this is a generalisation of the binary ROOT feature. Finally we need to distinguish complement clauses affected by an unbounded dependency and complement clauses in which there is no unbounded dependency. Remember that in HPSG, unbounded dependencies involve the SLASH feature, whose value is the empty set when there is no

[8]There is an implicit precedent for this idea in Sag (1997), who proposes that verbs and complementizers belong to a common 'verbal' part of speech, and that complementizers carry the VFORM feature.

dependency but non-empty when there is a dependency of some kind. Generally when an argument of some head has a non-empty SLASH value, the head has the same value, and so does its mother. Thus, we typically have structures of the form indicated in Figure 2. In the present context this means that *bod* and the clause it heads have a non-empty SLASH value when affected by an unbounded dependency and are [SLASH { }] when there is no dependency.

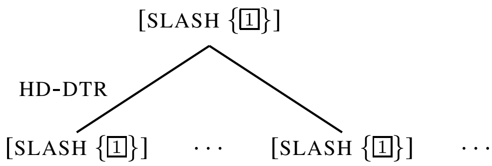


Figure 2: A typical situation of SLASH percolation.

We can now provide analyses for a range of examples. We assume following Borsley (1989) that post-verbal subjects are realizations of the first element of the COMPS list. Given this assumption, and those spelled out above, (7) will have the analysis in Figure 3.[9]

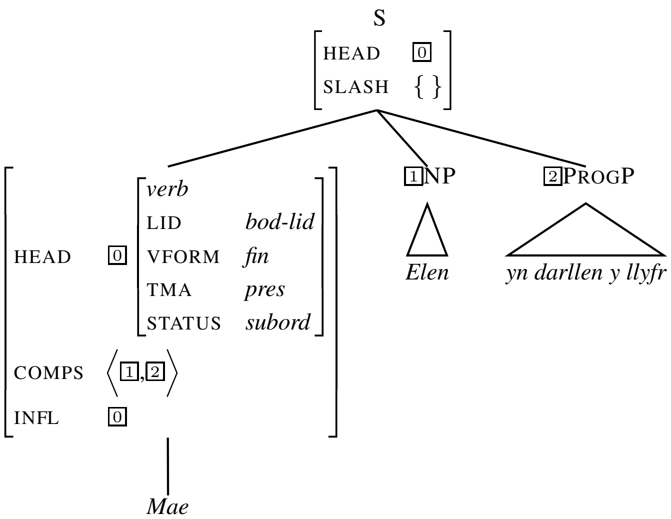


Figure 3: Analysis for example (7).

The crucial feature of this analysis is that HEAD and INFL have the same value. Thus, the verb is both syntactically and morphologically finite. For the complement clause in (16), which is affected by an unbounded dependency, we will have the analysis in Figure 4.

[9]Both NP and ProgP here will be [SLASH { }], but we omit this in order to keep the tree as simple as possible. We will also omit [SLASH { }] in later trees when there is no need to highlight.

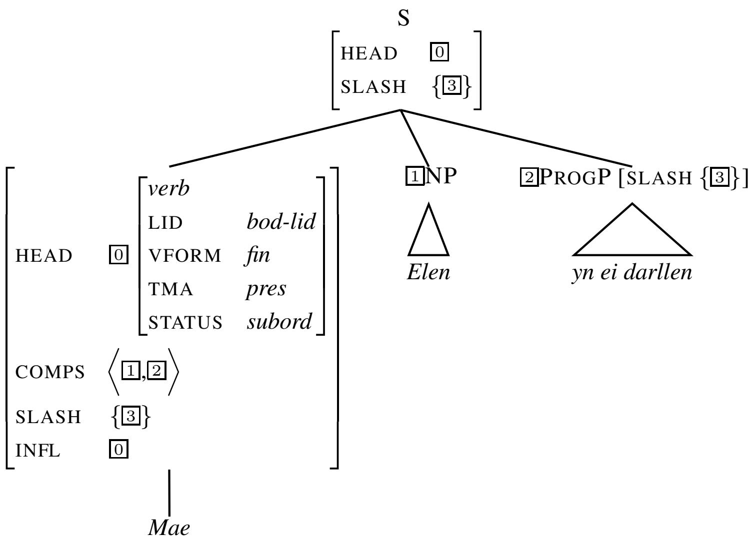


Figure 4: Analysis for example (16).

Here, the STATUS feature makes it clear that this is a complement clause, and the non-empty values for SLASH indicate that it is affected by an unbounded dependency. Again HEAD and INFL have the same value. Finally, we have the analysis in Figure 5 for the *bod*-clause in (13).

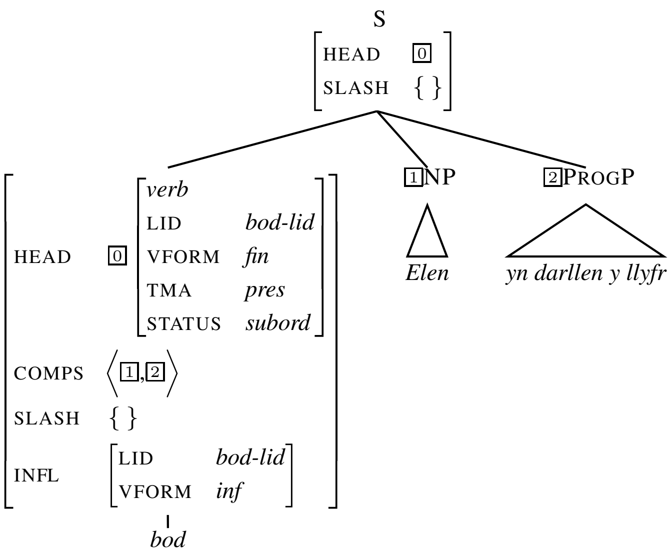


Figure 5: Analysis for example (13).

Here, HEAD and INFL have different values. We attribute this to the interface statement in (36):

$$(36)\quad \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{LID} & \textit{bod-lid} \\ \text{VFORM} & \textit{fin} \\ \text{STATUS} & \textit{subord} \\ \text{TMA} & \textit{pres} \end{bmatrix} \\ \text{SLASH} & \{\,\} \end{bmatrix} \rightarrow \begin{bmatrix} \text{INFL} & \begin{bmatrix} \text{LID} & \textit{bod-lid} \\ \text{VFORM} & \textit{inf} \end{bmatrix} \end{bmatrix}$$

In the absence of (36), constraint (35) would predict the use of *mae* as a head verb. However, by Pāṇinian strengthening, (36) overrides (35) and ensures that *bod* appears instead of ordinary present tense forms in a complement clause not affected by an unbounded dependency.[10] This is a case of morphosyntactic blocking. For speakers who also have *bod* instead of imperfects, we will have a constraint with [TMA *special*].[11]

We noted earlier that *bod* shows agreement in the form of a preceding clitic with a following pronoun. We assume that clitics are specifiers and we attribute their appearance to the constraint in (37a).[12] This applies to elements which are morphologically non-finite, hence both to ordinary non-finite verbs and *bod*. Ordinary finite verbs show agreement in the form of a suffix and do not show agreement. We attribute this to the constraint in (37b), which also ensures that nonfinite verbs do not combine with a clitic if there is no agreement trigger.

$$(37)\quad \text{a.}\quad \begin{bmatrix} \text{INFL} & \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \textit{inf} \\ \text{AGR} & \textit{index} \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \text{INFL} & \begin{bmatrix} \text{AGR} & \boxed{1} \end{bmatrix} \\ \text{SPR} & \left\langle \begin{bmatrix} \textit{agr-clitic} & \\ \text{IND} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$

$$\text{b.}\quad \begin{bmatrix} \text{INFL} & \begin{bmatrix} \textit{verb} \\ \text{VFORM}\,\textit{fin} \vee \text{AGR}\ \textit{none} \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \text{SPR} & \langle\,\rangle \end{bmatrix}$$

This licenses the analysis in Figure 6 for the subordinate clause in (14).

[10]More precisely, assuming for simplicity that (35) and (36) are the two only interface statements in Welsh grammar, Pāṇinian strengthening has the effect of turning (35) into the following constraint:

$$\text{(i)}\quad \neg \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{LID} & \textit{bod-lid} \\ \text{VFORM} & \textit{fin} \\ \text{STATUS} & \textit{subord} \\ \text{TMA} & \textit{pres} \end{bmatrix} \\ \text{SLASH} & \{\,\} \end{bmatrix} \rightarrow \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{INFL} & \boxed{1} \end{bmatrix}$$

[11]For speakers who have *bod* in complement clauses affected by an unbounded dependency, we can propose a version of (36) without the [SLASH { }] stipulation.

[12]This raises the question: what ensures that nouns and other heads have appropriate agreement features? Borsley (2009) proposes that this is a consequence of a constraint on order domains, but it would be possible to attribute it to a constraint on constituent structures.

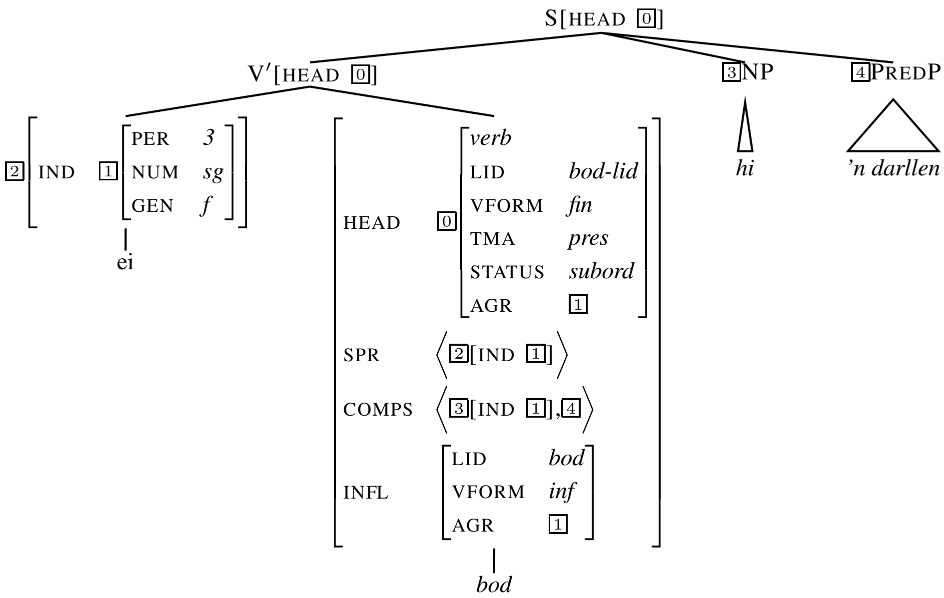


Figure 6: Analysis for example (14).

3.3 Finite *i*-clauses

Finite *i*-clauses are rather more challenging than *bod*-clauses. A satisfactory analysis needs to (a) capture the idea that they involve an unusual realization of a finite verb, and (b) treat them as similarly as possible to non-finite *i*-clauses. They appear to be a case of periphrasis, where instead of the expected inflected form of a lexeme, some other form appears together with some other element. The situation is broadly similar to Latin passive perfects, e.g. (38), where a participle and a form of the copula appears, albeit passive and perfect are otherwise realised by a synthetic form of the verb:

(38) monitus sum
advise.PASS-PART.MASC.SG be.PRES.1.SG
'I have been advised'

Following Bonami (2015), the element which looks like the lexeme (the verb in (21) and the participle in (38)) can be called the 'main' element and the other element (*i* in (21) and the copula in (38)) can be called the 'ancillary' element.

Two approaches to periphrasis within HPSG have been explored by Bonami and colleagues. On one, developed in Bonami (2015), the main element (the verb in the case of finite *i*-clauses) is the real realization of the lexeme. On the other, developed in Bonami & Webelhuth (2013) and Bonami & Samvelian (2015), the real realization of the lexeme is the ancillary element (*i* in the present case). If we adopted the first approach, we would have to say that the verb in a finite *i*-clause counts as the realisation of a finite cell in the paradigm, in fact a past tense cell, in spite of the fact that it looks like a nonfinite verb both in terms of its morphology and its syntax. In other words, we would need to recognize a third, 'paradigmatic' notion of finiteness in addition to the more directly observable notions of morphological and syntactic finiteness. This seems an undesirable position. Therefore, we will develop a new version of the second approach.[13]

Following Borsley (1999, 2009), we assume that non-finite *i*-clauses are CPs with a ternary branching analysis parallel to Sag's (1997) analysis of English *for-to* clauses. Thus, the complement clause in (22) has the schematic analysis in Figure 7. Note that the NP complement is identified as the subject of the VP complement. We will develop a fuller analysis shortly. We assume that finite *i*-clauses have the same basic structure, and hence we have the following schematic analysis for the complement clause in (21). Again the NP complement is identified as the subject of the VP complement. Again we will develop a more detailed analysis shortly.

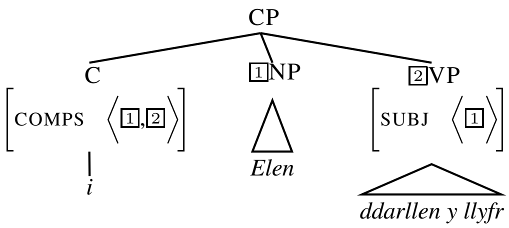


Figure 7: Schematic analysis for example (22).

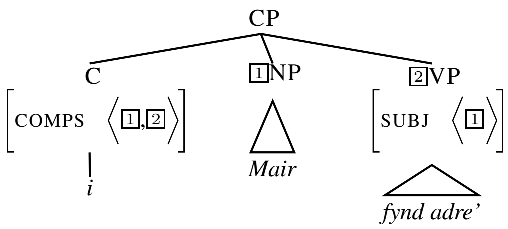


Figure 8: Schematic analysis for example (21).

Although they have slightly different syntactic properties, non-finite and finite *i* have the same basic complement selection properties. They also have the same morphology. As indicated in Table 2, like the preposition *i* and unlike most prepositions, non-finite and finite *i* only show agreement with a third person pronoun. This is captured by assuming that the INFL|LID value of both non-finite and finite *i* has a common supertype with that of the preposition *i*, as indicated in Figure 9. Rules of inflection happen to be formulated in terms of this supertype.

One important difference we assume between finite and nonfinite *i* is in terms of their HEAD|LID values, that is, the lexical identity information they project into syntax. Although non-finite *i* is just an ordinary word with its own lexical identity

[13] Although we focus here on one construction in Welsh, the approach outlined here is intended to be general and an improvement on both Bonami (2015) and Bonami & Samvelian (2015).

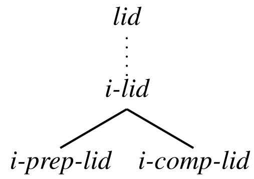


Figure 9: Partial hierarchy of LID values.

which it projects to phrase level, finite *i* has a dual lexical status: from the point of view of inflection it counts as a form of *i*, and hence is specified as INFL|LID *i*; but from the point of view of syntax it counts as a realisation of the main verb, and hence has the same value for the HEAD|LID feature as its VP complement.

It is often assumed that all realizations of a lexeme must be of the same part-of-speech. However, Spencer (2013) argues at length against this assumption. Hence, we see nothing wrong in treating complementizers as forms of a verb.

Is there any way to maintain the assumption that all realizations of a lexeme are of the same part-of-speech? Someone who favours this assumption might propose that both non-finite and finite *i* are not complementizers but verbs. A verbal analysis of finite *i* faces no obvious problems. However, a verbal analysis of non-finite *i* is implausible. If it were a verb, it would be unlike every other non-finite verb in taking a following subject. We think, then, that non-finite *i* must be a complementizer. As far as we can see, the only way to maintain the assumption that all realizations of a lexeme are of the same part-of-speech would be to propose that non-finite *i* is a complementizer with its own LID value while finite *i* is a verb sharing a LID value with its VP complement. However, as noted above, we want to treat finite *i*-clauses as similarly as possible to non-finite *i*-clauses. We think, then, that it is preferable to assume that both non-finite and finite *i* are complementizers.

With the assumptions just introduced, we can propose more detailed analysis in Figure 10 for the complement clause in (22).

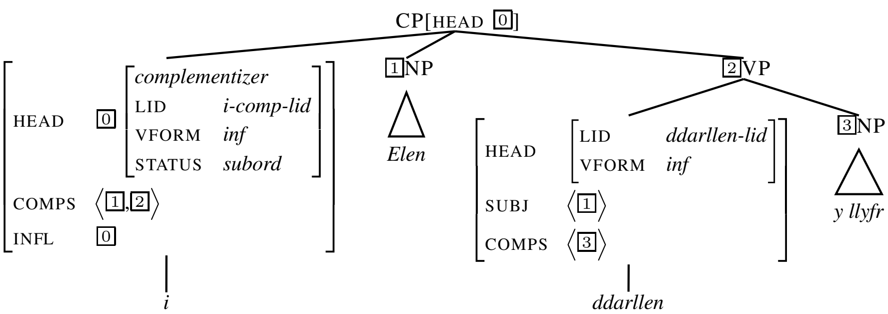


Figure 10: Detailed analysis for example (22).

Here, *i* is identified as non-finite, and both its HEAD and INFL features include [LID *i-lid*] in their value. The INFL value makes it clear that it has the morphological properties of the preposition *i*.

For the complement clause in (21), we can propose the more detailed analysis in Figure 11.

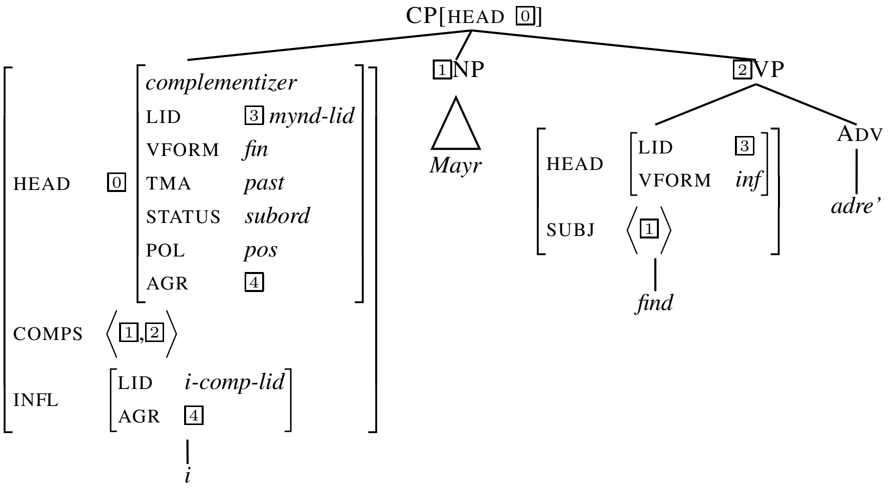


Figure 11: Detailed analysis for example (21).

Here *i* is identified as a positive past subordinate form. The LID value of its HEAD feature identifies it as a form of the following verb, and the INFL value makes it clear that it has the morphological properties of the preposition *i*.

Non-finite and finite *i* differ in some important ways. However, they also show important similarities. The similarities and the differences can be captured by treating them as subtypes of a single lexical type. We will call the supertype *i-clausal* and the two subtypes *i-inf* and *i-fin*, as indicated in Figure 12.

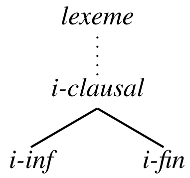


Figure 12: Partial hierarchy of lexeme types.

We assume that the types are associated with the contraints below:

(39) a. *i-clausal* →
$$\begin{bmatrix} \text{HEAD} & \textit{complementizer} \\ \text{INFL} & \begin{bmatrix} \text{LID} & \textit{i-comp-lid} \end{bmatrix} \\ \text{COMPS} & \left\langle \boxed{1}, \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{VFORM} & \textit{inf} \end{bmatrix} \\ \text{SUBJ} & \langle \boxed{1} \rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

b. *i-inf* → $\begin{bmatrix} \text{HEAD}|\text{VFORM} & \textit{inf} \end{bmatrix}$

c. $\textit{i-fin} \rightarrow \begin{bmatrix} \text{HEAD}|\text{LID} & \boxed{0} \\ \text{COMPS} & \left\langle [\,], \begin{bmatrix} \text{HEAD}|\text{LID} & \boxed{0} \end{bmatrix} \right\rangle \end{bmatrix}$

Constraint (39b) ensures that non-finite *i* is non-finite. Because nothing precludes it, the default syntax-morphology interface statement in (35) will apply to non-finite *i*, and ensure that it is just an ordinary complementizer; in particular, it has an ordinary specification [HEAD|LID *i-comp-lid*]. On the other hand, constraint (39c) states that finite *i* has the same LID value as its verbal complement. This entails that the default syntax-morphology interface statement in (35) can not apply to finite *i*, which is lexically specified as having distinct HEAD|LID and INFL|LID values.

The syntax-morphology interface statement in (40) captures the distribution of finite *i*:

(40) $$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{VFORM} & \textit{fin} \\ \text{STATUS} & \textit{subord} \\ \text{TMA} & \textit{past} \\ \text{POL} & \textit{pos} \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{complementizer} \\ \text{AGR} & \boxed{1} \end{bmatrix} \\ \text{INFL} & \begin{bmatrix} \text{LID} & \textit{i-comp-lid} \\ \text{AGR} & \boxed{1} \end{bmatrix} \end{bmatrix}$$

(40) ensures that, in any situation where a word is inflected in the positive past in a subordinate clause, this word is a complementizer with the morphological shape of *i*. Because *i-fin* is the only lexical type compatible with that description, finite *i* is the only word that can satisfy (40); hence the realization of any lexeme's positive subordinate past is finite *i* taking that lexeme's infinitive as a complement. Notice that it is crucial that, by morphosyntactic blocking, the existence of the statement in (40) prevents the use of the default statement in (35) for expression of the positive subordinate past — and hence prevents the use of synthetic inflection.[14]

We noted in section 2.3 that finite *i*-clauses cannot be negated with the negative verb *peidio* in the way that non-finite *i*-clauses can. One might ensure this explicitly by stipulating that the verbal complement of finite *i* is [POL *pos*]. However, this is unnecessary: as we stated above, any use of finite *i* has to be licensed by a special syntax-morphology interface statement, because finite *i*'s lexical entry is inherently incompatible with (35). But since (40) is the only such statement and presupposes positive polarity, the grammar provides no way of using *i* in a negative context.

We now have an analysis of finite *i*-clauses which has the two desirable features that we identified earlier: (a) it captures the idea that they involve an unusual

[14]More precisely, assuming that (35), (36) and (40) are the only interface statements in Welsh grammar, Pāṇinian strengthening has the effect of turning (35) into the following constraint:

(i) $$\left(\neg \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{LID} & \textit{bod-lid} \\ \text{VFORM} & \textit{fin} \\ \text{STATUS} & \textit{subord} \\ \text{TMA} & \textit{pres} \end{bmatrix} \\ \text{SLASH} & \{\} \end{bmatrix} \bigwedge \neg \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{VFORM} & \textit{fin} \\ \text{STATUS} & \textit{subord} \\ \text{TMA} & \textit{past} \\ \text{POL} & \textit{pos} \end{bmatrix} \end{bmatrix} \right) \rightarrow \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{INFL} & \boxed{1} \end{bmatrix}$$

realization of a finite verb, and (b) treat them as similarly as possible to non-finite *i*-clauses.

3.4 Conclusions

We have been concerned in this paper with two types of pseudo-non-finite clause that are a feature of Welsh grammar, one involving *bod* 'be', which we have called *bod*-clauses, and the other involving all verbs, which we have called finite *i*-clauses. There is good evidence that both types of clause are really finite despite their superficial appearance. They are similar in various ways but differ in that *bod*-clauses have a clearly finite internal structure whereas finite *i*-clauses have essentially the same internal structure as non-finite *i*-clauses. *Bod*-clauses are quite easy to incorporate into a formal analysis. They just require a distinction between syntactic and morphological finiteness, and we have developed this with the HEAD-INFL distinction. Finite *i*-clauses are more challenging. It would be difficult to claim that the verb is really a finite form given that it seems non-finite in both its morphological form and its syntactic properties. Instead, we have proposed that finite *i* is a finite form of the verb whose non-finite form heads its complement and have developed an analysis that embodies this assumption.

The analysis of finite *i*-clauses above constitutes a periphrastic analysis in the spirit of Bonami & Webelhuth (2013) and Bonami & Samvelian (2015): the head of the clause contributes to filling a cell in the paradigm of one lexeme (the verb) but is realized morphologically as another lexeme (the preposition *i*). Unlike those previous analyses, the present approach does not postulate valence-changing morphological rules and takes morphosyntactic blocking at face value. Unlike the approach of Bonami (2015), is does not rely on the hypothesis that a morphologically nonfinite verb fills a cell in a finite paradigm. Quite on the contrary, it implements rather directly Blevins's (forthcoming) notion of 'periphrasis as syntactic exponence': the head value of the whole clause is the locus of evaluation of what constitutes an extended paradigm. From a technical HPSG standpoint, the main innovation is to implement Pāṇinian competition explicitly at the morphology-syntax interface. The advantage of such a strategy is that both the morphology and the syntax of periphrastic constructions can be taken to be straightforward.

We believe that the analyses we have developed here capture both the similarities and the differences between the two types of pseudo-non-finite clause in a satisfactory way.

References

Andrews, Avery D. 1990. Unification and morphological blocking. *Natural Language and Linguistic Theory* 8. 507–557.

Blevins, James P. forthcoming. Periphrasis as syntactic exponence. In Farrell Ackerman, James P. Blevins & Gregory T. Stump (eds.), *Paradigms and periphrasis*, Stanford: CSLI Publications.

Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25. 63–110.

Bonami, Olivier & Berthold Crysmann. 2013. Morphotactics in an information-based model of realisational morphology. In Stefan Müller (ed.), *Proceedings of HPSG 2013*, 27–47. Stanford: CSLI Publications.

Bonami, Olivier & Pollet Samvelian. 2015. The diversity of inflectional periphrasis in Persian. *Journal of Linguistics* 51(2). 327–382.

Bonami, Olivier & Gert Webelhuth. 2013. The phrase-structural diversity of periphrasis: a lexicalist account. In Marina Chumakina & Greville G. Corbett (eds.), *Periphrasis: The role of syntax and morphology in paradigms*, 141–167. Oxford: Oxford University Press.

Borsley, Robert D. 1989. An HPSG approach to Welsh. *Journal of Linguistics* 25. 333–354.

Borsley, Robert D. 1999. Mutation and constituent structure in Welsh. *Lingua* 109. 267–300.

Borsley, Robert D. 2009. On the superficiality of Welsh agreement. *Natural Language and Linguistic Theory* 27. 225–265.

Borsley, Robert D. 2013. On the nature of Welsh unbounded dependencies. *Lingua* 133. 1–29.

Borsley, Robert D. & Bob Morris Jones. 2005. *Welsh negation and grammatical theory*. Cardiff: University of Wales Press.

Borsley, Robert D., Maggie Tallerman & David Willis. 2007. *The syntax of Welsh*. Cambridge: Cambridge University Press.

Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics in Information-Based Morphology. *Journal of Linguistics* 52(2). 311–374.

Koenig, Jean-Pierre. 1999. *Lexical relations*. Stanford: CSLI Publications.

Sadler, Louisa & Andrew Spencer. 2001. Syntax as an exponent of morphological features. In Gert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2000*, 71–96. Dordrecht: Kluwer.

Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33. 431–484.

Spencer, Andrew. 2013. *Lexical relatedness: a paradigm-based model*. Oxford: Oxford University Press.

Stump, Gregory T. 2006. Heteroclisis and paradigm linkage. *Language* 82. 279–322.

Stump, Gregory T. 2015. *Inflectional paradigms*. Cambridge: Cambridge University Press.

Tallerman, Maggie. 1998. The uniform case-licensing of subjects in Welsh. *The Linguistic Review* 15. 69–133.

Willis, David. 2000. On the distribution of resumptive pronouns and wh-trace in Welsh. *Journal of Linguistics* 531–573.

Willis, David. 2011. The limits of resumption in Welsh wh-dependencies. In Alain Rouveret (ed.), *Resumptive pronouns at the interfaces*, 189–222. Amsterdam: John Benjamins.

Aspectual object marking in Libyan Arabic

Kersti Börjars
The University of Manchester

Khawla Ghadgoud
The University of Manchester

John Payne
The University of Manchester

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 125–139

Keywords: aspectual object marking, Libyan Arabic, inside-out designators, anaphoric control

Börjars, Kersti, Khawla Ghadgoud & John Payne. 2016. Aspectual object marking in Libyan Arabic. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 125–139. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.7.

Abstract

In Libyan Arabic, the preposition *fi* 'in' has developed into a marker of continuous or habitual aspect. While structurally remaining a preposition which marks the objects of the non-tensed forms of dynamic transitive verbs, it serves to attribute an aspectual interpretation to the clause as a whole. We argue that this aspectual object marking is naturally modeled by an inside-out functional designator, and provide arguments that the aspectual value contributed by aspectual *fi* is best treated as an f-structure feature.

1 Introduction[*]

In Libyan Arabic, direct objects can be either plain or preceded by *fi*, which we will refer to as an aspectual object marker, as illustrated in (1):[1]

(1) a. Aħmed kle el-kosksi
Ahmed eat.PST.3MSG DEF-couscous
'Ahmed ate couscous.'

b. Aħmed yākil fi el-kosksi
Ahmed eat.NONT.3MSG FI DEF-couscous
'Ahmed eats/is eating couscous.'

The presence of *fi* is excluded when the object is governed by a tensed verb-form such as the past form, as in (2):[2]

---

[*] We are grateful to the audience at HEADLEX and to two anonymous referees for their comments, which have improved this paper and will also be helpful for further work on this topic.

[1] A similar use of *fi* has been noted in Cairo Arabic (Woidich 2006) and in Tunisian Arabic (Pallottino & Askri 2015). The analysis presented here is unrelated to earlier analyses. We will gloss *fi* in this use as FI throughout since its precise function is the focus of our investigation.

[2] Our glossing follows the Leipzig guidelines. Note that we use PST for the form frequently referred to in the literature on Modern Standard Arabic as 'perfective' and NONT 'non-tensed' for the form referred to as 'imperfective' (compare Ryding 2005). The designation 'non-tensed' for this latter form is intended to reflect the fact that it is not an absolute tense like the past form, but rather indicates temporal identity to some reference time. It occurs not only in main clauses, where it implicates a non-past reading through identity to the time of

(2)

* Aħmed	kle	fi	el-kosksi
Ahmed	eat.PST.3MSG	FI	DEF-couscous

‘Ahmed ate couscous.’

In a non-tensed environment, *fi* is obligatory if the governing verb is dynamic, as in (1b) above, but disallowed if the governing verb is stative, as in (3):

(3) a.

Aħmed	yħib	(*fi)	el-kosksi
Ahmed	like.NONT.3MSG	FI	DEF-couscous

‘Ahmed likes couscous.’

b.

Aħmed	yibbi	(*fi)	el-kosksi
Ahmed	want.NONT.3MSG	FI	DEF-couscous

‘Ahmed wants couscous.’

The aspectual interpretation of a clause containing a dynamic governing verb with a *fi*-marked object is either continuous, as in (1b) above and (4a), where the adverb *tawwa* ‘now’ forces the actual present reading, or habitual, as in (4b):

(4) a.

Aħmed	yākil	fi	el-kosksi	tawwa
Ahmed	eat.NONT.3MSG	FI	DEF-couscous	now

‘Ahmed is eating couscous now.’

b.

Aħmed	yākil	fi	el-kosksi	kol	youm
Ahmed	eat.NONT.3MSG	FI	DEF-couscous	every	day

‘Ahmed eats couscous every day.’

When, however, the interpretation is neither continuous nor habitual, *fi* is excluded. Two of the interpretations permitted in the event of a non-*fi*-marked object are either generic, as in (5a), or a scheduled future, as in (5b):

(5) a.

Aħmed	yākil	kosksi
Ahmed	eat.NONT.3MSG	couscous

‘Ahmed eats couscous.’ (i.e. he is a couscous-eater)

---

utterance, but also in COMP clauses dependent on a past-tense matrix verb, where it indicates past time. Both forms have a full paradigm of subject-agreement affixes and involve pronoun incorporation. We consider them both to be finite.

b. fi rijīmī ġoḍwa nākil kosksi
in diet.1SG.GEN tomorrow eat.NONT.1SG couscous
'In my diet, tomorrow I eat couscous.'

A third type of interpretation in which *fi* is omitted involves universal quantification over event tokens, as in (6):

(6) lamma nākil kosksi netfakker
when eat.NONT.1SG couscous remember. NONT.1SG

ħinn-āi
grandmother-1SG.GEN
'When I eat couscous, I remember my grandmother.'

To sum up, a *fi*-marked object occurs when the governing verb is dynamic, non-tensed, and has either a continuous or habitual interpretation. In all other cases, *fi* is excluded. In non-tensed clauses with transitive verbs, it is therefore solely the presence of a *fi*-marked object which indicates that the interpretation is habitual or continuous.[3] In section 2 below, we demonstrate that *fi* is best analysed as a preposition heading a PP which contains the object. Since the aspectual information which *fi*-marked objects contribute to the clause is structurally internal to the object, an analysis involving inside-out functional designators is proposed in section 3. An interesting complication is that the occurrence of aspectual *fi* is blocked in a clause in which a dynamic verb in the non-tensed form is subordinated to a verb which is itself stative. There is no apparent semantic reason why the presence of *fi* in such a clause should not be able to signal a continuous or habitual interpretation. In section 4, we discuss the relevant data and claim that this fact provides additional justification for the treatment provided, which invokes f-structural features rather than a direct mapping to semantic form.

2 Structural properties of Libyan *fi*

Other than as an aspect marker internal to the object, the form *fi* has two further functions in Libyan Arabic. These are both illustrated in (7):

[3] In a clause headed by a dynamic intransitive verb, therefore, there is no grammatical marking of aspect, and any of the aspectual interpretations are possible (depending on context).

(7) fi ʕasīr fi et̠-t̠alāja
exist juice in DEF-fridge
'There is juice in the fridge.'

The first *fi* is as an invariant form which introduces existential sentences, parallel to English 'there is'. The second *fi* is a locative preposition meaning 'in'. It is well-known that there may be a diachronic relation between predicates meaning 'live, exist' and the continuous, and secondly that continuous forms may develop into habituals (see Bybee at al 1994:158 and Heine & Kuteva 2002:127). Many languages also show a historical connection between the locative 'in' and the continuous (Heine & Kuteva 2002:178–9), although in this development, unlike in Libyan Arabic, the locative marker typically becomes a continuous marker of the whole predication including the verb, rather than a marker of just the object. We will not here provide historical data to trace the details of the historical relationship between the three *fi* elements in Libyan Arabic. Our goal is rather to demonstrate the conceptual and structural similarity between the locative preposition and the one which we refer to as the aspectual object marker. We hypothesise that the second developed from the first.

Our claim then is that locative *fi* and aspectual *fi* are conceptually similar, sharing the notion "interior". The observation that the continuous and the habitual might both be construed as "internal" aspects is due to Stassen (1997: 252): essentially continuous aspect portrays an activity as ongoing within a relatively short timespan, while habitual aspect portrays an activity as ongoing within a relatively long timespan.[4] Not only does aspectual *fi* have this semantic link with locative *fi,* however, it also shares the structural characteristics of a preposition.

Firstly, both locative *fi* and aspectual *fi* can be fronted together with their noun-phrase complements:

(8) a. fi London Aħmed yoskun
in London Ahmed live.NONT.3MSG
'It is in London that Ahmed lives.'

b. fi el-kosksi Aħmed yākil kol youm
FI DEF-couscous Ahmed eat.NONT.3MSG every day
'It is couscous that Ahmed eats every day.'

[4] The continuous and the habitual are not typically grouped together in formal semantic analyses of aspect, since the first reports a particular event while the second reports a generalization over events, and is therefore treated as a kind of generic (see Krifka et al. 1995:12). The conceptual similarity between continuous and habitual that Stassen identifies does, however, appear to be reflected in the historical development of habitual interpretations from continuous forms, and is further supported by the analysis presented here.

Or, alternatively, just the noun-phrase complement can be fronted, in which case both locative *fi* and aspectual *fi* take an oblique resumptive pronoun:

(9) a. London Aħmed yoskun fi-ha
London Ahmed live.NONT.3MSG in-3FSG.OBL
'It is London where Ahmed lives.'

b. el-kosksi Aħmed yākil fi-h kol youm
DEF-couscous Ahmed eat.NONT.3MSG FI-3MSG.OBL every day
'It is couscous that Ahmed eats every day.'

Secondly, as prepositions both locative *fi* and aspectual *fi* can take scope over co-ordinated noun phrases, as in (10):

(10) a. Aħmed yexdim fi Paris w London
Ahmed work.NONT.3MSG in Paris and London
'Ahmed works in Paris and London.'

b. Aħmed yākil fi el-kosksi w eṣ-ṣlāṭa
Ahmed eat.NONT.3MSG FI DEF-couscous and DEF-salad
'Ahmed eats/is eating couscous and salad.'

Or, alternatively, both locative *fi* and aspectual *fi* can be repeated before each noun phrase:

(11) a. Aħmed yexdim fi Paris w fi London
Ahmed work.NONT.3MSG in Paris and in London
'Ahmed works in Paris and in London.'

b. Aħmed yākil fi el-kosksi w fi eṣ-ṣlāṭa
Ahmed eat.NONT.3MSG FI DEF-couscous and FI DEF-salad
'Ahmed eats/is eating couscous and salad.'

These data suggest that, from a structural point of view, aspectual *fi* is analogous to locative *fi* and has the properties of a preposition heading a PP.

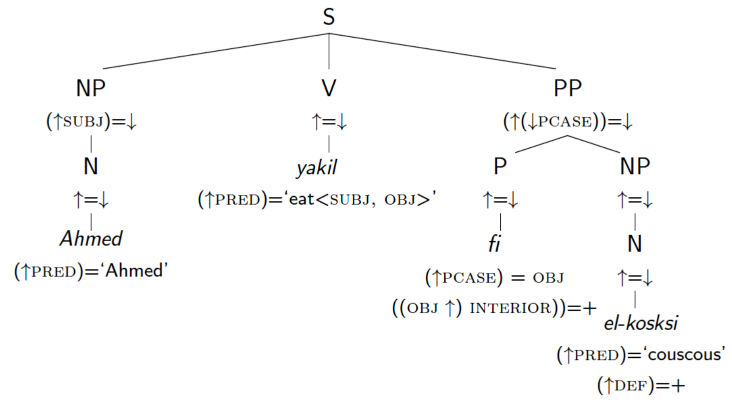
3 Analysis

As the basis for the analysis, we take it that continuous and habitual aspect cannot be distinguished grammatically in Libyan Arabic, and are represented as a single f-structure aspectual feature INTERIOR with value +. Non-tensed verb-forms will not carry any tense or aspect feature, and non-tensed forms of dynamic verbs are therefore compatible with any aspect including the continuous and habitual. On the other hand, tensed verb-forms of dynamic verbs such as the past are valued as INTERIOR with value −, and stative verbs, which are inherently incompatible with interior aspect, will be blocked from bearing any value of the feature INTERIOR at all. These assignments will account for the inability of aspectual *fi* to co-occur with past verb-forms of dynamic verbs, as shown in (2), and with stative verbs in general, as shown in (3).

Importantly, then, the information that a clause containing a non-tensed dynamic verb must have a continuous or habitual interpretation is contributed by aspectual *fi*, which heads a PP mapping to the object function. Inside-out functional designators (Nordlinger 1998, Dalrymple 2001: 143-146) are an ideal tool to allow *fi* to contribute the aspectual information contained in the object to the f-structure of the clause which dominates it. Note that this use of inside-out designators does not involve any uncertainty: it simply attributes the aspectual information within the object to the f-structure of the immediately dominating clause.

An annotated c-structure representation of (1b) is given in (12).[5]

(12)



As can be seen, we assume that Libyan Arabic has a flat clause structure in which clauses are represented as S. There are no special features that could be associated with an I projection, and no separate set of auxiliary verbs. The non-tensed verb-form *yākil* 'eat' has no features apart from its semantic PRED

[5] In (12), we use the approach to prepositions found in Dalrymple (2001:151–153); for an alternative approach, see Butt et al (1999:125–131). Note that we assume that the *fi*-marked NP does not change its status from OBJ to OBL simply by virtue of occurring within a PP. That is, we assume that *fi*-marking does not have a sufficient detransitivising effect for a change of function to be implicated.

feature, which contains inter alia the information that it requires an object, and its agreement features, which we omit here for simplicity.

The annotations on the PP do the work. In itself, the PP will be associated with the object function by virtue of the PCASE feature: that is, aspectual *fi* is lexically specified as creating an object. Most importantly, however, aspectual *fi* also contains the inside-out designator ((OBJ ↑) INTERIOR))=+, which will attribute the + value of the feature INTERIOR to the f-structure containing the OBJ function, i.e. to the f-structure corresponding to S. The resulting f-structure is as in (13):

(13)

$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{`Ahmed'}\end{bmatrix} \\ \text{PRED} & \text{`eat} < \text{SUBJ , OBJ} >\text{'} \\ \text{INTERIOR} & + \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`couscous} \\ \text{PCASE} & \text{OBJ} \\ \text{DEF} & + \end{bmatrix}\end{bmatrix}$$

Crucially, the grammatical aspect value of the clause comes from aspectual *fi*, and not from the verb.

4. Aspectual *fi* in complement clauses

We distinguish two cases when the clause containing aspectual *fi* is itself the complement of a higher lexical verb. Firstly, if the complement clause can in principle take a complementiser, the verb in the complement clause determines the presence or absence of *fi*. This is illustrated in (14):

(14) a.	ʔaʕtaqid	(ennah)	yākil	fi	el-kosksi
	think.NONT.1SG	that	eat.NONT.3MSG	FI	DEF-couscous

‘I think that he is eating couscous.’

b.	ʔaʕtaqid	(ennah)	yħib		el-kosksi
	think.NONT.1SG	that	like.NONT.3MSG		DEF-couscous

‘I think that he likes couscous.’

Note that the complementiser is optional. In (14a), we have a dynamic verb in the subordinate clause, which we take to be a COMP, and the interior aspectual interpretation of the subordinate clause derives from the presence of aspectual *fi*. In (14b), the verb in the subordinate clause is stative, and aspectual *fi* is predictably absent.

By contrast, if the interpretation of the subject of a complement clause is identified by what we take to be an anaphoric control mechanism, the control verb determines the presence or absence of *fi*. In this case, the complement clause, which we again take to be a COMP rather than an XCOMP because of the presence of the finite verb-form, cannot take a complementizer,

(15) a. yibbi yākil el-kosksi
want.NONT.3MSG eat.NONT.3MSG DEF-couscous
'He wants to eat couscous.'

b. *yibbi yākil fi el-kosksi
want.NONT.3MSG eat.NONT.3MSG FI DEF-couscous
Intended: 'He wants to be eating couscous.'

In (15), the matrix verb is stative, and this prevents the subordinate clause from containing *fi*, even though the dynamic verb *yākil* 'eat' in its non-tensed form in principle allows *fi*.

The lexical entry we propose for the matrix verb *yibbi* 'want' in its control use is given in (16):

(16) *yibbi* V (↑PRED) = 'want <SUBJ, COMP>'
(↑COMP SUBJ) = 'PRO'
$((\uparrow\text{COMP SUBJ})_\sigma$ ANTECEDENT$) = (\uparrow\text{SUBJ})_\sigma$
¬(↑COMP COMPFORM)
¬(↑INTERIOR)
¬(↑COMP INTERIOR)

The equations (↑COMP SUBJ) = 'PRO' and $((\uparrow\text{COMP SUBJ})_\sigma$ ANTECEDENT$) = (\uparrow\text{SUBJ})_\sigma$ require the subject of the complement clause to be a pronominal, and for its antecedent to be identified as the subject of the matrix clause.[6] In this case, the complement clause cannot contain any form of complementizer.[7] As a stative verb itself, *yibbi* is not permitted to have a value for the aspectual feature INTERIOR. The crucial work is however done by the final equation ¬(↑COMP INTERIOR). As a stative verb taking a COMP with an obligatorily anaphorically

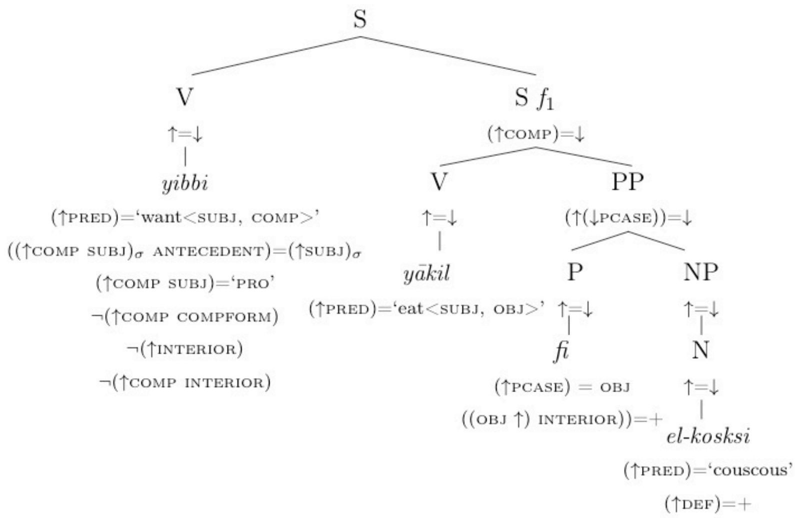
[6] We assume here the basic approach to obligatory anaphoric control constructions provided by Dalrymple (2001:323-338). The lexical entry provided is consistent with an analysis of the COMP clause either as a proposition or as a property. See Ash & Mortazavinia (2011) for a recent discussion of the status of finite COMPs in control constructions.

[7] The verb *yibbi*, just like English *want*, also occurs in a construction where there is no control, i.e. where its COMP contains a disjoint subject. In this case a complementizer can be inserted.

controlled subject, *yibbi* will be specified as also not permitting its COMP to have a value for INTERIOR. Since the COMP cannot have a value for the feature for INTERIOR, the value + assigned to it by aspectual *fi* in (15b) will result in the required ungrammaticality.

The annotated c-structure corresponding to the ungrammatical (15b) will then be (17a):

(17) a.



b. $(f_1 \text{ INTERIOR}) = +$

c. $\neg(f_1 \text{ INTERIOR})$

If we set the f-structure of the complement clause to be f_1, the inside-out designator on *fi* will result in the f-description equation (17b). The substitution by f_1 of (↑COMP) in the annotation ¬(↑COMP INTERIOR) on the control verb *yibbi* will result in the contradictory equation (17c). Since no f-structure can satisfy these equations, (15b) will be ungrammatical.

We emphasize that this is indeed an ungrammaticality, and not something which can be enforced simply by semantic principles. There is nothing semantically amiss with the interpretation which would be expected if *fi* were permitted, namely 'He wants to be eating couscous'. This interpretation simply cannot be achieved by (17b). We take this ungrammaticality to represent an argument in favour of the adoption of a featural approach to the role of aspectual *fi*.

In order to express the continuous or habitual past, a clause containing a non-tensed verb-form must be used as a complement of the verb *kān* 'be.PST'. In this case, however, the distribution of aspectual *fi* is unaffected by the presence of the matrix verb:

(18) a. kān yākil fi el-kosksi amis
be.PST.3MSG eat.NONT.3MSG FI DEF-couscous yesterday
'He was eating couscous yesterday.'

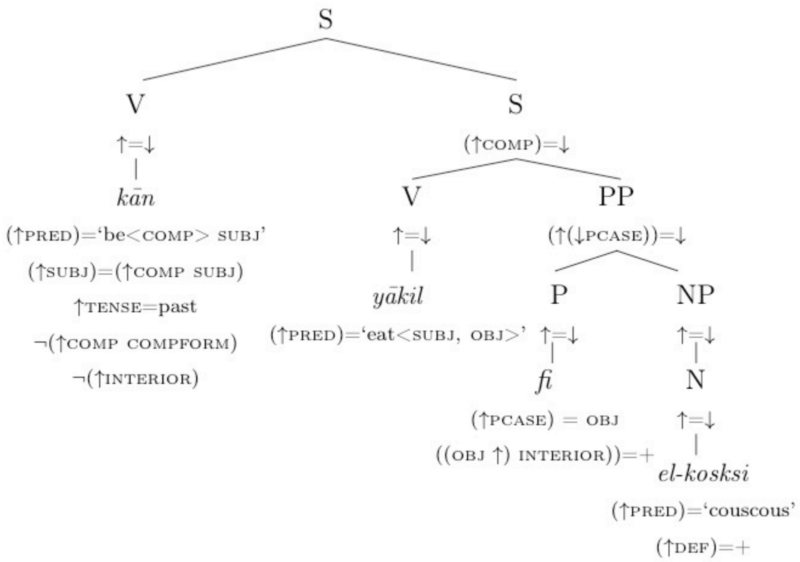
b. kān yākil fi el-kosksi kol youm
be.PST.3MSG eat.NONT.3MSG FI DEF-couscous every day
'He used to eat couscous every day.'

We take *kān* itself to be a stative verb, and hence specified as ¬(↑INTERIOR). In this case, analysing *kān* as a functional co-head would clearly conflict with the presence of aspectual *fi* in the complement (this means that we follow the approach taken by Dyvik (1999) rather than Butt et al (1999)). Instead, we take *kān* to be a raising verb taking a complementizer-less COMP, though unlike *yibbi* 'want' it will not enforce the absence of the feature INTERIOR on its complement. The lexical entry for *kān* in this function is then as in (19):

(19) *kān* V (↑PRED) = 'be <COMP> SUBJ'
(↑SUBJ) = (↑COMP SUBJ)
(↑TENSE) = past
¬(↑COMP COMPFORM)
¬(↑INTERIOR)

This lexical entry results in the tree (20a) and the f-structure in (20b) for the sentence in (18):

(20) a.



b.

$$\left[\begin{array}{ll} \text{PRED} & \text{`be } < \text{ COMP} > \text{ SUBJ'} \\ \text{SUBJ} & \left[\text{PRED} \quad \text{pro}_i\right] \\ \text{TENSE} & \text{past} \\ \text{XCOMP} & \left[\begin{array}{ll} \text{PRED} & \text{`eat } <\text{SUBJ, OBJ}> \text{'} \\ \text{SUBJ} & \\ \text{OBJ} & \left[\begin{array}{ll} \text{PRED} & \text{`couscous'} \\ \text{DEF} & + \\ \text{PCASE} & \text{OBJ} \end{array}\right] \end{array}\right] \end{array}\right]$$

The role of *kān* in (20) is essentially to add past tense, and it does this without affecting the licensing of aspectual *fi* in its complement.

The treatment of *kān* as a matrix verb is supported by examples in which a clause containing aspectual *fi* is the complement of the non-tensed counterpart to *kān*, i.e. *ykūn* 'be.NONT', which in its turn is the complement of a higher stative verb. This non-tensed form only occurs in complement clauses, and it has the effect of "blocking" the reach of the higher stative verb into the complement clause which it governs. This effect can be seen in (21):

(21) yibbi ykūn yākil fi el-koksi
want NONT.3MSG be.NONT.3MSG eat. NONT.3MSG FI DEF-couscous
'He would like to be eating couscous.'

The lexical entry for *ykūn* is given in (22):[8]

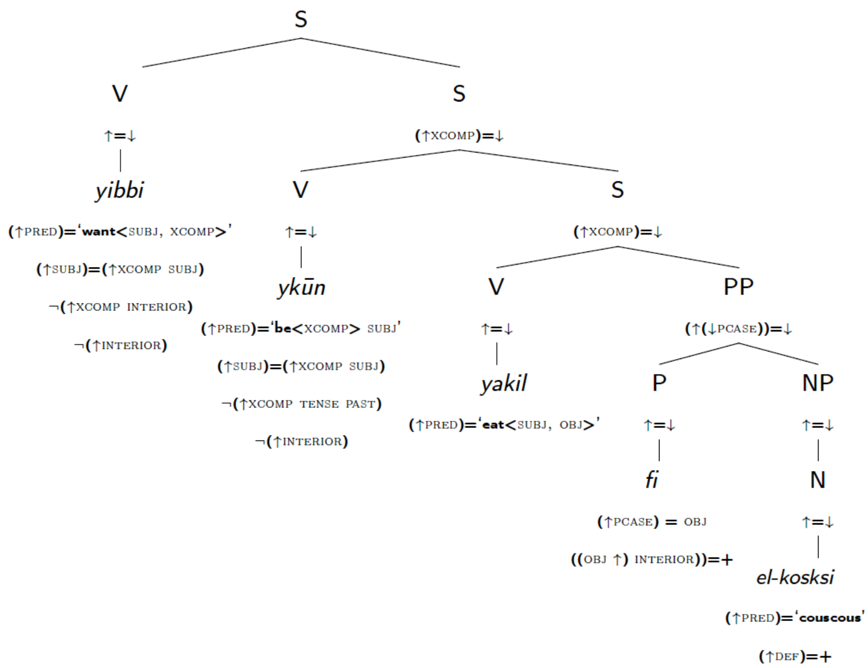
(22) *ykūn* V (↑PRED) = 'be <COMP> SUBJ'
(↑SUBJ) = (↑COMP SUBJ)
(COMP↑)
¬(↑COMP COMPFORM)
(↑INTERIOR)

As a non-tensed form *ykūn* lacks any specification for tense, but just like *kān* it does not bear an annotation enforcing the absence of any value for the INTERIOR feature within its subordinate clause.

The tree corresponding to (21) is given in (23):

[8] The form *ykūn* only occurs as a subordinate verb, and a restriction will need to be added to the lexical entry to enforce this. Assuming that *ykūn* must head a COMP, this could be done, as here, by the addition of the annotation (COMP↑).

(23)



The stative lexical verb *yibbi* 'want' requires its complement not to have a value for the feature INTERIOR, but *ykūn* as a stative verb satisfies this requirement. In its turn, *ykūn* itself imposes no requirement on the value INTERIOR of its complement, and this is compatible with the presence of aspectual *fi*. Crucially, this blocking effect is predicted by the analysis of *kān* / *ykūn* as a raising verb.

5 Conclusion

In this paper, we have argued for an analysis in which the aspectual object marker *fi* in Libyan Arabic is responsible for contributing an aspectual feature +INTERIOR to the f-structure of a clause headed by a non-tensed dynamic verb which contains it. The feature arises solely from the presence of aspectual *fi*, and not from the verb itself, which is compatible with other aspectual interpretations. The role of aspectual *fi* is naturally modelled by an inside-out functional designator.

Because of the inability of *fi* to occur in complement clauses subordinate to COMP-taking stative verbs, it is further argued that there are compelling reasons to analyse the contribution of aspectual *fi* in f-structure terms. There is no semantic incompatibility between a stative matrix verb and a complement which has interior aspect. Stative matrix verbs rather impose a grammatical requirement that their complements should not be marked by *fi*.

This grammatical requirement can be subverted by the intervention of the verb *ykūn* 'be.NONT', which, as long as it itself is syntactically analysed as a raising verb heading a separate clause, has a blocking effect.

The aspectual marking of objects that we see in Libyan Arabic has close affinities to the aspectual function of the partitive case in languages such as Finnish (Kiparsky 1998) or Estonian (Tamm 2006). Tamm indeed provides an analogous LFG analysis of the aspectual function of the partitive using inside-out functional designators. The aspectual object marking that is seen in Finnic appears, however, to indicate the non-boundedness of an event, that is imperfectivity more generally (including generic utterances). This contrasts quite strikingly with the narrow limitation to continuous and habitual aspect in Libyan Arabic. Equally, aspectual object marking in Finnic does not appear to be restricted to particular grammatical contexts (such as co-ocurrence with particular tenses), and is arguably therefore more semantically and less grammatically constrained than the aspectual object marking of Libyan Arabic. We leave a more detailed discussion of these comparisons for future research.

References

Ash, Asudeh & Marzieh Mortazavinia 2011. Obligatory control in Persian: implications for the syntax-semantics interface. Paper presented at *ICIL 4*, Uppsala University.

Dalrymple, Mary 2001. *Lexical Functional Grammar*. Syntax and Semantics vol 34. Academic Press.

Kiparsky, Paul 1998. Partitive case and aspect. In Miriam Butt and Wilhelm Geuder (eds.) *The projection of arguments: lexical and compositional factors,* 265-307. Stanford: CSLI.

Krifka, Manfred, Francis Jeffrey Pelletier, Gregory N. Carlson, Alice ter Meulen, Gennaro Chierchia and Godehard Link 2005. Genericity: an introduction. Chapter 1 of Gregory N. Carlson & Francis Jeffry Pelletier (eds.), *The generic book*, 1-124. Chicago and London: The University of Chicago Press.

Nordlinger, Rachel 1998. *Constructive case: evidence from Australian languages*. Dissertations in Linguistics series. Stanford University: CSLI Publications.

Pallottino, Margherita & Majid Askri. 2015. Aspectual "fi" in Tunisian Arabic. Paper presented at *Forum for Arabic Linguistics*, Essex University.

Ryding, Karin C. 2005. *A reference grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.

Stassen, Leon 1997. *Intransitive predication*. Oxford: Clarendon Press.

Tamm, Anne 2006. Estonian transitive verbs and object case. *Proceedings of the LFG06 conference*. Stanford: CSLI Publications.

Woidich, Manfred A. 2006. *fi*-objects in Cairo Arabic: the case for telic verbs. In Salah Mejri (ed.), *L'arabe dialectal: enquêtes, descriptions, interprétations. actes d' aida 6 (2004). travaux offerts au professeur taïeb baccouche.* Série Linguistique No 13, 473–486. Tunis: C.E.R.E.S.

Adverb agreement in Urdu, Sindhi and Punjabi

Miriam Butt
University of Konstanz

Sebastian Sulger
University of Konstanz

Mutee U Rahman
Isra University

Tafseer Ahmed
DHA Suffa University

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 140–160

Keywords: Urdu, Sindhi, Punjabi, agreeing adverbs, resultatives

Butt, Miriam, Sebastian Sulger, Mutee U Rahman & Tafseer Ahmed. 2016. Adverb agreement in Urdu, Sindhi and Punjabi. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 140–160. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.8.

Abstract

We discuss agreeing adverbs in Urdu, Sindhi and Punjabi. We adduce crosslinguistic evidence that is based mainly on similar patterns in Romance and posit that there is a close connection between resultatives and so-called pseudo-resultatives, which the agreeing adverbs appear to instantiate. We propose a diachronic relationship by which the originally predicative part of a resultative is reinterpreted as an adjunct that modifies the overall event predication, not just the result.

1 Introduction

We provide a first description and analysis of adverb agreement in three South Asian languages: Urdu, Sindhi and Punjabi. To the best of our knowledge, this has not been discussed before. While existing grammars have noticed the phenomenon in passing, they do not go into structural details and do not provide an explanation (cf. Trumpp 1872, Kellogg 1893, Cummings & Bailey 1912, McGregor 1972).

Examples (1)–(3) illustrate the core phenomenon with respect to the three languages.[1] The lexical item *sast-e/sast-a* 'cheap' in (1)–(3) is based on an adjective, but expresses adverbial manner modification of the overall event. Despite modifying the predicate, it agrees with the object 'clothes' in number and gender as a normal adjective would do.

In all of the examples the object is masculine plural and the adjective/adverb (AdjAdv below) agrees with the object. The verbal complex, on the other hand, agrees with the nominative subject as per the languages' normal agreement rules.

(1)	rɑvi	**kɑpṛe**	**sɑst-e**	bec-ta	hɛ
	Ravi.M.Sg.Nom	clothes.M.Pl.Nom	cheap-M.Pl	sell-Impf.M.Sg	be.Pres.3.Sg

'Ravi sells clothes cheap.' Urdu
(he sells them cheaply, the clothes are not inherently cheap)

(2)	rɑvi	**kɑpṛe**	**sɑst-e**	vec-da	ɛ
	Ravi.M.Sg.Nom	clothes.M.Pl.Nom	cheap-M.Pl	sell-Impf.M.Sg	be.Pres.3.Sg

'Ravi sells clothes cheap.' Punjabi
(he sells them cheaply, the clothes are not inherently cheap)

(3)	rɑvi	**kɑpɽ̣a**	**sɑst-a**	vik-ɳɛ	t^{h}o
	Ravi.M.Sg.Nom	clothes.M.Pl.Nom	cheap-M.Pl	sell-Impf.M.Sg	be.Pres.3.Sg

'Ravi sells clothes cheap.' Sindhi
(he sells them cheaply, the clothes are not inherently cheap)

[†] We thank the audience of the HeadLex16 Conference for helpful comments. We also gratefully acknowledge funding from the DAAD scheme on German-Pakistani cooperation which allowed Tafseer Ahmed and Mutee U Rahman to work on the syntactic annotation of Urdu and Sindhi in 2015 and 2016. Sebastian Sulger's work is supported by a grant from the Nuance Foundation on *Tense and Aspect in Multilingual Semantic Construction.*

[1] Example due to Rajesh Bhatt.

We discovered this phenomenon as part of on-going work on Urdu and Sindhi grammar development and syntactic annotation within the overall ParGram effort (Butt et al., 2002; Sulger et al., 2013). In seeking to understand how to handle it, we looked to a crosslinguistic comparison, which yielded the suggestion that the adverb agreement is an instance of a *pseudo-resultative*. We investigated this hypothesis and concluded that the phenomenon does indeed fall under this category. We propose an analysis that models a diachronic development from an original adjectival resultative to an adverbial pseudo-resultative via a reanalysis of an f-structural relationship (cf. proposals for understanding diachronic change via f-structure reanalysis in Butt & King (2001)).

2 Crosslinguistic Comparison and Pseudo-Resultatives

As indeclinable elements, adverbs are not expected to show agreement inflection (e.g., Anderson, 1985; Alexiadou, 1997; Evans, 2000). Examples which parallel our South Asian constructions quite closely can, however, be found in Romance.[2]

2.1 Romance

Agreeing adverbs are attested in Southern Italian dialects (Ledgeway, 2011, 2016; Silvestri, 2016). Southern Italian shows (at least) two ways of forming adverbs, one via the derivational morpheme *-ment* and one via adjectives that modify the manner of the event, but that tend to agree with one of the arguments of the clause.

(4a) and (4b) show agreement with the object ('health problems') and subject ('I'), respectively (Ledgeway, 2011, 10). In (4c), the adjective is in the masculine singular, which also functions as a default form indicating absence of agreement.

(4) a. tu li sa canusciri **buoni** li disturbi di **saluti**
you them.M know know.INF good.MPL the.MPL complaints.M of health
'you can recognize health problems expertly' Eastern Sicilian

b. havi tri ghiorna ca mi priparu, ma **bona** **nisciu**, averu?
it.has 3 days that me= I.prepare but good.FSG I.come.out true
'I've been preparing for 3 days, but I'll do alright, won't I?' Eastern Sicilian

c. Maria ma tu chi dici ca ficimu **bonu** [...]?
Maria but you what you.say that we.did good.MSG
'Maria, do you think we acted correctly?' Eastern Sicilian

As in Urdu, Sindhi and Punjabi, the agreeing adverbs are all based on adjectives, but the overall effect is one of adverbial manner modification of the event.

[2]Typological work has established that agreeing adverbs do exist in a few other languages as well, for example in the Daghestanian languages Avar and Archi (e.g., Evans, 2000). A close inspection of these examples show that these are not of the same type as our South Asian examples. They appear to involve true adverbs, not adverbs based on adjectives and are not confined to manner modification.

2.2 Undergoer Agreement and Variable Readings

The different Italian dialects investigated by Ledgeway and Silvestri show interesting microvariation in terms of agreement possibilities. However, Ledgeway (2011) shows that the main pattern with respect to agreeing adverbs is that they target the underlying UNDERGOER. In (5) and (4b) the agreement is with the subject, which, in both cases, is an undergoer (the subject of an unaccusative verb). In (4a) and (6) the agreement is also with the undergoer (the object of a transitive verb).

(5) **Campài** tantu tempu **mala**
I.lived so.much time bad.FSG
'I lived badly for such a long time.' Reggio Calabria

(6) Anna miscava **bone** **'i** **carte**
Anna shuffled good.FPL the.PL cards.F
'Anna shuffled the cards well.' (the cards are shuffled well) Cosenza

As already illustrated by (4c), the adjectives can also appear in the masculine singular default form. In this case, there is a failure of agreement and the interpretation is always that of a manner modification of the event, as shown particularly between the contrast in (6) and (7).

(7) Anna miscava **buonu** 'i carte
Anna shuffled good.MSG the.PL cards.F
'Anna shuffled the cards well.' (Anna shuffles well) Cosenza

Silvestri (2016) further points out that agreeing adverbs can actually have two different interpretations in Southern Italian dialects, as the examples in (8) show.

(8) a. Pietrə cusa na giacchetta **traviersə**
Pietro sews a jacket.FSG oblique.MSG
'Pietro is sewing a jacket haphazardly.' Northern Calabrian
'Pietro, sitting crookedly, is sewing a jacket.'

b. Pietrə cusa na **giacchetta traviersa**
Pietro sews a jacket.FSG oblique.FSG
'Pietro sews a malformed jacket.' Northern Calabrian
'Pietro sews a jacket crookedly.'

When there is no agreement, as in (8a), the interpretation is either a subject-oriented adverbial one or a manner modification of the event. When there is agreement with the object (the undergoer), as in (8b), it can be interpreted as straightforward adjectival modification or as a manner modification of the event.

The vast majority of the examples found in corpora or via speaker elicitation by Ledgeway and Silvestri show undergoer agreement. However, there are some which show agreement with agents in transitives and unergatives. An example with unergatives from Silvestri (2016) is shown in (9a). The usages are always subject-oriented and are ambiguous between an adverbial and an adjectival function, though these are difficult to distinguish, as also illustrated by (9), where the non-agreeing version in (9b) has only the adverbial manner modification reading.

(9) a. **Maria** parlava **segreta**
Maria talked secret.FSG
'Maria spoke enigmatically.' Northern Calabrian
(i.e., she was an enigmatic interlocutor)

b. Maria parlava **segretə**
Maria talked secret.MSG
'Maria was speaking secretly.' Northern Calabrian
(i.e., her talking was secret)

The ambiguity between subject-oriented and manner modification is systematic and is presumably due to the semantically overlapping readings effected by argument-oriented vs. manner modification, as discussed in the next section.

2.3 Pseudo-Resultatives

Washio (1997) and Levinson (2010) look at pairs as in (10).

(10) a. He tied the shoelaces loose/loosely.

b. Janet braided her hair tight/tightly.

Washio (1997) and Levinson (2010) note that these pairs are semantically almost indistinguishable; in particular, Washio (1997) provides a reason why one finds potential ambiguity in English with respect to *He tied the shoelaces loose.*

> Although it still seems possible to regard the adjective *loose* [in (10)] as specifying the state of the shoelaces, it can also be regarded as describing *the way* he tied his shoelaces, that is, he did it without much force. This is natural because in an activity like tying the shoelaces, the manner (with or without force) determines the resulting state (tight or loose); and it is typical of cases like this that the adjectives can alternate with adverbs with virtually no difference in meaning: . . .
> (Washio, 1997, p. 17)

This type of reasoning would also seem to apply to the Italian examples. Despite the considerable semantic overlap, however, there are some differences. While manner adverbs (*loosely, tightly*) are generally analyzed as predicates of events, Levinson (2010) calls the adjective version (*loose, tight*) a **pseudo-resultative** and argues that it is not a predicate of events, unlike the manner adverbs. She takes the adjective to be predicating of a created result/object and sees the pseudo-resultatives as only applying to root creation verbs, i.e., verbs which entail the creation of an entity denoted by the root of the verb.

Levinson (2010) also identifies such pseudo-resultatives in Finnish, Norwegian and Romance and shows that they have special morphology in those languages that sets them apart from both resultatives and manner adverbs.[3] We have more to say about the similarities between resultatives and agreeing adverbs in §4.

3 Adverb Agreement in Urdu, Sindhi and Punjabi

In this section we present the basic adverb agreement patterns found in Urdu, Sindhi and Punjabi.[4] As is predominantly the case in Romance, we find that undergoers are the target of agreement. As argued for by Levinson (2010), we find that these agreeing adjectives/adverbs seem to be related to resultatives, but are semantically distinct in the sense identified by Washio (1997).

3.1 Urdu

3.1.1 Standard Adverbs

Adverbs in Urdu take several different forms. They can be simple words like the temporal adverb *roz* 'daily' in (12a). Case marked NPs like *d^{h}ɪyan=se* 'with care' in (12b) can also function as adverbials. In fact, this is a very common strategy for the expression of event modification. Finally, adverbs can be based on a noun or adjective and be realized through masculine oblique inflection as shown in (12c) for the word *pahl-a/i* 'first'. In line with what is generally expected for adverbs, none agree with other elements of the clause.

(12) a. lɑṛki gaṛi roz cɑla-ti hɛ
girl.F.Sg.Nom car.F.Sg.Nom daily drive-Impf.F.Sg be.Pres.3.Sg
'The girl drives a/the car daily.'

b. lɑṛki gaṛi d^{h}ɪyan=se cɑla-ti hɛ
girl.F.Sg.Nom car.F.Sg.Nom care.M.Sg=Inst drive-Impf.F.Sg be.Pres.3.Sg
'The girl drives a/the car with care.'

c. lɑṛki gaṛi pahl-e cɑla-ti hɛ
girl.F.Sg.Nom car.F.Sg.Nom first.M.Sg.Obl drive-Impf.F.Sg be.Pres.3.Sg
'The girl drives a/the car first.'

[3]Levinson also shows that they are distinct from resultative adverbs as in (11) (cf. Geuder 2000).

(11) a. They decorated the room beautifully.
b. She dressed elegantly.
c. They loaded the cart heavily.

[4]Note that the major facts hold for Hindi as well, which is structurally almost identical to Urdu.

3.1.2 Agreeing Adverbs

The adverbs that are based on adjectives retain the gender and number agreement morphology of the adjective and agree with an argument of the clause. This can be seen in (13a) where *acc^h-i* 'good' modifies the verbal predication 'drive' but does not agree with the verb. Instead the AdjAdv agrees with the feminine object *gaṛi* 'car'. The AdjAdv is not able to agree with the (agentive) subject, see (13b), and when the object is non-nominative and therefore also not available for agreement as per the general verb agreement rules of the language, the AdjAdv does not agree with it, as shown in (13c). Instead, the AdjAdv shows default masculine singular.

(13) a. lɑṛka **gaṛi** **ɑcc^h-i** cɑla-ta hɛ
boy.M.Sg.Nom car.F.Sg.Nom good-F.Sg drive-Impf.M.Sg be.Pres.3.Sg
'The boy drives a/the car well.'

b. * lɑṛka gaṛi ɑcc^h-a cɑla-ta hɛ
boy.M.Sg.Nom car.F.Sg.Nom good-M.Sg drive-Impf.M.Sg be.Pres.3.Sg
'The boy drives a/the car well.'

c. lɑṛki gaṛi=ko **ɑcc^h-a** cɑla-ti hɛ
girl.F.Sg.Nom car.F.Sg=Acc good-M.Sg drive-Impf.F.Sg be.Pres.3.Sg
'The girl drives the car well.'

3.1.3 Verb Agreement

Agreement in Urdu is complex, with the verbal complex showing agreement for person, number and gender (Mohanan, 1994). When the subject is nominative (unmarked) as in (14a) then agreement is with the subject. When the subject is non-nominative and the object is nominative, as in (14b), agreement is with the object. When the subject and object both are overtly case marked, the verbal complex appears in the default masculine singular agreement as shown in (14c).[5]

(14) a. **lɑṛki** xɑt **lɪk^h-ti** **hɛ**
girl.F.Sg.Nom letter.M.Sg.Nom write-Impf.F.Sg be.Pres.3.Sg
'The girl writes a/the letter.'

b. lɑṛki=ne **xɑt** **lɪk^h-a** **hɛ**
girl.F.Sg=Erg letter.M.Sg.Nom write-Perf.M.Sg be.Pres.3.Sg
'The girl wrote a/the letter.'

c. lɑṛki=ne ciṭṭ^hi=ko lɪk^h-a hɛ
girl.F.Sg=Erg note.F.Sg=Acc write-Perf.M.Sg be.Pres.3.Sg
'The girl wrote the note.'

[5]Urdu shows Differential Object Marking with a semantically motivated nominative/accusative alternation by which the accusative marked object is specific (Butt, 1993).

The agreeing adverbs follow this agreement pattern in the sense that they only agree with nominative arguments (either subject or object). As shown below, a further constraint is that agreement only targets undergoers, as in Romance.

3.2 Sindhi

3.2.1 Standard Adverbs

The data for Sindhi mirror those for Urdu. Adverbs can be simple words as in (15a) or they can be realized via a case marked NP as in (15b). As in Urdu, this is a very common strategy for the expression of event modification. Adverbs may also be based on a noun or adjective, but bear masculine oblique inflection as in (15c). None of these adverbial types agrees with any element of the clause.

(15) a.

c^{h}lokiri	gaɗi	roz	hɑla-ɛ	t^{h}i
girl.F.Sg.Nom	car.F.Sg.Nom	daily	drive-Pres.3.Sg	be.Pres.F.Sg

'The girl drives a/the car daily.'

b.

c^{h}lokiri	gaɗi	d^{h}ɪyan=sañ	hɑla-ɛ	t^{h}i
girl.F.Sg.Nom	car.F.Sg.Nom	care.M.Sg=Inst	drive-Pres.3.Sg	be.Pres.F.Sg

'The girl drives a/the car with care.'

c.

c^{h}lokiri	gaɗi	pehr-eñ	hɑla-ɛ	t^{h}i
girl.F.Sg.Nom	car.F.Sg.Nom	first-M.Sg.Obl	drive-Pres.3.Sg	be.Pres.F.Sg

'The girl drives a/the car first.'

3.2.2 Agreeing Adverbs

Sindhi shows an identical overall pattern to Urdu. In (16a) *sʊʈ^h-i* 'good' is acting as an adverb in that it modifies the verbal predication. However, it agrees with the object and not the verb. (16b) shows that it cannot agree with the (agentive) subject and (16c) demonstrates the sensitivity towards overt case marking on the object. When the object is overtly marked, as in (16c), then the AdjAdv does not agree with it and instead is in the default masculine singular.

(16) a.

c^{h}okɪro	**gaɗi**	**sʊʈ^h-i**	hɑla-e	t^{h}o
boy.M.Sg.Nom	car.F.Sg.Nom	good-F.Sg	drive-Pres.3.Sg	be.Pres.M.Sg

'The boy drives a/the car well.'

b. *

c^{h}okɪro	gaɗi	sʊʈ^h-o	hɑla-e	t^{h}o
boy.M.Sg.Nom	car.F.Sg.Nom	good-M.Sg	drive-Pres.3.Sg	be.Pres.M.Sg

'The boy drives a/the car well.'

c.

c^{h}okɪri	gaɗi=k^{h}e	**sʊʈ^h-o**	hɑla-e	t^{h}i
girl.F.Sg.Nom	car.F.Sg=Acc	good-M.Sg	drive-Pres.3.Sg	be.Pres.F.Sg

'The girl drives the car well.'

3.2.3 Verb Agreement

Just like Urdu, Sindhi also shows agreement for person, number and gender within the verbal complex. Agreement patterns are identical to Urdu, although the morphology differs. Sindhi does not have an ergative case clitic, but ergative patterns are realized via the oblique morphological form of the noun. This can be seen in (17b) and (17c) where the oblique form of the noun *c^{h}okɪri-ɑ* 'girl' is used.

(17) a. **c^{h}okɪri** xatʊ **lɪk^{h}-e** **t^{h}i**
girl.F.Sg.Nom letter.M.Sg.Nom write-Pres.3.Sg be.Pres.F.Sg
'The girl writes a/the letter.'

b. c^{h}okɪri-ɑ **xatʊ** **lɪk^{h}-yo** **ahɛ**
girl.F.Sg-Obl letter.M.Sg.Nom write-Perf.M.Sg be.Pres.3.Sg
'The girl wrote a/the letter.'

c. c^{h}okɪri-ɑ gaɗi=k^{h}e hɑla-yo ahɛ
girl.F.Sg-Obl car.F.Sg=Acc drive-Perf.M.Sg be.Pres.3.Sg
'The girl drove the car.'

As in Urdu, the verb does not agree with the overtly marked subject and instead agrees with the unmarked object (17b). When the object is also overtly marked as in (17c), the verb is realized in the default masculine singular. The adverb agreement patterns shown in (16) are in line with the overall verb agreement patterns.

3.3 Punjabi

Punjabi shows the same patterns, modulo minor differences in the morphology and the lexicon. We nevertheless provide the data due to its under-researched nature.

3.3.1 Standard Adverbs

As in Urdu and Sindhi, standard adverbs can be realized as simple words (18a), case marked NPs (18b) or an adjective or noun in the invariant masculine oblique (18c). These adverbial expressions never agree with any element in the clause.

(18) a. kʊṛi gɑḍḍi roz cɑlañ-di ɛ
girl.F.Sg.Nom car.F.Sg.Nom daily drive-Impf.F.Sg be.Pres.3.Sg
'The girl drives a/the car daily.'

b. kʊṛi gɑḍḍi d^{h}ɪyan=nal cɑlañ-di ɛ
girl.F.Sg.Nom car.F.Sg.Nom care.M.Sg=Inst drive-Impf.F.Sg be.Pres.3.Sg
'The girl drives a/the car with care.'

c. kʊṛi gɑḍḍi pahl-añ cɑlañ-di ɛ
girl.F.Sg.Nom car.F.Sg.Nom first-M.Sg.Obl drive-Impf.F.Sg be.Pres.3.Sg
'The girl drives a/the car first.'

3.3.2 Agreeing Adverbs

(19a) and (19b) show that the adverb does not agree with an agentive subject (masculine), but instead agrees with the feminine object. When the object is overtly case marked, then the agreeing adverb carries default masculine singular morphology.

(19) a.

mʊṇḍa	**gɑḍḍi**	**cɑñg-i**	cɑlañ-da	ɛ
boy.M.Sg.Nom	car.F.Sg.Nom	good-F.Sg	drive-Impf.M.Sg	be.Pres.3.Sg

‘The boy drives a/the car well.’

b. *

mʊṇḍa	gɑḍḍi	cɑñg-a	cɑlañ-da	ɛ
boy.M.Sg.Nom	car.F.Sg.Nom	good-M.Sg	drive-Impf.M.Sg	be.Pres.3.Sg

‘The boy drives a/the car well.’

c.

kʊri	gɑḍḍi=nũ	**cañg-a**	cɑlañ-di	ɛ
girl.F.Sg.Nom	car.F.Sg=Acc	good-M.Sg	drive-Impf.F.Sg	be.Pres.3.Sg

‘The girl drives the car well.’

3.3.3 Verb Agreement

Again, the AdjAdv agreement is in line with the overall pattern of verb agreement whereby the verb only agrees with a nominative subject or object argument.

(20) a.

kʊṛi	xat	**lɪk^{h}-di**	**ɛ**
girl.F.Sg.Nom	letter.M.Sg.Nom	write-Impf.F.Sg	be.Pres.3.Sg

‘The girl is writing a/the letter.’

b.

kʊṛi=ne	**xat**	**lɪk^{h}-ya**	**ɛ**
girl.F.Sg=Erg	letter.M.Sg.Nom	write-Perf.M.Sg	be.Pres.3.Sg

‘The girl wrote a/the letter.’

c.

kʊṛi=ne	gɑḍḍi=nũ	cɑla-ya	ɛ
girl.F.Sg=Erg	car.F.Sg=Acc	drive-Perf.M.Sg	be.Pres.3.Sg

‘The girl drove the car.’

4 The Resultative Connection

The previous sections have established the basic patterns for Urdu, Sindhi and Punjabi and have surveyed existing cross-linguistic data. This section investigates follow-up questions resulting from the cross-linguistic comparison. For one, we demonstrate that the undergoer generalization established by Ledgeway for Romance also holds for our South Asian languages and that there is indeed a close connection between agreeing adverbs and resultatives. We further investigate whether there are systematic effects connecting targets of AdjAdv agreement with subject vs. object adverb orientation. We also investigate whether the presence of pseudo-resultatives always results in manner modification and whether the semantic connection posited by Washio (1997) holds.

Our data is based on native speakers, material found in grammars (Trumpp, 1872; Kellogg, 1893; Cummings & Bailey, 1912; McGregor, 1972), the Hindi-Urdu Treebank (HUTB; Bhatt et al., 2009; Palmer et al., 2009) and a new morphosyntactically annotated corpus for Urdu that is currently under development.[6]

4.1 Resultatives in Urdu, Sindhi and Punjabi

The syntactic equivalent of clear resultatives in Urdu, Sindhi and Punjabi are structurally very similar to the agreeing adverbs because resultative semantics in these languages are syntactically expressed via adjective-verb sequences (Ahmed et al., 2012).[7] Some examples are provided in (21)–(23).

(21) a. lɑṛke=ne **pani** **ṭ^{h}anḍ-a** k-iya
boy.M.Sg.Obl=Erg water.M.Sg.Nom cold-M.Sg do-Perf.M.Sg
'The boy made the water cool.' Urdu

b. lɑṛke=ne **caɪ** **ṭ^{h}anḍ-i** k-i
boy.M.Sg.Obl=Erg tea.F.Sg.Nom cold-F.Sg do-Perf.F.Sg
'The boy made the tea cool.' Urdu

(22) a. c^{h}okir-e **paɳi** **t^{h}ad^{h}-o** kɑ-yo
boy.M.Sg-Obl water.M.Sg.Nom cold-M.Sg do-Perf.M.Sg
'The boy made the water cool.' Sindhi

b. c^{h}okir-e **cañheñ** **t^{h}ad^{h}-i** kɑ-i
boy.M.Sg-Obl tea.F.Sg.Nom cold-F.Sg do-Perf.F.Sg
'The boy made the tea cool.' Sindhi

(23) a. mʊṇḍe=ne **paɳi** **ṭ^{h}anḍ-a** kit-a
boy.M.Sg.Obl=Erg water.M.Sg.Nom cold-M.Sg do-Perf.M.Sg
'The boy made the water cool.' Punjabi

b. mʊṇḍe=ne **ca** **ṭ^{h}anḍ-i** kit-i
boy.M.Sg.Obl=Erg tea.F.Sg.Nom cold-F.Sg do-Perf.F.Sg
'The boy made the tea cool.' Punjabi

---

[6]This corpus is being developed by Tafseer Ahmed and Mutee U Rahman as part of a DAAD funded German-Pakistan cooperation with the University of Konstanz. The corpus currently contains 8000 sentences, of which 1300 are annotated. The corpus is balanced across different genres.

[7]In Ahmed et al. (2012), some of these sequences are analyzed as complex predicates, others as syntactically biclausal resultatives; nothing hinges on this for the purposes of this paper.

Resultative adjectives always agree with the object, just as with the agreeing adjectives/adverbs. However, a crucial semantic difference is that the adjective only modifies the object and describes its resulting state.

We conclude that as in Romance, Urdu, Sindhi and Punjabi have agreeing adverbs that seem to be structurally connected to resultative adjectives. The agreeing adjectives/adverbs are morphosyntactically distinct from other manner adverbs and are also distinct from resultatives in their semantics.

4.2 Undergoer Sensitivity

In our South Asian languages, the target of AdjAdvs is limited to undergoers.[8] As shown in (24)–(26) for each of the languages, the AdjAdvs can in fact agree with a subject, but only with the subject of an unaccusative verb or the subject of a passivized clause. Crucially, the Adj/Adv does not agree with a subject of an unergative (see the c examples below) or a subject of an agentive transitive (cf. the examples in §3). Unlike what is reported by Ledgeway (2016) and Silvestri (2016), we have not found evidence of AdjAdv agreement with agentive subjects.

(24) a. **gaṛi** (laṛke=se) **ɑccʰ-i** cɑla-yi ja-ti hɛ
car.F.Sg.Nom boy.M.Sg.Obl=Inst good-F.Sg drive-Perf.F.Sg go-Impf.F.Sg be.Pres.3.Sg
'A/The car is driven well (by the boy).' Urdu passive

b. **roṭi** **ɑccʰ-i** pɑk-ti hɛ
bread.F.Sg.Nom good-F.Sg bake-Impf.F.Sg be.Pres.3.Sg
'(The) Bread bakes well.' Urdu unaccusative

c. ye laṛki **ɑccʰ-a** hɑns-ti hɛ
this girl.F.Sg.Nom good-M.Sg laugh-Impf.F.Sg be.Pres.3.Sg
'This girl laughs well/nicely.' Urdu unergative

(25) a. **gɑḍḍi** (mʊṇḍe=toñ) **cañg-i** cɑla-yi jañ-di ɛ
car.F.Sg.Nom boy.M.Sg.Obl=Inst good-F.Sg drive-Perf.F.Sg go-Impf.F.Sg be.Pres.3.Sg
'A/The car is driven well (by the boy).' Punjabi passive

b. **roṭi** **cañg-i** pɑk-di ɛ
bread.F.Sg.Nom good-F.Sg bake-Impf.F.Sg be.Pres.3.Sg
'(The) Bread bakes well.' Punjabi unaccusative

[8] Note that we use undergoer in the same broad sense as Ledgeway, i.e., in line with the notion of a Proto-Patient (Dowty, 1991) or the RRG notion of Undergoer (Van Valin & Polla, 1997).

c. e kʊṛi **cɑñg-a** hɑs-di ɛ
this girl.F.Sg.Nom good-M.Sg laugh-Impf.F.Sg be.Pres.3.Sg
'This girl laughs well/nicely.' Punjabi unergative

(26) a. (cʰokɪre=kʰã) **gaɗi** **sʊʈʰ-i** hɑla-ije tʰi
boy.M.Sg.Obl=Abl car.F.Sg.Nom good-F.Sg drive-Pass.Sg be.Pres.F.Sg
'The car is driven well (by the boy).' Sindhi passive

b. **mani** **sʊʈʰ-i** pac-e tʰi
bread.F.Sg.Nom good-F.Sg bake-Pres.3.Sg be.Pres.F.Sg
'The bread bakes well.' Sindhi unaccusative

c. hi cʰokɪri **sʊʈʰ-o** kʰil-e tʰi
this girl.F.Sg.Nom good-M.Sg laugh-Pres.3.Sg be.Pres.F.Sg
'This girl laughs well/nicely.' Sindhi unergative

4.3 Manner Modification

Overall, very few examples of agreeing adverbs can be found in existing resources and corpora. However, all the examples we have been able to find or think of involve manner modification. Examples (27)–(30) are representative.

(27) a. lɑrki **gana** **ɑccʰ-a** sun-ti hɛ
girl.F.Sg.Nom song.M.Sg.Nom good-M.Sg listen-Impf.F.Sg be.Pers.3.Sg
'The girl listens to the song very well.' Urdu

b. lɑrki **gɑzɑl** **ɑccʰ-i** sun-ti hɛ
girl.F.Sg.Nom song.F.Sg.Nom good-F.Sg listen-Impf.F.Sg be.Pers.3.Sg
'The girl listens to the gazal (type of song) very well.' Urdu

(28) a. **kʊri** **okʰ-i** ga-i
girl.F.Sg.Nom difficult-F.Sg go-Perf.F.Sg
'The girl walked with difficulty.' Punjabi

b. **kʊryã** **kaṭʰɪy-ã** ai-ỹa
girl.F.Pl.Nom together-F.Pl come-Perf.F.Pl
'The girls came together.' Punjabi

c. **gʰoṛi** **baṛ-i** **sohɳ-i** ṭʊr-di ɛ
mare.F.Sg.Nom big-F.Sg pretty-F.Sg walk-Impf.F.Sg be.Pres.3.Sg
'The mare walks [very] well/prettily.'[9] Punjabi
(Cummings & Bailey, 1912, Section II, p. 35)

[9]This example involves an additional agreeing modifier, *baṛ-i* 'big', which in turn agrees with the agreeing adverb. We note in passing that this modifier looks like an agreeing adverb itself, but do not go into detail concerning this additional agreement type.

(29) mʊṇḍa **kɑhaɳi** boht=i **cɑng-i** lɪk^{h}-da
boy.M.Sg.Nom story.F.Sg.Nom much=Foc good-F.Sg write-Impf.M.Sg
ɛ
be.Pres.3.Sg
'The boy writes a/the story very well.' Punjabi

(30) huɑ **xatu** ɗad^{h}o **sʊʈ^{h}-o** lik^{h}-ɛ
she.Nom letter.M.Sg.Nom much/very.M.Sg good-M.Sg write-Pres.3.Sg
t^{h}i
be.Pres.F.Sg
'She writes a letter very well.' Sindhi

4.4 Variation in Agreement and Available Readings

Recall that Italian showed some variation in agreement possibilities and that the variation coincides with interpretive possibilities (§2.2). In addition, in our South Asian languages, the position of the AdjAdv is the same as that of a resultative adjective. As such, a straight-forward resultative reading is also in principle possible in our language data. The following subsections investigate the relationship between agreement and interpretive possibilities for Urdu. The data for Sindhi and Punjabi are not provided, but are again parallel.

4.4.1 Same Agreement and Position but Semantic Variation

When an adjective is not modifying an argument, but is part of the verbal predication, the adjective always agrees with the undergoer. In examples like (31) this leads to ambiguity as posited by Washio (1997). The sentence can be interpreted either as a AdjAdv manner modification or as a resultative.

(31) **roṭi** **ɑcc^{h}-i** pɑk-i
bread.F.Sg.Nom good-F.Sg bake-Perf.F.Sg
'The bread baked well/nicely.' (AdjAdv) Urdu
'The bread came out well.' (resultative)

4.4.2 No Agreement: Only AdjAdv

In unergative clauses where there is no undergoer, adjectives can still be used to express adverbial modification, as in (32). In this case, the adjective is realized with default masculine singular morphology and the adverb is subject-oriented.

(32) amra ravi=se **ɑcc^{h}-a** bol-ti hɛ
Amra.F.Sg.Nom Ravi.M.Sg=Inst good-M.Sg speak-Impf.F.Sg be.Pres.3.Sg
'Amra speaks nicely to Ravi.' Urdu

4.4.3 Agreement Variation: Difference in Meaning

Recall that one possibility for the formation of adverbs was an adjective with oblique inflectional morphology (§3). In (33a) the oblique inflection signals an adverb expressing the manner of the overall baking. In (33b), on the other hand, the agreeing version gives rise to the AdjAdv manner modification.

(33) a. roṭi **ɛs-e** pɑk-i ke sɑb hɛran ho
bread.F.Sg.Nom such-M.Sg.Obl bake-Perf.F.Sg that all surprised be
gɑ-ɛ
go-Perf.M.Pl
'The bread baked in such a way that everyone was surprised.' Urdu
(manner of baking)

b. **roṭi** **ɛs-i** pɑk-i ke k^ha-i nɑ
bread.F.Sg.Nom such-F.Sg bake-Perf.F.Sg that eat-Perf.F.Sg not
gɑ-i
go-Perf.F.Sg
'The bread baked in such a way that (no one) was able to eat it.' Urdu
(AdjAdv modification primarily targeting the result state of the bread)

4.5 Summary

The undergoer generalization identified by Ledgeway for Italian holds for our South Asian languages. The close connection between AdjAdv (pseudo-resultatives in Washio and Levinson's terms) and resultatives suggested by Washio (1997) also holds. Furthermore, AdjAdv always express manner adverbs (cf. also Ledgeway 2011). There is variation in agreement patterns, but unlike in Southern Italian, we have not found agreement with agentive subjects. Our agreement effects involve true adverb vs. AdjAdv manner readings targeting the undergoer (§4.4.3) and subject-oriented adverbs with default masculine singular inflection (§4.4.2).

5 Analysis

In this section we develop an analysis of the South Asian facts. We suggest that the connection between resultative adjective and agreeing AdjAdv follows naturally from analyses previously posited for resultative adjectives in Urdu (Ahmed et al., 2012). We posit that the relationship is a diachronic one: the semantic overlap identified by Washio leads to the reinterpretation of a PREDLINK as a manner ADJUNCT over time. That is, an originally resultative construction with a fully-fledged adjective has its semantics shifted towards manner modification of an event. The original adjectival inflection is retained, though we assume that it will be lost over time or be subject to variation, as in Southern Italian (Silvestri, 2016). Our analysis differs from the proposals presented to date for pseudo-resultatives.

5.1 Synchronic Relationship via Movement

In providing an analysis for the Southern Italian data, Ledgeway (2011, 2016) posits a synchronic, derivational relationship between resultative adjectives and the corresponding agreeing adverbs.[10] For example, the agreeing adverb in (34a) is derived via movement from an underlying resultative construction as in (35).

(34) a. Anna miscava **bone** **'i** **carte**
Anna shuffled good.FPL the.PL cards.F
'Anna shuffled the cards well.' (the cards are shuffled well) Cosenza

b. Anna miscava **buonu** 'i carte
Anna shuffled good.MSG the.PL cards.F
'Anna shuffled the cards well.' (Anna shuffles well) Cosenza

(35) Anna miscava **bone**$_i$ ['i carte [$_{\text{RESULTP}}$ bone$_i$]]

Under this analysis, the agreeing adverbs are taken to show agreement because they entered a predicational relationship (ResultP) with the undergoer in the base position. The original adjective then moves further up in the tree to an adverbial position that modifies the event denoted by the VP. This position is responsible for the manner modification reading of the agreeing adverb.

Non-agreeing adjectives (34b), on the other hand, are taken to be base-generated in a higher adverb position. Because they never enter a direct relationship with the undergoer, there is no agreement relation.

(36) Anna miscava [**buonu** [$_{\text{VP}}$ 'i carte]]

Italian shows word order effects in that different positions of an adverb signal scopal relationships. Ledgeway's account makes crucial use of this. We have not found similar word order effects in our South Asian languages.

5.2 Direct Predication Modification

Levinson (2010) posits a different relationship by which pseudo-resultatives (our AdjAdv) can modify the root of a verb at a certain point in the derivation, thus resulting in the event modification reading. The target of agreement is assumed to be the closest NP in the tree. A crucial part of her account is that only root creation verbs can be involved as these show the right resultative semantics. However, the data from Urdu, Sindhi and Punjabi as well as Southern Italian show that agreeing adverbs in these languages are not confined to root creation verbs. Levinson (2010) also does not consider the undergoer connection noted by Ledgeway, which provides a broader generalization for the data.

[10] Silvestri (2016) broadly follows Ledgeway's analysis but works out more precise details accounting for agreement variation.

5.3 Our Proposal

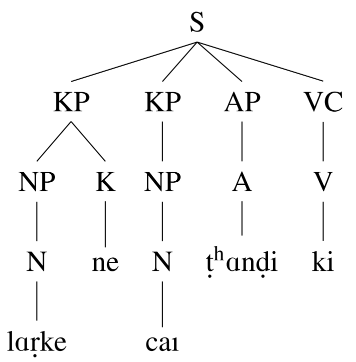
Ledgeway posits a synchronic derivational relationship between resultatives and pseudo-resultatives. We posit a diachronic relationship, but do note that the compilation of concrete diachronic evidence remains to be done.

5.3.1 Resultative Adjectives — PREDLINK

We see the resultative as the historically prior construction and follow Ahmed et al. (2012) in our analysis. They posit the c-structure in (38) and the f-structure in (39) for a resultative as in (37).[11]

(37) laṛke=ne **caɪ** **ṭʰanḍ-i** k-i
boy.M.Sg.Obl=Erg tea.F.Sg.Nom cold-F.Sg do-Perf.F.Sg
'The boy cooled the tea.' (lit. 'The boy did the tea cool.') Urdu

(38)



(39)

$$\left[\begin{array}{ll} \textsc{pred} & \text{`kar<}\textsc{subj obj predlink}\text{>'} \\ \textsc{subj} & \left[\begin{array}{ll} \textsc{pred} & \text{`laṛka'} \\ \textsc{num} & \text{sg} \\ \textsc{gend} & \text{masc} \end{array}\right] \\ \textsc{obj} & \left[\begin{array}{ll} \textsc{pred} & \text{`caɪ'} \\ \textsc{num} & \text{sg} \\ \textsc{gend} & \text{fem} \end{array}\right] \\ \textsc{predlink} & \left[\begin{array}{ll} \textsc{pred} & \text{`ṭʰanḍi'} \\ \textsc{atype} & \text{predicative} \\ \textsc{num} & \text{sg} \\ \textsc{gend} & \text{fem} \end{array}\right] \end{array}\right]$$

We assume a verb frame augmentation approach (cf. Christie 2010) by which the argument frame of semantically suitable (agentive, transitive) verbs is augmented with a predicative element in the right syntactic and semantic conditions (cf. Wechsler 1995, 2005). In the syntax, this predicative element is a PREDLINK containing the resultative adjective. Via the standard PREDLINK analysis (Butt et al., 1999), the adjective *ṭʰanḍi* 'cold' in (37) is analyzed as being predicated of *cai* 'tea'. Together with the verb *kar* 'do', this has the semantics of a resultative.

The agreement facts are accounted for by way of general agreement rules governing PREDLINKs. Whether and how a PREDLINK agrees with the entity it is predicating over is governed by language-dependent factors (Butt et al., 1999). In

[11] Adj-V combinations as in (34) have traditionally been seen as complex predicates (e.g., Mohanan 1994). However, there is little structural evidence for this. Unlike with standard complex predicates in Urdu, the adjective is not contained within the verbal complex (VC) and the adjective is not licensed by any other element in the clause, nor does it contribute to the overall argument structure of the clause. Also see the Complex Predicate Reference Bank: http://ling.uni-konstanz.de/pages/home/pargram_urdu/main/Resources.html#CP

Urdu, adjectival PREDLINKs agree in gender and number with the argument they are predicating over. In our South Asian languages there is thus agreement between the adjective and the predicated-of noun as a reflex of the general predication relation. Under the proposal of Ahmed et al. (2012), the PREDLINK in Urdu agrees with and predicates of the lowest grammatical function in the f-structure as per the grammatical function hierarchy in Bresnan (2001). In (39), this is the OBJ. Agreement is realized as part of the lexical entry of the resultative adjective, using standard inside-out functional constraints as discussed by e.g., Attia (2008).

Our analysis is in line with how predicatives are analyzed generally under the PREDLINK proposal (e.g., Attia 2008, Laczkó 2012). However, resultatives have traditionally been analyzed as containing an XCOMP rather than a PREDLINK within LFG (and the equivalent in HPSG), as first proposed by Simpson (2006) (e.g., Christie 2010, also Müller (2002, 2016) and references therein).

If predicatives are treated as a PREDLINK (Butt et al., 1999), then resultatives should also be analyzed in terms of a PREDLINK since they also instantiate a predication relation, albeit a specific one. Conversely, under the XCOMP proposal, both predicatives and resultatives would be treated as XCOMPs. However, under an XCOMP proposal, *cai* 'tea' in (39) would function as an embedded subject within the resultative XCOMP headed by *ṭ^{h}anḍi* 'cold'. At least for the South Asian languages examined here, there is no evidence for such embedded subjects, i.e., *cai* 'tea' in (37) shows no evidence of subjecthood (via the tests established for Urdu/Hindi by Mohanan (1994), cf. also Attia (2008) and Laczkó (2012)).

5.3.2 AdjAdvs (Agreeing Adverbs/Adjectives)

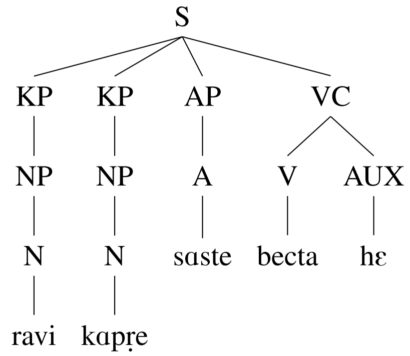
We analyze AdjAdvs much along the same lines, with the major difference being that the agreeing AdjAdv does not give rise to a PREDLINK but to a clause-level adjunct at f-structure. (41) and (42) provide a c-structure and an f-structure analysis for (1), repeated in (40). Under this analysis, the AdjAdv is treated at c-structure as an adjective directly attaching as a daughter to the clause. The difference lies in the f-structure, where the adjective corresponds to a clause level ADJUNCT.

(40)	rɑvi	**kɑpṛe**	**sɑst-e**	bec-ta	hɛ
	Ravi.M.Sg.Nom	clothes.M.Pl.Nom	cheap-M.Pl	sell-Impf.M.Sg	be.Pres.3.Sg

'Ravi sells clothes cheap.' Urdu
(he sells them cheaply, the clothes are not inherently cheap)

We posit that diachronic change has caused an f-structural reanalysis. The reanalysis is due to the overlapping semantics of a resultative and a manner adverb as described by Washio in that the manner of performing an action determines the resultative state. The diachronic change ultimately is predicted to also give rise to a c-structural reanalysis so that the original predicative adjective is reanalyzed as a manner adverb. However, at this stage, the f-structural function is one of clause level manner modification, but at c-structure the original part of speech is still retained. This explains why a manner modifier displays agreement with an argument

(41)



(42)

$$\begin{bmatrix} \text{PRED} & \text{‘bec<SUBJ OBJ>’} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘rɑvi’} \\ \text{NUM} & \text{sg} \\ \text{GEND} & \text{masc} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{‘kɑpṛa’} \\ \text{NUM} & \text{pl} \\ \text{GEND} & \text{masc} \end{bmatrix} \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PRED} & \text{‘sɑst’} \\ \text{MOD-TYPE} & \text{manner} \\ \text{NUM} & \text{pl} \\ \text{GEND} & \text{masc} \end{bmatrix} \right\} \end{bmatrix}$$

of the clause — the agreement is a reflex of the original resultative adjective agreement that targets undergoers. This agreement is predicted to be lost or regularized over time and to be subject to variation during the diachronic change, as in Southern Italian. In terms of the annotation of the agreement pattern, we can use the same inside-out functional constraints as with the resultatives, with the slight difference that the inside-out path starts out from an ADJUNCT set item instead of a PREDLINK function.

6 Summary and Conclusions

In this paper we have presented novel data from Urdu, Sindhi and Punjabi; we have argued that these languages contain agreeing adverbs that are originally based on predicational adjectives. The data parallels the patterns found in Southern Italian dialects to a great extent. Given the crosslinguistic similarities, it seems that the same syntactic and semantic forces are at play across languages. The overlap of meaning between specifying the result state of an undergoer and modifying the overall result subevent opens up the possibility for a manner adverbial meaning of originally resultative adjectives. In our analysis, this subtle but significant difference is captured at the level of f-structure in a difference of the functional contribution: PREDLINK vs. ADJUNCT. The agreement targeting undergoers by the AdjAdvs is taken to be a reflex of the original predicative structure and we predict that it will be lost over time and that the adjective will then be reanalyzed as a straightforward manner adverb at c-structure.

References

Ahmed, Tafseer, Miriam Butt, Annette Hautli & Sebastian Sulger. 2012. A Reference Dependency Bank for Analyzing Complex Predicates. In *Proceedings of LREC12*, Istanbul, Turkey.

Alexiadou, Artemis. 1997. *Adverb Placement: A case study in antisymmetric syntax*. Amsterdam: John Benjamins.

Anderson, Stephen. 1985. Inflectional morphology. In Timothy Shopen (ed.), *Grammatical categories and the lexicon*, vol. III Language Typology and Syntactic Description, 150–201. Cambridge: Cambridge University Press.

Attia, Mohammed. 2008. A Unified Analysis of Copula Constructions in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG08 Conference*, 89–108. CSLI Publications.

Bhatt, Rajesh, Bhuvana Narasimhan, Martha Palmer, Owen Rambow, Dipti Misra Sharma & Fei Xia. 2009. A Multi-Representational and Multi-Layered Treebank for Hindi/Urdu. In *Proceedings of the Third Linguistic Annotation Workshop*, 186–189. Suntec, Singapore: Association for Computational Linguistics.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Butt, Miriam. 1993. Object specificity and agreement in Hindi/Urdu. In *Papers from the 29th Regional Meeting of the Chicago Linguistic Society*, 80–103.

Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi & Christian Rohrer. 2002. The Parallel Grammar Project. In *Proceedings of the COLING-2002 Workshop on Grammar Engineering and Evaluation*, 1–7.

Butt, Miriam & Tracy Holloway King (eds.). 2001. *Time Over Matter: Diachronic Perspectives on Morphosyntax*. Stanford: CSLI Publications.

Butt, Miriam, Tracy Holloway King, María-Eugenia Niño & Frédérique Segond. 1999. *A Grammar Writer's Cookbook*. Stanford: CSLI Publications.

Christie, Elizabeth. 2010. Using Templates to Account for English Resultatives. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG10 Conference*, 155–164. Stanford: CSLI Publications.

Cummings, Thomas & T. Grahame Bailey. 1912. *Panjabi Manual and Grammar: A Guide to the Colloquial Panjabi of the Northern Panjab*. Calcutta: Baptist Mission Press.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Evans, Nicholas. 2000. Word classes in the world's languages. In Geert E. Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphology*, vol. 17 Handbooks of Linguistics and Communication Science, chap. X, 708–731. De Gruyter.

Geuder, Wilhelm. 2000. *Oriented Adverbs: Issues in the Lexical Semantics of Event Adverbs*: Universität Tübingen dissertation.

Kellogg, S. H. 1893. *Grammar of the Hindi Language*. Delhi: Munshiram Manoharlal Publishers Pvt. Ltd. Second Edition, reprinted 1990.

Laczkó, Tibor. 2012. On the (Un)Bearable Lightness of Being an LFG Style Copula in Hungarian. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference*, 341–361.

Ledgeway, Adam. 2011. Adverb Agreement and Split Intransitivity: Evidence from southern Italy. *Archivio glottologico italiano* 96. 31–66.

Ledgeway, Adam. 2016. Parameters in Romance Adverb Agreement. In Martin

Hummel & Salvador Valera (eds.), *Adjective–Adverb Interfaces in Romance*, Amsterdam: Benjamins. In Press.

Levinson, Lisa. 2010. Arguments for pseudo-resultative predicates. *Natural Language and Linguistic Theory* 8(1). 135–182. doi:10.1007/s11049-010-9089-x.

McGregor, R. S. 1972. *Outline of Hindi grammar: with exercises*. Oxford: Clarendon Press.

Mohanan, Tara. 1994. *Argument structure in Hindi*. Stanford: CSLI Publications.

Müller, Stefan. 2002. *Complex Predicates: Verbal Complexes, Resultative Constructions and Particle Verbs in German*. Stanford: CSLI Publications.

Müller, Stefan. 2016. Phrasal constructions, derivational morphology, constituent structure and (cross-linguistic) generalizations: A discussion of template-based phrasal LFG approaches and a lexical HPSG alternative. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway-King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, CSLI Publications.

Palmer, Martha, Rajesh Bhatt, Bhuvana Narasimhan, Owen Rambow, Dipti Misra Sharma & Fei Xia. 2009. Hindi Syntax: Annotating Dependency, Lexical Predicate-Argument Structure, and Phrase Structure. In *Proceedings of the 7th International Conference on Natural Language Processing (ICON)*, 259–268.

Silvestri, Giuseppina. 2016. Adverb agreement in the dialects of the Lausberg area. In Martin Hummel & Salvador Valera (eds.), *Adjective–adverb interfaces in Romance*, Amsterdam: Benjamins. In Press.

Simpson, Jane. 2006. Resultatives. In *Lexical semantics in LFG*, 149–161. Stanford: CSLI Publications. Originally published as part of *Papers in Lexical Functional Grammar*, 1983.

Sulger, Sebastian, Miriam Butt, Tracy Holloway King, Paul Meurer, Tibor Laczkó, György Rákosi, Cheikh Bamba Dione, Helge Dyvik, Victoria Rosén, Koenraad De Smedt, Agnieszka Patejuk, Özlem Çetinoglu, I Wayan Arka & Meladel Mistica. 2013. ParGramBank: The ParGram Parallel Treebank. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 550–560. Sofia, Bulgaria: Association for Computational Linguistics. http://www.aclweb.org/anthology/P13-1054.

Trumpp, Ernest. 1872. *Grammar of the Sindhi Language*. Osnabrück: Biblio Verlag. Reprinted 1970.

Van Valin, Robert D. & Randy J. La Polla. 1997. *Syntax: Structure, Meaning and Form*. Cambridge: Cambridge University Press.

Washio, Ryuichi. 1997. Resultatives, compositionality and language variation. *Journal of East Asian Languages* 6. 1–49.

Wechsler, Stephen. 1995. *The Semantic Basis of Argument Structure*. Stanford: CSLI Publications.

Wechsler, Stephen. 2005. Resultatives Under the Event-Argument Homomorphism Model of Telicity. In Nomi Erteschik-Shir & Tova Rapoport (eds.), *The Syntax of Aspect*, 255–273. Oxford: Oxford University Press.

Preliminary results from the Free Linguistic Environment project

Damir Ćavar
Indiana University

Lwin Moe
Indiana University

Hai Hu
Indiana University

Kenneth Steimel
Indiana University

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 161–181

Keywords: parser, Free Linguistic Environment, implementation, grammar engineering, natural language processing

Ćavar, Damir, Lwin Moe, Hai Hu & Kenneth Steimel. 2016. Preliminary results from the Free Linguistic Environment project. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 161–181. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.9.

Abstract

The Free Linguistic Environment (FLE) project focuses on the development of an open and free library of natural language processing functions and a grammar engineering platform for Lexical Functional Grammar (LFG) and related grammar frameworks. In its present state the code-base of FLE contains basic essential elements for LFG-parsing. It uses finite-state-based morphological analyzers and syntactic unification parsers to generate parse-trees and related functional representations for input sentences based on a grammar. It can process a variety of grammar formalisms, which can be used independently or serve as backbones for the LFG parser. Among the supported formalisms are Context-free Grammars (CFG), Probabilistic Context-free Grammars (PCFG), and all formal grammar components of the XLE-grammar formalism. The current implementation of the LFG-parser includes the possibility to use a PCFG backbone to model probabilistic c-structures. It also includes f-structure representations that allow for the specification or calculation of probabilities for complete f-structure representations, as well as for sub-paths in f-structure trees. Given these design features, FLE enables various forms of probabilistic modeling of c-structures and f-structures for input or output sentences that go beyond the capabilities of other technologies based on the LFG framework.

1 Introduction

Our main motivation to launch the FLE project was to get access to a free and open parser environment for research and educational purposes. It aims to develop a grammar engineering platform for Lexical-Functional Grammar (LFG) (Kaplan & Bresnan, 1982; Bresnan, 2001; Dalrymple, 2001) and related grammar frameworks. The long-term goal is to create an open and platform-independent system that facilitates the testing of algorithms and formal extensions of the LFG framework. Among others, our interests are to merge current technologies and approaches in Natural Language Processing (NLP) with an LFG parser environment, and to integrate semantic analysis in the resulting computational environment.

The FLE project is motivated by a variety of concerns. One is to experiment with new algorithms within the LFG framework that can facilitate probabilistic modeling, as suggested in Kaplan (1996) and elsewhere.[1] The FLE environment should make it possible to experiment with probabilistic models and extensions to the classical LFG framework, as described below. Consequently, probabilistic LFG models would allow us to extend the spectrum of application in NLP and HLT, to address new research questions, and to boost grammar development and engineering using machine learning strategies and treebanks.

[†]We are grateful to Ron Kaplan, Ken Beesley, Lionel Clément, Larry Moss, Mary Dalrymple, Agnieszka Patujek, Adam Przepiórkowski, Paul Meurer, Helge Dyvik, Annie Zaenen, Valeria de Paiva for many helpful suggestions, data sets, grammar samples, ideas, and comments.

[1]While a reviewer pointed out that there are many approaches to quantitative LFG, to our knowledge there is no parser platform that implements probabilistic c-structure and f-structure generation or processing.

Besides providing an environment to test different algorithms and approaches to parsing natural language sentences with LFG-based grammars, one purpose of the project is to create a grammar engineering platform that integrates better in common operating systems and computing environments.

For language documentation projects, in particular work on under-resourced and endangered languages, we need a platform that is not only usable on tablets and mobile thin-computers like Chromebooks, but also one that is easy to use for grammar engineers without strong technological skills.

While we see a need for a parser and grammar engineering environment that provide help to documentary linguists and grammar writers, we also see a need for efficient implementations that are scalable, parallelized and distributed. This might have been a major concern by many parser developers in the past. Given the ongoing changes in computing hardware, infrastructure, and environments, it is a permanent challenge to constantly adapt algorithms and code to be able to benefit from ongoing innovations. By providing a library of atomic functions, we hope to create an architecture that we, and others, can subsequently optimize with respect to these goals.

Our ultimate goal is to integrate semantic processing and computational components in some future version of the library and the resulting parsers.

2 Existing LFG Grammar Engineering Platforms

There are various grammar engineering platforms on the market. In the following we will concentrate only on the LFG related platforms and component software environments that could benefit from those.

The Xerox Linguistic Environment (XLE) is the most significant and complete implementation of the LFG framework in a grammar engineering environment. It is accompanied by a large amount of documentation in form of textbook sections (e.g. Butt et al. (1999)), online documents (e.g. Crouch et al. (2011)), and academic publications (e.g. Maxwell & Kaplan (1996a)).

The FLE-project aims at reaching compatibility with XLE. XLE is not freely available. This prevents us from studying and teaching the particular algorithms related to LFG-parsing. The graphical environment of XLE has a rather vintage appearance with limited grammar engineering functionalities. It is available for Unix-based and – in an older version also – for Windows operating systems.

Another LFG-based grammar engineering environment is XLFG (Clément, 2016; Clément & Kinyon, 2001). It comes as a web-based platform that can be used in a browser window. It is not openly available and it is accompanied by only sparse documentation. We have been discussing the possibility of integrating the grammar formalism supported by XLFG in the FLE environment.

There is a growing number of environments that are used for the application of NLP components, for example, tokenizers, parsers, named entity recognizers, etc. However, none of these, to our knowledge, has integrated LFG parsers.

The Unstructured Information Management Architecture (UIMA) (Ferrucci & Lally, 2004), for example, is a component software architecture that allows speech and language processing engineers to define a processing pipeline of NLP components for the analysis of data sets (e.g. texts or even audio recordings), and the aggregation and visualization of analysis results. All FLE components can be used within an UIMA-based application, if the necessary wrappers for the UIMA infrastructure are provided.

The General Architecture for Text Engineering (GATE) (The GATE Team, 2011) is another such environment that is not geared towards the engineering of morphologies, LFG-based grammars, or specific NLP components, but rather towards the application of such components on some textual data. The integration of FLE components in the GATE environment is possible.

Existing educational and experimental environments for NLP, like the Natural Language Toolkit (NLTK) (Loper & Bird, 2002; Bird et al., 2009) provide various algorithms and tools implemented in the Python programming language. There is no LFG parser or adequate morphological analyzer integrated in NLTK yet. NLTK components are coded in Python and not necessarily tuned for efficiency and use with large data sets. The NLTK license also excludes commercial use and thus significantly differs from the FLE license. It is possible, though, to create an FLE-module for Python that uses the compiled C++ classes that are part of FLE.

3 Architecture

While the development of FLE focuses on providing a collection of algorithms necessary to read LFG formalisms and parse with them, we do implement experimental parser pipelines or fragments of those to test our algorithms. Currently, there is one experimental setting and implementation of FLE that uses a pipeline architecture for processing which consumes an input sentence, tokenizes it, and syntactically parses it on the basis of morphological analyses of the lexical items using different kinds of chart parser implementations:

Input Sentence → Tokenizer → Morphological Analyzer →
Syntactic Parser → c-structure & f-structure

As discussed in the next subsections, this is a very common way to arrange linguistic processing components in most common NLP architectures that involve raw sentence input processing and syntactic parsing. Our goal is not to provide a component architecture for linguistic processing modules, but to provide the library functionalities that can be arranged in different general architectures that enable LFG type parsing of raw sentence input.

The pipeline architecture above is common in Natural Language Processing applications. For a more cognitively or psycho-linguistically adequate model we can also arrange the language processing components in a parallel fashion, as proposed in Jackendoff (1997, 2007). In this kind of architecture the individual language components generate representations that are mapped or synchronized with representations generated by other language processing components. The mapping constraints can be implemented as selection functions or constraints over possible representations. An implementation of such a parallel processing environment can be achieved using a blackboard or alternative message passing architecture.[2]

In the following sections, we will briefly describe the current state of integration of NLP components necessary for any kind of processing chain.

3.1 Grammar Formalisms

A prerequisite for any grammar-based parser is at least one grammar formalism that specifies the format for rules as supported by the parser. The grammar formalisms that FLE supports include a generic CFG-formalism, two different PCFG-formalisms, and the full set of XLE grammar components.

The CFG-formalism allows for the use of regular expression operators in the right-hand-side of rules, such as `*`, `+`, and `?`, round brackets for grouping of symbols, and curly brackets with the disjunction operator |, as in the following example:

```
S --> NP VP
NP --> {Art|Q} N
NP --> {Art+|Q*} N
VP --> ( Adv )* V
```

The use of regular expression operators is similar to the rule formalism used in XLE. The right-hand-sides of the rules are mapped on specific Finite State Machines in FLE that include cyclic paths or recursion, optionality, and disjunction. The XLE parser uses a similar CFG-backbone.

An extension of this formalism allows additional augmentation of rules with probabilities. This could be considered a version of a PCFG, in which the probability is associated with a complex right-hand side that may contain the regular expression operators as described above.[3]

```
1.0 S --> NP VP
0.7 NP --> {Art|Q} N
0.3 NP --> {Art+|Q*} N
1.0 VP --> ( Adv )* V
```

For compatibility reasons we have added the NLTK supported PCFG-formalism to FLE. This formalism uses CFG-rules in the Chomsky Normal Form (CNF). In this format, right-hand-sides of rules are augmented with their particular probability in square brackets. The right-hand-sides are optionally grouped to their corresponding left-hand-side symbol using the disjunction operator |.

[2]See Erman et al. (1980), Corkill (1991), or Hayes-Roth (1985) for some such solutions.

[3]Note that the interpretation of the probabilities in this formalism is open and can be used in many different ways. A discussion of some of the possibilities to interpret and use these probabilities would be beyond the scope of this paper.

```
S -> NP VP [1.0]
NP -> Det N [0.5] | NP PP [0.25] | 'John' [0.1]
NP -> 'I' [0.15]
Det -> 'the' [0.8] | 'my' [0.2]
N -> 'man' [0.5] | 'telescope' [0.5]
VP -> VP PP [0.1] | V NP [0.7] | V [0.2]
V -> 'ate' [0.35] | 'saw' [0.65]
PP -> P NP [1.0]
P -> 'with' [0.61] | 'under' [0.39]
```

As we will discuss in more detail, the inclusion of the two PCFG-type formalisms allows us to extend the back-end of FLE to include probabilities for c-structure representations in LFGs and the parsing algorithm.

3.1.1 XLE Grammar Formalism Parser

In addition to the existing grammar formalisms as described above, FLE provides parsers for all XLE-based grammar sub components. XLE grammars potentially consist of multiple specifications of sections that are labeled: CONFIG, FEATURES, LEXICON, MORPHOLOGY, MORPHTEXT, RULES, TEMPLATES. Each of these sections uses a specific formalism or language to specify configuration parameters for the parser, the features used in the lexicon, morphology, and syntax, or LFG rules. A description or explanation of all these sections is beyond the scope of this article. Please consult the XLE documentation online (Crouch et al., 2011) for more details. In the following sections, we will briefly discuss the parsers for the XLE RULES and FEATURES.

The following rules are taken from the toy-grammar that is part of the XLE documentation.

```
S --> e: (^ TENSE);
    (NP: (^ XCOMP* {OBJ|OBJ2})=!
         (^ TOPIC)=!)
     NP: (^ SUBJ)=!
         (! CASE)=NOM;
    { VP | VPaux }.
```

The following LBNF rules represent a segment of the LBNF-specification for the RULES section:

```
Grammar.        GRAMMAR ::= [RULE] ;
RuleS.          RULE    ::= LHS "-->" [RHS] RULEES ;
RuleS2.         RULE    ::= LHS "=" [RHS] RULEES ;
RuleEndSymbol. RULEES  ::= ".";
RHSSymbolOptional.    RHS ::= "(" RHSSYMBOL ")" ;
RHSymbolsDisjunction. RHS ::= "{" [ORHS] "}" ;
```

These arbitrarily selected LBNF rules display the specification of the `GRAMMAR` as a list of `RULE` symbols. Each rule is specified as a left-hand-side `LHS` followed by a production symbol/string “=” or “->” and a list of right-hand-side symbols RHS. Each rule is terminated by a RULEES symbol, the rule-end symbol. Some examples for right-hand-side symbols are given in the final two rules, i.e. an optional symbol or a list of disjoint symbols.

As described in the previous section, FLE allows for different types of grammar backbones. It can be used with a CFG or PCFG backbone. The PCFG-capability allows for the use of grammars that can be handcrafted or extracted and trained from common treebanks. The internal encoding of the grammar is implemented atop a Weighted Finite State Transducer (WFST) (Berstel & Reutenauer, 1988; Kuich & Salomaa, 1986; Salomaa & Soittola, 1978).[4]

The WFST-architecture of the grammar backbone provides a simple way to not only encode a probabilistic backbone, but also to encode and process independent lexical properties, feature specifications and related constraints in a probabilistic manner. It allows for extended probabilistic models to be applied to the transitions via weights, weight functions, or objects that encapsulate complex weight functions. For example, the weight functions can entail unification or even semantic operations.

In FLE, the parsed XLE RULES section of a grammar is mapped onto a WFST as shown in Figure 1. The upper tape in the WFST in Figure 1 is represented by the initial symbol preceding the first “/”, the lower tape corresponds to the embedded symbol between two “/”. In this example, an initial probability or weight of 1 is assigned to all transitions, as represented by the final element following the final “/”. These probabilities or weights can be changed, “trained”, and subsequently tuned, if the weights are not already provided by the grammar. For example, the weights are provided when a PCFG-backbone is utilized.

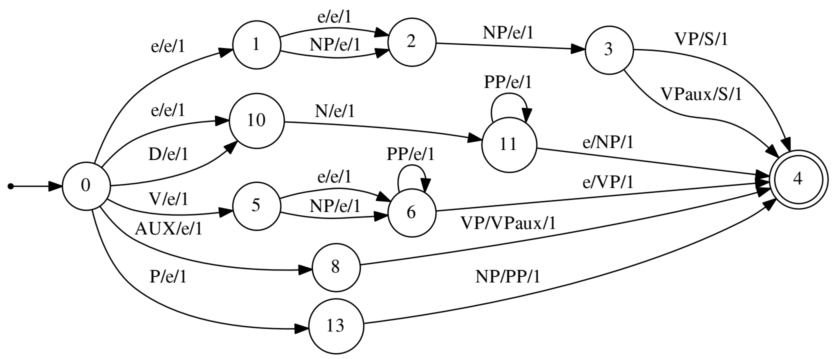


Figure 1: Mapping a CFG on a WFST

F-structures, as specified in the toy-grammar RULES section above, are not displayed in this WFST in Figure 1. These are left out for simplification reasons.

[4]See (Mohri, 2004) for an overview.

We discuss feature representations and unification in section 3.4 below.

Notice that the WFST in Figure 1 contains epsilon transition symbols on the upper tape, the lower tape, or on both tapes of the transducer. These are either specified in the grammar, as for example, in the sentence rule (“`S -> e (NP) NP { VP | VPaux }`”) above, or they are inserted by the grammar-to-WFST mapping algorithm to express, among other properties, optionality.

Given the compact representation of the grammar rules as a WFST, we are able to model LFG grammars for bottom-up, top-down, and probabilistic parsing or generation. That is, the upper or lower tape can be used for parsing or generation. The upper tape paths represent the right-hand-sides and the lower-tape – the left-hand sides of the rules from a CFG-backbone. A top-down parser would use the lower tape. A bottom-up parser would use the upper tape to match input sequences or edges and then replace them with their counterpart on the opposite tape. The lower tape could be used to replace left-hand-side symbols with the corresponding right-hand-sides to generate output c-structures and sentences (assuming an existing f-structure and corresponding mapping functions).

In addition to the flexibility of this grammar representation for parsing and generation, we add the capability of probabilistic modeling of c-structures. The transitions in the WFST could be weighted with probabilities that correspond to rule probabilities in PCFGs. Alternatively, they could be used as pure transitional probabilities for symbols and paths without any PCFG association that may be learned from corpora or parser application. In this way we can represent the CFG-backbone of the LFG grammar as a probabilistic model. Simply put, if we map PCFG rule-probabilities on transition probabilities in the WFST, we can calculate the probability of a c-structure as the product of the probabilities of the rules used to generate or parse it. There are various ways in which the weights in a WFST can be used or specified to enable probabilistic modeling of c-structures in LFGs. We will discuss various options in subsequent work.

Note that the WFST-based model provides more possibilities and potential extensions to the grammar backbone. As discussed above, weights in the WFST could be used as transitional probabilities or rule probabilities mapped on transition probabilities to estimate the likelihood of c-structures. Weights could also be complex functions and data-structures that, for example, combine probabilities with f-structures and unification operations. At the same time, the CFG backbone can be extended by allowing the lower tape, i.e. the CFG-rule’s left-hand-side to be extended. This data structure would allow for the formulation and use of context-sensitive rules by specifying contexts and even f-structures of left-hand-side symbols. We will have to postpone the discussion of such extensions to future work.

3.1.2 Formal Specification and Parser Generation

The grammar formalisms that FLE can read are specified using an extension of the Backus-Naur Form (BNF).[5] A BNF specification of a grammar contains a set of replacement rules. The rules in the sample below are interpreted as: a.) a `Grammar` consists of a list of `Rule` symbols, b.) a symbol surrounded by square brackets is interpreted as a list, c.) a `Rule` consists of a `LHS` (left-hand-side) symbol, d.) the production symbol `->` as a literal string, and e.) a list of `RHS` (right-hand-side) symbols. In our formal specification of grammar formalisms we used the the Labeled Backus-Naur Form (LBNF) formalism (Forsberg & Ranta, 2005). The LBNF formalism is used by the BNF Converter (BNFC) (Forsberg & Ranta, 2004) to generate parser code in a variety of programming languages. The labels here are prefixes to the rules, e.g. `Gram` and `Rul` separated from the rule by a "`.`". They are used for constructing a syntax tree that contains sub-trees that are defined by the non-terminals of the labeled rule.

```
Gram. Grammar ::= [Rule] ;
Rul.  Rule    ::= LHS "-->" [RHS] ;
```

The BNF Converter (BNFC) is independent software written in Haskell that generates code for the programming languages C, C++, Java and Haskell.[6] Many other languages can and surely will be added to the converter. The generated code constitutes a functioning syntactic parser for the grammar formalism that generates a parse tree for the grammar and rules in the specific formalism. These parse trees need to be mapped to data structures for processing in FLE. This is what we refer to as the semantics of the formalism parsers.

Since the FLE code-base is based on C and C++, we use the BNFC conversion to C++ for our grammar formalism parsers. The LBNF specification is converted to freely available lexer and parser generators *flex*[7] and *bison*.[8] The generated *flex* and *bison* code is wrapped in a C++ class using the Visitor Design Pattern.[9]

Specifying the grammar formalisms in LBNF and using BNFC for parser code generation has many advantages. The LBNF-specification of the formalisms for CFG, PCFG, and XLE grammar components provides an intuitive and easy understandable representation that allows for convenient adaptations, extensions, corrections and changes. The BNFC generated code and the Visitor Design Pattern that

[5]A discussion of the detailed specification of the different grammar formalisms that FLE supports would go beyond the scope of this paper. The formal specifications are available in the online Bitbucket repository of FLE at the following URL: https://bitbucket.org/dcavar/fle.

[6]BNFC also generates a LaTeXand PDF documentation of the formal language specification in a LBNF.

[7]As explained in Levine (2009), *flex* was written by Vern Paxson around 1987 in the programming language C. It generates code in C or C++ for lexical analyzers and tokenizers for parsers of formal languages that are based on Deterministic Finite Automata (DFA).

[8]As explained in Levine (2009), *bison* is a parser generator that was originally written by Robert Corbett in 1988. Given a CFG-specification for a formal language it generates parser code in the programming languages C, C++, or Java.

[9]See Gamma et al. (1995) for more details).

we utilized provide parser code in various programming languages and at the same time minimize the coding effort for the implementation of mappings to internal data structures and representations needed by different parsing algorithms. Given the openness and free licensing of all the necessary components, we contribute to the sustainability of the grammars that are based on the supported formalisms, including the XLE grammars.

3.2 Tokenizer

FLE can process tokenized input. It also provides a set of different tokenization approaches that can be integrated in the processing chain in various ways. A tokenizer provides a list of *tokens* for a given input sentence. The sentence *John reads a book.* is tokenized into the sequence of strings `["John", "reads", "a", "book", "."]`. The following tokenization strategies and components are implemented: a.) **C++ tokenizer subclass directly compiled into FLE.** The present system falls back to a simple whitespace tokenizer, should no other be provided; b.) **Foma-based Finite State tokenizer.** A tokenizer that makes use of the Foma library (Hulden, 2009); c.) **Conditional Random Field (CRF).** A tokenizer based upon CRF machine learning (see, for example, Wallach (2004)) which does segmentation for Chinese; d.) **Ucto-based tokenizers.** Tokenizers that use Ucto (Jain et al., 2012).

At this time, we are working on tokenizers for languages such as Burmese (mya) and Chinese (zho). We have developed FST-based tokenizers for Burmese (mya), a Tibeto-Burman language written in *abugida* scripts with no spaces indicating word boundaries, and for Chinese (zho). These tokenizers use wordlists and Foma regular expressions, providing a baseline left-to-right maximum segmentation for Burmese or Chinese text. For both Burmese and Chinese, we are experimenting with improved segmenting algorithms using Conditional Random Field (CRF) sequence tagging (Tseng et al., 2005), as described below.

Other Foma, Ucto, and CRF-based tokenizers will be available in the codebase for English (eng), German (deu), Croatian (hrv), Mandarin (cmn), Polish (pol) and other languages in the near future.

3.2.1 Segmentation of Chinese and Burmese

Unlike English and many other Indo-European languages, Asian languages, such as Chinese and Burmese, do not mark word boundaries with spaces. Thus, to parse Chinese and Burmese in FLE, we first need to segment the sentences into words so that they can be passed into the morphological analyzer (or even part-of-speech tagger) and subsequently to the parser.

In FLE, we use CRF for Chinese segmentation, because this method has achieved the best performance previously, and the dictionary-based longest match algorithm for Burmese because it has yielded competitive results using only very limited language resources. We will briefly discuss two approaches for tokenization for Chinese and Burmese.

Segmentation of Chinese Using Conditional Random Field Essentially, the segmentation task is seen as a sequence labeling problem, where the CRF model learns to give labels to each character as to whether it is a word boundary or not. To implement a CRF segmenter, we used the C++ library *dlib* (King, 2009).

As of now, our CRF segmenter uses three types of features as described in Zhao et al. (2010), i.e. character unigram features (previous character C_{-1}, current character C_0 and next character C_{+1}), character bigram features ($C_{-1}C_0$, C_0C_{+1}, $C_{-1}C_{+1}$) and character type features. The character type features denote which of the five types the previous, current and next character belongs to. We follow Zhao et al. (2010) to classify characters in Chinese into five types: numbers, characters referring to time (year, month, day, etc.), English letters, punctuation and other Chinese characters. The intuition is that the character type information of the neighboring characters will be helpful for the CRF segmenter to decide on word boundaries.

We used training data from the Second International Chinese Word Segmentation Bakeoff (Emerson, 2005) to train our model. Our current result for the test data of the bakeoff is **93.0%** on recall, **94.1%** on precision and an F score of **93.6%**, using the official scorer from the bakeoff. In the next step, we will further tune the parameters and experiment with other types of features to improve our results.

Segmentation of Burmese using Foma Our goal is a language independent system for low-resourced languages using openly available resources without relying on tagged corpora. We implemented a Finite State Machine for the longest match algorithm (Poowarawan, 1986) using Foma (Hulden, 2009) and a word list for Burmese (LeRoy Benjamin Sharon, 2016).

For an experiment, we tested the segmenter on a Wikipedia article with **3302** words in **77** sentences. The sentences were hand-segmented by a native Burmese speaker. The baseline performance of the dictionary-based segmentation gave **86.92%** precision and **92%** recall. We plan to integrate this with the CRF machine learning approach with morphology as discussed above for Chinese.

3.3 Morphology

In the current codebase of FLE we make use of Foma-based Finite State Transducer (FST) morphologies using Lexc and Foma regular expressions that are also compatible with the Xerox Finite State Toolkit (XFST) (Hulden, 2009; Beesley & Karttunen, 2003).

The FLE codebase includes an open English (eng) morphology. This morphology currently contains the all irregular verbs and nouns, most of the closed class lexicon, a large set of open class items, including a broad variety of named entities (e.g. toponyms, anthroponyms, and institutions and companies). We are also able to process multi-word expressions and unknown morphemes efficiently using a multi-word recognizer and unknown word guesser that wraps the Foma-based morphology. The guessing algorithms that we experiment with in the morphology includes lexical category and morphological feature guessing using Hidden Markov Models (HMMs). These can be trained to guess the detailed features required by the LFG parser. An integration of the guesser into the syntactic parser will be done in a future development phase. A detailed description of this approach to handle unknown words is beyond the scope of this paper.

In its current version, the English morphology contains 117,705 paths, which corresponds to the number of surface word forms with ambiguities. We estimate that it contains 38,964 morphemes, including inflectional and derivational suffixes and prefixes. The binary form of the FST is 1.9 MB large, with 88,783 states and 124,221 arcs. These numbers are expected to increase significantly before the final public release of the source and the binaries.[10]

The following example illustrates the output from our multi-word enabled English morphology for an input sentence like *Tim Cook, the CEO of Apple, works now for Google*.

```
0 1 Tim Tim+N+Sg+Masc+NEPersonName
1 2 Cook Cook+N+Sg+NEFamilyName
3 4 the the+D+Art+Def
4 5 CEO CEO+N+Sg+Abbrev
5 6 of of+P
6 7 Apple Apple+N+NEBusiness
8 9 works work+N+Pl, work+V+3P+Sg
9 10 now now+Adv+Temporal
10 11 for for+P, for+P+Time
11 12 Google Google+N+NEBusiness
0 2 Tim Cook Tim Cook+N+Masc+NEPersonName+NESTBusiness
```

The tab-delimited fields in the output consist of the index of the token span for the respective token sequence (that is from and to token positions), followed by the token sequence itself and a list of comma separated potential analyses. We extend common feature sets with named entities that include `NEPersonName`, `NEFamilyName`, `NEBusiness` (a business name), or `NESTBusiness` (a subtype of any named entity that is business related). We will be able to expand abbreviations like `CEO` to their full forms in the final release of the complete morphology.

Most of the lexical category and feature labels do not yet conform to any standard, nor are they synchronized with proposals from other relevant projects or best-practice recommendations. In the final release of the morphology we will take standardization into account and very likely change the category and feature labels of the existing morphologies accordingly.

Within the FLE environment we also provide partial implementations of Foma-based morphologies for other languages, including Burmese (mya), Mandarin Chi-

[10]The size increase would not affect the size of the binary FST-file in a significant way, rather the number of paths and covered morphemes and morphotactic regularities.

nese (cmn),[11] or Croatian (hrv).[12] From here, work is planned on further morphologies for various under-resourced and endangered languages. Additionally, we will extend the support to other binary formats and formalisms.

3.4 Features and Unification

F-structures in FLE are represented as Directed Acyclic Graphs (DAG). The Attribute Value Matrix in Figure 2 is a simplified structure representing basic morpho-syntactic features.

$$\begin{bmatrix} \text{AGR} & \begin{bmatrix} \text{NUM} & \text{PL} \\ \text{GEN} & \text{M} \end{bmatrix} \\ \text{TNS} & \text{PAST} \\ \text{ASP} & \text{PERF} \end{bmatrix}$$

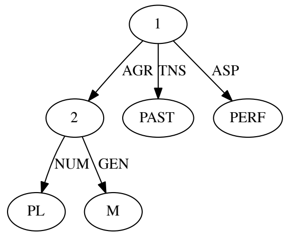


Figure 2: F-structure as Attribute Value Matrix (AVM) and DAG

In FLE, these AVMs are mapped to DAGs as in Figure 2. Note that, although the DAG shows the terminal values as states, the internal representation of DAGs that we use represents them as edges.

The graph is defined as a set G of edges that are represented as tuples $\{f_i, t_i, s_i, w_i\}$, where f is the start-state of the edge, t the target state, s the associated symbol (an attribute or value), and w a weight. In our implementation a DAG is a set G of such weighted edges that is associated with a general weight p for the entire DAG. This implementation of a f-structure allows us to associate weights with edges or entire paths in the DAG. It also allows us to represent the probability of a specific DAG and thus the encoded f-structure independently in the DAG weight. Such a weight could be estimated using a corpus of c-structures and f-structures, or applying parses to a corpus of sentences.

This implementation of f-structures allows us to model associations between f-structures and c-structures based on probabilities. A probability of a sentence could be described as the conditional probability $P(s_i) = P(f_n \mid c_n)$, where c is one possible c-structure n for the sentence s_i, and f is the corresponding f-structure for c. Probabilistic feature representations provide new possibilities for

[11] As one reviewer correctly points out, Mandarin Chinese is not a morphologically rich language, thus the morphology should be rather called a lexicon or FST-based word-list. This assessment is true, if we focus only on the *Han character*-based orthography. Word formation in Mandarin Chinese could indeed be described in terms of affixation and morphotactics, as described in Packard (2000). If we consider the Romanized orthography using Pinyin, it indeed appears plausible to describe word formation using a two-level morphology based approach.

[12] A Croatian morphological analyzer was compiled using a very different FST-based framework (Cavar et al., 2008). We are transferring the morphology to a Foma-based lexc and regular expression format.

robust unification or adaptation to speaker or domain specific feature properties.

Our graph-based implementation of unification is based on two specific constraints. All DAGs are stored in a uniform DAG-space. All edges that represent attributes and values in those DAGs are mapped uniquely. A synchronization object replaces all symbols (attribute names and concrete values) with unique numerical IDs across the entire DAG-space. All states and paths receive a uniform reference, i.e. in all DAGs an edge representing the Attribute-Value path `ROOT -> AGR -> NUM -> PL` will be mapped onto a uniform edge of numerical reference points `(1, 18, 21, 35)`, with 1 representing $ROOT$, 18 representing AGR, and so on.[13] A path is a set G of edges $\{f_i, t_i, s_i, w_i\}$. Terminal edges in G are specific. Their t-value is always 0. Our unification algorithm creates the union of all edges in two sets G_1 and G_2 and it only places a matching constraint on terminal-edges, i.e. unification fails if for two edges $f_1 = f_2$, $t_1 = t_2 = 0$, and $s_1 \neq s_2$. In general, two DAGs are unified by copying one and building the edge union of the copy with the other. Copying is necessary because unification might fail and the previous DAG/f-structure of the first is needed in subsequent unification steps or attempts.

Although this unification algorithm is very efficient, more complicated filters and conditions are necessary and will be implemented in future. In the current version it is sufficient for initial experiments. Future developments might result in revisions and changes in the data structures and related algorithms.

3.5 Parsing

The architecture of the parser depends partly on the particular grammar formalism that is used but also on the underlying linguistic components. The parsing strategy and the grammar properties determine the computational grammar representation and particular parsing algorithms. For example, using an Earley parsing algorithm (Earley, 1968) requires a specific mapping of a CFG-grammar to data structures that makes use of left-peripheral symbols of the right-hand-side of rules. Mapping a grammar to a Finite State Machine (FST) with or without a stack, using the OpenGrm Thrax environment, for example(Tai et al., 2011), requires a very different parsing algorithm and grammar data structure internally.

We have implemented a CFG and a PCFG formalism parser, the possible backbones for our LFG-grammars, using the LBNF formalism. These implementations make use of regular expression operators in the right-hand-side of rules, such as `*`, `+`, and `?`. We allow for the use of | as a disjunction operator, and grouping brackets "`(`" and "`)`" to simplify the rule sets. This is similar to the rule formalism used in XLE. Finite State Machines are ideal for mapping right-hand-sides of rules with such properties (e.g. recursion, optionality, and disjunction).

In one implementation of the mapping of grammars to parser-internal data structures, we make use of Finite State Transducers (FST), such that the one side

[13] The assignment of numerical values to symbols happens during run-time when parsing the grammar.

of CFG rules is read and the other written. That is, the left-hand-side of a CFG rule is emitted when a certain right-hand-side is processed.

As mentioned above, to be able to integrate probabilities in the grammar representation, we use a WFST (Allauzen et al., 2007). This allows us not only to store complex right-hand-sides of rules in an FST data-structure, but also to represent rule probabilities of a PCFG-type in the same data structure.

In addition to the mentioned CFG and PCFG formalisms, we have implemented an LBNF specification of grammar rules used in XLE. For core rule types of the XLE grammar formalism, we have also integrated the semantics, i.e. the mapping of rules to internal data-structures or functions in the parser in the following way.

The different grammar formalisms are parsed and mapped on one internal grammar representation using a WFST. This internal representation is based on our own implementation of a WFST. We intend to base future versions on the the OpenFST library. Our own implementation is simpler and more efficient than the OpenFST version, and initially we did not foresee a need for functions like minimization or ϵ reduction. Avoiding dependencies on external libraries and tools is our core strategy. An integration of OpenFST, however, would save us implementation of various FST-based operations and WFST features that are already integrated in it.

Our use of a WFST as a data structure for the parsing algorithm allows different grammar formalisms to be represented in similar data structures, rendering them compatible with a variety of parsing algorithms. The mapped grammars can also be stored persistently as binary files and exported, for example, to various other FST exchange or visualization formats.

In the previous section we have demonstrated how the grammar formalisms are mapped onto WFST representations internally. We have also mentioned that each of the current grammar formalisms is defined using the LBNF formalisms (Forsberg & Ranta, 2004), an extension of the common BNF formalism. The BNF Converter (Forsberg & Ranta, 2004) is used to generate the C++ code implementation of the syntactic parsers for these formalisms using an intermediate *flex*-based and *bison*-based code for lexer and parser generation (Levine, 2009) underlyingly. This generated code is extended with the semantics to map grammars to internal WFST representations. The effort of porting the grammar representations to other internal formats or even other programming languages is thus significantly reduced.

Note that the mapping of CFGs to WFSTs does not include a reduction of expressiveness of the underlying CFG formalism. While we do intend to experiment with a reduction of CFGs to FSTs by limiting the recursion depth of specific recursions, in this case the WFST is applied recursively by our parser without any effect on the complexity of the grammar as such.

The parsers for the (P)CFG backbone and the XLE RULES specification are implemented mapping to WFSTs. The parsing algorithms that we test include an optimized variant of the Earley Parser (Earley, 1968, 1970). In the implementation, we took into consideration several implementation and optimization suggestions discussed in (Aycock & Horspool, 2002), among others. The specific implementation of the parsing algorithm makes use of the WFST-encoded grammar. It recursively generates edges from the WFST representation based on the input tokens in a bottom-up fashion. Instead of positions in dotted rules, the edges are tuples that point to the span of input tokens and to a state in the WFST.

While we use feature representations in form of DAGs and a unification algorithm that is based on such DAGs (see for example Shieber (1985)), as described in the previous section, our ongoing evaluation is concerned with the question of applying unification after the possible c-structure representations are generated, or during the generation of edges while tracking paths through the WFST. In the implementation of WFSTs *weights* can be implemented as instances of objects[14] or functions. This allows us to associate not only a probabilistic metric to the transitions through the WFST, but also operations like unification and resulting f-structures.

Our goal is to integrate the algorithm for uncertainty, the unification algorithm, and other proposals discussed in Kaplan & Maxwell (1988); Maxwell & Kaplan (1996a,b, 1991, 1993) into the WFST architecture in a systematic way. We are able to handle probabilities in the c-structure representations in the form of weights in the WFST that could be mapped from a PCFG backbone or quantification of rule applications while processing corpora. As mentioned above, these probabilities can also be related to f-structures using DAG probabilities or products of path probabilities in the resulting DAGs. We have not performed any experiments related to c-structure and f-structure probabilities.

4 Development Plan

The FLE codebase and the entire environment with external libraries are coded in standard C++11 and C++14. They utilize exclusively open components, including the C++ Boost framework (Schling, 2011) and additional specialized libraries like Foma (Hulden, 2009), OpenFST (Allauzen et al., 2007), OpenGram Thrax (Roark et al., 2012), Ucto (Jain et al., 2012), and Dlib-ml (King, 2009). The parsers for formalisms in the FLE library are implemented using the Labeled Backus-Naur Form (LBNF). The syntactic parser code for these grammar formalisms is generated by the Backus-Naur Form Converter (BNFC) (Forsberg & Ranta, 2004), and for C and C++ the freely available lexer and parser generators *flex* and *bison* (see Levine (2009) and footnotes 7 and 8). The development environment requires CMake[15] and common working C++ compilers with at least C++11 support.

The entire FLE environment is released under the Apache License 2.0, as are most of the components that it uses.[16] This includes all code, morpheme collections and trained models that we created ourselves and that we legally can distribute. The Apache License appears to be more adequate to facilitate collaborative

[14] Any C++ object for example could function as a weight.

[15] See http://cmake.org/.

[16] See http://www.apache.org/licenses/LICENSE-2.0.

academic and industry projects. It allows one to freely download and use the FLE code or binaries for personal, company internal, or commercial purposes. The system can be integrated in third party systems and packages. Any modifications that are made to the FLE code do not have to be shared and redistributed, although this is encouraged. The license requires that any reused and distributed piece of the FLE source code or binary software has to contain proper attribution and that a copy of the license has to be included in any redistribution.

Most of the components and external libraries that are used in FLE are released under the Apache License 2.0. The one exception at the moment is the optional Ucto Unicode tokenizer library, which is released under GPL version 3.0.

The resulting codebase is tested to compile on common operating system platforms, e.g. Windows, Mac OS X, and various Linux distributions. The binaries will provide libraries and executables for the common operating systems and linked modules for some programming languages, e.g. Python.

Our development plan includes some of the following goals, without implying any priorities: 1.) Full compatibility with the grammars of the current XLE environment. While we finalized parsers for the different grammar sections of XLE grammars, not all features of the grammar have been semantically mapped onto internal data representations or functions; 2.) Integration of a graphical environment for the parser and grammar development; 3.) Development of a Python module interacting with the library and FLE components; 4.) Integration of initial semantic components (e.g. Glue Semantics); 5.) Integration of a parallelized processing chain with a blackboard architecture; 6.) Extension of the morphologies and grammars or grammar fragments to other languages than English, Burmese, Chinese, Croatian etc., by integrating specific tokenizers and morphological analyzers; 7.) Since we foresee the possibility for generating initial PCFGs from treebanks to bootstrap the grammar engineering process, we need to map tags from treebanks to morphological feature annotations and LFG-type feature representations.

To integrate and experiment with a PCFG backbone we developed initial PCFG extraction tools that can process treebanks of the Penn Treebank format (Marcus et al., 1993). For languages that have existing treebanks such as English and Mandarin Chinese,[17] we can generate a PCFG and compact the rules for the parser (Krotov et al., 1999).

Initial steps have been taken towards the implementation of the second goal. The other goals are scheduled for implementation and testing in 2017.

5 Conclusion

While much of the FLE environment is under development, many components can be considered ready and usable. This includes the specification of grammar formalisms for the LFG parser backbone and XLE formalism parsers, the integration

[17] See for example the Chinese Treebank Project (Xue et al., 2002).

of various types of tokenizers and morphological analyzers, the CFG/PCFG backbones (and XLE rules) to a WFST-based grammar representation, two different types of parsing algorithms, and DAG-based unification.

Some of the currently implemented algorithms and components might be useful to other projects: a.) The LBNF specifications of the XLE grammar formalism. With BNFC these formal specifications can be used to generate parsers for the grammar formalisms in various other programming languages, including Java, Haskell, or C#. b.) The mapping of treebanks to PCFGs has been exported to a standalone tool. c.) A morphology-based language independent multi-word analyzer has been isolated as a standalone program.

We have performed preliminary performance tests using various Foma morphologies and the first parser implementation without unification and feature logic. Currently the English morphology as specified in a previous section, can process more than 150,000 tokens per second on a personal computer with an Intel Core i7 CPU and a current Linux distribution using GCC/G++ 6.x. This performance includes only covered vocabulary with lexical ambiguities and no guesser.

The WFST backbone based syntactic parser was tested on a small grammar with structural and lexical ambiguities and preceding tokenization and morphological analysis of all tokens. It parses approx. 3,000 sentences per second with an average sentence length of 7 words using the same architecture as described above. Unification is not included in the performance tests yet. This suggests that an improved version of such a parser can be expected to perform even faster in a final release.

The code of FLE is split into two sections. We have released public code in the Bitbucket repository at: https://bitbucket.org/dcavar/fle/. The development repository is open to team members by invitation only. If you want to join the development team, please send us an email and we will share the Bitbucket repository with you.

References

Allauzen, Cyril, Michael Riley, Johan Schalkwyk, Wojciech Skut & Mehryar Mohri. 2007. OpenFst: A general and efficient weighted finite-state transducer library. In *Proceedings of the Twelfth International Conference on Implementation and Application of Automata, (CIAA 2007)*, vol. 4783 Lecture Notes in Computer Science, 11–23. Prague, Czech Republic: Springer.

Aycock, John & R. Nigel Horspool. 2002. Practical Earley parsing. *The Computer Journal* 45. 620–630. doi:10.1093/comjnl/45.6.620.

Beesley, Kenneth R. & Lauri Karttunen. 2003. *Finite state morphology*. CSLI Publications. http://www.fsmbook.com.

Berstel, Jean & Christophe Reutenauer. 1988. *Rational series and their languages*. Berlin, New York: Springer-Verlag.

Bird, Steven, Ewan Klein & Edward Loper. 2009. *Natural language processing with python: Analyzing text with the natural language toolkit*. O'Reilly Media.

Bresnan, Joan. 2001. *Lexical-functional syntax*. Blackwell.

Butt, Miriam, Tracy Holloway King, María-Eugenia Niño & Frédérique Segond. 1999. *A grammar writer's cookbook*. CSLI Publications.

Cavar, Damir, Ivo-Pavao Jazbec & Siniša Runjaić. 2008. Interoperability and rapid bootstrapping of morphological parsing and annotation automata. *Informatica* 33(1). 107–113.

Clément, Lionel. 2016. XLFG Documentation. Technical report LaBRI. https://hal.archives-ouvertes.fr/hal-01277648.

Clément, Lionel & Alexandra Kinyon. 2001. XLFG-an LFG parsing scheme for french. In Miriam Butt & Tracy Holloway King (eds.), *Proceeding of the LFG01 Conference*, CSLI Publications.

Corkill, Daniel D. 1991. Blackboard systems. *AI Expert* 6(9). 40–47.

Crouch, Dick, Mary Dalrymple, Ron Kaplan, Tracy King, John Maxwell & Paula Newman. 2011. XLE documentation. Online document at http://www2.parc.com/isl/groups/nltt/xle/doc/xle_toc.html.

Dalrymple, Mary. 2001. *Lexical functional grammar* (Syntax and Semantics 42). New York: Academic Press.

Earley, Jay. 1968. *An efficient context-free parsing algorithm*: Carnegie-Mellon University dissertation.

Earley, Jay. 1970. An efficient context-free parsing algorithm. *Commun. ACM* 13(2). 94–102. doi:10.1145/362007.362035.

Emerson, Thomas. 2005. The second international Chinese word segmentation bakeoff. In *Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing*, vol. 133, 123–133.

Erman, Lee D., Frederick Hayes-Roth, Victor R. Lesser & D. Raj Reddy. 1980. The Hearsay-II speech-understanding system: Integrating knowledge to resolve uncertainty. *Computing Surveys* 12(2). 213–253.

Ferrucci, D. & A. Lally. 2004. UIMA: An architectural approach to unstructured information processing in the corporate research environment. *Natural Language Engineering* 10(3-4). 327–348.

Forsberg, Markus & Aarne Ranta. 2004. The BNF converter. In *Proceedings of the 2004 ACM SIGPLAN Workshop on Haskell* Haskell '04, 94–95. New York, NY, USA: ACM. doi:10.1145/1017472.1017475.

Forsberg, Markus & Aarne Ranta. 2005. The labelled BNF grammar formalism. *Department of Computing Science, Chalmers University of Technology and the University of Gothenburg* .

Gamma, Erich, Richard Helm, Ralph Johnson & John Vlissides. 1995. *Design patterns: Elements of reusable object-oriented software*. Reading, Massachusetts: Addison-Wesley Publishing Company.

Hayes-Roth, Barbara. 1985. A blackboard architecture for control. *Artificial Intelligence* 26(3). 251–321. doi:10.1016/0004-3702(85)90063-3.

Hulden, Mans. 2009. Foma: A finite-state compiler and library. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, 29–32. Association for Computational Linguistics.

Jackendoff, Ray. 1997. *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 2007. A parallel architecture perspective on language processing. *Brain Research* 2–22.

Jain, Anil K., Lin Hong & Sharath Pankanti. 2012. Ucto: Unicode tokenizer version 0.5.3 reference guide. Tech. Rep. ILK 12-05 Induction of Linguistic Knowledge Research Group, Tilburg Centre for Cognition and Communication, Tilburg University Tilburg, The Netherlands. https://ilk.uvt.nl/ucto/ucto_manual.pdf.

Kaplan, Ronald. 1996. A probabilistic approach to lexical-functional grammar. Presentation at the LFG Colloquium and Workshops, Rank Xerox Research Centre.

Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-functional grammar: A formal system for grammatical representation. Cognitive Theory and Mental Representation, 173–281. Cambridge, MA: The MIT Press.

Kaplan, Ronald M & John T Maxwell. 1988. An algorithm for functional uncertainty. In *Proceedings of the 12th Conference on Computational Linguistics-Volume 1*, 297–302.

King, Davis E. 2009. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research* 10. 1755–1758.

Krotov, Alexander, Mark Hepple, Robert Gaizauskas & Yorick Wilks. 1999. Evaluating two methods for treebank grammar compaction. *Natural Language Engineering* 5(4). 377–394.

Kuich, Werner & Arto Salomaa. 1986. *Semirings, automata, languages* (EATCS Monographs on Theoretical Computer Science 5). Berlin, Germany: Springer-Verlag.

LeRoy Benjamin Sharon. 2016. Myanmar-Karen word lists. https://github.com/kanyawtech/myanmar-karen-word-lists.

Levine, John R.2009. *flex & bison*. Sebastopol, CA: O'Reilly Media.

Loper, Edward & Steven Bird. 2002. NLTK: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, vol. 1, 63–70. Association for Computational Linguistics.

Marcus, Mitchell P., Mary Ann Marcinkiewicz & Beatrice Santorini. 1993. Building a large annotated corpus of english: The Penn treebank. *Comput. Linguist.* 19(2). 313–330. http://dl.acm.org/citation.cfm?id=972470.972475.

Maxwell, John & Ronald M. Kaplan. 1996a. An efficient parser for LFG. In *Proceedings of the First LFG Conference*, CSLI Publications.

Maxwell, John T. & Ronald M. Kaplan. 1991. A method for disjunctive constraint satisfaction. In Masaru Tomita (ed.), *Current Issues in Parsing Technology*, vol. 126, 173–190. Kluwer Academic Publishers.

Maxwell, John T. & Ronald M. Kaplan. 1993. The interface between phrasal and functional constraints. *Computational Linguistics* 19(4). 571–590.

Maxwell, John T. & Ronald M. Kaplan. 1996b. Unification parser that automatically take advantage of context freeness. In *Proceedings of the First LFG Conference (Grenoble)*, Stanford: CSLI Publications.

Mohri, Mehryar. 2004. Weighted finite-state transducer algorithms. An overview. In *Formal Languages and Applications*, 551–563. Springer.

Packard, Jerome L. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge University Press.

Poowarawan, Yuen. 1986. Dictionary-based Thai syllable separation. In *Proceedings of the Ninth Electronics Engineering Conference*, 409–418.

Roark, Brian, Richard Sproat, Cyril Allauzen, Michael Riley, Jeffrey Sorensen & Terry Tai. 2012. The OpenGrm open-source finite-state grammar software libraries. In *Proceedings of the ACL 2012 System Demonstrations*, 61–66.

Salomaa, Arto & Matti Soittola. 1978. *Automata-theoretic aspects of formal power series*. New York: Springer-Verlag.

Schling, Boris. 2011. *The Boost C++ libraries*. XML Press.

Shieber, Stuart M. 1985. Using restriction to extend parsing algorithms for complex-feature-based formalisms. In *Proceedings of the 23rd Annual Meeting on Association for Computational Linguistics* ACL '85, 145–152. Stroudsburg, PA, USA: Association for Computational Linguistics. doi:10.3115/981210.981228.

Tai, Terry, Wojciech Skut & Richard Sproat. 2011. Thrax: An open source grammar compiler built on OpenFst. Paper presented at the ASRU 2011.

The GATE Team. 2011. *Text processing with GATE (version 6)*. University of Sheffield Department of Computer Science.

Tseng, Huihsin, Pichuan Chang, Galen Andrew, Daniel Jurafsky & Christopher Manning. 2005. A conditional random field word segmenter for SIGHAN bakeoff 2005. In *Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing*, 168–171.

Wallach, Hanna M. 2004. Conditional random fields: An introduction. Tech. Rep. MS-CIS-04-21 University of Pennsylvania.

Xue, Nianwen, Fu-Dong Chiou & Martha Palmer. 2002. Building a large-scale annotated Chinese corpus. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002)*, Taipei, Taiwan.

Zhao, Hai, Chang-Ning Huang, Mu Li & Bao-Liang Lu. 2010. A unified character-based tagging framework for Chinese word segmentation. *ACM Transactions on Asian Language Information Processing (TALIP)* 9(2). 5.

Transforming the AnCora corpus to HPSG

Luis Chiruzzo

Universidad de la República, Montevideo, Uruguay

Dina Wonsever

Universidad de la República, Montevideo, Uruguay

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 182–193

Keywords: HPSG, Spanish, Corpus, AnCora

Chiruzzo, Luis & Dina Wonsever. 2016. Transforming the AnCora corpus to HPSG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 182–193. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.10.

Abstract

We present the construction of a HPSG corpus for Spanish, based on the transformation of the AnCora Spanish corpus into a HPSG compatible format. We describe the transformation process and the evaluation of the resulting corpus.

1 Introduction

We describe the first phase of a currently ongoing project for building a statistical HPSG parser for Spanish. It consists in transforming the AnCora Spanish corpus from its CFG-style annotations to an HPSG (Pollard and Sag, 1994) compatible format, in a next stage we extract a lexicon and train a supertagger over the transformed corpus (Chiruzzo and Wonsever, 2015). Head-driven Phrase Structure Grammars (HPSG) are a strongly lexicalized grammar formalism. This family of grammars are very expressive, allowing the modeling of many linguistic phenomena and capturing syntactic and semantic notions at the same time. The rules used in an HPSG grammar are very generic, indicating how a syntactic head can be combined with its complements, modifiers (adjuncts) and specifier. The categories of the elements are organized in a type hierarchy and the parsing result is a tree whose nodes are typed feature structures (Carpenter, 1992).

Our work is inspired by Enju (Matsuzaki et al., 2007), a statistical HSPG parser for English that has high performance and language coverage. This parser was built based on the Penn Treebank corpus (Marcus et al., 1993). As the Penn Treebank was not annotated in an HPSG compatible format but rather in a CFG-style grammar, they built a set of rules to transform the Penn Treebank trees into a structure that is similar to HPSG (Miyao et al., 2005). The Enju parser is trained using the result of this transformation.

Other HPSG grammars for Spanish exist, the most relevant one being the Spanish Resource Grammar (SRG) (Marimon, 2010), a Spanish HPSG grammar built using the LinGO Grammar Matrix (Bender et al., 2002), a framework for building HPSG grammars for many languages. SRG can be used with the LKB development system (Copestake et al., 1999), as well as the PET runtime parser (Callmeier, 2000), and its results are very rich HPSG trees that include all of the constructions supported by the theory. Our objective is to build a new HPSG parser whose representations will not be as rich as SRG's, but we aim at making it more robust. Also, the statistical model, trained from the transformed corpus and the extracted lexical units, will compute directly the desired output instead of acting as a filter for the great number of output trees resulting from the grammar non-stochastic constraints as in (Marimon et al., 2014).

AnCora is a corpus for Spanish and Catalan (Taulé et al., 2008) that contains about half a million words in 17,000 sentences. The corpus has CFG-style annotations, but it is also enriched with attributes such as morphological information and predicate-arguments structure. Inspired by Enju, we aimed to transform this corpus into a treebank compatible with HPSG. There exists another Spanish treebank with HPSG annotations: the Tibidabo corpus (Marimon, 2015). However, this corpus seems not to be publicly available, and, more important to us, its structure does not suit to our purposes. Also, Tibidabo contains only 4000 sentences from the AnCora Spanish corpus which consists of 17000 annotated sentences. Each sentence in Tibidabo is represented by three graphs: a binary constituent tree, with atomic category names, a dependency tree annotated with syntactic label names and a MRS structure (Copestake et al., 2005), as shown in (Marimon, 2015). Another difference between our approach and Tibidabo is that we transform the whole AnCora corpus, using the corpus information to guide the transformation, while Tibidabo re-annotated some of the sentences of the corpus using SRG, but dropping the original annotation information.

On the other hand, as mentioned above, the trees annotated using SRG have richer structures than the ones we get after the transformation. In particular, the Tibidabo corpus maintains the MRS structures as its sentences semantic representation, which includes event variables, standard arguments naming and quantifier scopes, among other things. MRS is, in some sense, a meta-notation for first order logical forms that allows underspecification and thus packs scope ambiguities. Our approach to semantics is much simpler: using the information readily available in AnCora, we include features for representing the predicate-argument structure of the verbs (also for other predicates, e.g. deverbal nouns). The predicate-argument structure is annotated in PropBank style (Kingsbury and Palmer, 2002), so all our `SEM` feature needs is a set of features for each of the PropBank arguments (`ARG0`, `ARG1`, `ARGM`...). This simplified approach to semantics is similar to the one used in Enju.

2 Description of the grammar

This section describes the main aspects of the feature structure we used and the grammar rules.

2.1 Feature structure

The general feature structure for a lexical entry in our grammar is shown in figure 1. This feature structure tries to summarize all features that could be included in one of the lexical entries. The structure has morphological, syntactic and our simplified semantic features. In this feature structure, the feature `COMPS` can have a list of expressions, while the features `SPEC` and `MOD` are shown as lists because they might have zero or one expressions.

Figure 2 shows a concrete example of a lexical entry for a typical transitive verb. Notice that the subject of the verb (`SPEC`) is coindexed with the proto-agent argument (`ARG0`), and the only complement is coindexed with the proto-patient argument (`ARG1`). Thus, this lexical entry represents the active voice instance of

$$\begin{bmatrix} \text{word} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{n} \vee \text{v} \vee \text{a} \ldots & \\ \text{AGR} & \begin{bmatrix} \text{PER } 1 \vee 2 \vee 3 \\ \text{NUM sg} \vee \text{pl} \\ \text{GEN f} \vee \text{m} \\ \ldots \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPEC} & \left\langle \text{expr} \right\rangle \\ \text{COMPS} & \left\langle \text{expr} \right\rangle \\ \text{MOD} & \left\langle \text{expr} \right\rangle \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{ARGS} & \begin{bmatrix} \text{ARG0} & \text{expr} \\ \text{ARG1} & \text{expr} \\ \ldots & \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 1: Feature structure for a lexical entry

this transitive verb.

$$\begin{bmatrix} \text{word} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{v} & \\ \text{AGR} & \begin{bmatrix} \text{FORM} & \text{personal} \\ \text{MODE} & \text{i} \\ \text{PER} & 3 \\ \text{NUM} & \text{sg} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPEC} & \left\langle \boxed{1} \text{ SYN.HEAD n} \right\rangle \\ \text{COMPS} & \left\langle \boxed{2} \text{ SYN.HEAD n} \right\rangle \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{ARGS} & \begin{bmatrix} \text{ARG0} & \boxed{1} \\ \text{ARG1} & \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{TEXT} & \textit{come} \end{bmatrix}$$

Figure 2: Feature structure for transitive verb *"come"*, indicative third person singular form of the verb *"to eat"*

2.2 Grammar rules

The rules of the grammar we use are a simplified version of the ones used in (Pollard and Sag, 1994). The grammar has rules for combining a specifier, a complement or an adjunct to a head, and two rules for binarizing the coordinated constructions. There are also extra rules for simplifying the analysis of clitics and relative constructions, which will have further development in the future. Despite using these simplified rules, the grammar is able to deal with some interesting linguistics constructions.

2.2.1 Specifiers

We define two rules for combining a specifier with a head: `spec_head` and `head_spec` which apply the specifier to the left or to the right of the head respectively. In both cases the `HEAD` feature of the resulting phrase is coindexed with the `HEAD` feature of the head. These rules are used for applying the determiner of a noun phrase (only the `spec_head` rule in this case), and also for applying the subject of a sentence. The `SPEC` feature of the resulting phrase is cleared.

Notice that we allow for a specifier to be combined both to the left or to the right of the head. Although Spanish typology is generally regarded as SVO, there are plenty of exceptions to this rule. It is very common to find sentences in which the object is located before the verb, or the subject is located after the verb, for example: *"llegó el tren" / "the train arrived"*. The AnCora corpus contains many examples of these constructions. We chose this representation instead of using a `SLASH` feature and a head-filler rule because we consider it would be easier to extract statistics from the corpus on which verbs are usually combined with a subject to the left or to the right.

2.2.2 Complements and adjuncts

There are two rules for combining a complement with a head: `comp_head` and `head_comp` which apply the complement to the left or to the right of the head respectively. In both cases the `HEAD` feature of the resulting phrase is coindexed with the `HEAD` feature of the head. One of the expressions in the list of the `COMPS` feature is cleared. The expression that is cleared depends on the verb and the complement being addressed. This information has to be extracted from the corpus. Notice that these rules are binary, so in order to combine a head with multiple complements the rules have to be applied several times.

There are two different rules for combining an adjunct or modifier with a head: `mod_head` and `head_mod` which apply the adjunct to the left or to the right of the head respectively. In both cases the `HEAD` feature of the resulting phrase is coindexed with the `HEAD` feature of the head. The `MOD` feature of the adjunct is coindexed with the head.

In the AnCora corpus the distinction between complements and adjuncts is not always overtly annotated, we rely on a series of hand written rules that consider the category of the head, the category of the expression and several different annotation attributes the corpus includes. These rules were created by manually inspecting the corpus.

For example, the rules for detecting the complements of a verb in a subordinate sentence take into consideration the attribute `func` that might be present in the AnCora XML element that describes the constituent. This attribute represents the syntactic function of the constituent. In an ideal case, this attribute would be enough to detect if a constituent is a complement or not. However, the attribute is not always properly annotated in the corpus in all the constituents that should require it. Because of this, by manually inspecting the corpus, several other rules were added that consider other attributes and exceptional cases, in order to capture as many correct examples as possible. For example, if the `func` attribute is missing from the constituent we might make use of the attribute `arg`, which defines its role in the predicate argument structure. This distinction is not perfect in all cases, see section 4 for details about the performance of these rules.

2.2.3 Coordinations

In our grammar, coordinated structures need to be binarized, which is done using two rules: `coord_right` and `coord_left`. First the conjunction and the right expression are put together using the `coord_right` rule, then the resulting phrase and the left expression are put together using the `coord_left` rule. This is iterated for longer chains of coordinations, resulting in a chain of binary trees.

2.2.4 Clitics

Clitic pronouns need special attention in Spanish because they sometimes act as complements (in substitution of a complement that was already mentioned in the text) and sometimes both the real complement and the clitic are present at the same time (Pineda and Meza, 2005) (this is called *clitic doubling*). Because of this, we created a new rule for dealing with clitics, different from the rules for applying complements. This rule is `clitic_head`, which applies a clitic to the left of the head. During the transformation, we annotate all clitics using this rule but we do not perform any further analysis to recover the actual complement the clitic is refering to, should it be present. An appropriate handling of the clitic analysis would need a way of classifying the cases that deal with clitic doubling and providing a consistent analysis for this cases. This aspect has not been addressed yet.

2.2.5 Relative clauses

Relative pronouns introduce a subordinate sentence inside another sentence that acts as a modifier to a noun phrase and at the same time use the noun phrase as an argument (e.g. in *"el perro que me mordió"* / *"the dog that bit me"* the noun *"perro"* / *"dog"* is both modified by and the subject of the subordinate sentence). This is another kind of long distance dependency that is usually dealt with using `SLASH` features and filler rules in HPSG. Currently in our work we are not resolving this type of long distance dependency, so we created a new rule `head_rel` to mark these constructions. In the future these `head_rel` will be resolved using a `SLASH` feature or a similar construction.

2.2.6 Control verbs

Control verbs are verbs which govern over the arguments (subject or object) of another subordinate verb. In AnCora, the combination of a subject control verb (such as *"comenzar" / "to start"*) together with its subordinate verb, is generally annotated as a verb phrase structure. However, an object control verb (such as *"obligar" / "to oblige"*) is not annotated as a verb phrase structure together with its subordinate verb. In our corpus transformation, we resolve the subject of the subordinate verb in the subject control verb constructions, but further analysis is needed in order to resolve the correct coindexation in the object control verb constructions.

3 Transformation process

In a HPSG tree, it is necessary to know the syntactic head of every constituent and also the roles that the rest of the elements of the constituents have. This information is not directly available in AnCora, so we created a series of heuristics that exploit the information in the corpus (structure and attributes) in order to transform it to a HPSG compatible format. Figure 3 shows an example of a sentence annotated using the AnCora markup: *"El desarrollo, la integración y la cooperación fueron los asuntos protagonistas de esta reunión." (Development, integration and cooperation were the main matters of this meeting.)*

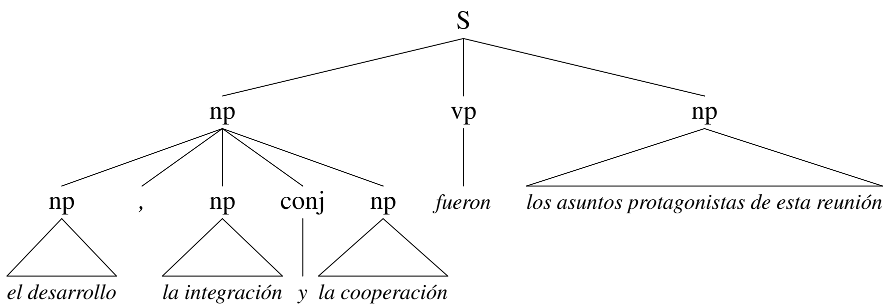


Figure 3: Sentence annotated with its syntactic structure in AnCora

If we consider the syntactic structures of AnCora as annotated in a CFG, the number of rules in this grammar would be very large. For example, there are 5,800 ways of writing a subordinate sentence, and 900 ways of writing noun phrases. Because of this, we tried to reduce the complexity of the problem using a transformation process which uses two stages: a top-down approach that works together with a bottom-up approach.

– `n` (noun)
"…Río_Bravo y Saltillo para la [[$_{\mathrm{H}}$ compañía] [francesa]]…"
– `grup.nom` (nested noun phrase)
"…y sobre [[$_{\mathrm{H}}$ transmisiones y retenciones] [de fondos de inversión]] ."
– `p` (pronoun)
"…obtuvo 19 diputados, [[$_{\mathrm{H}}$ dos] [más]] que en 1996…"
– `w` (date)
"…hundimiento del "Kursk" el [[pasado] [$_{\mathrm{H}}$ 12_de_agosto]] en aguas árticas…"
– `z` (number)
"…donde lograron el [[$_{\mathrm{H}}$ 71_por_ciento] [de los sufragios]] …"
– `a` (adjective)
"…quien cuestiona al entrenador es [[$_{\mathrm{H}}$ enemigo] [del Barça]] ."
– `v` (verb)
"…sobre todo en el [[$_{\mathrm{H}}$ capitulo] [de las infraestructuras]] …"
– `s.a` (adjective phrase)
"…y la [[$_{\mathrm{H}}$ segunda] [, mucho más potente,]] a las 07.30.42…"
– `participi` (participle)
"…el relato ZZadjNM de lo [[$_{\mathrm{H}}$ ocurrido] [en la sima de ZZlugar]] …"
– `S/clausetype=participle` (subordinate sentence of type participle)
"…en_lugar_del [[$_{\mathrm{H}}$ destituido] [Carlos_Sainz_de_Aja]] ."
– `S/clausetype=relative` (subordinate sentence of type relative)
"…incluidos los [[$_{\mathrm{H}}$ que él mismo ha hablado] [sobre sí mismo]] …"
– `S/clausetype=completive` (subordinate sentence of type completive)
"Al [[$_{\mathrm{H}}$ correr] [de los siglos]] se había manifestado un…"
– `sp` (prepositional phrase)
"aeropuerto de Miami, uno de los [[$_{\mathrm{H}}$ de mayor tráfico aéreo] [en EEUU]]…"
– `sn` (noun phrase, maximal projection)
"…el hotel (un [[$_{\mathrm{H}}$ cinco estrellas de gran lujo]])…"

Table 1: Rules for head detection inside a `grup.nom`

We define an *elementary HPSG tree* as a simple tree that consists of a syntactic head surrounded by elements that are directly related to the head (complements, modifiers, specifier). The top-down process tries to transform the most complex structures of the corpus into simpler trees. This means breaking up a node with too many children into a composition of elementary trees that preserve the original structure. The top-down process is in charge, among other things, of extracting quoted or punctuated blocks; marking clitics; extracting prepositional phrases, relative clauses and subordinate sentences; and binarizing sequences of coordinations.

The bottom-up approach assumes that the top-down process has dealt with all those complex structures and left only a set of homogeneous simpler structures, those structures will become elementary HPSG trees after the transformation. In order to transform these trees, we created head detection and arguments classification heuristics. For the English language there is a commonly used heuristic for finding the syntactic head of a phrase in the Penn Treebank corpus, as described in (Collins, 2003). As there is no equivalent for Spanish, and the grammatical differences between both languages make it impossible to apply the same rules, a set of head detection rules was manually crafted for the elements of Ancora. We defined lists of constraints that an element must match in order to be considered the head of a constituent. The constraints are written in a small language for rules that was created for this purpose. Table 1 shows some examples of the list of detection rules that is used to find the head of a noun phrase (elements of type `grup.nom` in AnCora).

After finding the syntactic head of a phrase, we proceed to analyze the elements that are directly to the left or to the right of the head, and apply a series of heuristics that try to classify the role of those arguments with respect to the head. The heuristics use information about the node such as its part of speech, but also the attributes of the element. The rules for classifying the elements are written in the same language as the rules for detecting heads. In total there are 70 head detection rules and 184 argument classification rules.

Besides these rules, there are specific transformation heuristics for verb phrases, because the verb phrases in AnCora behave different from other constituents and could not be reduced to elementary HPSG trees (see section 2.2.6).

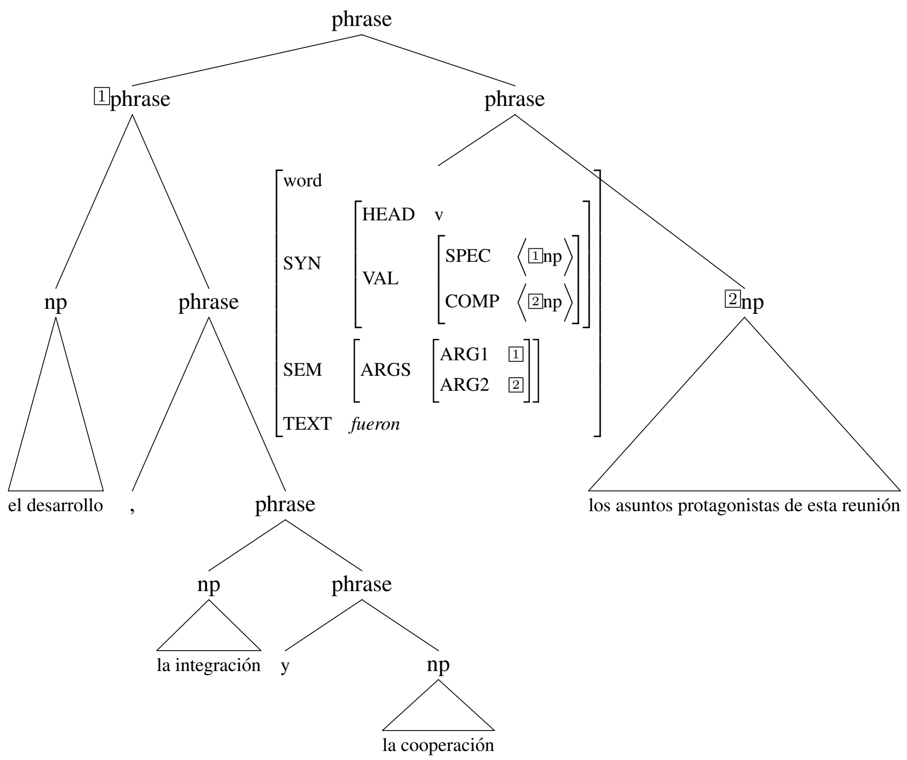


Figure 4: Sentence after the transformation process

Figure 4 shows what the sentence in the previous example looks like after the transformation. Notice that the coordination has been binarized; the head, complement and specifier have been identified; and the appropriate arguments are coindexed in the structure. Although for the sake of clarity only the feature structure for the word *"fueron"* is shown in the diagram, the transformation process creates feature structures for all the word and phrase nodes in the tree.

4 Evaluation

The transformed corpus contains only binary or unary constituents and all nodes indicate their syntactic head and the applied rule. AnCora has a total of 780950 constituents and almost all of them could be transformed. We evaluated the accuracy of the transformation heuristics in the following way: We took a random sample of 40 sentences (779 constituents) and manually identified the syntactic head of every constituent and the role of every other element with respect to the head (complement, modifier, specifier, clitic or punctuation mark).

We found that the head detection heuristics have a precision of 95.3%, which climbs to 98.7% if we do not consider the nodes with coordinations. Table 2 shows the precision of the head detection rules by constituent category, considering nodes with coordinations.

AnCora Category	Total	Correct	Precision
grup.a (adjectival group[1])	9	6	66.7%
grup.adv (adverbial group)	3	3	100.0%
grup.nom (noun group)	162	154	95.1%
grup.verb (verb phrase)	23	23	100.0%
infinitiu (infinitival verb phrase)	3	3	100.0%
relatiu (relative pronominal expression)	1	1	100.0%
S (subordinate sentence)	91	85	93.4%
s.a (adjectival phrase)	4	3	75.0%
sa (adjectival phrase[2])	1	1	100.0%
sadv (adverbial phrase)	7	7	100.0%
sentence (sentence)	40	35	87.5%
sn (noun phrase)	220	216	98.2%
sp (prepositional phrase)	207	204	98.6%
spec (determiner phrase)	8	1	12.5%

Table 2: Precision of head detection rules

The arguments classification heuristics have a precision of 92.5% on average, and the category which is the most difficult to classify is the complements (84.95% precision). Table 3 shows the confusion matrix for the arguments classification.

[1] In AnCora, a "group" in general is different from a "phrase" in that it cannot contain a specifier, though there are many examples that break this rule in the corpus.

[2] There are two types of adjectival phrases in AnCora: `sa` and `s.a`. In practice, there seems to be no difference between them as they are used in the corpus.

	Specifier	Complement	Modifier	Clitic	Punctuation
Specifier	279	3	3	0	0
Complement	6	333	53	0	0
Modifier	1	18	247	0	0
Clitic	0	0	0	19	0
Punctuation	0	0	0	0	155

Table 3: Confusion matrix for the arguments classification

5 Conclusions and future work

We described a transformation process that takes the AnCora Spanish corpus and transforms its CFG style annotations into HPSG compatible structures. The result of this process is a collection of trees annotated in HPSG style where the head of every constituent is marked; the arguments are classified; and all lexical entries include morphological, syntactic and semantic information.

The transformation process achieves a precision of 95.3% for head detection (98.7% without considering coordinations) and a precision of 92.5% for arguments classification. These are promising results, but there is still room for improvement, specially for the arguments classification. In order to improve performance we might need to refine the arguments classification heuristics. Furthermore, the transformed corpus is missing some interesting features such as the analysis of the Spanish clitics as arguments of the verbs and the analysis of long distance dependencies. This transformed corpus is used in a later stage to extract a lexicon of Spanish words and to train a supertagger for verbs, nouns and adjectives, with the aim of creating a statistical parser for Spanish.

References

E. M. Bender, D. Flickinger, and S. Oepen. The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In *Proceedings of the 2002 workshop on Grammar engineering and evaluation-Volume 15*, pages 1–7. Association for Computational Linguistics, 2002.

U. Callmeier. Pet–a platform for experimentation with efficient hpsg processing techniques. *Natural Language Engineering*, 6(01):99–107, 2000.

B. Carpenter. *The logic of typed feature structures*. Cambridge University Press, 1992.

L. Chiruzzo and D. Wonsever. Supertagging for a statistical hpsg parser for spanish. In *International Conference on Statistical Language and Speech Processing*, pages 18–26. Springer, 2015.

M. Collins. Head-driven statistical models for natural language parsing. *Computational linguistics*, 29(4):589–637, 2003.

A. Copestake, J. Carroll, R. Malouf, and S. Oepen. The (new) lkb system. *Center for the Study of Language and Information, Stanford University*, 1999.

A. Copestake, D. Flickinger, C. Pollard, and I. A. Sag. Minimal recursion semantics: An introduction. *Research on Language and Computation*, 3(2-3):281–332, 2005.

P. Kingsbury and M. Palmer. From treebank to propbank. In *LREC*. Citeseer, 2002.

M. P. Marcus, M. A. Marcinkiewicz, and B. Santorini. Building a large annotated corpus of english: The penn treebank. *Computational linguistics*, 19(2):313–330, 1993.

M. Marimon. The spanish resource grammar. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC*, pages 17–23, Valletta, Malta, 2010.

M. Marimon. Tibidabo: a syntactically and semantically annotated corpus of spanish. *Corpora*, 10(3):259–276, 2015.

M. Marimon, N. Bel, and L. Padró. Automatic selection of hpsg-parsed sentences for treebank construction. *Computational Linguistics*, 40(3):523–531, 2014.

T. Matsuzaki, Y. Miyao, and J. Tsujii. Efficient hpsg parsing with supertagging and cfg-filtering. In *IJCAI*, pages 1671–1676, 2007.

Y. Miyao, T. Ninomiya, and J. Tsujii. Corpus-oriented grammar development for acquiring a head-driven phrase structure grammar from the penn treebank. In *Natural Language Processing–IJCNLP 2004*, pages 684–693. Springer, 2005.

L. Pineda and I. Meza. The spanish pronominal clitic system. *Procesamiento del lenguaje natural*, 34:67–103, 2005.

C. Pollard and I. A. Sag. *Head-driven phrase structure grammar*. University of Chicago Press, 1994.

M. Taulé, M. A. Martí, and M. Recasens. Ancora: Multilevel annotated corpora for catalan and spanish. In *LREC*, 2008.

An underspecification approach to Hausa resumption

Berthold Crysmann
Laboratoire de linguistique formelle, CNRS and U Paris-Diderot

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 194–214

Keywords: Hausa, resumption, underspecification, islands

Crysmann, Berthold. 2016. An underspecification approach to Hausa resumption. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 194–214. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.11.

Abstract

Within recent work on the treatment of resumption in HPSG, there is growing consensus that resumptive unbounded dependency constructions (=UDCs) should be modelled on a par with gap-type UDCs (Alotaibi and Borsley, 2013; Borsley, 2010; Crysmann, 2012b; Taghvaipour, 2005), using a single feature SLASH for both types of dependencies, rather than separate features, as proposed by Vaillette (2001a,b). Yet, authors disagree as to where exactly in the grammar the resumptive function of pronominals should be established: while Crysmann (2012b, 2015) advances an ambiguity approach that has pronominal *synsem* objects being ambiguous between a resumptive and an ordinary pronoun use, Borsley (2010); Alotaibi and Borsley (2013), by contrast, treat all pronominals, resumptive or not, as ordinary pronouns and effect their resumptive use by means of tailoring the SLASH amalgamation principle to potentially include pronominal indices. While their decision provides a straightforward account of McCloskey's generalisation that resumptives always look like the ordinary pronouns of the language, it fails to capture the difference in semantics between ordinary pronominal and resumptive uses.

In this paper, I shall reexamine the evidence from Hausa and propose to synthesise the approaches put forth by Alotaibi and Borsley (2013) and Crysmann (2012b), and propose that the potential for pronominal and resumptive function (including their difference w.r.t. semantics and non-local features) is captured by means of underspecification, yet the decision as to canonical vs. non-canonical use is made at the level of the governing head (Borsley, 2010; Alotaibi and Borsley, 2013). I shall argue that this division of labour is sufficient to derive the correct gap-like semantics for resumptives, maintains standard deterministic SLASH amalgamation, and, finally, provides an answer to McCloskey's generalisation.

1 Gaps and resumptives in Hausa

Unbounded dependency constructions in Hausa provide evidence for both gap and resumptive strategies in the grammar of extraction. Hausa employs a resumptive strategy with extraction of possessors or complements of prepositions. As shown in (1), possessor resumptives are realised as bound pronominal affixes, whereas true prepositions make use of the independent pronoun set. Use of a gap strategy is illicit in either of these constructions.

[†]I am gratefully indebted to the audience of HEADLEX 2016, the Joint Conference on LFG and HPSG for their insightful and stimulating comments, in particular to Bob Borsley, Tracy King, Adam Przepiorkowski, and Louisa Sadler. The paper in its present form has also greatly benefitted from comments by the four anonymous reviewers of the original abstract and the highly detailed and extremely helpful comments provided by the two non-anonymous reviewers of the edited proceedings, Ash Asudeh and Doug Arnold. The work presented here has been partially supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-LABX-0083).

(1) a. wā̀ ka àuri 'ya *(-r -sà) ?
who 2.M.CMPL marry daughter(F) -of.F -3.S.M
'Whose daughter did you marry?' (Jaggar, 2001)

b. sàndā sukà dṑkē shì dà *(ita)
stick 3P.CPL beat 3S.DO with 3S.F
'It was a stick they beat him with.' (Jaggar, 2001)

As witnessed in (2), indirect objects display overlap between the two strategies, i.e. both gaps and resumptives may be used in principle. Again, the resumptive is a pronominal affix fused with the indirect object marker.

(2) mutā̀nên dà sukà ƙi sayar musù / wà ∅ dà àbinci sukà fìta
men REL 3.P.CPL refuse sell to.them / to with food 3.P.CPL left
'the men they refused to sell food to left.' (Jaggar, 2001)

Human direct objects show a clear preference for zero realisation (Newman, 2000), at least with very short extractions.

(3) Gā̀ yārinyàr$_i$ dà ka sanī ∅$_j$ / ??san tà$_i$
here.is girl REL 2.S.F.CPL know / know her
'Here's the girl that you know.'

However, with extraction out of embedded clauses, both gaps and overt resumptives appear equally acceptable, as illustrated in (4).

(4) mùtumìn$_i$ dà ɗā̀lìbai sukà san [cḗwā mālàma-r-sù tanā̀
man REL students 3P.CPL know COMP teacher-L.F-3P.GEN 3.S.F.CONT
sô-n-sà$_i$ / sô ∅$_i$]
like.VN-L-3.S.M.GEN / like.VN
'the man that the students know that their teacher likes' (Newman, 2000, 539)

Similarly, as shown in (16) and (17) below, resumptives are also attested for human direct objects with ATB extraction.

However, with extraction out of strong islands, e.g. relative clauses, use of an overt resumptive becomes obligatory, both for indirect (5) and human direct objects (6).

(5) Gā̀ tābōbîn$_j$ dà Àli ya san mùtumìn$_i$ dà ∅$_i$ zâi yī
here.is cigarettes REL Ali 3.S.M.CPL know man REL 3.S.M.FUT do
musù$_j$ / *wà ∅$_j$ kwālī
to.them / to ∅ box
'Here are the cigarettes that Ali knows the man that (he) will make a box for.' (Tuller, 1986)

(6) Gà̄ mùtumìn$_j$ dà ka ga yārinyàr$_i$ dà $\emptyset_i$ ta san shì$_j$ /
here.is man REL 2.S.M.CPL see girl REL 3.S.F.CPL know him /
*sanī $\emptyset_j$
know
'Here's the man that you saw the girl that knows.' (Tuller, 1986)

Turning to non-human direct objects, Hausa observes an intricate interaction with argument drop: as shown by the contrasts below, non-human direct objects (7) permit argument drop, whereas human direct objects do not (8). Subjects in Hausa equally undergo argument drop, as shown by the presence vs. absence of a lexical subject in examples (4) and (2) above.

(7) a. Kā ga littāfì-n Mūsa?
2S.M.CPL see book-of Musa
'Did you see Musa's book?'

b. Ī, nā gan shì. / Ī, nā ganī $\emptyset$
Yes 1.S.CPL see 3S.M Yes 1.S.CPL see
'Yes, I saw it.' (Tuller, 1986, 61)

(8) a. Kā ga ƙanè-n Mūsa?
2S.M.CPL see brother-of Musa
'Did you see Musa's brother?'

b. Ī, nā gan shì. / *Ī, nā ganī $\emptyset$
Yes 1.S.CPL see 3S.M Yes 1.S.CPL see
'Yes, I saw him.' (Tuller, 1986, 62)

As argued by Tuller (1986), Hausa permits long relativisation without an overt resumptive in exactly those cases where the language independent licenses pro-drop, i.e. for subjects (9) and non-human direct objects (10).

(9) mùtumìn$_i$ dà ka san littāfìn$_j$ dà $\emptyset_i$ ya rubū̀tā $\emptyset_j$
man REL 2S.M.CPL know book REL 3S.M.CPL write
'the man that you know the book (he) wrote' (Tuller, 1986)

(10) littāfìn$_i$ dà ka san mùtumìn$_j$ dà $\emptyset_j$ ya rubū̀tā $\emptyset_i$
book REL 2S.M.CPL know man REL 3S.M.CPL write
'the book that you know the man who wrote (it)' (Tuller, 1986)

The possibility for long relativisation out of strong islands generalises from relative clauses to wh-islands, as shown in

(11) mùtumìn$_i$ dà ka san [mḕ$_j$ $\emptyset_i$ ya rubū̀tā $\emptyset_j$]
man REL 2S.M.CPL know what 3S.M.CPL write
'the man that you know what (he) wrote' (Tuller, 1986, 80)

(12) littāfìn$_i$ dà ka san [wā̀$_j$ $\emptyset_j$ ya rubū̀tā $\emptyset_i$]
book REL 2S.M.CPL know who 3S.M.CPL write
'the book that you know who wrote (it)' (Tuller, 1986, 80)

Note, however, that while long relativisation out of relatives and embedded wh-clauses is possible, long wh-extraction out of these islands is not. This holds for both subjects (13) and direct objects (14).

(13) * wànè mùtûm$_i$ ka bā nì littāfìn$_j$ dà $\emptyset_i$ ya rubū̀tā $\emptyset_j$
which man 2S.M.CPL give me book REL 3S.M.CPL write
'Which man did you give me the book that wrote' (Tuller, 1986, 81)

(14) * wànè littāfī̀$_j$ ka san wā̀$_i$ $\emptyset_i$ ya rubū̀tā $\emptyset_j$
which book 2S.M.CPL know who 3S.M.CPL write
'which book do you know who wrote' (Tuller, 1986, 80)

Furthermore, where an overt resumptive is required in situ, its presence has no effect on the acceptability of long wh extraction (cf. (15)).

(15) a. wā̀$_j$ ka yi màganā̀ dà shī$_j$
who 2S.M.CPL do talking with 3S.M
'Who did you talk with?' (Tuller, 1986, 158)

b. * wā̀$_j$ ka san mā̀târ$_i$ [dà $\emptyset_i$ ta yi màganā̀ dà shī$_j$]
who 2S.M.CPL know woman REL 3S.F.CPL do talking with 3S.M
'Who do you know the woman that talked to him' (Tuller, 1986, 159)

Hausa permits mixing of gap and resumptive strategies in ATB extraction, as shown in (16):

(16) [àbōkī-n-ā]$_i$ dà [[na zìyartā̀ $\emptyset_i$] àmmā [bàn sā̀mē shì$_i$ à
friend-L-1.S.GEN REL 1.S.CPL visit but 1.S.NEG.CPL find 3.S.M.DO at
gidā ba]]
home NEG
'my friend that I visited but did not find at home' (Newman, 2000, p. 539)

(17) mùtumìn$_i$ dà na bā shì$_i$ aro-n bàrgō-nā àmmā
man REL 1.S.CPL give 3.S.M.DO lending-L blanket-L.1.S.G but
duk dà hakà $\emptyset_i$ yakḕ jî-n sanyī
in spite of that $\emptyset$ 3.S.M.CONT feel-L cold
'the man whom I lent my blanket but who still felt cold' (Newman, 2000)

This observation suggests that resumptive and gap strategies should be compatible in principle.

As stated above, resumptive function in Hausa is independent of the mode of realisation: it is equally attested with independent pronouns, found with e.g. true prepositions, bound pronominals and even zero pronouns. There are in principle two ways to capture this generalisation: either one can assimilate the syntactic representation of zero-pronominals and pronominal affixes to that of independent pronouns by postulating a phonetically empty pronominal in syntax (*pro*), as assumed widely in Transformational Grammar, or else one can make the representation of resumptives independent of the lexical sign, and represent it instead on the argument structure of the governing head. Fortunately, the language provides the necessary evidence to choose among the two options: Hausa verbs (and nouns, for that matter) inflect according to the mode of realisation of direct object complements (Parsons, 1960; Crysmann, 2005), establishing a maximally three-fold distinction between (i) phrasal complements in situ (18a), (ii) pronominal affixes (18b), and (iii) non-realisation, which includes gaps (18c), intransitives, and object pro-drop (18d).

(18) a. nā ga/*gan/*ganī àbōkī-nā
1s.CMPL see.C friend-POSS.1.SG
'I saw my friend.'

b. nā gan/*ga/*ganī -shì
1s.CMPL see.B -3s.M
'I bought/read it.'

c. àbōkī dà na ganī
friend-POSS.1.SG REL 1s.CMPL see.A
'the friend that I saw'

d. nā ganī
1s.CMPL see.A
'I bought/read it/*him/*her .'

If pronominal affixation, pro-drop and extraction equally involve valence reduction, a unified account follows directly. However, any account that relies on the presence of a phonetically null pronominal to model resumption with object drop will end up making the wrong prediction w.r.t. verbal inflection. Moreover, since frame alternation is arguably a lexical process the difference between zero and non-zero NP complements will not be detectable on the verb's valence lists (which specify *synsem* objects, to the exclusion of PHON and DTRS).

Synopsis

To summarise the main points of the empirical patterns, we observed that Hausa witnesses both resumption and gap strategies, showing considerable overlap in their use: in principle, both gaps and resumptives can foot long distance dependencies, independently of whether we are dealing with relativisation or rather wh/focus fronting. This functional similarity is further confirmed by the compatibility of gaps and resumptives in ATB extraction. Once island constraints come into play, however, we observe a marked contrast: while wh extraction and focus fronting may never escape strong islands, relativisation is island-insensitive, provided a resumptive at the bottom of the dependency. Typologically, this is an interesting finding: depending on the type of unbounded dependency construction, Hausa resumptives may either pattern with gaps (wh extraction), or rather show a markedly distinct behaviour (relativisation). In the terminology of Asudeh (2011, 2012), Hausa resumptives are of the syntactically active type, as far as relativisation occurs, thus patterning with Hebrew, yet of the syntactically inactive type, once we consider wh extraction (cf. e.g. Vata).

2 Analysis

2.1 Resumption in HPSG

HPSG practitioners working on resumption (Alotaibi and Borsley, 2013; Taghvaipour, 2005; Crysmann, 2012b) currently agree that this unbounded dependency should be analysed on a par with gap-type dependencies in terms of a non-local dependency, uniformly represented by means of SLASH feature percolation. In contrast to previous work by Vaillette (2001a,b), use of a single feature for both types of non-local dependency facilitates the analysis of ATB extraction where a single filler can be terminated simultaneously by a gap in one conjunct and a resumptive in the other.

Where views differ, however, is whether or not these two types of non-local dependencies should be differentiated by other means. On one side of the spectrum, Borsley (2010) and Alotaibi and Borsley (2013) categorically deny the need to distinguish resumptive and gap type dependencies along the SLASH percolation path, arguing that, e.g. island effects should be attributed to performance, rather than competence. See, however, section 3.2 for critical discussion of this claim.

On the other end of the spectrum, Taghvaipour (2005) proposes an elaborate system whereby information about the top of the unbounded dependency construction, differentiating wh-fillers from ordinary and free relatives, is passed down via SLASH, alongside the filler's LOCAL value, which enables him to account for the distribution of gaps and resumptives in Persian in a fine-grained way depending on properties of the construction the filler is in. However, his partitioning according to dependency type (wh filler vs. free relative vs. ordinary relative) fails to make the right distinctions to account for island effects in Hausa.

My own previous proposal (Crysmann, 2012b) roughly covers the middle ground between the two aforementioned perspectives, permitting some degree of differentiation on SLASH values, while abstaining from a full-blown encoding of construction-specific features. Rather, I distinguish members of SLASH with respect to the minimum amount of information to be percolated, which is minimally a referential index (for relatives/resumptives), or a full *local* value (for wh-fillers/gaps), a distinction I have previously employed to account for difference in locality with complement clause vs. relative clause extraposition in German (Crysmann, 2013).

I shall now briefly sketch the proposals by Borsley (2010) and Crysmann (2012b, 2015), assess their respective advantages and shortcomings, and, subsequently, propose a synthesis of the two lines of analysis that combines their strengths while minimising the weaknesses.

2.1.1 Borsley (2010); Alotaibi and Borsley (2013)

In their analyses of resumption in Welsh and Arabic, Borsley (2010) and Alotaibi and Borsley (2013) follow McCloskey (2002) and argue that the morphological identity of resumptives to their non-resumptive pronominal counterparts militates against an approach in terms of lexical ambiguity. Instead, they suggest that resumptive are just the ordinary pronouns of the language, i.e. they do not launch a non-local dependency by themselves. In order to capture the ATB facts and to relate the pronoun to the non-local filler (wh/topicalisation) or the antecedent noun (relativisation), they suggest to effect the resumptive function on the governing head. To this end, they revise the principle of lexical SLASH amalgamation (Ginzburg and Sag, 2001) to optionally introduce an element into SLASH whose INDEX is structure-shared with that of a pronominal argument. While this approach correctly launches the non-local dependency without having to postulate lexical ambiguity between resumptive and ordinary pronouns, it fails to provide an account of the difference in semantics between resumptive and ordinary pronoun uses: as a result, resumptive use will end up having the same argument role be instantiated simultaneously by the pronoun at the bottom of the dependency, and by the relation contributed by the filler, at the top of the dependency. Furthermore, their revision of SLASH amalgamation turns an originally deterministic constraint into a non-deterministic one.

As we have seen in our discussion of the Hausa facts, resumption and gap-type extraction differ crucially with respect to island effects. In order to exert tight control on the distribution of gaps vs. resumptives it appears necessary to distinguish non-local dependencies with a gap at the foot from resumptive ones. Faced with a similar situation in Modern Standard Arabic, Alotaibi and Borsley (2013) exploit case to achieve this goal. However, this approach will not scale up to Hausa, since case is essentially unattested in the syntax of this primarily head-marking language.

2.1.2 Crysmann (2012b, 2015)

Just like Alotaibi and Borsley (2013), Crysmann (2012b) takes the ATB facts as evidence to model both gap and resumptive dependencies via a single set-valued feature SLASH. However, in order to capture the difference w.r.t. island-sensitivity, I distinguished the elements of this set as to whether they are full *local* values (wh- and focus fronting) or rather impoverished *local* values, minimally containing INDEX information (cf. Figure 2.1.2). In essence, resumption is likened to an obligatory anaphoric process under this perspective (see Asudeh, 2011, 2012 and Sells (1984) for similar intuitions). In contrast to Alotaibi and Borsley (2013), however, constraints on weight can be imposed along the SLASH percolation path, offering a way to capture difference in island sensitivity, as detailed by the constraints regarding *weak-local* SLASH values in Figures 3 and 4.

At the bottom of the dependency, gaps enforce reentrancy with SYNSEM.LOC, coercing the element in SLASH to *full-local*, whereas resumptives only observe a minimal requirement for INDEX-sharing, thus being compatible with both relatives and

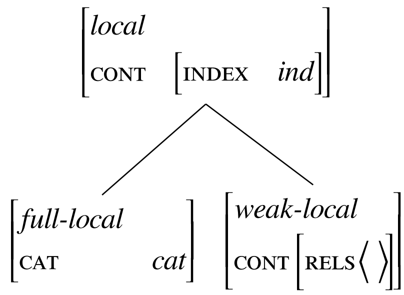


Figure 1: Hierarchy of *local*

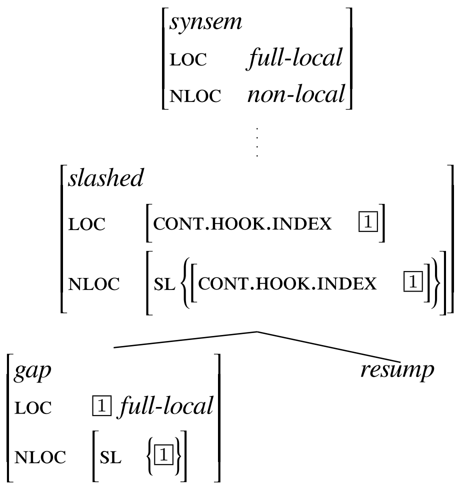


Figure 2: Hierarchy of synsem objects

wh-fronting, as shown in Figure 2.1.2. To generalise across bound and free pronominals, Crysmann (2012b) introduced disjunctive SLASH values for pronominal *synsem* objects. The implementation of this theory in Crysmann (2015), which also captures the semantic differences between resumptives and ordinary pronouns, employed lexical ambiguity. This not only led to doubling the number of lexical items for pronouns and pronominal affixation rules, but also failed to provide an account of McCloskey's generalisation, a rather sub-optimal solution.

At the top of the dependency, filler-head structures (Figure 3) impose full sharing of the filler's local value with an element in SLASH, thereby coercing this element's type to *full-local*. Relative complementisers (Figure 4), however, are content with index sharing, thus no coercion regarding the *local* sub-type will take place.

On the upside, the approach incorporated a treatment of island effects. To this end, retrieval sites, such as head-filler structures and relative complementisers constrain the set of SLASH values they pass on to be of the weaker anaphoric type. Since fillers and gaps are standardly subject to full sharing of local values, it follows that island effects ensue whenever a SLASH dependency features a gap at the bottom, or a filler at the top, i.e. only relatives footed by a resumptive are compatible with the con-

$$\begin{bmatrix} \textsc{ss} & \begin{bmatrix}\textsc{nloc}|\textsc{slash}\ \textit{set(weak-local)}\end{bmatrix} \\ \textsc{f-dtr} & \begin{bmatrix}\textsc{ss}|\textsc{loc}\ \boxed{l}\end{bmatrix} \\ \textsc{hd-dtr} & \begin{bmatrix}\textsc{ss}|\textsc{nloc} & \begin{bmatrix}\textsc{t-b}|\textsc{slash}\ \left\{\boxed{l}\right\}\end{bmatrix}\end{bmatrix} \end{bmatrix}$$

Figure 3: Filler-Head rule

$$\begin{bmatrix} \textsc{ss} & \begin{bmatrix}\textsc{loc}|\textsc{cat}|\textsc{head}|\textsc{mod}|\textsc{loc}|\textsc{cont}|\textsc{index}\ \boxed{i} \\ \textsc{nloc}|\textsc{slash}\ \textit{set(weak-local)}\end{bmatrix} \\ \textsc{hd-dtr} & \begin{bmatrix}\textsc{ss}|\textsc{nloc} & \begin{bmatrix}\textsc{t-b}|\textsc{slash}\ \left\{\begin{bmatrix}\textsc{cont}|\textsc{index}\ \boxed{i}\end{bmatrix}\right\}\end{bmatrix}\end{bmatrix} \end{bmatrix}$$

Figure 4: Relative complementiser

straint regarding *weak-local* , since neither end enforces full reentrancy, and therefore can escape islands. For illustration, consult the constraints imposed by head-filler structures and relative complementisers in Figures 3 and 4: since both relatives and head-filler structures (wh/focus fronting) are only transparent to *weak-local*, only those unbounded dependencies can cross where neither the top, nor the bottom of the dependency coerces the relevant SLASH element to full sharing. Finally, note that resumptives are not pretyped to *weak-local* (cf. Figure 2.1.2), but rather underspecified: thus, they are still compatible with *full-local* fillers, as long as no island constraints are imposed along the path.

2.2 A synthesis

In order to overcome the motivational problems associated with an ambiguity approach, I shall synthesise the respective proposals by Borsley and Crysmann. In essence, I propose that the potential to launch a non-local dependency vs. having pronoun semantics should be captured by way of underspecification. The decision on slashed realisation, however, is imposed on the argument structure of the governing head. As a net effect, this approach captures the semantic difference between ordinary pronominal and resumptive uses, keep the original deterministic formulation of SLASH amalgamation, and provide an explanation of McCloskey's generalisation.

To this end, I shall refine, in a first step, the type hierarchy of *synsem* objects along the lines of Figure 5. In essence, I propose a primary distinction between slashed and unslashed realisation, the former of which comprises *gap* and purely resumptive subtypes. Orthogonal to this distinction, I introduce pronominal synsem objects, which may resolve to either unslashed ordinary pronouns or slashed resumptives.

Having an underspecified common super-type for resumptive and ordinary pronoun uses directly avoids disjunctive specification in the representation of pronominals, regardless of whether they are free, bound or zero. Syntactic and semantic

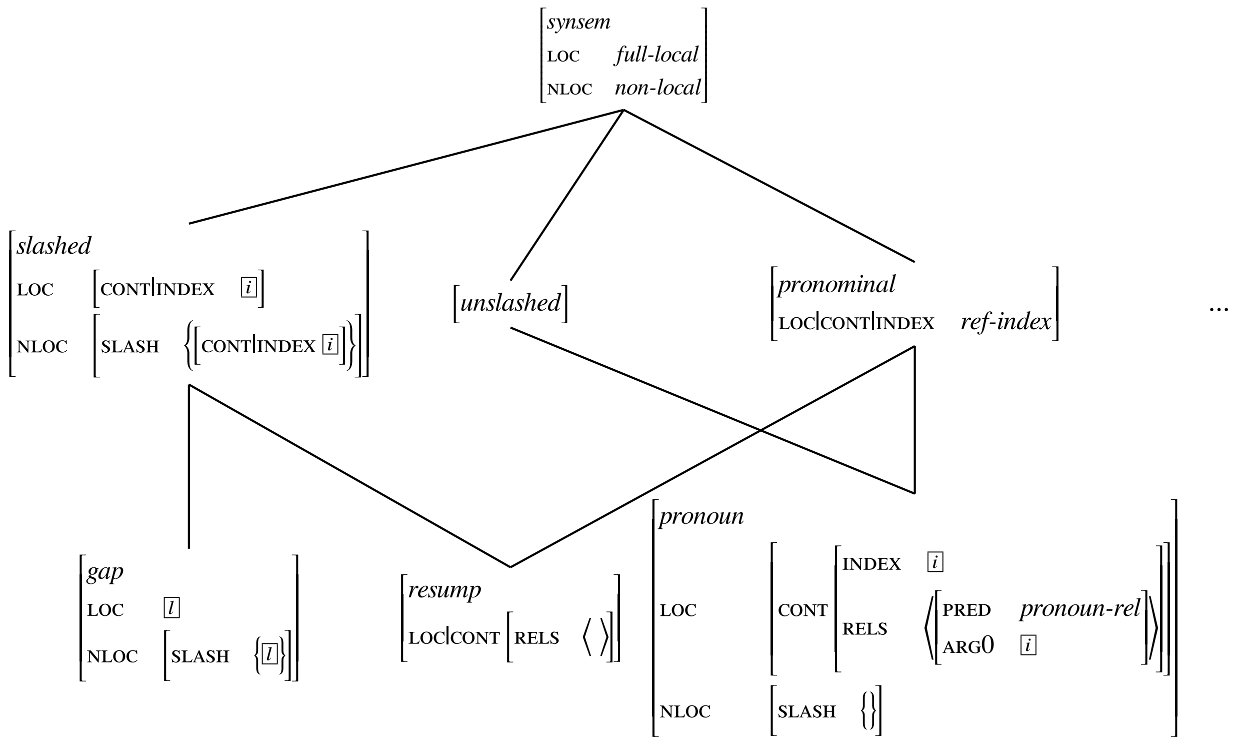


Figure 5: Hierarchy of synsem objects

differences are captured as latent constraints on the sub-types: if unslashed realisation is chosen, *pronominal(-synsem)* is specialised to *pronoun(-synsem)*, applying all constraints associated with this type (empty SLASH and non-empty semantics). If, by contrast, slashed realisation is chosen, *pronominal(-synsem)* is specialised to *resump*, enforcing a non-empty SLASH, yet empty semantics. Note that the constraints associated with *resump* only require minimal INDEX-sharing, following previous proposals by both Borsley and Crysmann.

Incorporating insights from Borsley, the ultimate decision on realisation type is associated with the governing head, i.e. crucially external to the pronominal itself: using a pair of lexical rules each, direct object (and subject) valencies are segregated into *slashed* and *unslashed*, i.e. the subject and the first complement are specialised to one of these two *synsem* sub-types.[1] Subsequent lexical rules of pronominal affixation or zero pronominal realisation have the desired effects owing to the intersection of types pertaining to the *slashed/unslashed* distinction with those relating to pronominal status. Similarly, syntactic combination with a free pronoun will result in either resumptive (*slashed* ∧ *pronominal* = *resump*) or ordinary pronominal use (*unslashed* ∧ *pronominal* = *pronoun*). Thus, in contrast to Crysmann (2015), this approach only ever needs a single pronominal affixation rule for any cell of the paradigm, or else a single lexical entry for each independent pronoun.

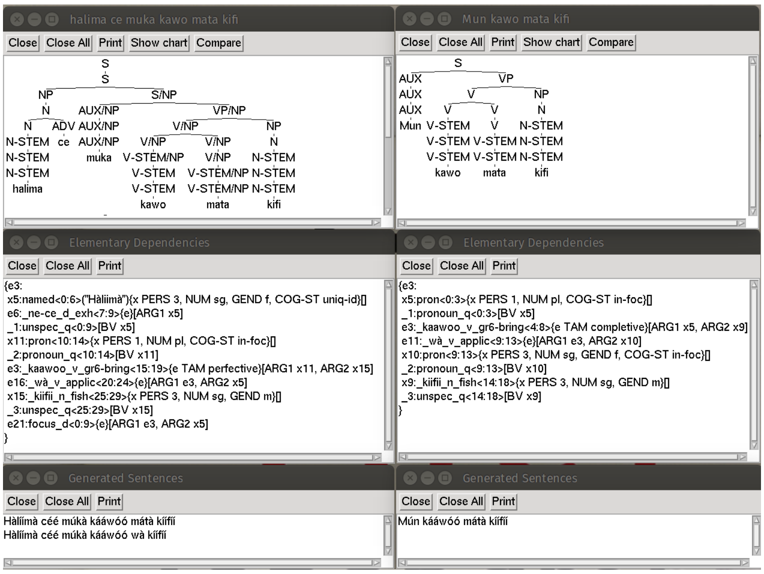
Given that pronominal arguments under the current account provide for the possibility of being slashed or not (in contrast to Borsley), standard HPSG SLASH amalgamation and head-driven propagation of NON-LOCAL features will ensure proper launching and percolation of gap and resumptive dependencies alike.

The synthesis of Crysmann and Borsley seamlessly integrates with the weight-based theory of island effects developed in my previous works. Since all I do here is relocate the decision between slashed and unslashed realisation from the dependent onto the governing head, the distinction between minimal sharing (for resumptives) and full sharing (for gaps) is fully maintained in the *synsem* type hierarchy. Together with the associated consequences regarding the weight of local values on SLASH, the selective transparency of islands for *weak-local* applies unmodified.

I have so far implicitly assumed that underspecification improves on lexical ambiguity not only in terms of economy of description, but that it is also instrumental in providing an answer to McCloskey's generalisation. To make this point fully explicit, let me summarise how the present approach accounts for the fact that in languages offering resumption, it is all pronouns, and only pronouns that do assume this function. The answer offered by the present approach is two-fold: as to the first clause (all pronouns), it is sufficient to assume that languages vary as to whether they include *pronominal* or only the more specific type *pronoun* in their descriptions of pronominals. The answer to the second part of the generalisation is slightly more complex: as suggested by the present approach, disambiguation according to resumptive vs. pronominal use requires statement of a semantic relation for non-resumptive uses.

[1] Although Hausa verbs may take both direct and indirect objects, the latter are complements of the applicative marker *wà* (Abdoulaye, 1992).

Since type hierarchies are static, a single, concrete relation needs to be provided. It so happens that pronouns are the prototypical elements that can provide a constant relation, yet still fill every cell of the paradigm, making them compatible at the INDEX-level with every potential antecedent. Thus, instead of postulating different principles to account for resumption, this approach merely postulates a more abstract representation of what constitutes a pronominal.



UDCs Pronoun

Figure 6: Sample analyses and generator results: Parse tree, MRS elementary dependencies and generator result

The take on the semantics of resumptive vs. ordinary pronoun use in terms of latent syntactic and semantic constraints differs from the one adopted by Asudeh (2011, 2012), who assumes that resumptive pronouns create a resource surplus (pronominal semantics) that is later consumed by a manager resource (contributed at the top of the dependency). While Asudeh's approach is certainly workable within the specific confines of LFG and Linear Logic (see the detailed discussion below), the present approach offers the further advantage of providing identical semantic representations for gaps and resumptives. Given the overlap of the two extraction strategies, uniformity of representation is a highly desirable property, since paraphrasis in generation falls out directly.

Figure 6 provides sample analyses of Hausa pronominals in both resumptive and ordinary pronoun function, as implemented in the computational grammar HaG (http://hag.delph-in.net/logon; Crysmann, 2012a): at the top, we see the simplified parse trees for *Halima ce muka kawo mata kifi.* 'It's Halima we brought fish for.' and *Mun kawo mata kifi.* 'We brought her fish.', respectively. The resumptive indirect object pronoun[2] *mata* on the left is characterised by a SLASH value that is amalgamated onto the verb, and the TAM/agreement auxiliary *mukà*. The ordinary indirect object pronoun *mata* on the right, by contrast, has an empty SLASH value.

Just below each parse tree you find the semantics (MRS in elementary dependency format): with resumptive function, the value of ARG2 (x5) of the applicative relation is reentrant with the proper name *Halima*, and there is no pronominal relation other than that for the first plural subject. This is indeed the exact same semantic representation as one would obtain with a gap in lieu of the resumptive: regenerating the surface strings from these semantics yields a gap-type realisation (*Halima ce muka kawo wa kifi.*) alongside the resumptive one. With ordinary pronoun function, as shown on the right, the value of ARG2 of the applicative (x10) is reentrant with a third singular feminine pronoun. Regenerating from these semantics only yields a single surface string, containing an overt pronominal.

2.2.1 Adjuncts

The current take on resumption follows Borsley in identifying argument structure as the locus where the decision between resumptive and non-resumptive function is placed. This move raises the obvious question how adjuncts will be integrated under this perspective. There are essentially two sub-questions to be addressed here: first, resumptives contained within adjuncts, and second, adjunct resumptives.

Empirically, resumptives contained within adjuncts are well-attested: they are found, inter alia, with certain "true" prepositions, such as *dàgà* 'from' and *dà* 'with'.

(19) a. sàndā sukà dṑkē shì dà ita
stick 3P.CPL beat 3S.DO with 3S.F
'It was a stick they beat him with.' (Jaggar, 2001)

b. * sàndā sukà dṑkē shì dà ∅
stick 3P.CPL beat 3S.DO with
'It was a stick they beat him with.'

In addition to "true" prepositions, genitive prepositions (also known as prepositional nouns), may take overt resumptive complements. In contrast to the former, yet parallel to verbs and verbal nouns, we also find zero realisation here.[3]

(20) a. àdakā̀ mukàn sâ kuɗi-n-mù ciki -n -tà
box 1PL.HAB put money-L-1P inside L 3S.F
'It's inside a box we usually put our money.'

[2] The grammar treats indirect objects as inflected forms of the applicative marker *wà*.

[3] While it is clear that prepositional nouns admit zero pronominal direct objects with non-human reference (cf. Tuller, 1986, p. 357), as well as long extraction out of relatives (cf. Tuller, 1986, p. 361), it remains open whether prepositional nouns support filler-gap dependencies as well, e.g. for extraction of non-human referents. Examples where the locatum is animate are rare in general and the examples provided in Tuller (1986) are, unfortunately, inconclusive with respect to gap status.

b. àdakā̀ mukàn sâ kuɗi-n-mù ciki ∅
box 1PL.HAB put money-L-1P inside
'It's inside a box we usually put our money.' (Jaggar, 2001)

We can conclude that adjunct status per se does not constitute an island in Hausa, at least not for resumptive dependencies (see footnote 3). As for launching the non-local dependency, note that the resumptives in question are local complements of the prepositional head, i.e. adjunct status is involved in SLASH passing, but not in SLASH introduction.

As shown by the data above, the necessity of permitting overt or covert resumptives within adjuncts is evident in Hausa. In order to integrate the possibility for an adjunct daughter to contribute to the mother's SLASH value, all it takes is to complement head-driven SLASH percolation with a specific constraint on head-adjunct structures that determines the mother's SLASH value on the basis of both the head and the adjunct daughter. There are several ways to accomplish that: in versions of HPSG that are based on the Generalised Head Feature Principle of Ginzburg and Sag (2001), head adjunct phrases will merely constitute a specific override of general default SYNSEM sharing. Note further that exceptional SLASH passing out of adjuncts has already been attested for English, e.g. in the context of parasitic gaps (Pollard and Sag, 1994). Similarly, the English Resource Grammar (ERG; Copestake and Flickinger, 2000) permits SLASH inheritance from adjuncts in order to account for preposition stranding.

The second central question regarding adjuncts is whether or not they give rise to resumptives themselves. As far as Hausa is concerned, this does not seem to be the case (Newman, 2000; Jaggar, 2001): either we find stranding of the preposition (with a gap or resumptive), or else the entire adjunct phrase is pied-piped, as illustrated in (21).

(21) a. à Kanṑ akà hàifē nì
at Kano 4.CPL give.birth 1S.DO
'It was in Kano I was born' (Jaggar, 2001)

b. dà sàndā sukà dṑkē shì
with stick 3P.CPL beat 3S.DO
'It was a stick they beat him with.' (Jaggar, 2001)

c. ciki -n àdakā̀ mukàn sâ kuɗi-n-mù
inside L box 1PL.HAB put money-L-1P
'It's inside a box we usually put our money.' (Jaggar, 2001)

Incidentally, this observation regarding adjuncts is replicated in an even stronger form in Coptic Egyptian: while arguments in this language only ever relativise by means of resumption, adjuncts constitute the only instance where we find a gap-type dependency in Coptic (Crysmann and Reintges, 2014). This asymmetry in the grammar of resumption is not entirely unexpected: as argued on the basis of the semantics of intersective modifier attachment (Levine, 2003), adjunct extraction is best conceived as syntactic, whereas a lexical account appears preferable for argument extraction (Pollard and Sag, 1994, ch. 9).

3 Discussion

3.1 Comparison to Asudeh (2011, 2012)

Within LFG, Asudeh (2011, 2012) has developed a theory of resumption that departs from the assumption that resumptive pronouns are always the standard pronouns of the language. In order to neutralise the semantic surplus contributed by the pronoun in the case of resumptive use (see our discussion in section 2.1.1 above), he invokes a so-called "manager resource" to consume the extra pronominal semantics. Asudeh's theory further distinguishes syntactically inactive resumptives, which are indistinguishable from gaps in terms of island-sensitivity and across-the-board extraction (inter alia), from syntactically active ones, which contrast with gaps in being island-insensitive. While the distribution and interpretation of syntactically active resumptives is captured entirely in terms of obligatory anaphoric binding between the top and the bottom of the dependency, i.e. in semantic structure, syntactically inactive resumptives are linked to the filler or relativiser by means of functional equality, just like filler-gap dependencies. However, in order to circumvent a violation of Functional Uniqueness, the functional equation relating the top of the dependency to the base needs to invoke feature restriction (Kaplan and Wedekind, 1993) to discard identity of PRED values in case of resumption, yet enforce identity otherwise.

Trying to apply Asudeh's theory to the data at hand faces the obvious question as to whether we are dealing with resumptives of the syntactically active or rather the inactive kind. As we have seen in the synopsis of section 1 on page 5 above, island sensitivity of Hausa resumptives is differentiated according to the type of unbounded dependency construction: while relativisation footed by a resumptive is island insensitive, giving rise to "long" relativisation, wh extraction clearly is island sensitive, treating resumptives and gaps on a par. One way of making sense of this situation in Asudeh's terms, is to assume that Hausa has both syntactically active and inactive resumptives, constructionally distinguishing between the two. As a consequence, wh and focus fronted fillers will employ functional equality (island sensitive), whereas relativisation will employ either functional equality or anaphoric binding, depending on the construction. Since relative unbounded dependencies license gaps in Hausa, functional equality must be included as one of the options. Similarly, as witnessed by (16), ATB extraction treats resumptives and gaps on a par, so we are likely to include functional equality under restriction of PRED as an option. Finally, for island-insensitive long relativisation, we will need to include anaphoric binding, additionally being restricted to long extraction, in order to avoid spurious ambiguity. As a result, the phrase structure rule introducing relative complementisers will be three-ways disjunctive.

In the context of the present paper, the most important question is as to how Asudeh's approach compares to the one proposed here on the technical as well as the conceptual level, and whether it may benefit current approaches in HPSG?

On a conceptual level, it seems that underspecification with latent semantics and equally latent SLASH specification displays some similarity to what is achieved by feature restriction. The main difference is, however, that our present analysis never retracts any information, but rather expands an underspecified type to either pronominal semantics or UDC syntax, depending on whether the pronominal happens to display ordinary pronoun or resumptive function. Furthermore, underspecification not only attacks the syntactic side of the problem, but also inserts pronominal semantics exactly as required, by means of type inference. As a result, a resource surplus never arises, obviating the need to stipulate special manager resources to handle it. Therefore, the present approach is compatible with a wide range of approaches to semantic representation, including standard MRS or LRS (Richter and Sailer, 2003), rather than being dependent on a particular logic.

Furthermore, the current approach captures filler–gap dependencies as well as syntactically active and inactive resumptive dependencies by one and the same mechanism (SLASH percolation), rather than two different ones, as is the case for Asudeh (2011, 2012). Here, differences pertaining to island sensitivity are captured exclusively by means of the constraints imposed along the extraction path, targetting the type of local values admissible on SLASH. As a direct benefit, both syntactically active and inactive resumptives are predicted to be compatible with gaps in ATB extraction, irrespective of their status regarding island-sensitivity: this situation is found e.g. with Hebrew resumptives, a situation which is not directly captured by Asudeh's account, where only syntactically inactive resumptives are assimilated to gaps.

On a technical level, differences become even more pronounced: first, HPSG does not recognise the existence of a separate level of f-structure containing semantic predicates yet being distinct from semantics proper. Second, as far as I am aware, restriction is not a commonly assumed operation on feature structures in HPSG theory.[4] Furthermore, the idea of setting aside parts of a feature structure not only runs counter to the spirit of HPSG, which rather exploits the feature geometry to abstract over sets of features, but it also conflicts severely with the appropriateness function of typed feature structures. To conclude, despite some similarities in basic intuitions, a literal adoption of Asudeh's LFG approach appears to be at odds, both technically and conceptually, with basic assumptions of HPSG.

3.2 The place of island constraints

A property that the current proposal shares with e.g. Asudeh (2011, 2012) but that crucially distinguishes it from Borsley (2010) and Alotaibi and Borsley (2013) pertains to the possibility of handling island-sensitivity grammar-internally.

For Welsh, where there is no difference between gaps and resumptives in terms of island sensitivity, Borsley (2010) rightfully concludes that there is no need to draw

[4] Restriction is actually used in processing with HPSG, in order to increase packing rates for local ambiguity factoring (Oepen and Carroll, 2000). However, during unpacking, features that have been restricted out are reconstructed deterministically and indiscriminately.

a distinction according to extraction strategy. However, Alotaibi and Borsley (2013) settle on the same conclusion for Modern Standard Arabic, despite the fact that in this language resumption facilitates wh extraction from strong islands, whereas gap-type extraction is illicit in these contexts. To work around this problem, they refer, inter alia, to Hofmeister and Sag (2010), claiming that the acceptability contrast might just as well be attributed to performance effects. Interestingly enough, though, Hofmeister and Sag (2010) do not discuss resumption at all. Moreover, Alotaibi and Borsley (2013) do not offer any processing constraint that may explain the contrast. This becomes even more difficult if the grammatical treatment does not draw any distinction in terms of the non-local dependency, which, in the case of Alotaibi and Borsley (2013) is uniform SLASH passing.

A study, however, that may shed some light on the question is Alexopoulo and Keller (2007): investigating (intrusive) resumptives in English, German, and Greek, they observe that use of resumptive elements improve acceptability with weak islands and deep nesting without island constraint violations, yet do not improve acceptability for strong islands, most notably extraction out of relative clauses. They explicitly correlate this difference with the competence/performance distinction, concluding that strong island effects should be handled by the grammar.

In Hausa, however, resumption improves acceptability even for strong islands, suggesting that we are dealing with truly grammatical, not intrusive, resumption (in the sense of Sells, 1984).[5] Yet, as shown in the data discussion above, Hausa still observes a marked contrast depending on the top of the dependency: while relativisation out of relatives or wh-islands is possible, wh-extraction out of these construction remains ungrammatical, regardless of the use of a resumptive. I therefore conclude that this selective insensitivity to strong islands calls for an treatment in grammatical terms, as offered, e.g. by the weight-based perspective I have proposed in Crysmann (2012b) and Crysmann (2013).

4 Conclusion

In this paper I have proposed a synthesis of the approaches by Borsley and Crysmann regarding the treatment of resumptive and gap dependencies in HPSG and applied it to the case of Hausa. I have argued more specifically that a proper account of resumptive vs. ordinary pronoun semantics can be provided in HPSG on the basis of underspecification in a type hierarchy of *synsem* values. In order to address McCloskey's generalisation, the present approach embraces Borsley's idea that the decision with respect to resumptive function should be associated with the governing head and its argument structure. Concerning the representation of pronominals, however, the present take favours an approach in terms of underspecification, in order to facilitate both compositional semantics and the treatment of SLASH propagation. In future work, I shall establish how the current proposal will scale up to the treatment

[5] This is further corroborated by the fact that some lexical heads, e.g. true prepositioins require the presence of a resumptive pronouns independently of the complexity of the extraction construction.

of Modern Standard Arabic or Irish.

References

Abdoulaye, Mahamane L. 1992. *Aspects of Hausa Morphosyntax in Role and Reference Grammar*. Ph. D.thesis, SUNY Buffalo, NY.

Alexopoulo, Theodora and Keller, Frank. 2007. Locality, Cyclicity, and Resumption: At the Interface between the Grammar and the Human Sentence Processor. *Language* 83(1), 110–160.

Alotaibi, Mansour and Borsley, Robert D. 2013. Gaps and Resumptive Pronouns in Modern Standard Arabic. In Stefan Müller (ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar, Freie Universität Berlin*, pages 6–26.

Asudeh, Ash. 2011. Towards a unified theory of resumption. In Alain Rouveret (ed.), *Resumptive Pronouns at the Interfaces*, pages 121–187, Amsterdam: John Benjamins.

Asudeh, Ash. 2012. *The Logic of Pronominal Resumption*. Oxford: Oxford University Press.

Borsley, Robert D. 2010. An HPSG Approach to Welsh Unbounded Dependencies. In Stefan Müller (ed.), *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar, Université Paris Diderot, Paris 7, France*, pages 80–100, Stanford, CA: CSLI Publications.

Copestake, Ann and Flickinger, Dan. 2000. An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the Second conference on Language Resources and Evaluation (LREC-2000)*, Athens.

Crysmann, Berthold. 2005. An Inflectional Approach to Hausa Final Vowel Shortening. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 2004*, pages 73–112, Kluwer.

Crysmann, Berthold. 2012a. HaG — an implemented grammar of Hausa. In Michael R. Marlo, Nikki B. Adams, Christopher R. Green, Michelle Morrison and Tristan M. Purvis (eds.), *Selected Proceedings of the 42nd Annual Conference on African Linguistics (ACAL 42)*, pages 321–337, Somerville, MA: Cascadilla Press.

Crysmann, Berthold. 2012b. Resumption and Island-hood in Hausa. In Philippe de Groote and Mark-Jan Nederhof (eds.), *Formal Grammar. 15th and 16th International Conference on Formal Grammar, FG 2010 Copenhagen, Denmark, August 2010, FG 2011 Lubljana, Slovenia, August 2011*, volume 7395 of *Lecture Notes in Computer Science*, pages 50–65, Springer.

Crysmann, Berthold. 2013. On the Locality of Complement Clause and Relative Clause Extraposition. In Gert Webelhuth, Manfred Sailer and Heike Walker (eds.), *Rightward Movement in a Comparative Perspective*, pages 369–396, Amsterdam: John Benjamins.

Crysmann, Berthold. 2015. Resumption and Extraction in an Implemented HPSG of Hausa. In *Proceedings of the ACL-IJNLP workshop on Grammar Engineering Across Frameworks (GEAF-2015), Beijing, China*, ACL.

Crysmann, Berthold and Reintges, Chris H. 2014. The polyfunctionality of Coptic Egyptian Relative Complementisers. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University of Buffalo*, pages 63–82, Stanford: CSLI publications.

Ginzburg, Jonathan and Sag, Ivan. 2001. *Interrogative Investigations: the Form, Meaning and Use of English Interrogatives*. Stanford: CSLI publications.

Hofmeister, Philip and Sag, Ivan A. 2010. Cognitive constraints and island effects. *Language* 86(2), 366–415.

Jaggar, Philip. 2001. *Hausa*. Amsterdam: John Benjamins.

Kaplan, Ron and Wedekind, Jürgen. 1993. Restriction and Correspondance-based Translation. In *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics (EACL 1993)*, pages 193–202.

Levine, Robert D. 2003. Adjunct valents: cumulative scoping adverbial constructions and impossible descriptions. In Jongbok Kim and Stephen Wechsler (eds.), *The Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar*, pages 209–232, Stanford: CSLI Publications.

McCloskey, James. 2002. Resumptives, Successive Cyclicity, and the Locality of Operations. In Samuel David Epstein and T. Daniel Seely (eds.), *Derivation and Explanation in the Minimalist Program*, pages 184–226, Oxford: Blackwell.

Newman, Paul. 2000. *The Hausa Language. An Encyclopedic Reference Grammar*. New Haven, CT: Yale University Press.

Oepen, Stephan and Carroll, John. 2000. Ambiguity packing in constraint-based parsing - practical results. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics*, pages 162–169, Seattle, WA.

Parsons, Fred W. 1960. The Verbal System in Hausa. *Afrika und Übersee* 44, 1–36.

Pollard, Carl and Sag, Ivan. 1994. *Head–Driven Phrase Structure Grammar*. Stanford: CSLI and University of Chicago Press.

Richter, Frank and Sailer, Manfred. 2003. Basic Concepts of Lexical Resource Semantics. In Arne Beckmann and Norbert Preining (eds.), *ESSLLI 2003 - Course Material I*, volume 5 of *Collegium Logicum*, pages 87–143, Wien: Kurt Gödel Society.

Sells, Peter. 1984. *Syntax and Semantics of Resumptive Pronouns*. Ph. D.thesis, University of Massachusetts at Amherst.

Taghvaipour, Mehran. 2005. *Persian Relative Clauses in Head-driven Phrase Structure Grammar*. Ph. D.thesis, University of Essex.

Tuller, Laurice A. 1986. *Bijective Relations in Universal Grammar and the Syntax of Hausa*. Ph. D.thesis, UCLA, Ann Arbor.

Vaillette, Nathan. 2001a. Hebrew relative clauses in HPSG. In Dan Flickinger and Andreas Kathol (eds.), *The Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar*, pages 305–324, Stanford: CSLI Publications.

Vaillette, Nathan. 2001b. Irish Gaps and Resumptive Pronouns in HPSG. In Frank Van Eynde, Dorothee Beermann and Lars Hellan (eds.), *The Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, pages 284–299, Stanford: CSLI Publications.

On the perfect tense-aspect in K'ichee'an Mayan: An LFG approach

Lachlan Duncan

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 215–235

Keywords: perfect tense-aspect, K'ichee'an Mayan

Duncan, Lachlan. 2016. On the perfect tense-aspect in K'ichee'an Mayan: An LFG approach. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 215–235. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.12.

Abstract

Previous accounts of the perfect tense-aspect in the K'ichee'an languages have concluded that the category or part-of-speech of the perfect is a verb, or less often, a participle. We believe otherwise. Empirical support is presented for the hypothesis that the perfect is expressed using either a deverbal participial adjective or a deverbal possessed nominal in the form of a detransitivized non-verbal predicate. Although the perfect is always expressed as a one-place intransitive, the perfect retains the capacity to express two argument roles. We argue that the perfect is, in fact, a perfect. We present the various semantic types of perfect, including the perfect of result and the experiential perfect, and also show temporal restrictions that constrain the perfect. The analyses are implemented using the syntactic architecture of LFG.

1 Introduction

The perfect 'tense-aspect' of the K'ichee'an languages has attracted only the most cursory attention in the descriptivist Mayan literature (Dayley 1985, Larsen 1988, Mondloch 1978).[1,2] The situation improves with Classic Mayan in epigraphy and historical linguistics (MacLeod 2004, Wald 2007), yet the only theoretical analysis of the perfect is Bohnemeyer's (2002, inter alia) semantics research on Yukatek Mayan. We attempt to contribute to this research by reporting on the perfect tense-aspect of K'ichee' Mayan using the formal apparatus of LFG. This remains a preliminary step only because a more thorough knowledge of the perfect can only be gained through an understanding of its semantic and pragmatic components.

The perfect is referred to here as a tense-aspect because of the difficulty in determining its formal status: tense, aspect, or hybrid (Comrie 1976, Hornstein 1990, Kibort 2009, Klein 1994, Ritz 2012)? A further complication is that the Mayan languages with few exceptions are acknowledged to be grammatically tenseless (Bohnemeyer 2002 for Yukatek, inter alia). Aspect/mood has been grammaticalized

1. Many thanks to Helen Charters & Frank van Eynde; to Oleg Belyaev, Tibor Laczkó, John Lowe, & Helge Lødrup for discussion at HeadLex16, Warsaw, Poland; and to the editors Miriam Butt & Tracy Holloway King. I am indebted to my K'ichee' Maya consultants from Totonicapán, Guatemala. K'ichee' data is from the author's field work, unless otherwise noted. All the usual disclaimers apply.

2. x = [−voi] alveopalatal fricative, j = [−voi] velar fricative, [*x*'/ '] = glottalized occlusive / glottal stop; (**x*) / *(*x*) / [*x*] = *x* is ungrammatical / obligatory / reconstructed; - / < space > = morpheme / word boundary. ABBREVIATIONS: first / second / third person = 1 / 2 / 3, absolutive / ergative = ABS / ERG, adjective = A, adjunct = ADJ, actor focus = AF, adverb = ADV, agreement marker = AM, antipassive = AP, causative = CAUS, completive / incompletive aspect = COM / INC, definite = DEF, determiner = DET / D, discourse function = DF, directional = DIR, distributive = DISTR, derived transitive verb in -j / ' = DTJ / DT', finite = FIN, grammatical function = GF, inchoative = INCH, interrogative = INT, intensifier = INTS, irrealis = IRR, negative = NEG, non-verbal predicate negative = NVP.NEG, nominalizer = NOML, numeral = NUM, participle = PART, passive / completive passive / stative passive = PASS / COM.PASS / STAT.PASS, perfect = PERF, present perfect = PP, plural = PL, genitive possessor = POSS, positional = POSL, possession = Poss, preposition = PREP / P, independent pronoun = PRO, sentence = S, stem-forming vowel = SFV, singular = SG / S, tense = TNS, transitive / intransitive / dependent phrase final marker = TPF / IPF / DPF, (in)transitive verb = IV / TV.

on finite verbs, but not tense. The consensus in the Mayanist literature is that the perfect tense-aspect is a verb (5), or less often, a participle (9). We reject these proposals and suggest alternate accounts. We contend that K'ichee'an perfects are not finite verbs, an argument based partly on the absence of prefixed aspect morphology, which, we suggest, represents the crucial diagnostic of the verb category.[3] And because K'ichee'an perfects are non-periphrastic, they cannot be finite verbs due to the absence of auxiliaries. We show that what is often claimed to be a perfect is a perfect. We propose that the K'ichee' and Tz'utujiil perfects are statives, and deploy exclusively as non-verbal predicates. We show that K'ichee' and Tz'utujiil perfects are expressed with two distinct parts-of-speech: a participle-like deverbal adjective and a deverbal possessed nominal. The adjective functions as an attributive while both adjective and nominal function as predicates. Our analysis resolves multiple inconsistencies that exist with previous approaches and predicts several outcomes. Most notably, that the predicative perfect is grammaticalized as an intransitive, irrespective of the transitivity of the root or stem from which it was derived.

The paper is organized along the following lines. Section 2 reviews the literature on the perfect for K'ichee' and Tz'utujiil. Section 3 examines the grammatical constructions used with the perfect, outlines the basic semantic types available to the perfect, and provides a discussion on the perfect. Section 4 develops an account of the perfect using the architecture of LFG. Section 5 concludes the paper.

2 Background

2.1 Introduction In the K'ichee'aan languages in general, the verb category is composed of a single agglutinating constituent.[4,5] More specifically, verbs are morphologically marked with prefixed aspect and mood markers and with suffixed mood markers, when required, but not tense markers. Verbs host person and number marking absolutive (ABS) and ergative (ERG) agreement markers (AM):

(1) x-ee-w-il-o TV
COM-3PLABS-1SERG-see-TPF
'I saw them (Duncan 2013).'

(2) x-ix-b'iin-ik IV
COM-2PLABS-walk-IPF
'You all walked.'

Non-verbal predicates (NVP) are zero-copular (Duncan 2013): a NV possessed nominal predicate in (3), and a NV adjectival predicate in (4):

3. One exception in K'ichee is the morphologically unmarked imperative. Evidence exists that this diagnostic consistently holds across the Mayan language family. Marginal variation occurs in Itzaj (Hofling 2001), while in Tzeltal (Shklovsky 2005) prefixed/pre-predicate aspect morphology with non-standard tense-like behaviour co-occurs with the perfect.

4. Kichee' data in Introduction of Background section 2.1.

5. This excludes the non-bound periphrastic continuous aspect constituent *(ka)tajin*.

Table 1 ERG, POSS, ABS AMs and PRO in K'ichee'

	ergative		*possessive*		*absolutive*		*pronoun*	
	sg	*pl*	*sg*	*pl*	*sg*	*pl*	*sg*	*pl*
1.	in/w	qa/q	nu/w	qa/q	in	uj/oj	in	oj
2.	aa/aw	ii/iw	aa/aw	ii/iw	at/aa	ix	at	ix
3.	u(u)/r	ki/k	u(u)/r	ki/k	ø	ee	are'	a'are'

(3) (aree) ø aw-achi'l
3sPRO 3sABS 2sPOSS-friend
'(S/he) [is] your friend.'

(4) sib'alaj ee jeb'al
very.much 3PLABS pretty
'They are very pretty.'

The ergative (set A), absolutive (set B), and possessive AMs and independent pronouns of K'ichee' are shown in Table 1.

2.2 Tz'utujil Tz'utujil's perfect "indicates an activity that was completed in the past but that has some relevance to the present." [6] The finite verb paradigm falls into two mutually exclusive divisions: the nonperfect and the perfect. The nonperfect verb (1)–(2) must begin with a tense, aspect, and/or mode prefix, and may require a suffix that is dependent on the verb class.The perfect verb never uses TAM prefixes but requires a suffix that changes according to verb class: *–naq* for intransitive verbs (5), and *–(o)on* / *–(u)un* / *–(a)an* for transitive verbs (7).

Some perfect intransitive verbs (5) also function as adjectives (6), with a meaning of having acquired the state indicated by the intransitive verb:

(5) ee war-naq
3PLABS sleep.IV-PERF
'They have slept (Dayley 1985:77).'

(6) ee war-naq
3PLABS sleep.ADJECTIVE-PERF
'They are asleep (Dayley 1985:77).'

(7) in ki-kuuna-an
1sABS 3PLERG-cure.DTJ-PERF
'They have cured me (Dayley 1985:76).'

The perfect stems of transitives may also be considered as past passive participial adjectives ("adjectival passives"). In these cases they have passive meanings and inflect only for patients, using the absolutive AMs (9). Compare the derived transitive verb (DT') in (8) to the passivized predicate adjective (A) in (9):[7]

6. Dayley (1985:74–79, 213–5, 343–4, 352–4) for Tz'utujiil section 2.2.
7. (9) can also be translated as 'You have all been loved' (see Dayley 1985:343).

(8) ix q-ajo'-oon
2PLABS 1PLERG-love.DT'-PERF
'We have loved you all (D 1985:78).'

(9) ix ajo'-oon
2PLABS love.ADJECTIVE-PERF
'You are all loved (Dayley 1985:79).'

Past passive participial adjectives, which are stative predicates, differ from verbal passives: the former emphasize the state that results from the transitive activity (10):

(10) ja ti nuu-chaaq' chaj-il jaay
DET little 1SPOSS-little.brother guarder.of.NOML house
ya'-oon kan (r-mal nuu-tee')
put-PERF staying 3SPOSS-by 1SPOSS-mother
'My little brother has been made house watcher (by my mother) (D 1985:344).'

Agent focus perfect participles are adjectives derived from transitive verbs (11). The agent is highlighted by being put in focus before the adjective. While the sentence is semantically transitive the agent focus adjective is a stative predicate:

(11) n-mama' ee tzuq-uyun ja meeb'a'-ii'
1SERG-grandfather 3PLABS feed.ADJECTIVE-PERF DET orphan-PL
'My grandfather is the one who has fed the orphans (Dayley 1985:353).'

2.3 K'ichee' The intransitive suffix *-inaq* derives perfect participles ("deverbal adjectives") from intransitive verbs that function as "special" NVPs.[8] When predicative, they can be translated as verbs in the perfect aspect (14a). With patient-like subjects (12), they can be considered stative NVPs (Larsen 1988:186, 193):

(12) e' kam-inaq
3PLABS die-PERF
'They have died, they are dead.'

(13) ø peet-inaq
3SABS come-PERF
'S/he is coming.'

The transitive perfect suffixes *–oom/–uum/–m* derive perfect passive participles from transitive verb roots, and are used as noun modifiers (16) and adjectival predicates (9). Indicating the perfect in a transitive clause, perfect passive participles represent their objects as absolutive AMs and their subjects as ergative AMs (14a).

However Larsen conjectures that the perfect's prefixed AMs might instead represent possessive pronouns, because the first person singular AM *–nu* represents the possessive AM, not the ergative AM *–in* (14b) (see Table 1). In the end, Larsen remains ambivalent, and seems to settle for the transitive participle approach (14a):

8. Larsen (1988:185–8, 207–8 fn. 15, 230, 234–8, 281 fn. 7) for K'ichee' in section 2.3.

(14) at nu-ch'ay-oom
2sABS 1sERG/1sPOSS-hit-PERF
(a) 'I have hit you (Larsen 1988:236).' *nu-* is 1sERG
(b) 'You are my one-who-has-been-hit (Larsen 1988:238).' *nu-* is 1sPOSS

Perfect participles can also be used attributively, both intransitive (15) and transitive (16), and nominally (17):

(15) jun kam-inaq tz'i'
one die-PERF dog
'a dead dog (Larsen 1988:187)'

(16) tzak-om saqmo'l
cook-PERF egg
'boiled egg (Larsen 1988:235)'

(17) nu-mok-oom
1sPOSS-ask.services.of-PERF
'my servant (Larsen 1988:236)'

(18) in b'iin-inaq
1sABS walk-PERF
'I (will) have/had walked (L 1988:185).'

Because participle-based NVPs are not marked with aspect markers or tense, they can be interpreted, depending on context, as present, past, or future perfect (18).

2.4 Nahualá K'ichee' The data in (19) demonstrate Mondloch's (1978:127) treatment of K'ichee's perfect as a verb.[9] Mondloch's interpretation is based on the word order of SVO: *lee nujii'* as the subject, *r-uk'a'm* as the ergative-marked perfect verb, and *lee chiim* as the object:[10]

(19) lee nu-jii'$_{subj}$ r-uk'a'-m$_{verb}$ lee chiim$_{obj}$
DET my-son.in.law 3sERG-carry-PERF DET bag
areetaq x-oopan chuwa w-o'ch
when COM-arrive at my-house
'My son-in-law was carrying the bag when he arrived at my house.'

In addition to the 'present,' the perfect aspect also occurs in the 'past' and 'future' (18). When the accompanying clause uses the completive aspect, the pluperfect aspect is used to translate the perfect. When the accompanying clause includes an incompletive aspect, the perfect is translated as the future perfect aspect.

2.5 Perfects in non-K'ichee'an Mayan In epigraphic studies, Wald (2007) investigates the Classic Mayan perfect, analogizing it to Tzeltal and Tzotzil Mayan perfects. He builds on MacLeod's (2004) insight that perfect morphology is represented in the Classic Mayan glyphs. MacLeod (2004:292) suggests that Classic Mayan perfects, used in the form of derived transitive verbs with the suffix **–VVj* < ***–V–ej*, originated as perfect participles. The perfect suffix is cognate with suffixes

9. Mondloch (1978:127, 130, 134; 1981:85, 89–90, inter alia) for Nahuala K'ichee' in section 2.4.
10. Mondloch uses the present continuous when glossing the K'ichee' perfect (19).

on gerunds and inflected perfect verbs in Tzeltalan and Tojolobal. Wald argues that all so-called perfect forms are verbs, but are not 'actional' perfects. Rather he claims that they are 'statal' perfects, or resultatives. Wald (2007:316-9) acknowledges little agreement with Kaufman (1971, 1972), who indicates that Tzeltal and Tzotzil Mayan perfects include a transitive verb in *–ej/–oj*, a perfect active participle in **–em*, and a perfect passive participle in **–b'il* (**–ab'* passive (Kaufman 1989)).

In Zinacantán Tzotzil, verbs in the indicative mood inflect for prefixed aspect markers except for the perfect aspect, which uses suffixes: intransitive stems use *–em* (20), transitive stems use *–oj* (21), and passive verbs use *–bil* (22).[11] Transitives mark subjects with prefixed ergatives (set A), while the perfect aspect and NVPs require the suffixed absolutives (set B): (Aissen 1987:43–4, 66)

(20) tal-em-on
come-PERF-1SABS
'I have come.'

(21) av-il-oj
2SERG-see-PERF
'You have seen it.'

(22) pech-bil-on
bind-PASS.PERF-1SABS
'I have been bound.'

The 3rd person pluralizer *–ik* in (23) agrees with the inanimate nominal *ak'u'ik* but only because the clause is a NVP – plural agreement with inanimates is not permitted with verbs. Aissen therefore claims, at least, that intransitive perfects are NVPs. Aissen surmises further that "[t]here is some evidence that perfects are not verbs, but A's, for they suffix –uk/–ik– under negation" (24).[12] All three perfects when negated suffix the NVP negative marker *–uk/ik*: (Aissen 1987:53, 59)

(23) jat-em-**ik** a-k'u'-ik
torn-PERF-3PL 2SERG-shirt-2PL
'Your (PL) shirts are torn.'

(24) mu k-il-oj-**uk**
NEG 3SERG-see-PERF-NVP.NEG
'I did not see it.'

According to Vinogradov (2014:42), Tzotzil finite verbs must be aspectually marked for completive or incompletive. Thus the perfect and posterior aspects, which are not permitted to use prefixed aspect morphology, are non-finite participles.

In Petalcingo Tzeltal, there is an agent-oriented transitive perfect verb in *–oj/–ej* (25), (26), a patient-oriented transitive perfect deverbal in *–bil* (27), and an intransitive perfect verb in *–em/–en* (28) (Shklovsky 2005:48–9, 58–61). The *–oj/–ej* perfect requires the ergative, but the *–bil* perfect does not permit it. The *–bil* and *–em* perfects can also be used attributively, but not the *–oj/–ej* perfect:[13]

(25) s-mil-oj-ø
3SERG-kill-PERF-3SABS
'He has killed him/her/it.'

(26) x-jel-oj-ø
INC.3SERG-change-PERF-3SABS
'He will have changed it.'

11. Aissen (1987:5, 13, 41–3, 59 fn 1/fn 7, 66, 96–7, 117–8).
12. The 'A' class includes all stems that can predicate, but that are not nouns or verbs. A's cannot inflect for aspect, nor can they combine with genitives. We assume A's mainly consist of adjectives.
13. Shklovsky (2005:59, 59, 60, 61) for (25), (26), (27), and (28). Note (26): *xjeloj* < *x-s-jel-oj*.

(27) te mut mil-bil ta ts'i'
DET chicken kill-PERF PREP dog
'The chicken was killed by the dog.'

(28) yahl-em te alal-e
fall-PERF DET child-CL
'The child has fallen.'

Tzotzil exhibits 'unmarked aspect': zero marking aspects depending on the verb's transitivity. Intransitives mark the incompletive with *x–*, while transitives mark the completive with *laj*. Therefore (25) represents a completive intransitive perfect, while (26) is an incompletive intransitive perfect.[14]

In Chol Mayan, stative predicates are not permitted to use aspect morphology, unlike eventive predicates which require an initial aspectual marker (Coon 2010:29, 38, 204–5). The suffix *–em*, used on intransitive roots/derived intransitive stems, forms stative predicates (29). *–em* can also be used on passives derived from transitive roots (30), and on positional roots: (Coon 2010:204, 205)

(29) jul-em-ety-ix
arrive.here-PERF-2SABS-already
'You arrived here already.'

(30) mejk'-em-oñ
hug.PASS-PERF-1SABS
'I've been hugged.'

In Itzaj Mayan, transitive perfect verbs (*–maj*) (31) do not permit aspectual prefixes (Hofling 2000:50, 55, 165–72, 368–71, 369 fn. 6). Perfect participles (*–maja'an*) (32) are formed from transitive stems, but have both active and passive meanings:[15]

(31) uy-il-maj-ech
3SERG-see-PERF-2SABS
'She has seen you (Hofling 2000:50).'

(32) b'o'ol-maj-a'an (Hofl 2000:170)
pay-PERF-PART
'has been a payer, has been paid'

The Mayan languages discussed here, except possibly Petalcingo Tzeltal, do not permit prefixed / pre-predicate aspect marking on perfects. Another issue is evident in Chol, where the *–em* perfect suffix marks both transitive and intransitive perfects. Crucially transitive stems are morphologically marked as passive and detransitivized. This correlates to the K'ichee'an where the transitive root/derived transitive stem is detransitivized as a *–Vm* perfect—but without overt passive morphology. Although Aissen is inconsistent about the perfect's category, she claims, correctly we argue, that Tzotzil perfects are likely not verbs but NVPs. This adds crucial support to the hypothesis presented in this paper. In light of the various Mayan data discussed here, we believe that strong correlations are evident with the K'ichee'an data in this paper, allowing for language-specific lexical and morphological variation.

14. If this is correct, then, why are the two argument intransitive perfects in (25) and (26) glossed as two place transitives? Because Shklovsky provides no substantive account for this puzzle, we assume that the *–oj/–ej* perfect is an intransitive, as originally marked, and a possessed nominal predicate. Therefore we gloss (25) as, 'S/he/it is his one-who-has-been-killed.'

15. We gloss (31) as the NVP, 'You are her one-who-has-been-seen.'

2.6 Perfects and resultatives in non-Mayan languages It is not difficult to show that K'ichee'an perfects are indeed perfects, and further, are not verbs. Whether K'ichee'an perfects include resultative properties is less clear, although our data suggest otherwise. Yet judging by examples (6), (9), and (12), it appears that the descriptivists implicitly accept that perfects and resultatives share properties.

Regarding resultatives and statives, the former are defined as "those verb forms that express a state implying a previous event" (Nedjalkov & Jaxontov 1988:5–6), while statives "may denote natural, primary states which do not result from any previous event." Notwithstanding Wald's conclusions, Nedjalkov & Jaxontov (1988:7) themselves admit that "it is not always easy to distinguish between statives and resultatives." Therefore they amend their definition of resultatives to include a narrow use (resultative) and a broad use (resultative + derived stative).

Furthermore, "[s]ince [resultatives and statives] are very close to each other, in languages that possess both categories they are often interchangeable" (Nedjalkov 2001:928). For example, Ancient Greek's active perfect is considered to be a stative by one author and a resultative by yet another. There are Russian dialects that have resultative converbs that are in the process of acquiring perfect features, like non-terminative verbs ('to know, to walk') (Nedjalkov 2001:936–7). Lithuanian has perfects with properties of resultatives: the perfect (33) can be used as the resultative (34) simply by including an adverb of duration — here, *jau metai*:[16]

(33) *Jis [yra] mir-ęs* PERF
'He has died.'

(34) *Jis [yra] mir-ęs jau metai* RES
'He is dead already for a year.'

And recall that, unlike the resultative, the perfect:

> can be derived from any verb, either transitive, or intransitive, either terminative or durative, including those verbs that denote situations which do not change the state of any participant (e.g., verbs meaning 'to work', 'to sing', 'to laugh', etc.) (Nedjalkov & Jaxontov 1988:15)

Typological and detailed studies of European languages demonstrate just how unstable the present perfect, in particular, is, how the perfect can shapeshift along a gradient of 'perfect-ness'—result to simple past—and how language-specific that variability can be. This is a natural by product of the volatile diachrony of the periphrastic perfect, in the European languages at least, from statives to resultatives to perfects of result to indefinite past perfects to preterites and the simple past (see Bybee et al.:1994). Comrie (1976:53, 61) advises caution in adopting 'perfect' designations because in languages like Latin, in some Romance languages (French, Italian, Romanian) "the so-called Perfect covers both perfect and non-perfect meaning" and "[g]iven that the perfect partakes of both present and past, it is possible for languages to differ over just how present or past their perfect forms are."

Languages vary in their restrictions on the possible meaning types of perfect. A Brazilian Portuguese dialect greatly restricts meanings of the present perfect

16. Lithuanian data from Nedjalkov (2001 citing E. Geniušienė p. c.)

(PP) (Kibort 2009 citing Laca, Cabredo-Hofherr, & Carvalho 2007). The universal meaning of the PP is allowed. But the experiential, resultative, and recent anteriority meanings and adverbs *ja* 'already' and *ainda não* 'still not' are not permitted.

Languages also vary to the extent perfects and simple past tenses share properties. Dutch allows property sharing, which is transparent in the analysis of the 'aspectual' and 'past time' PPs and the latter's close semantic relationship with the simple past (Van Eynde 2000:231–49).[17] The aspectual PP allows combinations of durational and temporal adverbs, while the past time PP allows only past and durational temporal adverbs. The past time PP is more similar semantically to the simple past than to the aspectual perfect, with the English equivalent of the past time PP translated as the simple past. Yet the Dutch past time PP and the simple past are not identical. The former has a semelfactive interpretation, whereas the latter an habitual interpretation (Van Eynde p. c.).

3 Grammar and meaning of the K'ichee' perfect

3.1 Grammatical forms The perfect is very productive in K'ichee'an, being used with an array of grammatical constructions. They include the following intransitive and transitive roots and stems: completive passives/mediopassives (*–taj*), antipassives (*–n*), causatives (*–isa*), positionals (*–l*) (see (44)), reflexives (reflexive pronoun), pseudo noun-incorporated (NP), actor focus (*–Vw/–n*), directionals (*ee–*).

The verb stem (35) is composed of a verb root and the detransitivizing completive passive suffix (*–taj*). The verb stem in (36) consists of a nominal root, a stem-forming vowel, and the detransitivizing antipassive suffix (*–n*):

(35)	b'an-taj-inaq	lee	ja
	make-COM.PASS-PERF	DET	house
	'The house has gotten done/finished.'		

(36)	at	kun-a-n-naq
	2SABS	cure-SFV-AP-PERF
	'You have cured.'	

With the transitive verb in (37), the reflexive pronoun *–iib'* shows subject agreement. However the reflexive pronoun in (38) shows agreement with the perfect's possessor, not the grammatical subject; the perfect's subject is the null set (ø):[18]

(37)	ka-**ki**-chaj-i-j	**k**-iib'
	INC-3PLERG-guard-SFV-DTJ	3PLPOSS-self:REFL
	'They take care of themselves.'	

(38)	laa	ø	**ki**-tij-(o)-om	**k**-iib'/*r-iib'?
	INT	3SABS	3PLPOSS-teach-SFV-PERF	3PLPOSS-self/3SPOSS-self
	'Have they taught themselves?'			

17. The past time perfect is *not* a pluperfect.

18. (38) 'Is it their ones-who-have-been-taught [by] themselves?'

The verb stem in (39) consists of multiple derivational morphemes. The *–ir* suffix derives the inchoative intransitive verb *b'aqir* from the adjective *b'aq*, while the *–(i)sa* suffix derives the transitive causative stem *b'aqirsa* from the inchoative:

(39)	ee	b'aq-ir-sa-an	ja	ch'uu'	(r-umaal	ja	q'iiq')
	3PLABS	thin-INCH-CAUS-PERF	DET	fish	3SERG-by	DET	wind

'The fish have been made thin (by the wind) (Dayley 1985:344).'

The verb stem in (40) is a transitive root in a pseudo noun-incorporating construction. The bare nominal *k'ax* is the subject of the perfect possessed nominal NVP:[19]

(40)	le	kaqulja'	sib'alaj	u-b'an-om	(*k'i)	k'ax
	DET	storm	very.much	3SPOSS-make-PERF	many	bad

'The storm has done a lot of damage.'

The verb stem in (41) consists of a transitive root and the detransitivizing actor focus suffix (*–Vw/–n*). The grammatical subject, which must also be the agent or experiencer, must immediately precede the absolutive AM (ABS). It is curious that the actor focus perfect, which requires root transitives or derived transitive stems, uses the *–inaq* perfect suffix, not the anticipated *–Vm* perfect suffix. The verb root in (42) is *war* 'sleep' with the directional prefix *ee* 'go' (Dayley 1985:98):

(41)	aree	in	il-ow-inaq	in	(42)	in	ee-war-naq	TZ'UTUJIL
	3SPRO	1SABS	see-AF-PERF	1SPRO		1SABS	go-sleep-PERF	

(41) 'S/he (is the one who) has seen me.' (42) 'I have gone and slept (1985:98).'

3.2 Semantic types of the perfect The various types of perfects available in the K'ichee'an languages are outlined here. They include the perfect of result, the experiential perfect, and the perfect of the extended now (XN)/universal perfect.

A perfect of result is shown in (43). Comrie (1976:56) claims that in a result perfect "a present state is referred to as being the result of some past situation":

(43)	ee	b'ee-naq	lee	ixoq-iib'	chi	pa	Nawala
	3PLABS	go-PERF	DET	woman-PL	already	PRED	N.

'The women have already gone to Nahualá.'

Experiential, or indefinite past, perfects with atelic predicates are shown in (44)–(45). They indicate that "a given situation has held at least once during some time in the past leading up to the present" (Comrie 1976:58):[20]

19. 'It is the storm's thing-that-has-been-done-bad'/'It is the storm's badness-that-has-been-done.'
20. 'Great Grandfather Maximon (Tied-Up One) is my one-who-has-been-seen in Santiago Atitlán.'

(44) ee k'oo-l-inaq Tzolola
3PLABS exist-POSL-PERF Tz.
'They have been to Sololá.'

(45) w-il-om Rii Laj Mam Maximon pa Santiago Atitlán
1SERG-see-PERF DET INTS grandfather M. PREP S.
'I have seen Great Grandfather Maximon in Santiago Atitlán.'

The universal perfect, or the extended now perfect (XN), which requires an atelic predicate, is shown in (46)–(47). The universal perfect describes "a situation that started in the past but continues (persists) into the present" (Comrie 1976:60):[21,22]

(46) pa taq lajuj haab' ee k'oo-l-inaq pa Santa Cruz
PREP DISTR ten year 3PLABS exist-POSL-PERF PREP S.
'They have lived in Santa Cruz for ten years.'

(47) y k'oo jun laj u-baaq r-uk'a'-m b'iik
and exist NUM small 3SPOSS-bone 3SPOSS-receive-PERF DIR
'And he had a small bone he was carrying along (Mondloch 1978:194).'

3.3 Perfects? We demonstrate here the temporal constraints on the K'ichee'an perfect, and that it is, in fact, a perfect, not a resultative. Temporal adverb phrases, like *ojeer* 'before' (48) and *myeer* 'earlier' (49), can be used with the perfect:[23]

(48) **ojeer** k'oo-l-inaq pa wee k'oo-l-ib'al
before:ADV be-POSL-PERF PREP DET be-POSL-LOC
'He has been at this place before.'

(49) la **myeer** at ul-inaq
INT earlier 2SABS arrive-PERF
'Did you arrive earlier?'

Using a past temporal adverb with definite time is not permitted in the perfect:

(50) ***iwiir pa taq a las quatro** ee b'ee-naq pa ja
yesterday PREP DISTR P DET four 3PLABS go-PERF PREP house
(*'Yesterday at four o'clock they have gone home.')

Future temporal adverbs are allowed when the accompanying verb is incompletive:

21. The perfect *ruk'a'm b'iik* in (47) is the predicate of a headless relative clause.
22. The perfect *ruk'a'm* in (58) is atelic and an extended now perfect.
23. It is not possible to translate (49) into English as a perfect.

(51) ee ok-inaq chi lee ee ixoq-iib' pa k'ayib'al
3PLABS enter-PERF already DET PLU woman-PL PREP market

areetaq **k**-ee-q-il **chwe'q**
when INC-3PLABS-1SERG-see tomorrow

'The women will have already entered the market when we see them tomorrow (Mondloch 1978:134).'

Adverbs of motion can be used with K'ichee'an perfects (52), (53). Motion adverbs target the perfect's activity, and accordingly, are ungrammatical with resultatives:[24]

(52) **no'jimal** ee b'-inaq pa ch'iich' pa San Pedro la Laguna
slowly 3PLABS go-PERF PREP car PREP S.
'They have gone slowly by car to San Pedro la Laguna.'

(53) **aninaq** uj b'iin-inaq pa nim-alaj k'ay-i-b'al
quickly 1PLABS walk-PERF PREP big-INT buy-SFV-LOC
'We have walked quickly to the main market.'

Other linguistic pointers can assist in determining whether the K'ichee'an perfect is a perfect. Bybee et al. (1994:54) suggest that anteriors (perfects) are "often accompanied by the relational adverbs 'already' and 'just'." Michaelis (1996) claims that the adverb phrases "*already* and *still* express the presence of a state at a given reference time," and that "the contrast between *already* and *still* [i]s one involving the presence versus absence of a transition at some time prior to reference time."

The adverb phrase *chik* 'already' is routinely used with the K'ichee'an perfect, particularly with the 'pluperfect,' where it is used to distinguish the pluperfect from the 'present,' and is used even with the 'future perfect.'

In addition, the adverb *still* (54) is an established test for distinguishing resultatives from perfects (Nedjalkov & Jaxontov 1988:15–6):

(54) a. 'She is (still) gone.' RESULTATIVE
b. 'She has already/just/*still gone.' PERFECT

The adverb *maja'an* 'still not' tests for perfects, and *k'a* 'still' tests for resultatives:

(55) maja'an ee b'ee-naq
NEG.still 3PLABS go-PERF
'They have still not gone.'

(56) *k'a ee b'ee-naq
still 3PLABS go-PERF
(*'They have still gone.')

We infer from data in section 3.3 that the perfect is a perfect, not a resultative.

24. *b'eenaq/b'inaq* 'gone' < b'ee 'go', –inaq perfect.

4 Towards an analysis

4.1 Contesting previous proposals We discuss why the approach advocated in this paper is preferred to previous accounts of the perfect. Some of Dayley's (1985) claims are either contradictory or improbable (see (8) & (9)). For one, the perfects host subjects marked with absolutive AMs, except for the solitary case of the 'transitive' perfect, where the object is supposedly referenced by the absolutive AM. Prima facie, it is counterintuitive and inconsistent, particularly when possessed perfects have been shown to be nominals (17). Next the *–oon* perfect suffix marks both an intransitive non-verbal adjectival predicate (8) and a transitive verb (9). This configuration is highly unlikely in K'ichee'an where differentiating transitivity from intransitivity is paramount. Dayley (1985:73) himself states that:

> [i]n Tz'utujil there is a very important morphological distinction between *intransitive verbs* (IVs) and *transitive verbs* (TVs) with respect to their inflection as well as to their derivational possibilities.

The *–Vm* perfect licences a by-phrase (57a). We argue that (57a) and (57b) are identical in truth-conditional terms, although not in syntactic nor in informational structural terms. The proposition that (57a) is intransitive while (57b) is transitive simply because of the addition of the prefixed possessive AM is again improbable: they are one-place intransitives. These data show that the *–Vm* perfect is syntactically indistinguishable from the morphologically marked *–tal* stative passive (62):

(57) a.

at	il-om	(w-umal	**in**)
2SABS	see-PERF	1SPOSS-by	1SPRO

'You are the *person-who-has-been-seen* (by **me**).'

b.

at	**w**-il-oom	(*[w-umal	**in**])
2SABS	1SPOSS-see-PERF	1SPOSS-by	1SPRO

'You are **my** *person-who-has-been-seen* (*[by me]).'

Although the descriptivists interpret perfects such as (57b) and (58) as binary GF transitives (cf. (19)), we contend they instead represent a one-place intransitive non-verbal predicate with two argument roles. The matrix clause's subject in (58), *lee chiim*, is referenced by the third person absolutive AM, $ø_j$. The clause-initial DP, *lee nu-jii'*, is both a possessor and an external topic, an example of "external possession" (Aissen 1999:178). The DP agrees with the third person singular possessive pronoun *r–*, prefixed to the perfect-marked nominal *–uk'a'–m*:

(58)

lee	nu-jii'$_i$	$ø_j$	r$_i$-uk'a'-m	lee	chiim$_j$
DET	1SPOSS-son.in.law	3SABS	3SPOSS-carry-PERF	DET	bag

areetaq	x-oopan	ch-u-wa	w-o'ch
when	COM-arrive	PREP-3SPOSS-at	1SPOSS-house

'As for my son-in-law$_i$, the bag$_j$ is his$_i$ *thing-that-has-been-carried* when he$_i$ arrived at my house.'

Mondloch's analysis of the grammatical functions of (58) is shown in *a*., while the analysis proposed in this paper is shown in *b*.:

(59) *a*.	*lee nujii'* = subject,	*ruk'a'm* = verb,	*lee chiim* = object
b.	*lee chiim* = subject,	*ruk'a'm* = possessed noun,	*lee nujii'* = ext possessor

The grammatical subject, *lee ak'alaab'* 'the children' in (60), agrees with the absolutive AM *ee*, and is the only possible subject in the sentence. Compare (60) with (51): the prefixed possessive AM qa$_j$– is not the subject:[25]

(60)	ee$_i$	qa$_j$-mul-im	chi	lee	ak'al-aab'$_i$	pa	ja
	3PLABS	1PLPOSS-gather-PERF	already	DET	child-PL	PREP	house

areetaq	k-ee-ul	lee	ki-naan
when	INC-3PLABS-arrive	DET	3PLPOSS-mother

'We will have already gathered the children in the house when their mothers arrive (Mondloch 1978:127).'

4.2 Passives We review the K'ichee' passive in light of our claim that the *–Vm* perfect is a passive. K'ichee' has three morphologically marked passives: standard passive, vowel ablaut/*–(V)x* (61); completive passive, *–(i)taj* (35); and stative passive, *–(i)tal* (62).[26] While the standard and completive passives are verbs, the stative passive is a participial adjective, and uses the same NVP syntactic configuration as the perfect. All three passives license *–umaal* by-phrases and intransitive phrase final suffixes (IPF), the latter normally seen only on verbs and positional adjectives (*–l/–r*).[27] Note that the completive passive, a verb, and the stative passive, a NVP, both use the same passive morphology (*–ta*). Stative passive participial adjectives are not syntactically, not morphologically, not semantically the same as verbal passives. Though the data below gloss similarly, the stative passive represents the state achieved by the verb's action, whereas the passive focuses on the action itself:

(61)	k-uj-iil	k-umaal	(62)	uj	il-ital	k-umaal
	INC-1PLA-see:PASS	3PLPOSS-by		1PLABS	see-STAT.PASS	3PLPOSS-by
	'We are seen by them.'			'We are seen by them.'		

4.3 The Proposal We show that the K'ichee'an perfect is a perfect. We contend that the K'ichee'an perfect has been grammaticalized as a stative non-verbal predicate, but one that, nonetheless, retains a verb's argument structure. This is not surprising because the proto-Indo-European (PIE) perfect (Beekes 1995) and the

25. (60) 'The children$_i$ are our$_j$ ones$_i$-who-will-have-already-been-gathered in the house when their$_i$ mothers arrive.'

26. *–taj* < *–ta* passive, *–j* inchoative; *–tal* < *–ta* passive, *–l* positional (Larsen 1988).

27. The *–umaal* by-phrase, formally a possessed noun, is traditionally interpreted as a preposition syntactically. We suggest it is better analysed as a secondary predicate, a possessed nominal NVP.

English perfect (Katz 2003), amongst others, are considered statives. The absolutive AM is always cross-referenced as the grammatical subject in all non-verbal predicates, including perfects and positionals. The perfect consists of two different parts-of-speech: a deverbal participial adjective, and a deverbal possessed nominal.

Participial adjectives, which include the positional adjectives (Duncan 2013), differ from standard adjectives in that participial adjectives are excluded from the derivational paradigm of adjectives (Dayley 1985, Larsen 1988). That is, abstract nouns, intransitive inchoatives, and causatives (see (39)) can be derived from standard adjectives, but not from participial adjectives. The latter also never take the attributive suffix used on many attributives, nor the degree suffix (Larsen 1988).

It is generally held that participles in matrix clauses cannot act as finite verbs without a supporting infrastructure.[28] One LFG analysis of auxiliary verbs assumes that auxiliaries lack the PRED attribute, and provide only inflectional TAM attributes. Thus the participle of the analytic construction is understood as the 'finite' PRED-supplying verb (Bresnan 2001:78, King 1995:225–8). Moreover, if we accept that verb-derived deverbal participles can themselves be verbs, V_{Part} (Bresnan 1982:23), or alternatively, are category-neutral for attributive/predicate adjectives and verbs $[V_{Part}]_{A/V}$ (Kibort 2005), then the K'ichee'an perfect cannot be a participle.

The relationship between the derived adjectival and nominal perfects can be accounted for in several ways. A lexical process operates on the adjective to derive a nominal, while a second lexical process operates on the nominal to add a POSS function to its argument list. An alternate approach is to assume that the perfect in its uninflected derived form remains underspecified as an adjective or nominal.[29] The particular category is then realized according to whether the derived form is possessed or unpossessed. Finally the perfect is expressed using two different morphemes: *–inaq* and *–Vm*. In the former, themes map to the grammatical subject. In the latter, themes remap to grammatical subjects, and optional agents remap to the perfect's possessive pronoun, or else, to the governed object in the by-phrase.

4.4 Discussion The literature on the perfect tense-aspect is vast. We review definitions of the perfect, and then propose an explanation of the perfect in the context of discourse. The standard definition of the perfect is a " form that expresses an action (process, or state) in the past which has continuing relevance for the present" (Nedjalkov & Jaxontov 1988:15). McCoard (1978) introduced the notion of current relevance, which identifies the time of relevance of the perfect. Bybee et al. (1994:54) claim that the perfect is relational and "signals that the situation occurs prior to reference time and is relevant to the situation at reference time."

Viewpoint aspect reflects the "different ways of viewing the internal temporal consistency of a situation" (Comrie 1976:3). This raises the issue that our English glosses of Mayan data are not aspectual, but tensed. This approach is traditional in Mayan studies because of the awkwardness of aspectual glosses in English.

28. We assume also that hosting a subject is not a sufficient condition for 'finiteness.'

29. Based on Kibort's (2005) proposal for the adjective–participle conversion rule.

In his temporal model of tense-aspect, Reichenbach (1947) innovated the concept of reference time (R) including it with speech time (S) and event time (E). In his model, the perfect is construed as anterior $E < R$ and posterior $R < E$, such that $E \neq R$ is true (Kibort 2007). A concern with Reichenbach's approach is that its account of viewpoint aspect is inadequate. A more recent and popular semantic framework is Klein's (1994), whose conception of temporality and tense-aspect is refactored to a set of relations between intervals of time. TT, as an interval of time, is the declaration or proposition of a sentence, whereas TSit represents an interval of time during which the event itself takes place. Consequently Klein interprets viewpoint aspect–as opposed to Vendler's (1957) lexical aspect–as a relation between topic time (TT) and situation time (TSit). Thus Klein's anterior perfect is construed as $\text{TSit} < \text{TT}$, and posterior perfect as $\text{TT} < \text{TSit}$. Notwithstanding this, the semantics and pragmatics of the perfect are extremely complex with multiple avenues of analysis, and will not be pursued here.

Nevertheless we address the narrative use of the K'ichee'an perfect here. The English PP, well-known not to permit past temporal modifiers, also does not permit a sequence of perfects in discursive narrative (Bybee et al. 1994, Nedjalkov 1988). The default use of the perfect is to insert a stative into discourse (Nishiyama & Koenig 2004, Parsons 1990). Whereas event predicates advance narrative, perfects do not, rather perfects function as backgrounding devices. The perfect is a rhetorical device whose role is to improve the connectedness and cohesion of discourse.

We contend that the K'ichee'an perfect mirrors this behaviour in narrative. In our analysis, the perfect in a single sentence occurs either as an isolated predicate, or else, in combination with a verb, but never in a locally adjacent sequence of perfects. Consider the bi-clausal sentence in (58). The first clause is headed by the possessed nominal perfect, and the second headed by a verb marked for completive aspect. The perfect, which introduces the state of 'bag-carrying,' provides a background to the processual event of the sentence, the arrival of the son-in-law at the house.

Functionally the default configuration of the predicative perfect's lexical subject is that it agrees with absolutive AM, has properties of proto-patients, and immediately follows the perfect (58). But the lexical subject can also precede the perfect (47), and, on occasion, there is no lexical subject (46). The default configuration of the predicative perfect's lexical possessor is that it agrees with the possessive AM, precedes the perfect often as external topic, and has properties of proto-agents (58). But the lexical possessor may follow the perfect (38), and, on occasion, there is no lexical possessor (14). In discourse, the perfect's lexical possessor tracks the protagonist or primary participant of the current sentence or adjacent sentence.

4.5 Analysis We propose argument structures for the K'ichee'an perfect using Kibort's (2007, inter alia) revised Lexical Mapping Theory (LMT). The argument structure of the root intransitive *–inaq* perfect (53) is shown in (63), and of the completive passive *–inaq* perfect (35) is shown in (64):

(63)

	theme
	|
b'iin-inaq	arg_1
	−o
	|
	SUBJ

(64)

	agent	theme
	|	|
b'antaj-inaq$_{\text{COMPASS}}$	arg_1	arg_2
	−o +r	−r
	|	|
	(OBL_{AGT})	SUBJ

The argument structure of the DTJ antipassive *–inaq* perfect in (36) with an optional demoted patient in a dative prepositional phrase is shown in (65). The argument structure of the transitive stem actor focus *–inaq* perfect in (41) is shown in (66):[30]

(65)

	theme	patient
	|	|
kunan-naq$_{\text{AP}}$	arg_1	arg_4
	−r	−o
	|	|
	SUBJ	(OBL_{DAT})

(66)

	theme	patient
	|	|
ilow-inaq$_{\text{AF}}$	arg_1	arg_2
	−o	−r
	|	|
	SUBJ	FN_{Θ}

The argument structure of the passive *–Vm* perfect with an optional remapped demoted agent in a by-phrase from (57a) is shown in (67), and with a remapped agent to the possessive pronominal prefix is shown in (68):

(67)

	agent	theme
	|	|
il-om$_{\text{PASS}}$	arg_1	arg_2
	−o +r	−r
	|	|
	(OBL_{AGT})	SUBJ

(68)

	agent	theme
	|	|
wil-om$_{\text{PASS}}$	arg_1	arg_2
	−o	−r
	|	|
	POSS	SUBJ

The single-tier analysis of predicates (Nordlinger & Sadler 2007) is rejected because the possessed predicate perfect would require the complex functor 'carry-*thing/person*'. And because the NVP perfect is intransitive subcategorizing for a single GF in the semantic form, the double-tier approach with the closed complement PREDLINK is also rejected. We adopt the double-tier approach, using a constructional analysis (Dalrymple et al. 2004:192), with annotated phrase-structure rules, virtual copula ε, XCOMP open complement, and (raised) subject (69):

(69) S →

DP	ε	A ∨ N
(↑ SUBJ) = ↓	(↑ PRED) = 'ø-be⟨↑ XCOMP⟩SUBJ'	(↑ XCOMP) = ↓
	(↑ SUBJ) = (↑ XCOMP SUBJ)	((↑ XCOMP POSS) = (↑ ADJ))

30. For FN_{Θ}, see Duncan (2013:Fig. 8)

Simplified c-structure (Fig. 1) & f-structure (Fig. 2) represent the data in (58).[31,32,33]

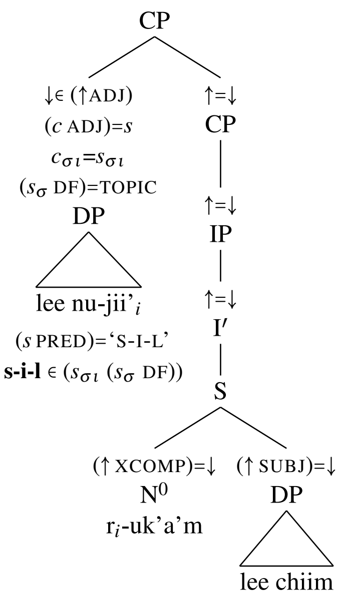


Figure 1 *lee nujii' ruk'a'm lee chiim*

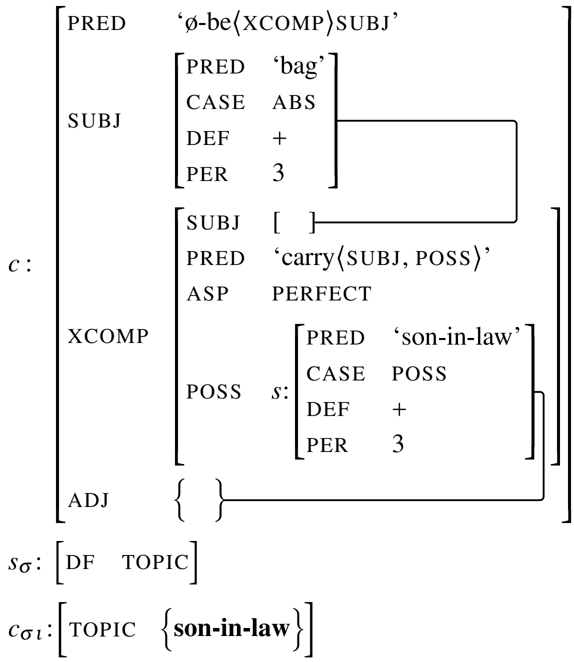


Figure 2 *f-str: lee nujii' ruk'a'm lee chiim*

Partial lexical entries of *–inaq* perfect (18) / *–inaq* suffix are shown in (70):[34,35]

(70) *b'iininaq*,	A	(↑ PRED) = 'walk⟨SUBJ⟩'			
		(↑ SUBJ PRED) = 'Pro'	*–inaq*,	Suff	(↑ ASP) = PERF
		(↑ SUBJ NUM) = 1			@NOTV(STEM)
		(↑ SUBJ CASE) $=_c$ ABS			(↑ FIN) = −
		$\text{STEM}_{\text{IV}} = +$			

Partial lexical entries of the *–Vm* perfect (58) / *–Vm* perfect suffix are shown in (71).[36] The possessor is optional as is the demoted agent in the passive's by-phrase:

(71) *r-uk'a'm*,	N	(↑ PRED) = 'carry⟨SUBJ, POSS⟩'			
		(↑ SUBJ PRED) = 'bag'	*–Vm*,	Suff	(↑ ASP) = PERF
		(↑ SUBJ CASE) $=_c$ ABS			@NOIV(STEM)
		(↑ POSS PRED) = 'son-in-law'			(↑ FIN) = −
		$\text{STEM}_{\text{TV}} = +$			({ (↑ POSS)
					| (↑ OBL_{AGT}) })

31. For analysis of the c-structure of the NVP, see Duncan (2013).
32. Information structure 'semantic' analysis of TOPIC is based on Dalrymple & Nikolaeva (2010).
33. The c-structure in Fig. 1 is indexed according to constituent indexing in (58).
34. @NOTV(STEM): 'no transitive verb stem – except AF.'
35. We exclude the non-predicative LEs of the perfect because they are tangential to our argument.
36. @NOIV(STEM): 'no intransitive verb stem.'

5 Conclusion

Previous accounts of the perfect tense-aspect in the K'ichee'an languages have concluded that the category or part-of-speech of the perfect is a verb, or less often, a participle. We have presented empirical support for the contention that the perfect is expressed using either a deverbal participial adjective or a deverbal possessed nominal in the form of a detransitivized non-verbal predicate. It has been shown that the perfect is always expressed as a one place intransitive but that it, nonetheless, retains the capacity to express two argument roles. It has also been shown that the K'ichee'an perfect includes various semantic types, including the perfect of result and the experiential perfect. Temporal restrictions that constrain the perfect have also been included. We have shown that the perfect is a perfect, not a resultative. Analyses have been implemented using the syntactic architecture of LFG. Future research should undertake semantic and pragmatic analyses of the perfect.

Bibliography

Aissen, Judith. 1987. *Tzotzil Clause Structure*. Dordrecht: Reidel, Kluwer.

Aissen, Judith. 1999. External Possessor and Logical Subject in Tz'utujil. In *External Possession*, I. Barshi & D. L. Payne, (eds.), pp. 451–485. Amsterdam: John Benjamins.

Beekes, Robert S. P. 1995. *Comparative Indo-European linguistics*. Amsterdam: John Benjamins.

Bohnemeyer, Jürgen. 2002. *The Grammar of Time Reference in Yukatek Maya*. Munich: LINCOM.

Bresnan, Joan. 1982. The Passive in Lexical Theory. In *The Mental Representation of Grammatical Relations*, Joan Bresnan (ed.), pp. 3–86. Cambridge, MA: The MIT Press.

Bresnan, Joan. 2001. *Lexical Functional Syntax*. Oxford: Blackwell.

Butt, Miriam, Tracy Holloway King, Niño María-Eugenia, and Frédérique Segond. 1999. *A Grammar-Writer's Cookbook*. Stanford, CA: CSLI Publications.

Bybee, Joan, Revere Perkins, and William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press.

Comrie, Bernard. 1976. *Aspect*. Cambridge University Press.

Coon, Jessica. 2010. Complementation in Chol (Mayan): A Theory of Split Ergativity. Ph.D. MIT.

Dahl, Östen. 2000. *Tense and Aspect in the Languages of Europe.* Walter de Gruyter.

Dalrymple, Mary, Helge Dyvik, and Tracy Holloway King. 2004. Copular Complements: Closed or Open? In *Proceedings of the LFG08 Conference*. Miriam Butt & Tracy Holloway King (eds.). Stanford, Ca.: CSLI Publications.

Dalrymple, M. & I. Nikolaeva. 2011. *Objects and information structure*, (No. 131). Cambridge U.P.

Dayley, Jon P. 1985. *Tz'utujil Grammar*. Berkeley, CA: University of California Press.

Duncan, Lachlan. 2013. Non-verbal predicates in K'ichee' Mayan: an LFG approach. In *Proceedings of the LFG13 Conference*, Miriam Butt & Tracy Holloway King (eds.). Stanford, CA: CSLI.

Hofling, Charles Andrew. 2000. *Itzaj Maya Grammar*. Salt Lake City: University of Utah Press.

Hornstein, Norbert. 1990. *As Time Goes By: Tense and Universal Grammar*. Cambridge, CA: MIT.

Jespersen, Otto. 1951. *The philosophy of grammar*. London: George Allen & Unwin, Ltd.

Katz, George. 2003. On the Stativity of the English Perfect. In *Perfect Explorations*, A. Alexiadou, M. Rathert, & A. von Stechow (eds.), pp. 205–234. Berlin: Mouton de Gruyter.

Kaufman, T. S. 1971. *Tzeltal phonology and morphology*, Vol. 61. Berkeley: University of California.

Kaufman, Terrence S. 1972. *El Proto Tzeltal-Tzotzil: fonología comparada y diccionario reconstruido*, Vol. 5. México: Universidad Nacional Autónoma de México.

Kibort, Anna. 2005. The Ins and Outs of the Participle-Adjective Conversion Rule. In *Proceedings of the LFG05 Conference*, Miriam Butt & Tracy Holloway King (eds.). Stanford, CA: CSLI Publications.

Kibort, Anna. 2007. Extending the Applicability of Lexical Mapping Theory. In *Proceedings of the LFG07 Conference*, Miriam Butt & Tracy Holloway King (eds.). Stanford, CA: CSLI Publications.

Kibort, Anna. 2009. Modeling 'the perfect,' a category between tense and aspect. In *Current Issues in Unity and Diversity of Languages*. Seoul: The Linguistic Society of Korea.

King, Tracy Holloway. 1995. *Configuring Topic and Focus in Russian*. Stanford: CSLI Publications.

Klein, Wolfgang. 1994. *Time in Language*. London: Routledge.

Larsen, Thomas W. 1988. *Manifestations of Ergativity in Quiché Grammar*. Ph. D dissertation, University of California. Berkeley, CA.

MacLeod, Barbara. 2004. A World in a Grain of Sand: Transitive Perfect Verbs in the Classic Mayan Script. In *Linguistics in Maya Writing*, Søren Wichmann (ed.), pp. 291–325. University of Utah Press.

McCawley, James D. 1971. Tense and time reference in English. In *Studies in Linguistic Semantics*, C. Fillmore & T. Langendoen (eds.), pp. 96–113. New York: Holt, Rinehart and Winston.

McCoard, Robert W. 1978. *The English perfect: Tense-choice and pragmatic inferences*. Elsevier.

Michaelis, L. 1996. On the Use and Meaning of "Already." *Linguistics and Philosophy* 19(5):477-502.

Mondloch, James Lorin. 1978. *Basic Quiché Grammar: 38 Lessons*. Publication No. 2. Albany, NY: Institute of Mesoamerican Studies, SUNY at Albany.

Nedjalkov, Vladimir P. 2001. Resultative constructions. In *Language typology and language universals: An international handbook*, Vol. 2, pp. 928–940. Berlin: de Gruyter.

Nedjalkov, Vladimir and Sergej Jaxontov. 1988. The Typology of Resultative Constructions. In *The Typology of Resultative Constructions*, Vladimir P. Nedjalkov (ed.), pp. 3–62. John Benjamins.

Nishiyama, Atsuko and Jean-pierre Koenig. 2004. *What is a Perfect State?* In *WCCFL 23 Proceedings*, B. Schmeiser, V. Chand, A. Kelleher, & A. Rodriguez (eds.), pp. 595–606. Somerville: Cascadilla.

Nordlinger, Rachel and Louisa Sadler. 2007. Verbless Clauses: Revealing the Structure within. In *Architectures, Rules, and Preferences: A Festschrift for Joan Bresnan*, Zaenen et al. (eds.), pp. 139–160. Stanford, CA: CSLI Publications.

Parsons, Terence. 1990. *Events in the Semantics of English*. Cambridge: The MIT Press.

Reichenbach, Hans. 1947. *Elements of symbolic logic*. New York: MacMillan.

Ritz, Marie-Eve. 2012. Perfect Tense and Aspect. In *The Oxford Handbook of Tense and Aspect*, Robert I. Binnick (ed.), pp. 881–907. New York: Oxford University Press.

Shklovsky, Kirill. 2005. Person Marking in Petalcingo Tzeltal. B. Arts. Reed College.

Van Eynde, Frank. 2000. A Constraint-based Semantics of Tenses and Temporal Auxiliaries. In *Grammatical Interfaces in HPSG*, Cann, Grover, & Miller (eds.), pp. 231–49. Stanford: CSLI.

Vendler, Zeno. 1957. Verbs and Times. *The Philosophical Review* 66:143–160.

Vinogradov, Igor. 2014. Aspect Switching in Tzotzil (Mayan) Narratives. In *Oklahoma Working Papers in Indigenous Languages*, Vol. 1:39–54.

Wald, R. F. 2007. The Verb Complex in Classic-Period Maya Hieroglyphic Inscriptions: Its Implications for Language Identification and Change. Ph. D dissertation. University of Texas at Austin, TX.

How much structure is needed: The case of the Persian VP

Pegah Faghiri
Université Sorbonne Nouvelle Paris 3

Pollet Samvelian
Université Sorbonne Nouvelle Paris 3

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 236–254

Keywords: Persian, phrase structure, Differential Object Marking (DOM), empirical/usage-based methods

Faghiri, Pegah & Pollet Samvelian. 2016. How much structure is needed: The case of the Persian VP. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 236–254. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.13.

Abstract

The aim of this paper is to tease apart two available views of the VP in Persian. The prevailing view of the Persian VP initially suggested in generative studies assumes a hierarchical structure with two object positions, mainly motivated by the existence of differential object marking in Persian. Building on quantitative studies, we revisit this hierarchical view and show that it is not born out by the data. A flat structure view of the VP, on the contrary, is in line with the data.

1 Introduction

In this paper we address the issue of the syntactic structure of the VP in Persian, an SOV language with mixed head direction (e.g. head-initial in NP, PP and CP), free word order in the clausal domain[1] and null pronouns.[2]

Previous generative studies on Persian VP have suggested a hierarchical structure that is motivated by the existence of differential object marking (DOM), which requires marking of definite direct objects (DOs) by the enclitic =*rā* (Karimi, 1990; Browning & Karimi, 1994; Ghomeshi, 1997; Karimi, 2005; Ganjavi, 2007). These studies have claimed that *rā*-marked and unmarked DOs display several syntactic and semantic asymmetries, for which this hierarchical view provides a straightforward account. Following insights from studies such as Diesing (1992), unmarked DOs have been assumed to be VP internal while *rā*-marked DOs are VP external. The higher syntactic position of the latter explains word order preferences in di-

[†]We would like to thank the audience at the HeadLex16 Conference (Polish Academy of Sciences, Warsaw, 24-29 July, 2016) for their questions, suggestions and insightful comments. We would also like to thank the two anonymous reviewers for their helpful comments on the earlier version of this paper. This work is supported by a public grant funded by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-LABX-0083).

[1]While SOV is the canonical word order in Persian, all other possibilities can also occur:

a. Puyān Sepide=rā did
   Puyan Sepideh=DOM see.PST.3SG
   'Puyan saw Sepideh.'
b. Sepide=rā Puyān did (OSV)
c. Puyān did Sepide=rā (SVO)
d. Sepide=rā did Puyān (OVS)
e. did Puyān Sepide=rā (VSO)
f. did Sepide=rā Puyān (VOS)

[2]The following pair of examples illustrates the possibility to have covert arguments in subject and object positions in Persian, contrary to languages like English.

a. Puyān Sepide=rā did?
   Puyan Sepide=DOM see.PST.3SG
   'Did Puyan see Sepideh?'

b. na na-did
   No NEG-see.PST.3SG
   'No, he did't see her.'

transitive constructions. It also accounts for other asymmetries concerning binding and scope.

Recently, experimental and corpus-based studies have established that the most essential argument on which this view is built, namely ordering preferences, does not hold (Faghiri & Samvelian, 2014; Faghiri et al., 2014). Building on these studies, we will show that some other commonly accepted asymmetries also turn out to be dubious. These findings lead us to question the almost uncontroversially admitted hierarchical view of the Persian VP and to suggest a flat structure in line with Samvelian (2001) and Bonami & Samvelian (2015). We claim that differences between different types of DOs can be accounted for by semantics, information structure and universal functional principles, such as a "salient-first" preference, without resorting to a hierarchical syntactic structure.

2 The Persian VP: Prevailing Analyses

In (formal) Persian, there is no overt marker for definiteness, ex. *xarguš* 'the rabbit'. By contrast, indefiniteness is overtly marked by the enclitic =*i*, the cardinal *ye(k)* or both, ex. *yek xarguš=i* 'a rabbit'. In the DO position, however, NPs with a definite reading are differentially marked by the enclitic *rā*, pronounced *(r)o* in the colloquial register, as in ex. (1-a). Moreover, indefinite NPs can be marked by =*rā* to receive an indefinite specific reading as in ex. (1-b).[3]

(1) a. Sara xarguš*(=rā) did
Sara rabbit=DOM see.PST.3SG
'Sara saw the rabbit.'
b. Sara xarguš=i(=rā) did
Sara rabbit=INDEF(=DOM) see.PST.3SG
'Sara saw a (certain/particular) rabbit.'

It should also be noted that in Persian, bare nouns, that is, nouns without any determination or quantification like *xarguš* in (2), are not specified for number and therefore can yield a mass reading. Bare objets have either an existential, ex. (2-a), or a kind-level/generic reading, ex. (2-b). Indefinite objects on the other hand are always specified for number and have an existential reading, as in (1-b) above.

(2) a. Sara xarguš did
Sara rabbit see.PST.3SG
'Sara saw a rabbit/rabbits.'
b. Sara xarguš dust dār-ad
Sara rabbit friend have.PRS-3SG
'Sara likes rabbits.'

[3]This explains why many authors have accounted for =*rā* in terms of a binary specificity feature (Karimi, 1990; Browning & Karimi, 1994; Karimi, 2003; Rasekhmahand, 2004), but see Lazard (1982), Dabir-Moghaddam (1992), Meunier & Samvelian (1997) and Ghomeshi (1997).

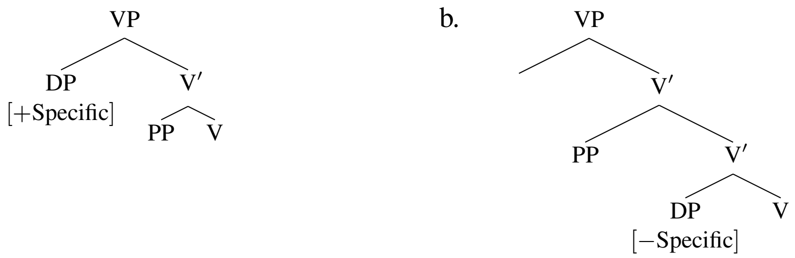
Note that DOM is a complex phenomenon and that definiteness and/or specificity are not the only features triggering *rā*-marking which can also be triggered by topicality, or more generally, discursive salience. Furthermore, other semantic features such as humanness are also shown to favor the presence of =*rā*.

2.1 The "Two Object Position Hypothesis" (TOPH)

Several studies claim that *rā*-marked DOs (definite or indefinite) and unmarked DOs (bare or indefinite) occur in two distinct syntactic positions (at spell out), whether base-generated or as a result of a movement (Karimi, 1990; Browning & Karimi, 1994; Ghomeshi, 1997; Karimi, 2005; Ganjavi, 2007; Modarresi, 2014). Some studies assume different base-generated positions for *rā*-marked and unmarked DOs as in (3-a) and (3-b) respectively, while others assume that both DOs are generated in the same position, as in (4-a), but *rā*-marked DOs must move to the Specifier position to receive interpretation, as illustrated in (4-b).

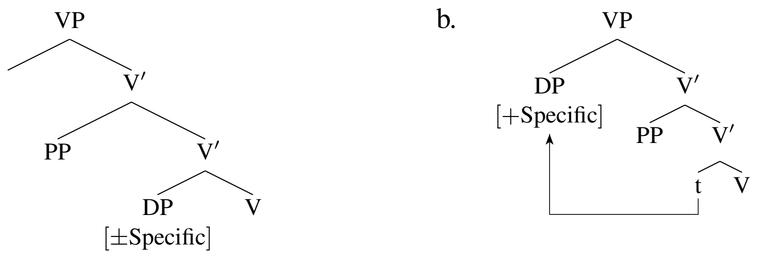
It should be noted that these studies formulate their claim in terms of features, often a binary one, such as specificity (cf. footnote on page 2), assumed to trigger *rā*-marking, and rarely in terms of DOM itself. However, given the ongoing debate on the analysis of =*rā* and the fact that there is no satisfactory account of *rā*-marking in terms of a binary feature, including specificity, we stick to the formal definition of these two DO types and refrain from referring to any semantic properties.

(3) a. b.



Adapted from Karimi (2003, p. 105)

(4) a. b.



Adapted from Karimi (2005, pp. 108–109)

Following Karimi (2003), who provides the most exhaustive argumentation in favor of this analysis, we refer to this claim as the "Two Object Position Hypothesis" (TOPH).

According to Karimi, the fact that *rā*-marked DOs occupy a higher syntactic position than unmarked DOs provides a straightforward account for several syntactic and semantic asymmetries between these two types of DOs. Indeed, the phrase structures in (3) (or in (4)), reflect two different neutral word orders for each type of DO and suggest that only *rā*-marked DOs c-command the indirect object (IO), since unmarked DOs are in a lower position. In the following subsections, we will present these so-called asymmetries. Note however, that we do not agree with a part of Karimi's grammaticality judgments. The following section provides a discussion of the data on which the TOPH is based.

2.2 The Relative Order with Respect to the IO

According to a widespread hypothesis put forward in theoretical studies as well as (some) grammars, the neutral (unmarked) word order between the direct and indirect objects in ditransitive constructions depends on markedness (Browning & Karimi, 1994; Mahootian, 1997; Rasekhmahand, 2004; Ganjavi, 2007; Windfuhr & Perry, 2009; Roberts et al., 2009, among others). *Rā*-marked DOs are assumed to precede while unmarked DOs follow the IO:

(5) a. (S) DO=rā IO V
b. (S) IO DO V

The following examples are provided by Karimi (2003) in support of this claim. The author furthermore claims that unmarked DOs can only be separated from the verb if they are contrastively focused.

(6) a. Kimea aqlab barā mā še'r mi-xun-e
Kimea often for us poem IPFV-read-3SG
'Kimea often reads poetry for us.'
b. Kimea aqlab barā mā ye še'r az Hafez mi-xun-e
Kimea often for us a poem from Hafez IPFV-read-3SG
'Kimea often reads a poem by Hafez for us.'
c. Kimea aqlab hame=ye še'r-ā=ye tāza=š=**ro** barā mā
Kimea often all=EZ poem-PL=EZ new=3SG=DOM for us
mi-xun-e
IPFV-read-3SG
'Kimea often reads all her new poems for us.'
d. Kimea aqlab ye še'r az Hafez=**ro** barā mā mi-xun-e
Kimea often a poem from Hafez=DOM for us IPFV-read-3SG
'Kimea ofte reads a (particular) poem by Hafez for us.'
Adapted from Karimi (2003, p. 91)

2.3 Semantic Fusion with the Verb

Karimi considers *rā*-marked DOs as (independent) participants of the event described by the verb and hence semantically autonomous. Unmarked DOs, by contrast, are assumed to be a part of the predicate and semantically non-autonomous and hence correspond to a [N+Ving] interpretation. This entails that ex. (7), in which the DO is unmarked, can be an appropriate answer to the question 'What does Kimea do every night ?' (2003, p. 100).

(7) Kimea har šab (ye) sib mi-xor-e
Kimea every night (a) apple IPFV-eat.PRS-3SG
'Kimea eats apples / does (an) apple eating every night.'

According to Karimi this difference also accounts for the fact that sentences containing unmarked DOs can only receive an activity/process reading, ex. (8-a), while those containing a marked DO have an eventive reading, ex. (8-b). Note that Karimi borrows this "durative adverbial test" from Ghomeshi & Massam (1994, pp. 190–191), who provide the grammaticality judgments in ex. (8-a) to support the analysis of bare objects (and not for all unmarked DOs) as a case of noun incorporation. Note that Ghomeshi & Massam's analysis in terms of noun incorporation only applies to bare objects. Indefinite unmarked objects are not concerned by this analysis.

(8) a. (man) *dar do daqiqe / barāye yek sāat sib xord-am
I in two minute / for one hour apple eat.PST-1SG
'I ate apples for one hour.'
b. (man) dar do daqiqe / *barāye yek sāat sib=rā xord-am
I in two minute / for one hour apple eat.PST-1SG
'I ate the apple in two minutes.'

Furthermore, Karimi (2003) claims that the semantic fusion of unmarked DOs with the verb explains why the latter, contrary to marked DOs, cannot take wide scope (and hence cannot trigger scope ambiguity), enter binding relations, and license parasitic gaps. Below, we will present Karimi's data in support of these claims.

2.3.1 Scope Ambiguity

As illustrated by the pair of examples in (9) and (10), according to Karimi (2003) only *rā*-marked (indefinite) DOs trigger scope ambiguity when scrambled to the left of a DP quantified by a universal quantifier. Note that according to Karimi scope ambiguity can only result from scrambling in Persian (also see Karimi, 2005).

(9) a. **har** dānešju=i **ye** še'r=ro bāyād be-xun-e
every student=INDF a poem=DOM must SUBJ-read.PRS-3SG
'Every student has to read one poem (out of a specific set).' ($\forall > \exists$)

b. **ye** še'r=ro$_i$ **har** dānešju=i t$_i$ bāyād be-xun-e ($\forall > \exists$; $\exists > \forall$)

(10) a. **har** dānešju=i **ye** še'r bāyād be-xun-e
every student=INDF a poem must SUBJ-read.PRS-3SG
'Every student must read a poem.' ($\forall > \exists$)

b. **ye** še'r$_i$ **har** dānešju=i t$_i$ bāyād be-xun-e ($\forall > \exists$)

Adapted from Karimi (2003, p. 103)

Karimi argues that unmarked indefinite DOs as part of the predicate can never take wide scope over the IO. That is, only (9-b) allows for both the wide or the narrow scope of existential quantifier over the universal quantifier (2003, p. 103).

2.3.2 Binding Relations

With respect to binding relations, Karimi provides the pair of examples in (11)[4] in support of the claim that only *rā*-marked DOs are able to bind an anaphor in the IO position.

(11) a. man [se=tā bačče-hā=ro]$_i$ be hamdige$_i$ mo'arrefi kardam
I three=CLF child-PL=DOM to each other introduction do.PST-1SG
'I introduced three children to each other.'

b. *man [se=tā bačče]$_i$ be hamdige$_i$ mo'arrefi kardam

Adapted from Karimi (2003, p. 102)

Karimi (2003) argues that unmarked DOs as part of the predicate cannot enter binding relations.

2.3.3 Licensing Parasitic Gaps

Likewise, with respect to parasitic gaps, Karimi claims that unmarked DOs, contrary to *rā*-marked DOs cannot license parasitic gaps as part of the predicate. This claim is illustrated by the grammaticality judgments provided by the author in (12).

[4] Note that the plural marker, *i.e.* *-hā*, is not compatible with the indefinite NP *se(=tā) bačče* in (b), since this marker implies a definite reading. Accordingly, in the DO position, the plural marker goes in pair with =*rā*-marking.

(i) se=tā bačče(-hā) xandid-and
three=CLF child(-PL) laugh.PST-3PL
'(The) three children laughed.'

(12) a. Kimea in ketāb=o$_i$ [qablaz in-ke –$_i$ be-xun-e] t$_i$ be
Kimea this book=DOM before that SUBJ-read.PRS-3SG to
man dād
me give.PST.3SG
'Kimea gave me this book before reading (it).'

b. *Kimea ketāb$_i$ [qablaz in-ke –$_i$ be-xun-e] be man t$_i$ dād

Adapted from Karimi (2003, p. 116)

2.4 Coordinate Structures

The last argument put forward in support of the TOPH is the claim that *rā*-marked and unmarked DOs cannot appear together in a coordination, as illustrated by Karimi's grammaticality judgments in (13).

(13) a. man diruz [in aks=ro] va [in ketāb=ro] xarid-am
I yesterday this picture=DOM and that book=DOM buy.PST-1SG
'Yesterday, I bought this picture and that book.'

b. man diruz [aks] va [ketāb] xarid-am
I yesterday picture and book buy.PST-1SG
'Yesterday, I bought pictures and books.'

c. *man diruz [in aks=ro] va [ketāb] xarid-am
I yesterday this picture=DOM and book buy.PST-1SG

Karimi (2003, p. 103)

According to Karimi (2003) this is a clear argument highlighting the fact that *rā*-marked and unmarked DOs do not occur in the same syntactic position.

3 Getting the Facts Right

The main problem with the TOPH is the fact that the data on which it is built are empirically dubious. The asymmetries either do not hold or are best represented as a cline and certainly not in terms of a dichotomy. More specifically, by adopting a dichotomous view of DOs in terms of *rā*-markedness, this hypothesis aligns indefinite DOs with bare DOs, which are commonly taken to represent all unmarked DOs. This is somewhat expected, albeit misleading, given that bare DOs display the lowest level of specificity and definiteness among unmarked DOs. Yet, bare DOs differ from non-bare unmarked DOs in many non-trivial respects (Ghomeshi & Massam, 1994; Samvelian, 2001; Ghomeshi, 2003; Ganjavi, 2007; Modarresi, 2014).[5] However, even when these differences are acknowledged and mentioned by scholars, they are put aside and ignored when dealing with the syntactic configuration of the VP (e.g. Ghomeshi, 1997; Ganjavi, 2007).

[5]In the generative frameworks, some scholars have suggested that bare nouns are inserted as N^{0}s while indefinites have maximal projections, and are built as NumPs (e.g. Ganjavi, 2007).

In what follows, we will assess the data put forward in the literature, especially in Karimi's studies, in favor on the TOPH.

3.1 Word Order Preferences in Ditransitive Sentences (Revisited)

Recent empirical studies on word order variations in Persian (Faghiri & Samvelian, 2014; Faghiri et al., 2014, 2016) invalidate the generalization in (5) above, clearly showing that unmarked DOs do not display a homogeneous behavior with respect to word order and should be divided into different categories.

Adopting a quantitative approach to word order variations (e.g. Wasow, 1997; Stallings et al., 1998; Yamashita & Chang, 2001; Wasow, 2002; Bresnan et al., 2007), these studies investigate the relative order between the two objects in ditransitive constructions in Persian and examine the effect of functional factors such as the relative length and animacy on ordering preferences. More precisely, Faghiri & Samvelian (2014) explore the predictions of the DOM criterion on the relative order between the DO and the IO (cf. (5)) in a sample of 905 occurrences extracted from the Bijankhan corpus.[6] Their study is based on a fine-grained typology of DOs with respect to their degree of determination or definiteness: *rā*-marked, indefinite/quantifier (unmarked), bare-modified and bare (single-word) DOs (2014, pp. 222–224). Their data contradict the predictions of the DOM criterion for indefinite (unmarked) DOs, since these DOs occur in the DO-IO-V order at the rate of 77%, grouping with *rā*-marked DOs instead of bare (unmarked) DOs (with 90% and 19.3% of DO-IO-V order respectively). Moreover, interestingly, bare modified DOs, that is, bare DOs carrying modifiers, are shown to have a significantly less stronger preference to appear adjacent to the verb than bare single-word DOs (33.3% vs. 15.8%).

Faghiri et al. (2014) ran a follow up sentence completion (web-based) questionnaire to study the ordering preference of indefinite (unmarked) DOs in a controlled experiment[7] and arrived at similar results. In their data, the mean rate of sentences like ex. (14) in which the participants placed the DO before the IO (68%) was significantly greater than the opposite order predicted by the DOM criterion.

(14) ... [$_{\text{DO}}$ yek livān (šarbat=e sekanjebin=e tagari)] [$_{\text{IO}}$ be moštari-hā (ke
a glass syrup=EZ (mint=EZ icy) to customer-PL (that
az garmā kalāfe bud-and)] dād
from heat frustrated were) gave
'... (he) gave a glass of (icy mint) syrup to the customers (that were frustrated from the heat).'

Adapted from Faghiri et al. (2014, p. 230)

Accordingly, these studies suggest that the DOM criterion should be revisited

[6]A freely available corpus of more than 2.6 million tokens, extracted from the *Hamshahri* daily newspaper contains, manually annotated for part-of-speech information (Faghiri & Samvelian, 2014, p. 222).

[7]For details see Faghiri & Samvelian (2014, pp. 229–232).

in such a way to account for the ordering preferences between the two objects on the basis of the degree of determination of the DO instead of its markedness. In other words, ordering preferences between the DO and the IO rather than being dichotomous, as predicted by the TOPH, follow a continuum based on the degree of determination of the DO : the more determined the DO, the more it is likely to precede the IO. Building on the fact that the degree of determination of the DO is strongly related to its discourse accessibility (cf. e.g. the *(Referential) Givenness Hierarchy*, Gundel et al., 1993), they argue that these ordering preferences reflect a "(discourse-)prominent-first" preference with bareness strongly favoring the IO-DO-V order and *rā*-markedness the inverse.

Furthermore, both Faghiri & Samvelian (2014) and Faghiri et al. (2014) find a significant effect of the relative length corresponding to the "long-before-short" preference. Following Yamashita & Chang (2001), these studies account for this preference in terms of the conceptual accessibility hypothesis. In other words, the "salient-first" preference, which assumes that longer constituents – being lexically richer – are conceptually more accessible than shorter ones (also see Faghiri et al., 2016).

In sum, these quantitative studies show that while *rā*-marked DOs do have a strong preference for the DO-IO-V order, only bare single-word unmarked DOs have a comparable preference to appear adjacent to the verb. Crucially, they show that indefinite (unmarked) DOs group with marked DOs in preferring overall the DO-IO-V order, but show a less strong preference for this order.[8] Consequently, the empirical findings of Faghiri & Samvelian (2014) and Faghiri et al. (2014) drastically undermine the TOPH, whose backbone argument is the ordering asymmetries between specific (*rā*-marked) DOs on the one hand and unmarked (non-specific) DOs on the other hand.

3.2 Semantic (In)dependence from the Verb (Revisited)

Karimi's (2003) claim on the semantic fusion of unmarked DOs with the verb faces the same problem as word order preferences, since it similarly builds on the assumption that all unmarked DOs behave in the same way, which is not the case. While bare objects are highly cohesive with the verb, leading some studies to consider them as semantically incorporated, non-bare non-*rā*-marked DOs are inarguably referential NPs and hence are construed as (independent) entities undergoing the event described by the verb rather than being a part of it.

This explains why the "durative adverbial test" argument mentioned by Ghomeshi & Massam (1994) applies only to bare objects. The authors claim that bare DOs are non-referential and as such cannot delimit the event described by the verb and hence are only compatible with adverbials denoting a process. Indefinite unmarked DOs on the other hand are compatible with adverbials denoting an event, ex. (15) .

[8] It should be noted that these differences between the four DO types in their ordering preferences with respect to the IO are replicated in other experiments conducted in the same paradigm as in Faghiri et al. (2014), see Faghiri (2016, ch. 5).

(15) Maryam dar do daqiqe čand sib=e bozorg xord
Maryam in two minute a-few apple=EZ big eat.PST.3SG
'Maryam ate a couple of big apples in two minutes.'

3.3 Scope Ambiguity (Revisited)

Karimi's (2003) judgments and/or interpretations with respect to scope are not straightforward and uncontroversial. In general, scope ambiguity is a complex matter sensitive to functional factors such as discourse and lexical factors. Furthermore, it is shown that scope interpretations display a certain amount of variation among speakers:

> Quantifier scope is a delicate phenomenon. It is not simply a matter of ambiguity vs. nonambiguity, but a continuum, and the judgments on a given sentence often fluctuate from speaker to speaker.

Kuno et al. (1999, p. 110)

We believe that any attempt for accounting for scope properties of DOs in Persian is in vain without solid data based on experimental investigations. Meanwhile, crucial to the issue at stake here, not all studies exclude scope ambiguity for unmarked DOs. For instance, Modarresi (2014) assumes both wide and narrow scope for the indefinite DO *film=i* in ex. (16), over the universal quantifier in the subject position. Likewise, Ghomeshi (1997) affirms that the indefinite DO *ye ketāb* in (17) can take both wide and narrow scope over the universal quantifier. Note that the universal quantifier *har* 'each', used in Karimi's (2003) examples above, favors a distributive reading entailing a wide scope, while the universal quantifier *hame* 'every' favors a collective reading.

(16) hame film=i did-and
everybody movie=INDF watch.PST-3SG
'Everybody watched a movie.' ($\forall > \exists$; $\exists > \forall$)
Adapted from Modarresi (2014, p. 30)

(17) hame=ye mo'allem-ā ye ketāb=i entexāb kard-and
all=EZ teachers-PL a book=INDF selection make.PST-3SG
'All the teachers selected a book.' ($\forall > \exists$; $\exists > \forall$)
Adapted from Ghomeshi (1997, p. 140)

3.4 Binding Relations (Revisited)

With respect to binding – building on an argument mentioned by Karimi herself (1999, p. 707) –, we have previously suggested, accounting for the ungrammaticality of (11-b) (repeated below in (18-b)) in terms of the semantic mismatch between the pronoun and its antecedent (Faghiri & Samvelian, 2015). Namely, the non-specific NP is not a felicitous antecedent for *hamdige*. Meanwhile, examples like

(19), which are by no means rare as illustrated by the attested example in (20),[9] show that any generalization based on a binary specific vs. nonspecific feature is too strong. Indeed, these counterexamples show that, contrary to Karimi's claim, an unmarked (that is, nonspecific) DO can bind an anaphor in the IO position.

(18) a. man [se=tā bačče-hā=ro]$_i$ be hamdige$_i$ mo'arrefi
I three=CLF child-PL=DOM to each other introduction
kardam
do.PST-1SG
'I introduced three children to each other.'
b. *man [se=tā bačče]$_i$ be hamdige$_i$ mo'arrefi kardam

(19) [čand varaq kāqaz]$_i$ be hamdige$_i$ mangane mi-kon-e
a-few sheet paper to each other staple IPFV-do.PRS-1SG
'(S)he staples a few sheets of paper together (lit. to each other).'

(20) ... mi-bin-am [čand=tā sandoq]$_i$ kenār=e hamdige$_i$
IPFV-see.PRS-1SG a-few=CLF box next-to=EZ each other
gozāšt-an ...
put.PST-3PL
'... I saw that they have put a few box next to each other...'

A comparison between (19) and (20) on the one hand and (18) on the other hand indicates, in accordance with the line of argumentation pursued here, that plausibly other factors are involved, e.g., humanness of the antecedent, the strength of the distributive reading implied by the predicate (e.g. Dalrymple et al., 1998). Indeed, in the unacceptable example provided by Karimi, ex. (18), contrary to the counterexamples provided here, the referent of the DO is human. Furthermore, in (18), the strong reciprocal predicate ('introduce each one to the other') implies a distributive reading (e.g. Dalrymple et al., 1998), while (19) and (20) are compatible with a collective reading.[10] Interestingly, these factors coincide with those favoring *rā*-marking. As pointed out by several studies, contra Karimi, *rā*-marking is not triggered by a single binary feature. It allows for a certain degree of variability and is sensitive to functional and discourse-related factors (e.g. Samvelian, 2017). The distributive reading implied by the strong reciprocal predicate, the humanness of the referent and the telicity of the event are different converging factors favoring *rā*-marking in (18).

Interestingly, however, the following attested example[11] shows that in the proper context, that is, a context favoring a collective reading, even an unmarked DO with

[9] http://www.noandishaan.com/forums/thread66514.html [consulted on 07/06/2016]

[10] More precisely, (19) favors a collective reading (compare 'staple a few sheets of paper together' with 'staple each sheet to the other'), and (20), while not particularly favoring a collective reading, does not impose a distributive reading in the sense that the sentence does not necessarily require each box to be next to another box.

[11] http://romanparisa.blogfa.com/post/10 [consulted on 07/06/2016]

a human referent can bind anaphora in the IO position.[12]

(21) Lidya yeki=ro mi-šnās-e ke [doxtar pesar]$_i$ [be
Lidya someone=DOM IPFV-know.PRS-3SG that girl boy to
ham$_i$] mo'arrefi mi-kon-e
each other introduction IPFV-do.PRS-3SG
'Lidya knows someone who introduces boys and girls to each other.'

In any case, further investigation, namely usage-based and experimental studies, is necessary to explore various parameters involved here. Meanwhile, the counterexamples provided here show that the asymmetry claimed by Karimi (2003) on the basis of *rā*-markedness (or specificity for that matter) is flawed.

3.5 Licensing Parasitic Gaps (Revisited)

Contrary to Karimi's claim, unmarked DOs can license parasitic gaps in proper contexts, ex. (22). The problem with Karimi's example in (12-b) (repeated below in (23-b)) results from the incompatibility between the referential properties of *ketāb* and the aspectual properties of the verb in the matrix clause. The telicity of the latter is incompatible with the cumulative reading implied by the bare DO (e.g. Krifka, 1992).

(22) man tā be hāl na-šode šalvār$_i$ [bedun=e inke –$_i$ porov
I until to now NEG-become.PP pants without=EZ that try
kon-am] –$_i$ be-xar-am
do.PRS-1SG SBJV-buy.PRS-1SG
'It has never happened that I buy pants without trying.'

(23) a. Kimea in ketāb=o$_i$ [qablaz in-ke –$_i$ be-xun-e] t$_i$ be
Kimea this book=DOM before that SUBJ-read.PRS-3SG to
man dād
me give.PST.3SG
'Kimea gave me this book before reading (it).'
b. *Kimea ketāb$_i$ [qablaz in-ke –$_i$ be-xun-e] be man t$_i$ dād

Our claim is supported by an acceptability rating experiment, on a Likert scale from 1 (absolutely unacceptable) to 7 (completely acceptable). This experiment was conducted via a web-based questionnaire conducted on the *Ibex Farm* platform (Drummond, 2013),[13] with 70 Persian native participants. To give an example, the sentence in (22) has received a mean rate of 5.7 ($CI_{95\%}$: $\pm$ 0.45).

[12]Note that *ham* is the simplified form of the reciprocal. Indeed, such structures are not fully acceptable with the complex form. But this intolerance has received a functional explanation (cf. Faghiri, 2016, ch. 7).

[13]This platform proposes free hosting for linguistic experiments that can be carried out via on-line questionnaires (http://spellout.net/ibexfarm).

Further experiments are needed in order to pin down parameters involved in favoring parasitic gaps with unmarked DOs. In this respect, Goldberg's discussion on island constraints and the role of the information structure also seems highly relevant (2006, ch. 7). Indeed, the left-extraction of an element needs to be justified on the discursive ground. That is, some discourse saliency is necessary to license the left-extraction of an element. This seems to be the reason why these constructions are more common with *rā*-marked DO, given that *rā*-markedness implies discourse salience. It should be noted that in Persian the enclitic =*rā* can also be used as a marker of topicality for other non-subject constituents extracted towards the initial position (e.g. Lazard, 1982; Dabir-Moghaddam, 1992).

3.6 Coordination between Marked and Unmarked DOs (Revisited)

Here again, contra Karimi, we claim that the coordination between unmarked and *rā*-marked DOs is perfectly grammatical. Our claim is supported by an acceptability rating experiment, similar to the one presented above, completed by 70 native speakers of Persian via a web-based questionnaire on *Ibex Farm*, with 20 target items and 40 fillers.[14]

In this experiment, examples involving a coordination, such as in ex. (25), received an overall mean rate of 5.4 ($CI_{95\%}$: ± 0.14), on a scale of 1 to 7, while the control sentence with no coordinate structure, ex. (24), received a mean rate of 6.0 ($CI_{95\%}$: ± 0.18). While the difference between the two rates is significant, the mean rate of sentences with a coordinate construction remains high enough to dismiss any doubts on their acceptability.[15]

(24) barāye sabtenām kāfi ast [form=e takmil-šode=rā] barāye mā
for registration enough is form=EZ completed=DOM for us
ersāl kon-id
send do.PRS-2PL
'To register you only need to send us the completed form.'

(25) a. ... [yek qat'e aks va form=e takmil-šode=rā] ....
... a piece photo and form=EZ completed=DOM
'To register you only need to send us a photo and the completed form.'

b. ... [form=e takmil-šode=rā va yek qat'e aks] ....
... form=EZ completed=DOM and a piece photo
'To register you only need to send us the completed form and a photo.'

Interestingly, the order between the coordinates is shown to be a relevant fac-

[14]Note that for convenience, we are only presenting a simplified version of this experiment here (see Faghiri, 2016, ch. 7, for details)

[15]Note that the mean rate of uncontroversially unacceptable sentences, included in the same questionnaire as fillers, is 2.4 ($CI_{95\%}$: ± 0.11).

tor. Examples such as (25-a)[16] were rated higher than their counterparts in the reverse order, as in (25-b) : 5.8 ($CI_{95\%}$: ±0.19) vs. 5.0 ($CI_{95\%}$: ±0.20). Further experiments are needed to explore this difference.

4 Less Structure, More Functional/Cognitive Principles

The data presented in the previous section shows that there is no conclusive empirical evidence in favor of the TOPH. If a hierarchical analysis is to be maintained, it should either posit more than two positions, or group unmarked non-bare DOs with *rā*-marked one. None of these solutions is satisfactory, given, among other things, that different types of DOs can be coordinated and that different groupings occur according to the phenomenon considered. Based on this body of evidence, we dismiss this consensual hierarchical analysis and adopt instead a flat structure for the Persian VP, in line with Samvelian (2001) and Bonami & Samvelian (2015).

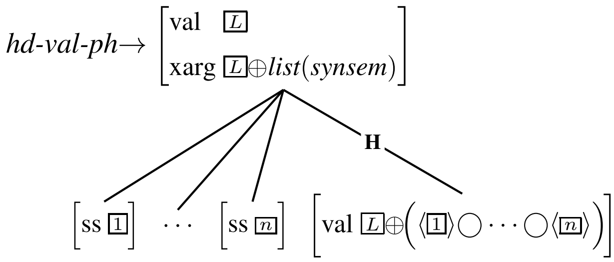
We claim that differences in the behavior of different types of DOs, which constitute a cline rather than a categorical distinction, can be accounted for in terms of universal functional principles. On an unrelated issue, Bonami & Samvelian (2015) claim that word order facts and constituency tests provide no motivation for a VP/S asymmetry in Persian, since subjects and phrasal complements may be freely reordered. We are in line with their view, represented by the head-valence phrase schema given in Figure (26). The schema realizes multiple dependents of the head in the same local tree without constraining their relative order.

Under this view, word order preferences for different DO types can be explained via a set of cross-linguistically valid interacting factors, such as discourse accessibility, definiteness, relative length (or grammatical weight) and animacy, and stated in terms of the principle of "prominent-first", pointed out for other SOV languages, such as Japanese (Yamashita & Chang, 2001).

[16]One can consider these examples misleading for the issue at stake, since the enclitic =*rā* is placed on the right periphery and can theoretically scope over both DOs. To avoid this confusion, we have prepared our stimuli in a way that in each item the indefinite DO would not be (semantically) felicitous if *rā*-marked. In other words, the *rā*-marked variant of the indefinite DO in these sentences yields an awkward utterance:

(i) ??barāye sabtenām kāfi ast [yek qat'e akse=rā] barāye mā ersāl kon-id
for registration enough is a piece photo=DOM for us send do.PRS-2PL
'To register you only need to send us a particular photo.'

Nonetheless, the mean rate of (b) sentences, 5.0 ($CI_{95\%}$: ±0.20), remains high enough, since it is significantly higher than 4 (t=9.71, df=279). Recall that the mean rate of uncontroversially unacceptable sentences in the same questionnaire is 2.4 ($CI_{95\%}$: ±0.11).

(26) 

5 Conclusion

In this paper we showed that the behavior of the DOs in Persian cannot be accounted for in terms of a hierarchical phrase structure, since the differences between different types of DOs are a matter of cline rather than a dichotomous opposition. Trying to account for these empirical facts by adding more structure, as theoretically appealing as it may seem, not only does not provide an appropriate modeling of data but also makes wrong predictions. In contrast, a simplified structure accompanied by few functional principles constitues a more satisfying option to explore.

References

Bonami, Olivier & Pollet Samvelian. 2015. The diversity of inflectional periphrasis in Persian. *Journal of Linguistics* 51(2). 327–382.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald Baayen. 2007. Predicting the dative alternation. In G. Boume, I. Kraemer & J. Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Royal Netherlands Academy of Science.

Browning, Marguerite & Ezat Karimi. 1994. Scrambling to object position in Persian. In N. Corver & H. van Riemsdijk (eds.), *Studies on scrambling: Movement and non-movement approaches to free word-order phenomena*, 61–100. Mouton de Gruyter.

Dabir-Moghaddam, Mohammad. 1992. On the (in) dependence of syntax and pragmatics: Evidence from the postposition-ra in Persian. *Cooperating with written texts*, 549–573.

Dalrymple, Mary, Makoto Kanazawa, Yookyung Kim, Sam Mchombo & Stanley Peters. 1998. Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy* 21(2). 159–210.

Diesing, Molly. 1992. *Indefinites*. Linguistic Inquiry Monographs. Cambridge, Mass: MIT Press.

Drummond, Alan. 2013. Ibex farm. Version 0.3. www.spellout.net/ibexfarm/.

Faghiri, Pegah. 2016. *La variation de l'ordre des constituants dans le domaine préverbal en persan : approche empirique*, PhD dissertation, Université Sorbonne Nouvelle .

Faghiri, Pegah, Barbara Hemforth & Pollet Samvelian. 2016. Length effects in an OV language with differential object marking and mixed head direction (poster). In *The 29th annual conference on human sentence processing (CUNY 2016)*, University of Florida, Gainesville, Florida, USA.

Faghiri, Pegah & Pollet Samvelian. 2014. Constituent ordering in Persian and the weight factor. In Christopher Pinon (ed.), *Empirical issues in syntax and semantics 10 (EISS 10)*, 215–232. CNRS.

Faghiri, Pegah & Pollet Samvelian. 2015. How much is determined by the syntax? an empirical approach to the position of the direct object in Persian. In *Proceedings of the 9th Iranian conference on linguistics, Allameh Tabataba'i University, Tehran, feb. 24-25 2015*, 1419–1435. Allameh Tabataba'i University.

Faghiri, Pegah, Pollet Samvelian & Barbara Hemforth. 2014. Accessibility and word order: The case of ditransitive constructions in Persian. In Stefan Müller (ed.), *Proceedings of HPSG 2014*, 217–237. Stanford: CSLI Publications.

Ganjavi, Shadi. 2007. *Direct objects in Persian*. PhD dissertation, University of Southern California.

Ghomeshi, Jila. 1997. Topics in Persian VPs. *Lingua* 102(2). 133 – 167.

Ghomeshi, Jila. 2003. Plural marking, indefiniteness, and the noun phrase. *Studia linguistica* 57(2). 47–74.

Ghomeshi, Jila & Diane Massam. 1994. Lexical/syntactic relations without projection. *Linguistic Analysis* 24(3-4). 175–217.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford University Press.

Gundel, Jeanette K, Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

Karimi, Simin. 1990. Obliqueness, specificity, and discourse functions: Ra in Persian. *Linguistic Analysis* 20. 139–191.

Karimi, Simin. 1999. A note on parasitic gaps and specificity. *Linguistic Inquiry* 30(4). 704–713.

Karimi, Simin. 2003. On object positions, specificity, and scrambling in Persian. In Simin Karimi (ed.), *Word order and scrambling*, 91–124. Blackwell Publishing Ltd.

Karimi, Simin. 2005. *A minimalist approach to scrambling: Evidence from Persian* Studies in Generative Grammar. Mouton De Gruyter.

Krifka, Manfred. 1992. Thematic relations as links between nominal reference and temporal constitution. In Ivan A. Sag & Anna Szabolcsi (eds.), *Lexical matters*, 29–53. Stanford: CSLI Publications.

Kuno, Susumu, Ken-ichi Takami & Yuru Wu. 1999. Quantifier scope in English, Chinese, and Japanese. *Language* 75(1). 63–111.

Lazard, Gilbert. 1982. Le morphème rā en persan et les relations actancielles. *Bulletin de la Société de Linguistique de Paris Paris* 77(1). 177–208.

Mahootian, Shahrzad. 1997. *Persian.* New York: Routledge.

Meunier, Annie & Pollet Samvelian. 1997. La postposition rā en persan : son rôle dans la détermination et sa fonction discursive. *Cahiers de Grammaire* 25. 187–232.

Modarresi, Fereshteh. 2014. *Bare nouns in Persian: Interpretation, grammar and prosody*, PhD dissertation, Humboldt-Universität zu Berlin.

Rasekhmahand, Mohammad. 2004. Jāygāh maf'ul mostaqim dar fārsi [the position of the direct object in Persian]. *Nāme farhangestān* 6(4). 56–66.

Roberts, John R, Behrooz Barjasteh Delforooz & Carina Jahani. 2009. *A study of Persian discourse structure* Studia Iranica Upsaliensia, 12. Acta Universitatis Upsaliensis.

Samvelian, Pollet. 2001. Le statut syntaxique des objets nus en persan. *Bulletin de la Société de Linguistique de Paris* 96(1). 349–388.

Samvelian, Pollet. to appear. Specific features of the Persian syntax the Ezāfe Construction, Differential Object Marking and Complex Predicates. In Anousha Sedighi & Pouneh Shabani-Jadidi (eds.), *Oxford handbook of persian linguistics*. Oxford: Oxford University Press.

Stallings, Lynne M, Maryellen C MacDonald & Padraig G O'Seaghdha. 1998. Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language* 39(3). 392–417.

Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change* 9(1). 81–105.

Wasow, Thomas. 2002. *Postverbal behavior.* Stanford: CSLI Publications.

Windfuhr, Gernot L & John R Perry. 2009. Persian and Tajik. In Gernot L Windfuhr (ed.), *The iranian languages* Routledge Language Famliy Series, 416–544. London: Routledge.

Yamashita, Hiroko & Franklin Chang. 2001. "Long before short" preference in the production of a head-final language. *Cognition* 81(2). B45–B55.

The prepositional passive in Lexical Functional Grammar

Jamie Y. Findlay
University of Oxford

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 255–275

Keywords: passive, syntax-semantics interface, argument linking

Findlay, Jamie Y. 2016. The prepositional passive in Lexical Functional Grammar. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 255–275. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.14.

Abstract

The aim of this paper is to provide an adequate analysis in LFG of the prepositional passive, e.g. *That problem has been dealt with*, *My pen has been written with*. This construction has been examined in LFG before by Bresnan (1982), Lødrup (1991), and Alsina (2009), but empirical and theoretical problems, some well-documented, some new, mean that such proposals cannot be maintained. Instead, I offer an account couched in recent work on the mapping between grammatical functions and arguments (Asudeh et al., 2014; Findlay, 2014a) that treats the defining characteristic of the prepositional passive not as purely syntactic, but rather as being located at the interface between syntax and semantics.

1 Introduction

The *prepositional passive* (also *pseudopassive*) is much like the regular passive, except that the subject in the prepositional passive corresponds not to the object of the *verb* in the active, but to the object of a *preposition*:

(1) a. Scott relies [on Logan].
    b. Logan is relied on (by Scott).

This construction is typologically highly restricted—it is attested in only about half a dozen languages, mostly in the Germanic family (Truswell, 2008).[1] Nevertheless, in the languages in which it occurs, including English, it is a common and perfectly standard part of the grammar. Section 2 surveys the data around the prepositional passive in more detail, and asks what a theory which deals with this phenomenon must account for. In Section 3, I argue that previous LFG accounts of the prepositional passive have been inadequate, and show how improvements can be made by making use of recent work in argument linking. In the final part of this section, I also address the question of argumenthood, and offer some thoughts on how this notion interacts with Pustejovsky's (1995) qualia structure. Section 4 concludes.

[†]This paper is based on work from my MPhil thesis (Findlay, 2014b). I am indebted for that work to my supervisor Mary Dalrymple, and to many others from the Oxford community for helpful discussions and comments, especially Ash Asudeh, Anna Kibort, John Lowe, and Louise Mycock. This is a far better paper for their input, although they bear no responsibility for any errors that remain. My thanks also to the two anonymous reviewers for their useful comments and suggestions. As an MPhil student, I was the recipient of a UK Arts and Humanities Research Council studentship (grant reference AH/K503198/1), which I gratefully acknowledge.

[1]Specifically: Norwegian, Swedish, English, Vata (Koopman, 1984), Gbadi (*ibid.*), some North American varieties of French, especially those spoken on Prince Edward Island (King & Roberge, 1990), and potentially Papiamentu (Abels, 2003).

2 Explananda

2.1 Type I vs. Type II

For verbs like *rely (on)*, the existence of the prepositional passive is perhaps not so surprising. After all, the preposition is idiosyncratically selected and semantically inert, and the whole verb+preposition complex has a semantically unified, transitive meaning. Since transitivity is undoubtedly related to passivisation, it makes sense that such a relation should participate in the passive alternation, and since the preposition is part of the expression of this relation, it makes sense that it should remain with the verb and be left stranded.

However, the preposition involved need not be (uniquely) specified by the verb. First of all, we have prepositional passives with semantically contentful argument PPs:

(2) a. Scott spoke to/about Jean.
    b. Jean was spoken to/about.

Given that the only alternation here is between the prepositions, and given that this changes the meaning, it seems clear that the prepositions themselves bear some meaning.[2]

What is more, there are prepositional passives whose subjects appear to originate in adjunct PPs:

(3) When I'm on the bus I don't like being sat next to. [Locative]
(https://twitter.com/spencernickson/status/654923013285126144)

(4) Charles Dickens' quill pen has been written with by me. [Instrumental]
(http://www.bustle.com/articles/117731-10-amazing-margaret-atwood-quotes-from-the-2015-texas-book-festival-from-the-future-library-project)

(5) To come back, and not get turned around for.... [Benefactive]
(Will.i.am on an episode of The Voice, 7 Feb. 2014)

Here there can be no doubt that the prepositions are both meaningful and not selected by the verb, at least not in the traditional sense of subcategorisation. What these examples show is that it is not only idiomatically combining pairs of verbs and prepositions that allow for passivisation in this way.

Huddleston & Pullum (2002, 1433–1434) describe the cases where the preposition is idiosyncratically selected as Type I prepositional passives, and the cases where the preposition is contentful as Type II. In our analysis, we would ideally like an explanation which carries over to both, since it seems that they are two sides of the same coin, rather than totally separate phenomena.

[2] Whether they merely align their object with a certain thematic role or actually carry truth-conditional content themselves is not important; the point is that they are not semantically inert in the same way as *on* is in *rely on*.

2.2 Relation to the regular passive

The prepositional passive "has all the features of a canonical passive construction, except for one" (Alsina, 2009, 45), *viz.* that the subject corresponds to a prepositional, not verbal, object in the active. It is not, in other words, particularly exceptional (and for this reason I prefer the term 'prepositional passive' to 'pseudopassive', with its implications of inauthenticity).

Morphologically, the prepositional passive is identical to the regular passive; the verb in the prepositional passive has the same form as any other passive, namely the perfect participle of the verb. We also observe the same range of auxiliary verb possibilities: prepositional passives occur with *be* as well as *get*, just like the regular passive in English.

Furthermore, the prepositional passive is very productive, being used with verb-preposition combinations that have surely not been lexicalised, including recent neologisms:

(6) a. We can't bring you everything that is being blogged about.
(COCA[3])

b. Sean was tweeted at by Molly Mesnick.
(http://hollywoodlife.com/2013/03/12/catherine-giudici-sean-lowe-secret-engagement/)

c. This will definitely be facebooked on!
(http://thenaturalnutritionist.com.au/coconut-oil-the-scoop/)

This productivity argues that the prepositional passive should not be treated as a lexically idiosyncratic phenomenon, but as a general property of English grammar.

2.3 Adjacency

There is one property, however, which by its nature the prepositional passive cannot share with the regular passive. It is commonly observed that the verb and preposition must be adjacent in the prepositional passive, even though this restriction does not hold in the active:

(7) a. We rely increasingly on David.

b. * David is relied increasingly on.

Certainly, taken in isolation, the contrast in (7) seems clear enough. However, others have pointed out that this simplistic formulation is inadequate (e.g. Tseng, 2006). A wide variety of intervening adverbs and PP specifiers are in fact attested:

(8) a. I've stood there [on the London Underground], heavily pregnant (and obviously so), and been looked **straight** through.
(https://londondigitalmum.wordpress.com/category/commuting-2/)

[3] The Corpus of Contemporary American English (Davies, 2008–).

b. Yerba mate is consumed in nearly 100% of Argentine households. It is considered a staple food, and is relied **especially** on by poor Argentines when food is scarce.
(https://www.tni.org/files/download/47_bowles_2013_0.pdf)

c. Gas turbines have rapid transient response capabilities and, thus, will be relied **increasingly** on in markets with large intermittent sources.
(Lieuwen et al., 2013, 1311)

Indeed, example (8c) involves precisely the adverb-verb pairing ruled ungrammatical in (7b)! It seems that context is crucially important in determining the acceptability of such sentences, which means that purely invented examples are difficult to come up with.

Adverbs can apparently intervene between the verb and preposition, then. But what about direct objects? These are generally very bad:

(9) a. We put some books on the table.
b. * The table was put some books on (by them).

Of course, however, there are exceptions. Firstly, direct objects are perfectly acceptable when they form part of an idiom or a light verb construction:

(10) a. You have been taken advantage of.
b. Russia was declared war on (by Germany).
c. I've been made a fool of!

But in fact, given an appropriate context, direct objects are acceptable more widely. Bolinger (1975, 65) goes so far as to claim that "the only real restrictions are clarity and intent", which are clearly not syntactic constraints (see also Ziv & Sheintuch, 1981, from which the following examples are drawn):

(11) a. That city has been fought many a battle over.
b. He has been burned, stuck pins in, beheaded—all in effigy, of course.
c. To be whispered such dirty innuendos about was enough to break any girl's heart.
d. I don't like to be told lies about.

Once again, it seems that context is crucial, and that there cannot be a narrow syntactic rule in operation here. Following Tseng (2006), then, I assume that there is nothing in principle, and certainly nothing in the syntax, ruling out the presence of direct objects or other intervening material in the prepositional passive.

2.4 Semantic-pragmatic constraints

While there might not be uniquely syntactic constraints on the prepositional passive, there is a large amount of literature delimiting the class of predicates which can participate in the prepositional passive via non-syntactic restrictions. This work usually focusses on properties borne or ascribed to the subject.

2.4.1 Affectedness

The first of these is *affectedness*, based on the claim by Bolinger (1977, 67) that "the subject in a passive construction is conceived to be a true patient, i.e. to be genuinely affected by the action of the verb". This applies as much to the regular passive as to the prepositional passive, and is appealed to in order to explain such contrasts as the following:

(12) a. * I was approached by the train.
b. I was approached by the stranger. (Bolinger, 1977, 68)

Bolinger's argument is that in (12a), the subject is not an affected patient, but merely some kind of 'terminus', and therefore the passive is ruled out. In (12b), by contrast, the subject is affected by the actions of the stranger, being interacted with in some way: perhaps s/he is a panhandler, perhaps s/he is propositioning me, etc.

Such a contrast can be observed in the prepositional passive as well:

(13) a. * Seoul was slept in by the businessman last night.
b. This bed was surely slept in by a huge guy last night. (Kim, 2009)

The argument runs parallel here: Seoul is not affected by being slept in by a single individual, but a bed is—it becomes dirty, unmade, etc.

Now, it is clear that this notion of affectedness is not very formally explicit. It interacts with aspects of world knowledge, for one thing: it is significant that the man sleeping in the bed is *huge*, for example, since this, we know, will affect the bed more. And what exactly counts as 'affecting' something is far from clear. Bolinger intends the definition to cover "all sorts of physical, psychological and metaphorical effects" (Riddle & Sheintuch, 1983, 538), to the point where the term has been accused of being stretched so broadly as to be essentially meaningless (see the exchange in Householder, 1978 and Bolinger, 1978 for more on this).

Despite these accusations, however, there are cases which suggest that the definition is in fact not broad enough:

(14) a. And my brother simply cannot be disagreed with.
b. Such a dress can't be sat down in.
c. There the mistakes were, in their houses, pervading their lives, having to be sat with at every meal and slept with every night.
(Riddle & Sheintuch, 1983, 538)

It is far from obvious how my brother is affected by being unable to be disagreed with, and in (14b–c), it would seem more natural to claim that the subject is, if anything, the thing *doing* the affecting, not the thing being affected. The dress is preventing the wearer from sitting down, and the mistakes are causing discomfort in people while they sit at table and sleep in their beds. So, while we might wish to retain some notion of affectedness as being important in licensing passives, it clearly cannot be the whole story.

2.4.2 Role Prominence

Riddle & Sheintuch (1983) claim that the discriminating factor in (14) is not affectedness but *role prominence*. In their formulation, "all and only NPs whose referent the speaker views as being role prominent in the situation described by the passive clause occur as subjects of passive verbs" (p. 546). Role prominence is explained by Schachter (1977, 282) as belonging to the NP whose referent the speaker views as "being at the center of events".

In some sense this explains the existence of (14) accurately: it does seem that the subjects are "at the center of events", and that they are what the speaker is choosing to focus on from a discourse perspective. (14c) is 'about' the mistakes, for example, rather than the people suffering from having made them.

However, role prominence is perhaps even more vague a concept than affectedness. Riddle & Sheintuch (1983, 559) themselves note that "it is not possible to offer an algorithm for determining what causes some entity or concept to be viewed as role prominent", but regard this as no weakening of their account, claiming that role prominence is first and foremost a psychological notion, and one that we clearly have the ability to access, which means it is therefore, derivatively, accessible to the grammar. This may be so, but it does mean that their theory cannot, as it stands, make predictions about acceptability/grammaticality, and thus can only ever be offered as a *post hoc* explanation of the data, which considerably reduces its appeal.

2.4.3 Characterisation

One final property which has been discussed in the literature is that of *characterisation*. The role of characterisation is illustrated in the following examples:

(15) a. * Seoul was walked around by his father.
b. Seoul can be walked around in a day.

(16) a. * This statue was stood beside by John.
b. No statue should be stood beside in this park. (Kim, 2009)

In the (b) sentences, the VP gives what is a general or characteristic property of the subject—in other words, it characterises it. In the (a) sentences, no such relation holds, and they are therefore illicit.

The relationship between characterisation and role prominence is an interesting one. Notice that the crucial examples for both involve modality and/or negation. Do we need both conditions? On the face of it, characterisation seems to offer an acceptable account of the problematic sentences in (14) which motivated the appeal to role prominence, whereas role prominence alone cannot account for the ungrammaticality of e.g. (15a).

If it does prove possible to do away with role prominence, we would have a simple disjunction of necessary conditions on the passive: the subject must either be affected by or characterised by the predicate. This would also suggest that characterisation is a property of the regular passive as much as the prepositional passive, although it is only in terms of the latter that it has usually been discussed. This offers an explanation for one class of passives which ought not to be permitted if affectedness were the only constraint on passive subjects, namely those where the active voice object is not a patient:

(17) Many people fear spiders. $\sim$ Spiders are feared by many people.

We cannot straightforwardly say that spiders are affected by being feared by many people, but it does seem to characterise them.

2.5 Summary

In summary, the prepositional passive is like the regular passive in most ways: morphologically, syntactically, and in terms of semantic-pragmatic constraints. Any analysis of the prepositional passive ought to be an extension to the analysis of the regular passive, therefore, and not a replacement for it.

3 Analysis

The essential property of the prepositional passive (especially where the preposition is contentful, i.e. Type II) is that the clause's subject in the syntax corresponds to the stranded preposition's internal argument in the semantics. What a formal analysis has to do, therefore, is provide a mechanism for passing the subject's referent to the prepositional meaning in the semantics. This is a question of the mapping from f- to s-structure. Previous LFG analyses have treated the prepositional passive as a purely syntactic phenomenon, and thereby miss this most basic formulation of the problem. I turn now to two such analyses, and the problems they face.

3.1 Previous LFG analyses

3.1.1 Reanalysis

The 'canonical' theory of the prepositional passive in LFG remains the reanalysis account of Bresnan (1982). This involves a lexical rule of V-P incorporation, given in (18), which morphologically incorporates the verb and preposition, and merges their valency frames. (19) gives an example for *rely*:

(18) **V–P Incorporation:**
Operation on lexical form: $(\text{P OBJ}) \mapsto (\text{OBJ})$
Morphological change: $\text{V} \mapsto [\text{V P}]_{\text{V}}$

(19) 'rely $\langle \text{SUBJ}, \text{ON OBJ} \rangle$' V $\mapsto$ 'rely on $\langle \text{SUBJ}, \text{OBJ} \rangle$' $[\text{V P}]_{\text{V}}$

Reanalysis accounts have been popular outside of LFG as well (e.g. van Riemsdijk, 1978; Hornstein & Weinberg, 1981), and provide an obvious way of accounting for the behaviour of prepositional verbs in the passive: as we noted above, *rely on* acts in many ways as a unit, and so it makes sense to unify it at some level of representation. Once this is done, it behaves just like any other transitive verb, and thus undergoes passive perfectly normally. Unfortunately, such accounts are not without their problems. Postal (1986) and Baltin & Postal (1996) have argued at length that the issues facing such an approach are insurmountable, and that, despite its appeal, the reanalysis account is ultimately untenable. Space precludes a full discussion of the issues here, but I will briefly illustrate two erroneous predictions made by the rule in (18).

Firstly, it predicts that the object of a preposition in a V+P sequence should (at least optionally) behave like the direct object of a normal transitive verb. However, this does not appear to be the case. Consider data from heavy NP shift, for example:

(20) a. I discussed $\_\__1$ with Lorenzo [the problems he was having with deliveries]$_1$.
b. * I argued with $\_\__2$ about such problems [the drivers' union leader]$_2$.
(Baltin & Postal, 1996, 129)

The same lack of parallelism is observed in subdeletion (Bresnan, 1973):

(21) a. Jane saw more of these people than Sally saw __ of those people.
b. * Jane spoke to more of these people than Sally spoke to __ of those people. (after Baltin & Postal, 1996, 131)

In neither case do the prepositional objects behave in the same way as the direct objects, contrary to the predictions of the reanalysis account.

The second prediction of the reanalysis account is that the V+P complex ought to behave like a single morphological word; but the preposition displays a high degree of syntactic mobility not expected if it is morphologically incorporated:

(22) a. The bridge was flown (both) over and under.
b. Communism was talked, argued, and fought about.
c. The bridge was flown over and then, but only then, under.
d. Fascism was fought for by Goebbels and (then) against by De Gaulle.
e. Fascism was fought for by Goebbels and then, but I assure you, only then, against by De Gaulle.
(Postal, 1986, 223, fn. 14)

Similarly, incorporation requires absolute adjacency between the verb and preposition, which, as we have seen, is not the correct characterisation of the data. For these reasons, any reanalysis-based explanation of the prepositional passive is simply unable to account adequately for the data, and so must be abandoned.

3.1.2 Structure sharing

The only alternative theory on the prepositional passive within LFG that I am aware of is the structure-sharing account of Lødrup (1991) and Alsina (2009). Although they differ in formal details, they both work on the principle that the prepositional passive should be analysed as a structure-sharing relation between the subject and the object of the oblique:

(23) *The bed was slept in.*

$$\left[\begin{array}{ll} \text{PRED} & \text{`sleep'} \\ \text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`bed'}\end{array}\right] \\ \text{OBL}_{\text{LOC}} & \left[\begin{array}{ll}\text{PRED} & \text{`in'} \\ \text{OBJ} & \end{array}\right] \\ \text{VOICE} & \text{PASSIVE}\end{array}\right]$$

This of course does not follow from anything about the passive, and thus requires further formal machinery: for Lødrup, the prepositional passive also involves the addition of a control equation $(\uparrow \text{SUBJ}) = (\uparrow \text{OBL}_\theta\ \text{OBJ})$, while for Alsina, other aspects of his more elaborated theory of structure sharing (Alsina, 2008) force the appropriate structure.

The structure sharing proposal in some ways captures the claim that the clause's subject corresponds to the preposition's argument, which is, after all, canonically its object. However, I do not believe it is right to ignore the role of the mapping between syntax and semantics here, instead framing this as a purely syntactic phenomenon.

Firstly, the structure-sharing account makes the wrong predictions with respect to case-marking. In the proposed structure-shared relation, we would expect case identity between the two positions, since the f-structures which are shared must be (token) identical. However, this is not what we observe:

(24) I rely on him.
NOM ACC

(25) a. He is relied on.
NOM

b. * Him is relied on.
ACC

The subject position requires nominative case, while the prepositional object requires accusative; whatever case is assigned to the subject of the prepositional passive will lead to a clash, therefore, and it turns out that it is the nominative case of the subject that is actually attested, providing no reason to suppose there is any structure sharing here.

We could escape this unwanted case clash by making use of the restriction operator (Kaplan & Wedekind, 1993), and rewriting the control equation used by Lødrup as below:

(26) $(\uparrow \textsc{subj})/\textsc{case} = (\uparrow \textsc{obl}_\theta\ \textsc{obj})/\textsc{case}$

This solves the problem for Lødrup, although we may note disapprovingly that it adds another layer of stipulation, but there remain formal problems for Alsina. Since the Theory of Structure-Sharing of Alsina (2008) does away with the need for control equations, there is nowhere to add the restriction operator, and the structure sharing must therefore, I presume, be total.

Secondly, on a more (meta-)theoretical level, we might object that such an approach makes the prepositional passive very different from the regular passive. In fact, it starts to look quite transformational: the subject is 'really' the object of the preposition, but has been displaced into the subject position. Given such an approach, we might wonder, for example, why the regular passive does not look like (27):

(27) *The cake was eaten.*

$$\left[\begin{array}{ll} \textsc{pred} & \text{'eat'} \\ \textsc{subj} & \left[\begin{array}{ll}\textsc{pred} & \text{'cake'}\\ \textsc{spec} & \left[\begin{array}{ll}\textsc{pred} & \text{'the'}\end{array}\right]\end{array}\right] \\ \textsc{obj} & \\ \textsc{voice} & \textsc{passive}\end{array}\right]$$

This is perhaps a little unfair, as Alsina would no doubt respond that *bed* in (23) bears no thematic relation to the predicate *sleep*, while *cake* does bear such a relation to *eat* in (27), thus accounting for the difference. I do not wish to overstate the charge on this count, therefore. But what I do want to emphasise is that such a disparity in analyses between the regular passive and the prepositional passive is not desirable. We noted above that the prepositional passive is identical to the passive in most respects, and so it seems to me that we should strive for a parallel analysis if at all possible.

3.2 Proposal

In this section I give my own proposal for the best way to represent the prepositional passive in LFG. I begin in the next section by outlining the underlying machinery I assume, before turning to the account of Type I and Type II prepositional passives in turn.

3.2.1 Machinery

For the mapping between arguments and grammatical functions (GFs), we use the model of Asudeh et al. (2014) (see also Asudeh & Giorgolo, 2012, Findlay, 2014a).

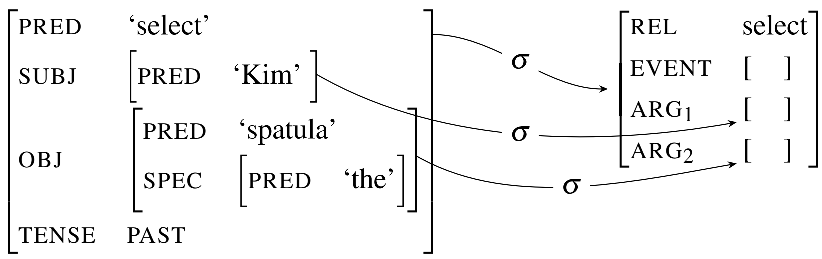


Figure 1: Mapping from f-structure to a connected semantic structure for *Kim selected the spatula.*

In this theory, such mapping is handled via various functional descriptions, primarily through defining equations like (28), ultimately to be provided by some version of Lexical Mapping Theory (LMT: Bresnan & Kanerva, 1989; Kibort, 2007; Findlay, 2014a).

(28) $(\uparrow \text{OBJ})_\sigma = (\uparrow_\sigma \text{ARG}_2)$

These define the possible links, via the sigma projection function, between the values of GF features in the f-structure and argument positions in a connected s-structure. The latter represent resources to be used in the Glue Semantics (Dalrymple, 1999). Such a mapping is illustrated for the active voice sentence *Kim selected the spatula* in Figure 1.

For the passive, we use the model of Kibort (2001), whereby the highest argument of a predicate, ARG_1, is marked as semantically restricted.[4] In the present model this means it must appear as an OBL_θ if it is realised syntactically.

The regular passive can thus be described via the following template:[5]

(29) PASSIVE :=
$(\uparrow \text{VOICE}) = \text{PASSIVE}$
@ADDMAP(PLUSR, ARG_1)

$(\lambda P \exists x.[P(x)] : [(\uparrow_\sigma \text{ARG}_1) \multimap \uparrow_\sigma] \multimap \uparrow_\sigma)$

The first line provides the relevant f-structural information regarding the VOICE feature. The ADDMAP template is responsible for restricting the first argument in the manner just described. Its definition is given below:

(30) ADDMAP(D, A) :=
$\{(\uparrow \text{D})_\sigma = (\uparrow_\sigma \text{A}) | (\uparrow_\sigma \text{A})_{\sigma^{-1}} = \varnothing\}$

[4] In LMT terms, it is marked as $[+r]$.

[5] A template is just a bundle of functional information given a name. They are 'called' or 'invoked' in a lexical entry or annotated c-structure rule by prefixing the name with the @ symbol. On the potential theoretical import of templates in forming generalisations, see Dalrymple et al. (2004). This version of the passive template is based on that in Asudeh & Giorgolo (2012).

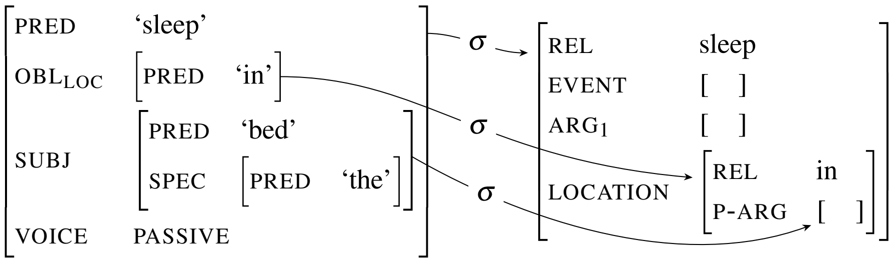


Figure 2: Mapping from f-structure to s-structure for *The bed was slept in*.

This template is used generally for adding mapping constraints, and says that either the feature disjunction D maps to the argument position A, or nothing maps to A; in other words, an argument must be mapped to by one of a specified pair of grammatical functions unless it is syntactically unrealised.[6] For more on this approach to mapping theory, see Asudeh et al. (2014) and Findlay (2014a).

Finally, the meaning constructor in the third line of (29) existentially closes the first argument of the passive predicate. It is optional because in the long passive it will not be needed: there the resource corresponding to ARG_1 will be provided by the *by*-phrase. The resource sensitivity of Glue Semantics will ensure that the meaning constructor in (29) is used if and only if it is required.

3.2.2 Expanding the passive template

In keeping with the claim that the prepositional passive is by and large identical to the regular passive, we would like the former to be an augmentation of the latter, rather than a replacement for it. In that case, what information must we add to the regular passive template in (29)? Firstly, we must include the fundamental information that the subject of the clause at f-structure is the argument of the stranded preposition at s-structure. That is, we want to arrive at the structure given in Figure 2, where P-ARG is the name of a feature at s-structure standing for the internal argument of the preposition.

If we assume for the moment that only OBLiques can be involved (I turn to the apparent adjuncts shortly), then we need an equation like the following, which maps the subject to the preposition's internal argument:

(31) $(\uparrow \text{SUBJ})_\sigma = ((\uparrow \text{OBL})_\sigma \text{ P-ARG})$

However, we also need to limit the OBL in question to the *nearest* PP to the verb, to account for contrasts like (32):

(32) a. Victor has been spoken to about this.
b. * Victor has been spoken about this to.

[6]The LMT features are reconceptualised as disjunctions of grammatical functions, so that e.g. $\text{PLUSR} \equiv \{\text{OBJ}_\theta | \text{OBL}_\theta\}$.

In order to achieve this, we use the following, more detailed, description:

(33) $(\uparrow \text{OBL}_\theta) = \%\text{STRDD-PREP}$
$\neg(\uparrow \text{OBL}_\theta) <_f^{\text{VP}} \%\text{STRDD-PREP}$
$(\uparrow \text{SUBJ})_\sigma = (\%\text{STRDD-PREP}_\sigma\ \text{P-ARG})$

OBL_θ represents a disjunction over all OBL functions, i.e. (34):

(34) $\text{OBL}_\theta \equiv \{\text{OBL}_{\text{GOAL}}|\text{OBL}_{\text{LOC}}|\ldots|\text{OBL}_{\text{TO}}|\text{OBL}_{\text{ON}}|\ldots\}$

The first line of (33) therefore picks some oblique and, using a local variable (Crouch et al., 2012), names it %STRDD-PREP. The second line then requires of this oblique that no other oblique f-precedes it within the VP.[7] In the last line, we then include the mapping information from (31), now relativised to the correct oblique.

Including all of this, the passive template is thus augmented as follows:

(35) PASSIVE :=

$$(\uparrow \text{VOICE}) = \text{PASSIVE}$$
$$@\text{ADDMAP}(\text{PLUSR}, \text{ARG}_1)$$
$$\left(\begin{array}{l} (\uparrow \text{OBL}_\theta) = \%\text{STRDD-PREP} \\ \neg(\uparrow \text{OBL}_\theta) <_f^{\text{VP}} \%\text{STRDD-PREP} \\ (\uparrow \text{SUBJ})_\sigma = (\%\text{STRDD-PREP}_\sigma\ \text{P-ARG}) \end{array}\right)$$
$$(\lambda P \exists x.[P(x)] : [(\uparrow_\sigma \text{ARG}_1) \multimap \uparrow_\sigma] \multimap \uparrow_\sigma)$$

The prepositional passive information is optional, because in the regular passive it will not be used. In the prepositional passive, however, it will have to be selected, or else there will be no appropriate analysis of the sentence: the preposition has no object to map to its internal argument, and so there will be a resource deficit in the semantics if the subject mapping equation is not selected.

[7]F-precedence is essentially the f-structure reflex of c-structure precedence (i.e. linear precedence): it is the image of c-precedence under the ϕ function from c-structure to f-structure (Kaplan & Zaenen, 1989). Relativised f-precedence (Zaenen & Kaplan, 1995, 236) further restricts this relation so that the corresponding c-structure nodes (reached via the inverse of the ϕ function, ϕ^{-1}) must all be dominated by the same, specified category:

(i) For two f-structure elements f_1 and f_2, and a category X,
f_1 *f-precedes* f_2 *relative to X* ($f_1 <_f^X f_2$) iff for all $n_1 \in \phi^{-1}(f_1)$ and for all $n_2 \in \phi^{-1}(f_2)$, n_1 c-precedes n_2 and n_1 and n_2 are co-dominated by X.

This more restrictive definition of f-precedence is needed here because fronted OBLs, which f-precede the stranded preposition in the general sense, where c-structure co-domination is not required, do not result in ungrammaticality:

(ii) About this, Victor has been spoken to.

Now, as we saw in Section 2, there are a number of other potential semantic and pragmatic constraints on the passive in general. Whatever the ultimate consensus on their exact nature, they can easily be accommodated in the present approach, simply by adding the requisite meaning constructor(s) or functional constraints to the passive template:

(36) $\textsc{Passive} :=$

$$(\uparrow \text{VOICE}) = \text{PASSIVE}$$
$$@\textsc{AddMap}(\text{PLUSR}, \text{ARG}_1)$$
$$\begin{pmatrix} (\uparrow \text{OBL}_\theta) = \%\text{STRDD-PREP} \\ \neg(\uparrow \text{OBL}_\theta) <^{\text{VP}}_{f} \%\text{STRDD-PREP} \\ (\uparrow \text{SUBJ})_\sigma = (\%\text{STRDD-PREP}_\sigma\ \text{P-ARG}) \end{pmatrix}$$

$$\lambda P \lambda x \lambda e. P(e,x) \wedge [\mathit{affected}(e,x) \vee \mathit{characterised}(e,x)] : \\ [(\uparrow \text{SUBJ})_\sigma \multimap (\uparrow_\sigma \text{EVENT}) \multimap \uparrow_\sigma] \multimap \\ (\uparrow \text{SUBJ})_\sigma \multimap (\uparrow_\sigma \text{EVENT}) \multimap \uparrow_\sigma$$

$$(\lambda P \exists x.[P(x)] : [(\uparrow_\sigma \text{ARG}_1) \multimap \uparrow_\sigma] \multimap \uparrow_\sigma)$$

Of course, the meaning constructor in (36) is intended as a placeholder only. For one thing, as it stands, it suggests that sentences like **Seoul was walked around by my father*, where the relevant constraints do not apply, is unacceptable because it is *false*, not because of any linguistic ill-formedness. This is surely not right. What we really need is a constraining equation which requires certain properties but does not provide them itself. Exactly what features, at which level of representation, are to be constrained, however, remains an open question.[8]

Let us turn now to how each of the two types of prepositional passive can be analysed under this approach.

3.2.3 Type I

Type I prepositional passives are the less problematic of the two, since the esoteric information can all be encoded locally, in a single lexical entry.

(37) *rely* V $(\uparrow \text{PRED}) = \text{`rely'}$

$$(\uparrow_\sigma \text{ARG}_2) = ((\uparrow \text{OBL}_{\text{ON}})_\sigma\ \text{P-ARG})$$

$$\lambda y \lambda x.\mathit{rely\_on}(x,y) : \\ (\uparrow_\sigma \text{ARG}_2) \multimap (\uparrow_\sigma \text{ARG}_1) \multimap (\uparrow_\sigma \text{EVENT}) \multimap \uparrow_\sigma$$

[8] If affectedness really is patient-hood, then this may be one piece of evidence that pleads for the presence of thematic role information in the grammar, *contra* the motivations of Kibort (2007) and Findlay (2014a), who argue that a grammar that makes no reference to thematic roles is preferable on theoretical grounds.

In the default case, the preposition's object will map to the P-ARG, while if the passive template is selected, it will be the clause's subject. Either way, that argument is identified as the second argument of the verb *rely* and passed to its semantics appropriately, where the verb now behaves exactly like any other transitive verb.

3.2.4 Type II

At present, the equations in (33) refer to the nearest oblique. If the PPs which participate in Type II prepositional passives are obliques, then no more need be said: the analysis will hold of them directly. But as we have seen, it looks as though NPs can be promoted to subject out of adjuncts, and so if we want to say that these are really obliques, contrary to appearances, we have some more work to do.

Given that the passive is an *argument* alternation, it would be surprising to find that it allowed non-arguments to participate, but only when they were PPs. There are no regular passives from adjuncts, for example:

(38) They smiled last night. ∼ *Last night was smiled (by them).

Adam Przepiórkowski (p.c.) argues that this shows little, however, since there are equally no passives from obliques (of the form **In the bed was slept*, for example), and thus we are begging the question by assuming a distinction between obliques and adjuncts in the first place. Indeed, Przepiórkowski sees the lack of contrast between arguments and adjuncts in the prepositional passive as further evidence that there is no distinction between these two categories in general (on which see Przepiórkowski, 2016). A full discussion of such a proposal would take us too far afield here, but I do note approvingly that the distinctions that seem relevant for the prepositional passive are semantic-pragmatic rather than syntactic, and gradient rather than categorical, which would seem entirely in keeping with an approach where the distinction between dependent-types is collapsed in the syntax.

In his discussion of this problem, Alsina (2009, 55) advocates that we "assume that certain verbs can augment their argument structures with a locative or instrumental argument". This is very reminiscent of the suggestion by Needham & Toivonen (2011) that certain classes of PP can be added as 'derived' arguments to a verb's argument structure, rather than being true adjuncts. It would be nice, then, if the class of apparent adjunct PPs which participate in the prepositional passive were a subset (proper or otherwise) of Needham & Toivonen's list of derived arguments. But this is not the case: instrumentals and benefactive *for*-phrases are productive sources of prepositional passives, and listed as derived arguments by Needham & Toivonen, but displaced themes and directionals, also on the authors' list, are not. And locatives, which account for large numbers of prepositional passives, are not mentioned in the list of derived argument types.

It is certainly not true, then, that being a derived argument is a sufficient condition for prepositional passive subjecthood. But this is perhaps not surprising: as we saw in Section 2, there are other constraints on the passive which still obtain. In fact, these constraints seem to be the ultimate arbiters of whether or not a given PP can participate in the prepositional passive. Recall the case of *This bed has been slept in*. It is not true that all locative PPs are automatically permitted here:

(39) a. This bed has been slept in.
b. * This bed has been slept under.

Rather, whether or not a particular type of PP will enter into the prepositional passive is heavily dependent on contextual factors. For example, passives with *slept under* are perfectly acceptable given an appropriate context:

(40) This heavy sheet is designed to be slept under.

What is happening here, then, is not purely linguistic. To return to the examples in (39), one suggestion for the source of the contrast is the Affectedness Condition: beds are affected by being slept in but not by being slept under. But this is inextricably linked to our knowledge of the world—of how beds work and what happens to them when people sleep in them. This kind of real-world knowledge that has linguistic effects is precisely the sort of information discussed by Pustejovsky (1995) in relation to *qualia structure*, where information is stored in lexical entries which relates to the canonical relations associated with particular expressions.

For example, the qualia structure for *book* will include the information that the prototypical relations it enters into are those of *reading* or *writing*. This, so Pustejovsky argues, allows us to correctly interpret sentences like *Tim began the book* as meaning *Tim began reading the book* (or *writing*, if we know he is an author), since the relations in the qualia structure of *book* are available during composition (what he refers to as co-composition).

If this approach is extended throughout the lexicon, then verbs and nouns will specify which kinds of relations are particularly associated with them, and it could well be that those are precisely the relations which are more argument-like when they are used, and therefore which will be realised as OBLs rather than ADJs. For example, it is clearly part of our understanding of sleeping that, for humans, it now normally happens *in* things, and usually in beds, at that. Similarly, it is part of our knowledge of beds that they are usually slept *in*—and of sheets that they are slept *under*. Whether this knowledge is properly *linguistic* knowledge is not so clear, however, and it may be preferable to integrate it in some way that avoids encoding it in the lexicon.[9]

Of course, we can imagine scenarios where other relations are appropriate: take a situation in which we are all sleeping in a dormitory, but the beds are all full, and so some of us are sleeping under the beds. Then, if a newcomer were looking for somewhere to bed down, it would seem perfectly sensible to say (41):

[9]There are large question marks over the desirability of bringing such real-world knowledge into our lexical representations in the manner advocated by Pustejovksy. A reviewer points out that Pustejovsky's qualia structure is not computationally implementable, for example, which is a major concern in a well-formalised theory like LFG.

(41) That bed is being slept under already.

But the point is that this requires a more marked context to come off successfully. Without details of the context, it does indeed seem odd, in a way that *That bed is being slept in already* does not. Really, relations that are available to the prepositional passive are simply those which are contextually relevant; it's just that some such relations are taken as the default, available even in the null context, which allows them to be used without a richer contextual background.[10] Notions like affectedness or prominence, and, perhaps, the distinction between arguments and adjuncts, may in fact be epiphenomena, the result of Gricean-style inference interacting with knowledge about the kinds of relationships with the world which the referents of words enter into.

4 Conclusion

Let us recap the main points raised in this paper. Firstly, the prepositional passive is only minimally different from the regular passive, and subject to the same semantic and pragmatic constraints. We should, therefore, strive for a parallel analysis as far as possible. Existing LFG analyses are either empirically inadequate or obscure this similarity.

A minimally sufficient analysis can be incorporated into existing theories of the passive by simply adding the crucial information that sets the prepositional passive apart, namely the mapping from f- to s-structure. Any additional constraints, unique to the prepositional passive or otherwise, can be added and elaborated on as and when needed, owing to the modular nature of the approach, but of course ideally would be reducible to more basic principles. I suggested, finally, what such a reduction might involve: namely, making use of qualia structure or some other mechanism to reduce many of the constraints to questions of relevance, although of course a good formalisation of this notion, as with so many higher-level cognitive processes, remains elusive.

[10]Berthold Crysmann (p.c.) has kindly brought to my attention another example of real-world knowledge having an impact on syntactic acceptability. In German, when a PP argument is separated from its noun predicator, the acceptability of the sentence is affected by whether the embedding verb and the noun form a contextually unmarked meaning or not (Grewendorf, 1989):

(i) Über Syntax hat Hans ein Buch ausgeliehen.
about syntax has Hans a book borrowed
'Hans has borrowed a book about syntax.'

(ii) * Über Syntax hat Hans ein Buch geklaut.
about syntax has Hans a book stolen
'Hans has stolen a book about syntax.' (De Kuthy, 1998)

Borrowing a book is more usual, and thus less contextually marked, than stealing one, and so (i) is grammatical while (ii) is not. Such a contrast equally pleads for the integration of real-world knowledge into the grammar, either via qualia structure or some other mechanism.

References

Abels, Klaus. 2003. Successive cyclicity, anti-locality and adposition stranding. Doctoral dissertation, University of Connecticut.

Alsina, Alex. 2008. A theory of structure-sharing: Focusing on long-distance dependencies and parasitic gaps. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG08 Conference*, 5–25. Stanford, CA: CSLI Publications.

Alsina, Alex. 2009. The prepositional passive as structure-sharing. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG09 Conference*, 44–64. Stanford, CA: CSLI Publications. http://www.stanford.edu/group/cslipublications/cslipublications/LFG/14/papers/lfg09alsina.pdf.

Asudeh, Ash & Gianluca Giorgolo. 2012. Flexible composition for optional and derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference*, 64–84. Stanford, CA: CSLI Publications. http://www.stanford.edu/group/cslipublications/cslipublications/LFG/17/papers/lfg12asudehgiorgolo.pdf.

Asudeh, Ash, Gianluca Giorgolo & Ida Toivonen. 2014. Meaning and valency. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG14 Conference*, 68–88. CSLI Publications. http://web.stanford.edu/group/cslipublications/cslipublications/LFG/19/papers/lfg14asudehetal.pdf.

Baltin, Mark & Paul M. Postal. 1996. More on reanalysis hypotheses. *Linguistic Inquiry* 27(1). 127–145.

Bolinger, Dwight. 1975. On the passive in English. In Adam Makkai & Valerie Becker Makkai (eds.), *The first LACUS Forum 1974*, 57–77. Columbia, SC: Hornbeam Press.

Bolinger, Dwight. 1977. Transitivity and spatiality: the passive of prepositional verbs. In A. Makkai, V. B. Makkai & L. Heilmann (eds.), *Linguistics at the crossroads*, Lake Bluff, IL: Jupiter Press.

Bolinger, Dwight. 1978. Passive and transitivity again. *Forum Linguisticum* 3. 25–28.

Bresnan, Joan. 1973. Syntax of the comparative clause construction in English. *Linguistic Inquiry* 4(3). 275–343.

Bresnan, Joan. 1982. The passive in lexical theory. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 3–86. Cambridge, MA: MIT Press.

Bresnan, Joan & Jonni M. Kanerva. 1989. Locative inversion in Chichewa: A case study of factorization in grammar. *Linguistic Inquiry* 20(1). 1–50.

Crouch, Dick, Mary Dalrymple, Ron Kaplan, Tracy King, John Maxwell & Paula Newman. 2012. *XLE documentation*. Palo Alto Research Center (PARC), Palo Alto, CA. http://www2.parc.com/isl/groups/nltt/xle/doc/xle.html.

Dalrymple, Mary (ed.). 1999. *Semantics and syntax in Lexical Functional Grammar: the resource logic approach*. Cambridge, MA: MIT Press.

Dalrymple, Mary, Ronald M. Kaplan & Tracy Holloway King. 2004. Linguistic generalizations over descriptions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG04 Conference*, 199–208. Stanford, CA: CSLI Publications. http://www.stanford.edu/group/cslipublications/cslipublications/LFG/9/lfg04dkk.pdf.

Davies, Mark. 2008–. *The Corpus of Contemporary American English: 450 million words, 1990–present*. http://corpus.byu.edu/coca.

De Kuthy, Kordula. 1998. Splitting PP arguments from NPs: an argument raising approach and its interaction with lexical semantics. In Tibor Imre Kiss & Walt Detmar Meurers (eds.), *Proceedings of the ESSLLI workshop 'Topics in Constraint-Based Germanic Syntax'*, Saarbrücken.

Findlay, Jamie Y. 2014a. Mapping Theory without Argument Structure. Unpublished ms., University of Oxford. Submitted to *Journal of Language Modelling*, September 2016. https://www.academia.edu/9055653/Mapping_Theory_without_Argument_Structure.

Findlay, Jamie Y. 2014b. The prepositional passive: A lexical functional account. MPhil thesis, University of Oxford.

Grewendorf, Günther. 1989. *Ergativity in German*. Dordrecht, NL: Foris.

Hornstein, Norbert & Amy Weinberg. 1981. Case theory and preposition stranding. *Linguistic Inquiry* 12(1). 55–91.

Householder, Fred W. 1978. Review of *The First LACUS Forum*. *Language* 54(1). 170–176.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge, UK: Cambridge University Press.

Kaplan, Ronald M. & Jürgen Wedekind. 1993. Restriction and corresponence-based translation. In *Proceedings of the 6th EACL Conference*, 193–202.

Kaplan, Ronald M. & Annie Zaenen. 1989. Long-distance dependencies, constituent structure, and functional uncertainty. In Mark Baltin & Antony Kroch (eds.), *Alternative conceptions of phrase structure*, 17–42. Chicago, IL: Chicago University Press.

Kibort, Anna. 2001. The Polish passive and impersonal in Lexical Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG01 Conference*, 163–183. Stanford, CA: CSLI Publications. http://www.stanford.edu/group/cslipublications/cslipublications/LFG/6/lfg01kibort.pdf.

Kibort, Anna. 2007. Extending the applicability of Lexical Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG07 Conference*, 250–270. Stanford, CA: CSLI Publications. http://www.stanford.edu/group/cslipublications/cslipublications/LFG/12/papers/lfg07kibort.pdf.

Kim, Jong-Bok. 2009. Three types of English pseudo-passives: A lexicalist perspective. *Korean Journal of Linguistics* 34(1). 1–24.

King, Ruth & Yves Roberge. 1990. Preposition stranding in Prince Edward Island French. *Probus* 3. 351–369.

Koopman, Hilda. 1984. *The syntax of verbs: From verb movement rules in the Kru languages to Universal Grammar* (Studies in Generative Grammar 15). Dordrecht, NL: Foris.

Lieuwen, Tim, Michael Chang & Alberto Amato. 2013. Stationary gas turbine combustion: Technology needs and policy considerations. *Combustion and Flame* 160(8). 1311–1314.

Lødrup, Helge. 1991. The Norwegian pseudopassive in lexical theory. *Working Papers in Scandinavian Syntax* 47. 118–129.

Needham, Stephanie & Ida Toivonen. 2011. Derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 Conference*, 401–421. http://www.stanford.edu/group/cslipublications/cslipublications/LFG/16/papers/lfg11needhamtoivonen.pdf.

Postal, Paul M. 1986. *Studies of passive clauses*. Albany, NY: State University of New York Press.

Przepiórkowski, Adam. 2016. How *not* to distinguish arguments from adjuncts in LFG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 539–559. Stanford, CA: CSLI Publications.

Pustejovsky, James. 1995. *The generative lexicon*. Cambridge, MA: MIT Press.

Riddle, Elizabeth & Gloria Sheintuch. 1983. A functional analysis of pseudo-passives. *Linguistics and Philosophy* 6(4). 527–563.

van Riemsdijk, Henk C. 1978. *A case study in syntactic markedness: the binding nature of prepositional phrases*. Lisse, NL: The Peter de Ridder Press.

Schachter, Paul. 1977. Reference related and role related properties of subjects. In Peter Cole & Jerrold M. Sadock (eds.), *Syntax and semantics 8: Grammatical relations*, New York, NY: Academic Press.

Truswell, Robert. 2008. Preposition stranding, passivisation, and extraction from adjuncts in Germanic. *Linguistic Variation Yearbook* 8. 131–178.

Tseng, Jesse. 2006. English prepositional passives in HPSG. In Gerhard Jäger, Paola Monachesi, Gerald Penn & Shuly Wintner (eds.), *FG-2006: Proceedings of the 11th Conference on Formal Grammar*, 147–159. Málaga, ES: CSLI Publications Online Proceedings. http://w3.erss.univ-tlse2.fr/textes/pagespersos/tseng/Pubs/FG06-tseng.pdf.

Zaenen, Annie & Ronald M. Kaplan. 1995. Formal devices for linguistic generalizations: West Germanic word order in LFG. In Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell, III, & Annie Zaenen (eds.), *Formal issues in Lexical-Functional Grammar*, 215–240. Stanford University: CSLI Publications.

Ziv, Yael & Gloria Sheintuch. 1981. Passives of obliques over direct objects. *Lingua* 54. 1–17.

Exhaustive object control constructions in Greek: An LFG/XLE treatment

Alexandra Fiotaki
University of Ioannina

Katerina Tzortzi
University of Crete

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 276–296

Keywords: exhaustive object control, na clauses, PRO, anaphoric control, Greek

Fiotaki, Alexandra & Katerina Tzortzi. 2016. Exhaustive object control constructions in Greek: An LFG/XLE treatment. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 276–296. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.15.

Abstract

In this paper we propose an LFG/XLE treatment of Exhaustive Object Control (EOC) constructions in Greek *na* clauses. We draw on data retrieved from the Hellenic National Corpus (HNC) in order to define the verbs that allow EOC. We treat EOC using anaphoric control. We take the subject of the subordinate *na* clause (controllee) to be a PRO marked with nominative case that is anaphorically related to the object of the matrix clause (controller). We implement this analysis in our LFG/XLE Grammar by adding the new feature ANAPH_C_BY.

1. Introduction

Control is a dependency between an unexpressed subject (the controlled element[1]) and an expressed or unexpressed constituent (the controller; Bresnan 1982). Control constructions in Greek *na* subordinate clauses have been widely discussed in the literature and they still remain a controversial topic (Iatridou 1993, Varlokosta 1994, Philippaki-Warburton and Catsimali 1999, Landau 2002). In this paper, we study exhaustive object control (EOC) in *na* subordinate clauses focusing on the verbal predicates illustrated in (1).

(1) mathainw 'teach', voithw 'help', peithw 'persuade', empodizw 'prevent', protrepw 'urge', epitrepw 'allow', apagoreuw 'forbid'.

In the analysis of EOC in English the subject of the infinitive is functionally controlled by the object of the matrix verb (Bresnan 1982). In the corresponding structure in Greek, the subordinate clause lacks an infinitival verb form but surfaces as a *na* clause, exemplified by (2) (Triadafillidis 1941, Philippaki 2004, Roussou 2009). *Na* complements differ from infinitives, among others, in that they show person and number agreement and in that in combination with certain control verbs, they license overt subjects (3)[2].

---

1. In constructions like "This book is tough to finish" the controllee could be an object. Dalrymple and King (2000) propose an anaphoric control analysis for these constructions.
2. A detailed anayles is not demonstrated for (3) since the differences between infinitives and *na* complements are not the point of this study.

(2) O Nikos apagoreuei ths Marias [*na*
The-DEF Nikos-NOM forbid-3SG the-DEF Maria-GEN [to-COMPL
erthei].
come-3SG][3]
'Nikos forbids Maria to come.'

(3) O Nikos apagoreuei *na* erthei **h**
The-DEF Nikos-NOM forbid-3SG to-COMPL come-3SG the-DEF
Maria.
Maria-NOM
'Nikos forbids Maria to come.'

As a result, the standard analysis of English EOC does not extend to Greek. We provide a theoretical analysis of the phenomenon. This analysis is also tested in the framework of the Greek LFG/XLE Grammar Development and adequacy is attained for these constructions.

In the cases at hand, the object of the matrix clause is always overt and functions as the controller of the subject of the *na* complement. In the literature, there is a general agreement that *na* subordinate clauses display the semantic properties characteristic of control infinitives. Varlokosta (1994) demonstrated, for one, that the subject of *na* clauses systematically is assigned de se readings, just like control subjects in English. However, there is no consensus on how to define the verb class licensing control constructions (Iatridou 1993, Varlokosta 1994, Alexiadou and Anagnostopoulou 1999, Philippaki and Catsimali 1999, Spyropoulos 2007, Kotzoglou and Papangeli 2007, Beys 2007, Roussou 2009). We pursue this issue in a corpus study based on the Hellenic National Corpus (HNC; Hatzigeorgiu et al., 2000). All these predicates are exhaustive control verbs.

In Greek the object of the matrix clause is always overt and functions as the controller of the subject of the *na* complement. The object controller can be marked either by different cases or it can be embedded within a PP[4]. In (4a) the object of the matrix verb is expressed in accusative case ('th Maria') while in (4b) in genitive case ('ths Marias'). In (4c) the object of the matrix verb is embedded within a se-PP ('sth Maria'), which is considered to be an oblique argument (OBL-TO). As can be observed there is no featural identity between the controller and the controllee which is always covert and marked with nominative case ('h Maria').

3. Nouns and determiners in Greek should have number, gender and case agreement on them. We only gloss them for case since number and gender are not important for our study.
4. Its EOC verb allows for different subcategorization frames. See table 3 for a detailed picture of the structures supported by each verb.

(4) a. O Kostas mathainei th Maria *na*
The-DEF Kostas-NOM teach-3SG the-DEF Maria-ACC to-COMPL
milaei (h Maria) Agglika.
speak-3SG (the –DEF Maria-NOM) English-ACC
'Kostas teaches Maria to speak English.'

b. O Kostas mathainei ths Marias
The-DEF Kostas-NOM teach-3SG the-DEF Marias-GEN
na milaei (h Maria) Agglika.
to-COMPL speak-3SG (the –DEF Maria-NOM) English-ACC
'Kostas teaches Maria to speak English.'

c. O Kostas mathainei sth Maria
The-DEF Kostas-NOM teaches-3SG se-PREP Maria-ACC
na milaei (the-DEF Maria- NOM) Agglika.
to-COMPL speak-3SG (h Maria) English-ACC.
'Kostas teaches Maria to speak English.'

In the following section, we discuss some of the properties of the *na* clauses. In §3 we illustrate the corpus retrieved data and the annotation schema followed in this study. §4 presents how control constructions are treated within the LFG Framework. In §5 we present our analysis of EOC in Greek and the implementation of this analysis in our LFG/XLE Grammar. Finally, in §6 we draw our conclusions.

2. The case of *na* clauses

We study control constructions in the case of *na* subordinate clauses. These clauses are associated with controversial linguistic issues such as the syntactic nature of *na*, the subjunctive and the lack or presence of an infinitive in Greek. Firstly, there is no unanimity as to whether *na* is a complementizer or not. Veloudis & Philippaki-Warburton (1984) and Terzi (1992) have analyzed *na* as a subjunctive marker while Tsimpli (1990) analyzed *na* as a modality marker that selects agreement and untensed phrases. On the other hand, Agouraki (1991) and Tsoulas (1993) claim that *na* is a complementizer and its meaning depends on the time reference of the main verb. A recent view that reconciles the two approaches has been proposed by Roussou (2000), within a Split-CP framework. Fiotaki (2014) treats *na* as a complementizer that introduces main and subordinate clauses expressing different modalities.

In general it is not clear whether a uniform semantics for *na* clauses is possible, and this raises multiple questions not only for Greek but crosslinguistically (Quer 2009). In Modern Greek, the indicative and the subjunctive mood have no different morphological endings although the moods exist (Mozer 2009). In the traditional Greek grammar the subjunctive mood can be found in Simple Present (e.g *na* paizw), Simple Past (e.g *na* paiksw) and Present Perfect (e.g *na* echw paiksei). The above verb types can also form indicative (5; Triadafullidhs 1941).

(5) a. O Panos mathainei to paidi
The-DEF Panos-NOM teach-3SG the-DEF child-ACC
tou ***na*** **diavazei.**
his-GEN POSS **to-COMPL study-3SG SUBJUNCTIVE**
'Panos teaches his child to study.'

b. O Panos **diavazei** ena
The-DEF Panos-NOM **study-3SG INDICATIVE** a-INDEF
vivlio.
book-ACC
'Panos studies a book.'

If we study *na* subordinate clauses from a syntax-semantics point of view, we have to deal with the dependent verbal form (e.g. paiksw) with no formal mood features, the so called PNP (Holton et al. 2012, Tsangalidis 2002, Giannakidou 2009, Lekakou & Nilsen 2009). The distribution of PNP triggers debate among linguists. PNP is not annotated by default in the feature TENSE (morphological tense) but it instantiates the combination of perfective and non past (Tsangalidis 1999, Giannakidou 2009, Iatridou et al. 2002). This verb form can occur under the subjunctive marker *na,* but also under the future/modal particle *tha*, the conditional *an*, the optative *as* and sometimes under some temporal connectives, for instance *prin 'before', otan 'when'* (6; Giannakidou, 2007). All of these are able to shift forward the evaluation time of the verb they embed.

(6) a. O Panos epeise to Giorgo
The-DEF Panos-NOM persuade-3SG the-DEF George-ACC
na paiksei-PNP volleu
to-COMPL play-3SG PNP volleyball-ACC
'Panos persuaded George to play volleyball.'

b. O Giorgos tha paiksei-PNP
The-DEF Giorgos-NOM will-PART play-3SG PNP
volleu.
volleyball-ACC
'George will play volleyball.'

All the above issues raise the question that regards the presence of tense in the verb head of *na* clause (Tsimpli 1990, Aggouraki 1991, von Stechow 1995). Controlled *na* subordinate clauses are generally accepted to be untensed (Fiotaki and Markantonatou 2004).

In our study we follow Fiotaki and Markantonatou (2004) in that:

- *Na* is a complementizer.
- Verb head of *na* clause is marked with indicative mood.
- Verb head of *na* clause is untensed and is marked with the feature TENSE by default.

3. The Corpus study

In the literature there is not a recorded list of verbs that allow EOC in Greek. Trying to define these verbs we studied at first 18 verbs (7) that in general are considered to take part in control constructions (Iatridou 1993, Varlokosta 1994, Alexiadou and Anagnostopoulou 1999, Philippaki and Catsimali 1999, Spyropoulos 2007, Kotzoglou and Papangeli 2007, Beys 2007, Roussou 2009).

(7) lew 'tell', epitrepw 'allow', sumvouleuw 'advice', upochrewnw 'obligate', diatazw 'order', entharrunw 'encourage', mathainw 'teach, peithw 'persuade', dokimazw 'try', aphhnw 'let', apagoreuw 'forbid', empodizw 'prevent', deichnw 'show', thumamai 'remember', voithw 'help', protrepw 'urge', susthnw 'recommend', parakolouthw 'watch', chairomai 'be glad'.

The data were drawn from the HNC which is a balanced corpus of written Modern Greek texts developed by the Institute for Language and Speech Processing (ILSP). It currently contains about 50.000.000 words and is constantly being updated. HNC consists of texts from several media which provide evidence for the current use of Modern Greek since texts rich in idiomatic or dialectic forms are excluded (Hatzigeorgiu et al., 2000). It allows lemma searches. For every lemma it returns up to 2000 sentences. It also gives the user the ability to make queries for specific words, lemmata, parts of speech and up to three combinations of all the above in which users can specify the distance among lexical items.

In our study we searched the verbs mentioned above as lemmas combined with the particle *na*. The specified distance between the verb and *na* was defined as up to 5 words. For all the above verbs HNC provided us with 19.998 sentences in total. We examined these sentences to find which ones contained the structure we were interested in. We came up with 7 verbs that allow EOC constructions (1)[5] HNC returned 9054 sentences for these verbs, out of which 4705 contained the relevant structure (V + OBJ + *na* clause). Table 1 shows the precise number of the data retrieved for each verb.

Verb	Sentences in total	Sentences with the relevant structure
empodizw (prevent)	959	768 (80%)
protrepw(urge)	316	281 (88,9%)
epitrepw (allow)	2000	1390 (69,5%)
apagoreuw (forbid)	708	344 (48,6%)
peithw (persuade)	1249	660 (52,8%)
voithw (help)	2000	1153 (57,6%)
mathainw (teach)	1822	109 (5,98%)

Table 1. HNC data

Since this study aims to enrich the Greek XLE Grammar that is being developed, we decided to follow the unified analysis of the tense system and the subjunctive mood as it is proposed by Fiotaki and Markantonatou (2014). According to them, the traditional analysis cannot capture the entirety of the Greek verb types, so a multilevel analysis is needed. Their proposal provides a Greek verbal tense system that models tense usage in main clauses and *na* subordinate ones. The tense system was adopted in the spirit of Reichenbach (1947) who introduces three abstract time points: Speech time (S), Event time (E), Reference time (R). The features of this tense system are described below:

- Linguistic Time (LING_TIME; TENSE as proposed by ParGram) models the relation between S and R with values +/- PAST.
- Time Frame (T_FR) encodes the relation between R and E with values N(ot)IDEN(tical) and IDEN(tical).
- Anticipation (ANTIC; FUTURE as proposed by ParGram) models the presence of the particle *tha*.
- Telicity expresses the grammatical aspect with values Perfective (PE) and Imperfective (IP).

The overall system of features is presented in Table 2. In the second column all the verb types attested in main declarative clauses are presented along

5. HNC did not provide us enough data for the verbs parakolouthw 'watch' and voithw 'help'. So for the time being we cannot make a certain claim for these verbs, although as native speakers we tend to assume that these verbs do not allow EOC, since we can easily come up with counterexamples.

with the English gloss (column 3). The next column provides the traditional analysis of the tense system in Greek by assigning the value (+/- Past) for each verb type. The next four columns capture the features of the tense system as described above.

Verb type	Greek Form	English Gloss	Tr.Anal	LING_TIME	T_FR	TEL	ANTIC
1	paizw	play/be playing	–PAST	–PAST	IDEN	IP	–
2	epaiza	was playing	+PAST	+PAST	IDEN	IP	–
3	epaiksa	played	+PAST	+PAST	IDEN	PE	–
4	tha paizw	will be playing	–PAST	–PAST	IDEN	IP	+
5	tha paiksw	will play	–PAST	–PAST	IDEN	PE	+
6	tha epaiza	would play	Ø	+PAST	IDEN	IP	+
7	echw paiksei	have played	+PAST	+PAST	NIDEN	PE	–
8	eicha paiksei	had played	+PAST	+PAST	NIDEN	PE	–
9	tha echw paiksei	will have played	Ø	–PAST	NIDEN	PE	+
10	tha eicha paiksei	would have played	Ø	+PAST	NIDEN	PE	+

Table 2. Analysis of verbal tenses in Modern Greek

Having in mind the above tense system, we created annotation labels for the verb of the matrix and the subordinate clause. We also needed labels for the type and the case of the matrix object as this is the controller. So, the data retrieved were annotated following the schema below:

- The labels **NON_PAST**, **PAST**, **FUTURE**, **FUTURE_+PAST** and **PNP** are used for both the verbs of the matrix and the *na* subordinate clause. These labels correspond to the temporal properties of the verb types based on the value of the feature LING_TIME (Table 2). Future tenses needed to be distinguished (labels FUTURE and FUTURE_+PAST) since the complementizer *na* stands in complementary distribution with the future particle *tha* (see Table 2 feature Anticipation). The label PNP was used for all the verb types corresponding to 'na paiksw' (section 2).

  Table 3 represents the labels and their correspondent temporal properties according to Table 2. The first column presents the labels used in the annotation schema. The second and third column present the temporal properties of each label as described above.

LABELS	LING_TIME	ANTIC
NON_PAST	–PAST	–
PAST	+PAST	–
FUTURE	–PAST	+
FUTURE_+PAST	+PAST	+

Table 3. The labels and their correspondence to the temporal properties of the verb types.

- The labels **ACC (OBJ)**, **GEN (OBJ)** and **PP (OBL-TO)** are used for the object of the matrix clause.

(8) represents an annotated example.

(8) mas **(ACC(OBJ))** empodizei **(NON_PAST)** *na*
us-OBJ prevents-3SG to-COMPL
epituchoume **(NON_PAST)** tous stochous.
achieve-3sg the-DEF goals-ACC
'It prevents us from achieving the goals.'

As you can see, there is not a direct correspondence between the tense features described in Table 2 and the labels used in our annotation schema. This is due to the fact that we aimed in designing functional templates for our XLE/LFG grammar. So, we decided to generalize the features that describe the temporal properties of the verbs. This generalization made the process of annotation faster without leading to ambiguity or loss of information from the adopted tense system. Also, this simplest form of the tense system can be used in the future for the annotation of examples concerning various phenomena with the exception of phenomena affected from the temporal property "Telicity" since this feature is not included in this schema.

This process of annotation gave us a clear picture of the structures supported by each verb. In Table 4 the three more frequent structures supported by each verb are presented. The first column presents the annotated verbs along with their english translation. Next the structures supported by each verb are given (column 2). The temporal properties of the main verb (column 4) and the subordinate verb (column 5) follow. Finally, the overall percentage of each structure is given.

Verb	Syntactic Structure	Temporal properties of the main verb	Temporal properties of the *na* clause verb	Percentage
apagoreuw (forbid)	A. Vmain + ACC (OBJ) / GEN (OBJ) / PP (OBL-TO) + *na* clause	1. NON PAST	1. NON PAST	1. 30,8%
	B. Vmain + GEN (OBJ) / PP (OBL-TO) + *na* clause	1. NON PAST 2. PNP	1. PNP 2. PNP	1. 22,4% 2. 3,5 %
epitrepw (allow)	A. Vmain + GEN (OBJ) / PP (OBL-TO) + *na* clause	1. NON PAST	1. PNP	1. 31%
	B. Vmain + ACC (OBJ) / GEN (OBJ) / PP (OBL-TO) + *na* clause	1. NON PAST 2. FUTURE	1. NON PAST 2. PNP	1. 22% 2. 11,6%
empodizw (prevent)	Vmain + ACC (OBJ) + *na* clause	1. NON PAST 2. PAST 3. NON PAST	1. PNP 2. PNP 3. NON PAST	1. 30,6% 2. 26,9% 3. 15,5%
mathainw (teach)	Vmain + ACC (OBJ) / GEN (OBJ) / PP (OBL-TO) + *na* clause	1. PAST 2. PNP 3. NON PAST	1. NON PAST 2. NON PAST 3. NON PAST	1. 40,3% 2. 28,5% 3. 16,5%
peithw (persuade)	Vmain + ACC (OBJ) + *na* clause	1. PNP 2. PAST 3. PNP	1. PNP 2. PNP 3. NON PAST	1. 65,6% 2. 15,3% 3. 6,8%
protrepw(urge)	Vmain + ACC (OBJ) + *na* clause	1. NON PAST 2. PAST 3. NON PAST	1. PNP 2. PNP 3. NON PAST	1. 38,8% 2. 36% 3. 30,2%
voithw (help)	Vmain + ACC (OBJ) + *na* clause	1. PNP 2. PAST 3. NON PAST	1. PNP 2. PNP 3. PNP	1. 27,9% 2. 22,8% 3. 20,3%

Table 4. Structures supported by each verb.

4. Exhaustive Object Control in the LFG Framework

In this section we study control constructions based on the criterion of the featural identity between the controller and the controllee.

LFG uniformly treats control constructions involving featural identity between controller and controllee as manifestations of functional control (Bresnan 1982). In this case the two functions, the controller and the controllee are allowed to have the same f-structure as their value (Falk 2001). This analysis can treat EOC in English (9) and Greek subject control constructions (10).

(9) Frank persuaded Mary to leave.[6]

(10) H Zwh emathe *na* kolumpaei.
the-DEF Zwh-NOM learned-3SG to-COMPL swim-3SG
'Zoi learned to swim.'

In this analysis, the infinitive 'to leave' in (9) and the embedded clause 'na kolumpaei' (10) are treated as an XCOMP argument of the matrix verb. This is the case where the person who was persuaded by Frank and the person who left must be one and the same person (9).

Contrary to the EOC in English (9) in Greek EOC, the controller and the controllee differ in Case features (2). In (9) the controller ('Mary') and the controllee (unexpressed subject) are both in accusative case, while in (2) the controller ('ths Marias') is in genitive case and the controllee (unexpressed subject) bears nomitave case (see section 5).

(2) O Nikos apagoreuei ths Marias [***na***
The-DEF Nikos-NOM forbid-3SG the-DEF Maria-GEN [to-COMPL
erthei].
come-3SG]
'Nikos forbids Maria to come.'

In this case there is an anaphoric link between the unexpressed subject of the *na* clause and the object 'ths Marias'. This control relation is called anaphoric control (Bresnan 1982, Falk 2001). Such occurrences of control are argued to occur with COMP. Under this analysis, the subject of the COMP is a PRO anaphorically related with the object of the matrix clause (Bresnan 1982, Darlymple 2001).

Another way of treating control phenomena is subsumption, which is a way of modelling asymmetric information. Zaenen and Kaplan use subsumption to treat partial VP fronting in German (2002) and subject

6. Examples like (9) can also be treated using anaphoric control (Dalrymple 2001, Falk 2001).

inversion in French (2003). Sells (2006) models forward and backward control and raising structures using subsumption. Subsumption allows us to constrain the information in the f-structure level: 'subsumption establishes an ordering relation between two units of information, stating that the one subsuming the other contains less information (or is less specific or more general) than the one that is subsumed' (11; Zaenen and Kaplan 2002)

(11)

Definition of Subsumption: $f \sqsubseteq g$ iff
f and g are the same symbol or semantic form, or
f and g are both f-structures, $\mathrm{Dom}(f) \subseteq \mathrm{Dom}(g)$, and $(f\ a) \sqsubseteq (g\ a)$ for all $a \in \mathrm{Dom}(f)$, or
f and g are both sets and every element of $f \sqsubseteq$ some element of g

$$f = \begin{bmatrix} \mathrm{A} & \begin{bmatrix} \mathrm{C} & + \end{bmatrix} \end{bmatrix} \sqsubseteq \begin{bmatrix} \mathrm{A} & \begin{bmatrix} \mathrm{C} & + \\ \mathrm{D} & - \end{bmatrix} \\ \mathrm{B} & \mathrm{E} \end{bmatrix} = g$$

Subsumption deals with case mismatch by making use of the restriction mechanism. Although using subsumption seems a viable solution for modeling EOC in Greek, for the time being we think that these cases are better treated using anaphoric control since Greek is a language which uses extensive morphological case marking. The restriction of case may affect the expressivity and the efficiency of our LFG/XLE Grammar. Also, by definition (11) subsumption contrasts with the one of the basic points of our proposal (section 5). According to our analysis PRO and the controller are two different semantic forms and thus PRO cannot be subsumed by the controller since their f-structures contain different elements. We follow Falk (2001:141) in that the controller and the controllee are "both considered to be thematic arguments of their respective verbs, and so they must be two distinct D-structure elements".

5. Our proposal

We propose to treat EOC as an instance of anaphoric control in the sense of Bresnan (1982), hence to analyze *na* subordinate clauses as implicating COMP functions. Given that COMP is also used as a formal device to model partial control one could hypothesize that EOC in Greek should admit partial control (Landau 2013; Pearson 2015). This prediction is not at all confirmed, as shown by the ill-formedness of the partial control structure in (12) which is not allowed with an EOC predicate (in this case mathainw 'teach'). (13) demonstrates that partial control is attested with subject control predicates.

(12) *O Giannis mathainei th Maria
the-DEF John-NOM teaches-3SG the-DEF Maria- ACC
na mazeuei mazi tomates/
to-COMPL pick up-3SG together- ADV tomatoes-ACC/
na sunanththoun.
to-COMPL meet-3PL
'John teaches Maria to pick up tomatoes together/to meet.'

(13) O Giannis proteine *na* sunanththoun.
the-DEF John-NOM proposed-3SG to-COMPL meet-3PL
'John proposed to meet.'

Interestingly though, the absence of collective readings for the EOC subject correlates with the absence of a second property which has been found to be characteristic of partial control predicates, i.e. temporal independence of the embedded clause. Combining a future oriented embedded adverbial with past matrix predicate leads to ill-formed results in the case of EOC verbs (14a), while these structures are acceptable when combined with partial control verbs (14b; on the relation between partial control and tense see Landau 2013; Pearson 2015).

(14) a. *Chthes o Giannis emathe th Maria
Yesterday the-DEF John-NOM taught-3SG the-DEF Maria- ACC
na grafei aurio.
to-COMPL write-3SG tomorrow- ADV.
'*Yesterday, John taught Maria to write tomorrow.'

b. *Chthes o Giannis proteine ths Marias
Yesterday the-DEF John-NOM suggested-3SG the-DEF Maria-GEN
na fugei aurio.
to-COMPL leave-3SG tomorrow- ADV.
'*Yesterday, John suggested Maria to live tomorrow.'

As mentioned above, we analyze EOC as an instance of anaphoric control. Bresnan (1982) argues that anaphoric control requires the presence of PRO, which is expressed only in f-structure. We propose that in Greek EOC constructions the subject of the *na* subordinate clause is a PRO[7] anaphorically controlled by the object of the matrix verb (15).

7. Although written in caps, this PRO is not identical to GB's big pro.

(15) O Kostas mathainei th Maria
The-DEF Kostas-NOM teach-3SG the-DEF Maria-ACC
[*na* PRO milaei Agglika].
[to-COMPL PRO speak-3SG English-ACC]
'Kostas teaches Maria to speak English.'

We follow Bresnan (1982) in that PRO is a semantic form and thus should be introduced in the lexicon. Specifically, it is introduced in the lexical entry of the governing verb. We claim that since PRO is the subject of the *na* subordinate clause it is case marked with nominative since:

i. In non-control cases the subject of the *na* subordinate clause is overtly expressed and bears nominative case (16). As we can conclude the covert subject of the *na* subordinate clause (the cotrollee) always bears nominative case. As Landau points out 'whenever a language provides means to detect the case of PRO it is identical to the case that a lexical DP would have been in the same position' (Landau 2013:104).

(16) O Giorgos eipe *na* kleisei
The-DEF Giorgos-SUBJ/NOM tell-3SG to-COMPL close-3SG
o Dimitris to parathuro.
the-DEF Dimitris-SUBJ/NOM the-DEF window-ACC
'George told Dimitris to close the window.'

ii. The embedded subject modifier of the covert subject appears in nominative case and not in accusative (17; Spyropoulos 2007, Kotzoglou and Papangeli 2007, Beys 2007).

(17) H Maria epeise to Gianni
The-DEF Maria-NOM persuade-3SG the-DEF Gianni-ACC
na fugei teleutaios / *teleutaio.
to-COMP leave-3SG last-MOD-NOM / * last MOD-ACC
'Maria persuaded John to leave last.'

iii. Although there is a controllee, an overt pronoun in nominative case can be licensed in *na* subordinate clause along with the conjunction "kai" for emphasis (18). In Greek this is a standard way to do emphasis. This pronoun is coreferential with the object of the matrix verb.

(18) Epeisa to Gianni *na* erthei
Persuade-3SG the-DEF Giannni-ACC/OBJ to-COMPL come-3SG
kai autos sto parti.
and-CONJ he-PRN/NOM to-PREP the-DEF party-ACC
'I persuaded John to (he) come to the party.'

To sum up, we propose that the subject of the embedded *na* clause in EOC is a semantic PRO anaphorically related to the object of the matrix verb that bears nominative case.

This anaphoric relation between PRO and its antecedent (object or oblique) must be expressed in the f-structure. For this reason, we introduce the new feature “Anaphorically_controlled_by” (ANAPH_C_BY) with the value OBJ (4a-b) or OBL-TO (4c), signaling that the predicates in (11) are not only marked for anaphoric control but also include a lexically required feature restricting arguments to a particular type of *na* complements (de se properties). As a result, there are two subtypes of COMP, one for clausal arguments that have their own overt or non overt subject and one for clausal arguments that have their subject anaphorically controlled by the object of the matrix verb.

(19) is a representative example of how EOC is treated in our LFG/XLE Grammar. In our effort to parse the corpus retrieved examples we defined two templates (20) that assign the allowed syntactic structures. The lexical entry for the verb *apagoreuw* ‘forbid’ is (21). The output of the parsed example is illustrated in (22).

(19) H Maria apagoreuei sto
the-DEF Maria-NOM forbid-3SG se-PREP the-DEF
paidi ths *na* paizei mpala.
child-ACC her-GEN POSS to-COMPL play-3SG ball-ACC
‘Mary forbids her child to play ball.’

(20) Templates

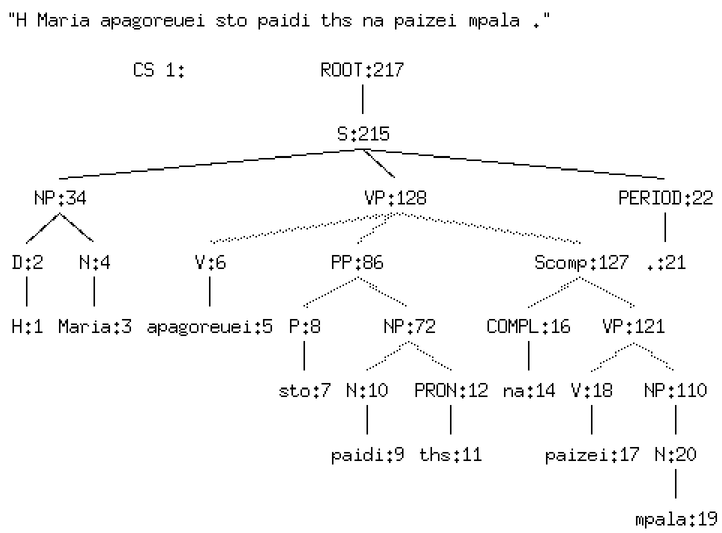
a. V-SUBJ-OBJ-COMP(P) = "closed comp verbs with subject anaphorically controlled by object"
(^ PRED) = 'P<(^ SUBJ)(^OBJ)(^ COMP)>'
(^ COMP SUBJ PRED) = 'PRO'
(^ COMP SUBJ CASE) = NOM
(^ COMP SUBJ ANAPH_C_BY)= OBJ.

b. V-SUBJ-OBL-TO-COMP(P) = "closed comp verbs with subject anaphorically controlled by OBL-TO"
(^ PRED) = 'P<(^ SUBJ)(^ OBL-TO)(^ COMP)>'
(^ COMP SUBJ PRED) = 'PRO'
(^ COMP SUBJ CASE) = NOM
(^ COMP SUBJ ANAPH_C_BY)= OBL-TO.

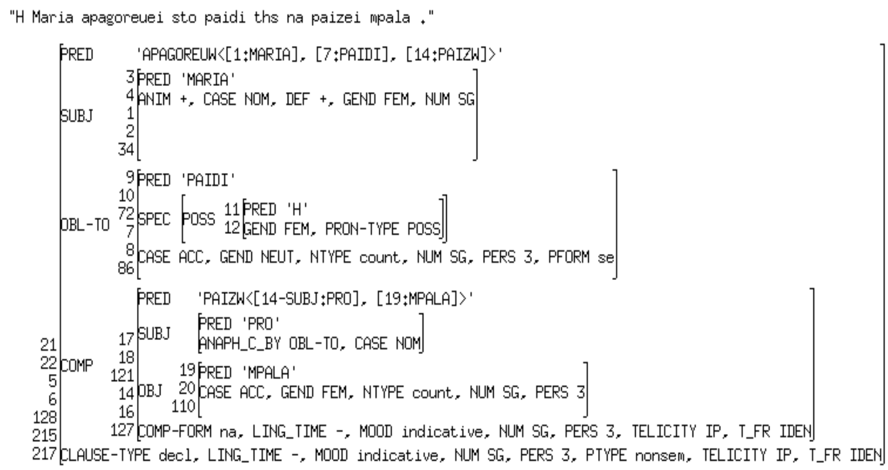
(21) Lexical entry

apagoreuei V * {@(OPT-TRANS APAGOREUW)
|@(V-SUBJ-OBJ-COMP APAGOREUW)
(^ COMP-FORM) = na
|@(V-SUBJ-OBL-TO-COMP APAGOREUW)
(^ COMP-FORM) = na}
@(TENSE -)
@(T_FR IDEN)
@(TELICITY IP)
@(MOOD indicative)
@(PERS 3)
@(NUM SG).

(22a) c-structure



(22b) f-structure



We have integrated the presented analysis of EOC into the fragment of the LFG/XLE Greek grammar. In order to measure the grammar efficiency of the proposed analysis we created a test suite which derived from the annotated corpus and contains 50 sentences per verb. Out of the 350 sentences of the test suite 236 are parsed. Some of the sentences are not parsed because they contain complex constructions not yet covered by our grammar and not due to flaws of our proposed analysis.

6. Conclusions

This paper demonstrates how EOC in Greek can be formalized in the LFG/XLE Grammar using anaphoric control. The proposed analysis allows for the case of the controller and the controllee to be overtly expressed in the f-structure. This expressivity is important for the flexibility of our newly developed grammar which should take into consideration other phenomena in which case seems to play an important role such as coordination.

The described annotated corpus can be used for the study and grammar modeling of the problematic issues related to *na* clauses such as PNP structures. Furthermore, this data could be the base for the study of coordination in *na* clauses.

References

Agouraki, Yoryia. 1991. "A Modern Greek complementizer and its significance for UG." UCL Working Papers in Linguistics 3:1–24.

Alexiadou, Artemis and Anagnostopoulou, Elena. 1999. Raising without infinitives and the nature of agreement. In Proceedings of the 18th WCCFL.

Beys, Michael. 2007. Subject-to-Object Raising Constructions in Modern Greek. In *Proceedings of* the 18th International Symposium on Theoretical and Applied Linguistics. Thessaloniki, Greece.

Beys, Michael. 2007. Subject-to-Object Raising Constructions in Modern Greek. In *Proceedings of* the 18th International Symposium on Theoretical and Applied Linguistics. Thessaloniki, Greece.

Bresnan, Joan. 1982. Control and complementation. In The mental representation of grammatical relations, ed. Joan Bresnan, 282–390. Cambridge, MA: MIT Press.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. Volume 34 of Syntax and Semantics. New York: Academic Press.

Dalrymple, Mary and King, Tracy Holloway. Missing-object constructions. In *Proceedings of the LFG00 Conference*, eds. Miriam Butt and Tracy Holloway King. Stanford: CSLI Publications.

Falk, Yehuda. 2001. *Lexical-Functional Grammar*. Stanford: CSLI Publications.

Fiotaki, Alexandra. 2014. Study of *na* and *oti* complementizer in Modern Greek: an LFG/XLE treatment. Master Thesis. University of Athens.

Fiotaki, Alexandra and Markantonatou, Stella. 2014. "Modern Greek Tense in main and να- subordinated clauses: an LFG/XLE treatment". In Miriam Butt & Tracy Holloway- King (eds), *Proceedings of the LFG14 Conference.* Stanford: CSLI Publications, p.221-240.

Giannakidou, Anastasia. 2009. The dependency of the subjunctive revisited: temporal semantics and polarity. *Lingua*, special issue on *Mood* ed. by Josep Quer.

Hatzigeorgiou, Nick, Gavrilidou, Maria, Piperidis, Stelios, Carayannis, George, Papakostopoulou, Anastasia, Spiliotopoulou, Athanassia et al.

(2000). “Design and implementation of the online ILSP corpus”. Proceedings of the Second International Conference of Language Resources and Evaluation (LREC), vol. 3, 1737-1740.

Holton, David, Mackridge, Peter, Philippaki-Warburton, Irene, Spyropoulos, Vasileios. 2012. Greek: A Comprehensive Grammar, second ed. Routledge, London/ New York.

Iatridou, Sabine. 1988/ 1993. On nominative case assignment and few related things. *MITWPL* 19.

Kotzoglou, George and Papangeli, Dimitra. 2007. Not really ECM, not exactly control: The ‘quasi-ECM’ construction in Greek. In *New horizons in the analysis of control and raising*, ed. W. Davies and S. Dubinsky: 111–132. New York: Springer

Landau, Idan. 2002. The scale of finiteness and the calculus of control. Master’s thesis, Ben Gurion University.

Landau, Idan. 2013. *Control in Generative Grammar: A research companion*. Cambridge: Cambridge University Press.

Lekakou, Marika, and Nilsen, Øystein, 2009. What aspect can tell us about the future of MUST. In: Baltazani, M., Giannakis, G., Tsangalidis, T., Xydopoulos, G. (Eds.), *Proceedings of 8th ICGL*. University of Ioannina, Ioannina. pp. 261--276

Mozer, Amalia. 2009. Apopsi kai xronos stin istoria tis Ellinikis [Aspect and Tense in the History of Greek]. Parousia Monograph n. 77. Athens: National and Kapodistrian University of Athens.

Pearson, Hazel. 2015. The semantics of partial control. Natural Language & Linguistic Theory. Online publication date: 3-Nov-2015.

Philippaki-Warburton, Irene and Catsimali, Georgia. 1999. On control in Greek . *Studies in Greek Syntax*, ed. Artemis Alexiadou, Geoffrey Horrocks & Melita Stavrou, 153-168. Dordrecht: Kluwer.

Philippaki-Warburton, Irene. 2004. Peremfatikotita ke ipotaktiki: anazitondas to xameno aparemfato se dhomes elegxou. In *Proceedings of* the 6th International Conference of Greek Linguistics. Crete: University of Crete.

Quer, Josep. 2009. Twists of Mood: The Distribution and Interpretation of indicative and subjunctive. In *Lingua* 119(2): Special issue on Mood, ed. Josep Quer, 1779-1968.

Reichenbach, Hans. 1947. The Tenses of Verb. In Elements of Symbolic Logic. New York: Macmillan Co.

Roussou, Anna. 2000. On the left periphery; modal particles and complementizers. *Journal of Greek Linguistics* 1: 63-93.

Roussou, Anna. 2009. In the mood for control. *In Lingua* 119: Special Issue *Mood*, ed. J. Quer.

Sells, Peter. 2006. Using subsumption rather than equality in functional control. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of LFG06*, Stanford: CSLI Publications.

Spyropoulos, Vasileios. 2007. Finiteness and control in Greek. In *New horizons in the analysis of control and raising*, ed. W. Davies and S. Dubinsky: 159–183. New York: Springer.

Terzi, Arhonto. 1992. PRO in Finite Clauses. A Study of the Infectional Heads of the Balkan Languages, PhD. CUNY Graduate Center.

Tsoulas, George. 1993. Remarks on the structure and interpretation of na-clauses. Stud. Greek Linguist. 14, 191- 206.

Tsangalidis, Anastasios. 1999. Will and tha: a comparative study of the category future. Thessaloniki: University Studio Press.

Tsangalidis, Anastasios. 2002. On the ‘tenses’ of the Modern Greek verb. Studies in Greek Linguistics: 22, 647-658.

Tsimpli, lanthi -Maria. 1990. The clause structure and word order of Modern Greek. UCL Working Papers in Linguistics, 2, 220−258.

Triadafillidis, Manolis. 1941. *Neoellhniki Grammatiki*. Athina

Varlokosta, Spyridoula. 1994. Issues on Modern Greek Sentential Complementation. Doctoral dissertation, University of Maryland.

Veloudis, John. Philippaki-Warburton, Irene, 1983. I ipotaktiki sta Nea Ellinika. Stud. Greek Linguist. 4, 151-168.

von Stechow, Armin. 1995, ‘On the Proper Treatment of Tense’, in M. Simons and T. Galloway (eds.), Proceedings of SALT V, CLC Publications, Cornell University, Ithaca, NY, 362-386.

Zaenen, Annie and Kaplan, Ron. 2002. Subsumption and equality: German partial fronting in LFG. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of LFG02*, Stanford: CSLI Publications.

Zaenen, Annie and Kaplan, Ronald M. 2003. Subject inversion in French. In Claire Beyssade, Olivier Bonami, Patricia Cabrede Hofherr and Francis Corblin (eds.), *Emprical Issues in Syntax and Semantics*, volume 4, pages 190–205, Paris: Presses Universitaires de Paris Sorbonne.

Second-position clitics and the syntax-phonology interface: The case of Ancient Greek

David M. Goldstein
University of California, Los Angeles

Dag T. T. Haug
University of Oslo

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 297–317

Keywords: Ancient Greek, second-position clitics, syntax-prosody interface, multiple context-free grammars

Goldstein, David M. & Dag T. T. Haug. 2016. Second-position clitics and the syntax-phonology interface: The case of Ancient Greek. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 297–317. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.16.

Abstract

In this paper we discuss second position clitics in ancient Greek, which show a remarkable ability to break up syntactic constituents. We argue against attempts to capture such data in terms of a mismatch between c-structure yield and surface string and instead propose to enrich c-structure by using a multiple context free grammar with explicit yield functions rather than an ordinary CFG.

1 Introduction

Second-position (2P) clitics have proven notoriously challenging for syntactic theory, because their distribution requires reference to both syntactic and prosodic constituents, as illustrated by the following example from ancient Greek ('=' marks prosodic dependence):

(1) *2P distribution*

(apò taúteːs)$_{\omega}$***=gár=sp***^{h}***i*** *tẽːs* *mák*^{h}*eːs* ...
from MED.F.GEN.SG=EXPL=3PL.DAT ART.F.GEN.SG battle.F.GEN.SG
kateúk^{h}*etai* *ho* *kéːruks*
pray.PRES.IND.MP.3SG ART.M.NOM.SG herald.M.NOM.SG
ho *At*^{h}*eːnaĩos* *háma* *te*
ART.M.NOM.SG Athenian.M.NOM.SG together.ADV CONJ
At^{h}*eːnáioisi* *légoːn*
Athenian.M.DAT.PL speak.PTCP.PRES.ACT.M.NOM.SG
gínest^{h}*ai* *tà* *agat*^{h}*à* *kaì*
happen.INF.PRES.MP ART.N.ACC.PL good.N.ACC.PL CONJ
Plataieũsi.
Plataean.M.DAT.PL

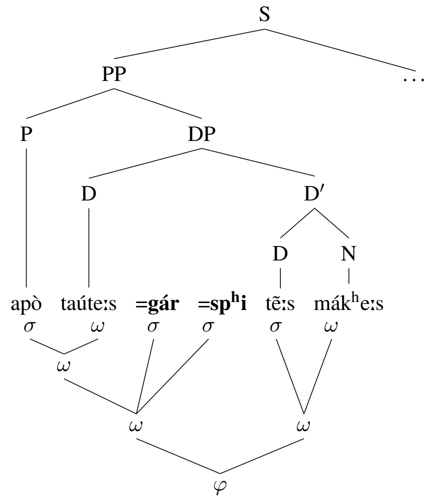
'Since this battle ..., the Athenian herald prays that good things befall the Athenians and Plataeans together, when the Athenians conduct their sacrifices at the festivals that occur every four years.' (Hdt. 6.111.2)[1]

The enclitics *gár* and *sp*^{h}*i* are hosted by the first prosodic word in the clause, *(apò taúteːs)*$_{\omega}$ 'from this'. As the preposition and demonstrative do not form a syntactic constituent, a mismatch between syntax and prosody results:

[†]For helpful discussion and critical feedback, we are grateful to the audience at HeadLex2016, in particular Ron Kaplan and Tracy King, as well as the critical feedback of two anonymous reviewers.

[1]The data for this paper come primarily from Herodotus' *Histories*, a corpus of 189,489 tokens written in the Ionic dialect of classical Greek in the 5th century BCE. Section 2.2 introduces some metrical data from fifth century BCE Attic drama. Our transliteration of the Greek is graphic, not phonological/phonetic.

(2) *Syntax-prosody mismatch*



This mismatch raises fundamental questions about the architecture of the grammar. First, what is the division of labor between syntax and phonology—can syntax see phonological properties? Below we argue that no information passing between syntax and prosody is needed beyond the ordinary interaction of projections in the LFG architecture, which means that a well-formed sentence must simultaneously satisfy syntactic and prosodic constraints. Syntax must therefore position the clitic where it gets a prosodically acceptable host.

Second, what are the capacities of c-structure and f-structure? The root of the problem in example (2) is that clitics appear in surface positions where they cannot be assigned a grammatical function. This inability is a direct consequence of the linguistic and formal differences between c- and f-structure. Linguistically, c-structure deals with word order and constituency, whereas f-structure deals with abstract syntactic relations. Formally, c-structure can only handle phenomena within the locality domain of a context-free grammar (CFG), i.e. the one level tree corresponding to a rule whereas f-structure can handle phenomena at an unbounded distance. What this amounts to is the claim that there are no non-local word order or constituency facts. But 2P clitics involve precisely non-local constituency.

To deal with non-local constituents, previous accounts have surrendered core assumptions of LFG by introducing idiosyncratic constituents, such as CL and CCL (i.e. syntactic categories 'clitic' and 'clitic cluster', Bögel et al. 2010, Ćavar & Seiss 2011, Lowe 2011); by relying on cross-derivational comparison (i.e., Output-Output Correspondence, Lowe 2015) under Optimality Theory; by permitting mismatch between the c-structure and the prosodic or syntacic string (Bögel 2015, Lowe 2015); or by allowing prosodic markers into syntax Bögel et al. (2009, 2010). We return to some of these proposals in greater detail below in section 4.

By contrast, we argue that the best way to capture the empirical facts and to maintain the spirit of the LFG architecture is to modify the division of labor between the c- and f-structures. Specifically, we increase the power of the c-structure

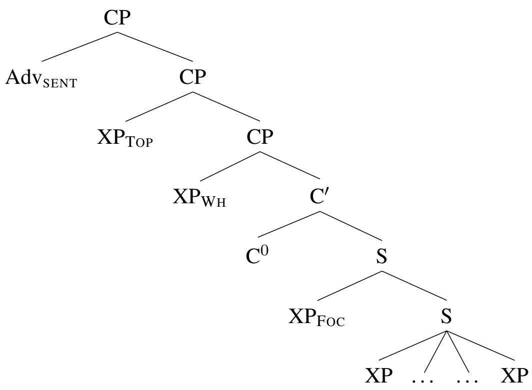


Figure 1: AG clause structure

by moving to a 2-multiple context-free grammar (2-MCFG). This move enables us to capture two crucial insights into the behavior of 2P clitics in ancient Greek. First, there are no clitic specific syntactic rules. Second, 2P clitics do not occupy dedicated c-structure positions. This results from the prosaic fact that syntactic constituents need not map onto identical prosodic constituents. 2P clitics only subcategorize for a syntactic domain and host properties.

2 Data

2.1 Background on ancient Greek

Ancient Greek is one of the earliest attested Indo-European languages. It is a language with rich nominal and verbal morphology and "free" word order, which includes remarkable discontinuities (Devine & Stephens 1999, Agbayani & Golston 2010a). Whether or not Greek has an underlying configurational word order has been debated for over a century, with the main contenders being OV and VO. What is clear is that Greek relies heavily on surface word order for encoding pragmatic properties of the clause, such as information structure (Dik 1995). Since no correlation between surface position and grammatical function has been demonstrated, we assume an exocentric S constituent for the basic clause. Various phrases encoding information structure functions can be adjoined to CP or to S, as shown in Figure 1, which is based on and deviates slightly from the clause structure proposed by Goldstein (2016).

Beginning at the top of the tree, adverbials with sentential scope adjoin at the highest level, with topicalized phrases occuring next. Both phrases are adjoined to CP. We assume that *wh*-words occupy Spec,CP. A focus projection can be adjoined to S, which broadly speaking encodes contrastive focus (Goldstein 2016 refers to it as NON-MONOTONIC FOCUS). This gives us the c-structure rules in (3).[2]

[2] The order of adjunctions to CP in Figure 1 does not follow from these rules: a sentential adverb

(3) *C-structure rules*

Clause-level rules

$$
\begin{array}{lcccc}
\mathrm{CP} & \rightarrow & \underset{(\uparrow \textsc{udf})=\downarrow}{\mathrm{XP}} & & \underset{\uparrow=\downarrow}{\mathrm{C}'} \\
\mathrm{C}' & \rightarrow & \underset{\uparrow=\downarrow}{\mathrm{C}^0} & & \underset{\uparrow=\downarrow}{\mathrm{S}} \\
\mathrm{S} & \rightarrow & \underset{(\uparrow\textsc{gf})=\downarrow}{\mathrm{XP}^*} & , & \underset{\uparrow=\downarrow}{\mathrm{V}^*}
\end{array}
$$

Adjunction to clausal categories

$$
\begin{array}{lccc}
\mathrm{CP} & \rightarrow & \underset{\uparrow=\downarrow}{\mathrm{AdvP}} & \underset{\uparrow=\downarrow}{\mathrm{CP}} \\
\mathrm{CP} & \rightarrow & \underset{\substack{(\uparrow(\textsc{gf}))=\downarrow \\ (\uparrow_\sigma \textsc{df}) = \textsc{topic}}}{\mathrm{XP}} & \underset{\uparrow=\downarrow}{\mathrm{CP}} \\
\mathrm{S} & \rightarrow & \underset{\substack{(\uparrow(\textsc{gf}))=\downarrow \\ (\uparrow_\sigma \textsc{df}) = \textsc{focus}}}{\mathrm{XP}} & \underset{\uparrow=\downarrow}{\mathrm{S}}
\end{array}
$$

Lexical phrases

$$
\begin{array}{lcccc}
\mathrm{PP} & \rightarrow & \underset{\uparrow=\downarrow}{\mathrm{P}} & & \underset{(\uparrow \textsc{obj})=\downarrow}{\mathrm{DP}} \\
\mathrm{DP} & \rightarrow & \left(\begin{array}{c}\mathrm{D} \\ \uparrow=\downarrow\end{array}\right) & & \underset{\uparrow=\downarrow}{\mathrm{NP}} \\
\mathrm{NP} & \rightarrow & \underset{(\uparrow\textsc{adj}) \in \downarrow}{\mathrm{A}^*} & , & \left(\begin{array}{c}\mathrm{N} \\ \uparrow=\downarrow\end{array}\right) \\
\mathrm{AdvP} & \rightarrow & \underset{\uparrow=\downarrow}{\mathrm{Adv}} & &
\end{array}
$$

The most prominent feature of these rules is the absence of any reference to clitics. In contrast to other accounts (see section 4 below), our analysis posits no clitic-specific syntactic rules. So the grammar handles pronominal clitics just as it does their stressed counterparts, and clitic discourse particles with sentential scope are treated just like sentential adverbials.

Determining the category of second-position clitics is no easy matter. Several recent LFG analyses rely on a category CL (for 'clitic'), while minimalist analyses often posit a category between that of a syntactic head and phrase (e.g., D/P for pronominal clitics). Here we take a different tack and analyze 2P clitics in AG as non-projecting heads (Toivonen 2003). However, unlike Toivonen, we assume that what is special about non-projecting words is simply that they do not project, which correctly predicts that they cannot, e.g., be the targets of adjunction.

2.1.1 Clitics versus postpositives

The philological literature standardly divides the inventory of second-position items into two classes: clitics and postpositives (Chandler 1881, Fraenkel [1933] 1964, Probert 2003). Second-position items without an orthographic accent are clitics (e.g., the third person singular accusative pronoun *min*), while those with an orthographic accent are postpositives, e.g., the modal particle *án*. The idea behind this division seems to be that of true phonological clitics versus "syntactic clitics," that is, words that despite bearing an accent are nevertheless restricted to second position (Devine & Stephens 1994, 303, 352, Dik 1995, 37–38, Lowe 2014).

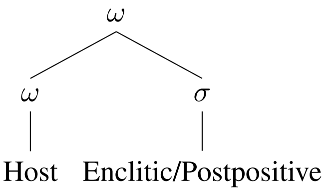
Although we believe that the graphic distinction found among postpositives and enclitics does reflect something prosodically real, we reject the traditional view

must always go higher than a topic. We assume this is related to scope, but it is notoriously difficult to capture adjunction scope in LFG.

for two reasons. First, metrical evidence demonstrates that both postpositives and enclitics exhibit phonological dependence on a host (see further Goldstein 2016, 52–53, 61–65). So they are not syntactic clitics. Second, there is no distributional difference that correlates with the presence or absence of a graphic accent (cf. Taylor 1990, 119, Fortson 2010, 161).

Instead, we argue that the prosodic distinction between postpositives and enclitics is limited to secondary stress assignment. When enclitics and postpositives incorporate with a prosodic word (i.e., their host), they uniformly project a secondary prosodic word (see Anderson 2005 for an overview, including earlier literature):

(4) *Recursive prosodic word*



The secondary prosodic word can trigger a secondary stress, the position of which is determined in one of two ways. The first is via a secondary accentual calculus, whose precise details do not concern us here. Suffice it to say that enclitic incorporation can trigger secondary stress on the host, the clitic, or none at all:

(5) *Secondary stress patterns*

i. *Secondary stress on the host*

(ánt^{h}roːpoi)$_\omega$+tines → *((ánt^{h}roːpoí)$_\omega$=tines)$_\omega$*
man.M.NOM.PL+INDF.C.NOM.PL

ii. *Secondary stress on the clitic*

p^{h}íloi+tines → *((p^{h}íloi)$_\omega$=tinés)$_\omega$*
friend.M.NOM.PL+INDF.C.NOM.PL

iii. *No secondary stress*

(pánta)$_\omega$+sp^{h}i → *((pánta)$_\omega$=sp^{h}i)$_\omega$*
all.N.NOM/ACC.PL+3PL.DAT

The crucial point for our account is that the surface effect of the incorporation of an enclitic is variable. It is this variety that distinguishes enclitics from postpositives. Incorporation of the latter always triggers secondary stress, which is uniformly located on the postpositive itself:

(6) *Fixed secondary stress*
(taúteːs)$_\omega$+gár → ((*taúteːs*)$_\omega$=*gár*)$_\omega$

Secondary stress will always occur on the postpositive, regardless of the prosodic shape of the host.

Before moving on, we call attention to one context in which the combination of a host plus enclitic does not project a secondary prosodic word:

(7) *hoi+gár* → *(hoi=gár)*$_\omega$
ART.M.NOMPL+EXPL

Examples of this type are remarkable because the host of the postpositive is a sub-prosodic word. The definite article is standardly assumed to be proclitic, but together with the postpositive it projects a prosodic word. Evidence that a string such as *hoi=gár* forms a prosodic word comes from its ability to host clausal clitics (see example 8i below; cf. also Goldstein 2016, 76–78). Similar behavior is known from Bilua, a Papuan language of Solomon Islands (see further Anderson 2012).

2.1.2 Clitic domains

The clitic lexicon of AG is larger than that of any other archaic IE language. It encompasses pronouns, verbs, conjunction, and discourse and modal particles. There is no single "second" position in which they all occur (cf. Hale 1987a,b on Sanskrit clitics).[3] Instead, clitics subcategorize for particular syntactic domains, as detailed in Table 1.[4]

DOMAIN	MEMBERS
SENTENCE	{*dé*, *mén*}—*gár*—*õːn*—{*déː*, *dẽːta*}
CLAUSE	*án*—{*kote*, *kou*, *koː*, *koːs*, *kéː(i)*}—*ára*—ACC—DAT—{*eimí*, *p*^{h}*eːmí*}$^?$
PHRASE	*te*—{*dé*, *mén*}—*ge*

Table 1: Clitic domains and the internal ordering of their members

Sentence clitics are invariably discourse connectives marking intersentential relations: we assume they are $\widehat{\text{Adv}}$. Clausal clitics realize grammatical features of the clause: they can be $\widehat{\text{Adv}}$, $\widehat{\text{D}}$ and $\widehat{\text{V}}$. Phrasal clitics realize grammatical features of sub-clausal XPs, and will be ignored here. A clitic can be a member of more than one domain. The clitics *dé* and *mén*, for instance, exhibit both sentential and phrasal scope (as in example 10 below, where it scopes over a topicalized phrase).

Clitic domains mirror clitic scope: CP for sentential clitics, S for clausal clitics, and sub-clausal XPs for phrasal clitics. This follows from their syntactic categories, since sentential adverbs (including clitic $\widehat{\text{Adv}}$) must be adjoined to CP whereas argument DPs/$\widehat{\text{D}}$s must be daughters of S to get the correct GF assigned. When there is no topicalized or focalized element (adjoined to CP or S respectively), sentential clitics immediately precede clausal clitics. When there is a topicalized or focalized element, then SPLAYING results, that is, the sentential clitic and clausal clitic are not adjacent (see example 10 in the next section). We take this to mean that in principle the whole CP with the core clause S forms one prosodic domain (an IntP); but whenever there is adjunction to CP or S for topicalization or focalization

[3]Some accounts have failed to incorporate this insight, e.g., Agbayani & Golston (2010b).

[4]We abstract away here from the motivation for the order of clitics within each domain, and leave open the question of whether it results from the phonology, morphology, or syntax, or some combination thereof.

purposes, this creates a prosodic break so that the core S is one IntP and the material to its left is another IntP.

Sentential and clausal clitics differ in how discriminating they are of their host. Clausal clitics are almost always hosted by prosodic words, while sentential clitics are routinely hosted by both prosodic words and sub-prosodic words:

(8) *Host variability of sentential clitics*

i. *Sub-prosodic word host*

*(hoi=**gár**)$_\omega$=**me*** *ek* *tẽːs*
ART.M.NOM.PL=EXPL=1SG.ACC from ART.F.GEN.SG
kóːmeːs *paĩdes* … *estéːsanto*
village.F.GEN.SG child.C.NOM.PL appoint.PFV.IND.MID.3PL
basiléa
king.M.ACC.SG
'For the children from the village… appointed me their king.' (Hdt. 1.115.2)

ii. *Prosodic word host*

(tà *toiaũta)$_\omega$=**gàr*** *érga* *ou* *pròs*
ART.N.ACC.PL such.N.ACC.PL=EXPL deed.N.ACC.PL NEG by
toũ *hápantos* *andròs*
ART.M.GEN.SG all.C.GEN.SG man.M.GEN.SG
nenómika *gínesthai...*
think.1SG.PERF.ACT.IND happen.PRES.ACT.INF
'For I have thought that not each man is capable of such deeds, but ...' (Hdt. 7.153)

In example (8i), the postpositive *gár* is hosted by the definite article *hoi*, which is a sub-prosodic word. In example (8ii), by contrast, the selfsame particle is hosted by the prosodic word (*tà toiaũta*)$_\omega$. This variation in host selection further distinguishes postpositives from enclitics.

2.2 Second-position distribution

The basic distributional generalization is that clitics occur in the leftmost position possible modulo their lexical entries. When no material is adjoined to CP or S, sentential and clausal clitics are hosted by the first prosodic word in CP or S. They will furthermore be adjacent, with sentential clitics preceding the clausal clitics. This is illustrated by example (8i), which we repeat here for convenience:

(9) *Adjacent clitics*

*(hoi=**gár**)$_\omega$=**me*** *ek* *tẽːs* *kóːmeːs*
ART.M.NOM.PL=EXPL=1SG.ACC from ART.F.GEN.SG village.F.GEN.SG
paĩdes … *estéːsanto* *basiléa*
child.C.NOM.PL appoint.PFV.IND.MID.3PL king.M.ACC.SG

'For the children from the village... appointed me their king.' (Hdt. 1.115.2)

The sentential clitic *gár* immediately precedes the clausal clitic *me*, because no constituents are adjoined to S. Adjunction to CP or S results in splaying if both sentential and clausal clitics are present:

(10) *Splaying*

[*tè:n*=***mèn*=*gàr*** *protére:n he:mére:n*],
ART.F.ACC.SG=PTCL=EXPL previous.F.ACC.SG day.F.ACC.SG
pánta=***sp^{h}i*** *kakà ék^{h}ein.*
everything.N.ACC.PL=3PL.DAT bad.N.ACC.PL have.INF.PRES.ACT
'[For on the previous day], everything was bad for them.' (Hdt. 1.126.4)

The DP [*tè:n protére:n he:mére:n*] '(on) the previous day' is adjoined to CP. The sentential clitic *gàr* is accordingly hosted by the first prosodic word within CP. (The particle *mèn* here is a phrasal clitic that together with adjunction to CP signals the topicalized status of the DP.) The pronominal clitic *sp^{h}i* is hosted after the first prosodic word within S.

As mentioned above, we assume that *wh*-words occupy Spec,CP. In both direct and embedded questions, clausal clitics are hosted by the first prosodic word of the *wh*-phrase:

(11) i. *Kroĩse, tís*=***se*** *ant^{h}ró:po:n*
Croesus.M.VOC.SG WH.C.NOM.SG=2SG.ACC person.M.GEN.PL
anégno:se epì gẽ:n tè:n
persuade.AOR.IND.ACT.3SG against land.F.ACC.SG ART.F.ACC.SG
emè:n strateusámenon polémion
my.F.ACC.SG campaign.PTCP.AOR.MID.M.ACC.SG enemy.M.ACC.SG
antì p^{h}ílou emoì katastẽ:nai?
instead friend.M.GEN.SG 1SG.DAT be.set.INF.AOR.ACT
'Croesus, what person persuaded you to stand against me as an enemy instead of with me as my ally, and campaign against my land.' (Hdt. 1.87.3)

ii. *Dareĩos epì tẽ:s heo:utoũ ark^{h}ẽ:s*
Darius.M.NOM.SG on ART.F.GEN.SG 3SG.M.GEN reign.F.GEN.SG
kalésas Hellé:no:n toùs
call.PTCP.AOR.ACT.M.NOM.SG Greek.M.GEN.PL ART.M.ACC.PL
pareóntas eíreto (epì
be.around.PTCP.PRES.ACT.M.ACC.PL ask.IMPF.IND.MP.3SG for
kóso:i)$_\omega$=***àn*** *k^{h}ré:mati*
how.much.WH.N.DAT.SG=MOD money.N.DAT.SG
bouloíato toùs patéras
want.PRES.OPT.MP.3PL ART.M.ACC.PL father.M.ACC.PL
apot^{h}né:iskontas katasitéest^{h}ai.
die.PTCP.PRES.ACT.M.ACC.PL eat.INF.PRES.MP

‘During his reign Darius summoned the Greeks who were around and asked (them) at what price they would eat their fathers after they had died.’ (Hdt. 3.38.3)

In example (11i), the *wh*-interrogative hosts the pronominal clitic *se* ‘you’, while in (11ii), the modal particle *án* is hosted by the first prosodic word of the interrogative phrase, (*epì kósoːi*)$_\omega$. This is a significant pattern, which reveals that clausal clitics cannot be analyzed as adjoined to S (or on other analyses, TP/IP) in c-structure (if that were the case, *án* would be hosted by *k^hréːmati*), which is what many analyses assume. Below in section 3 we show how we handle this pattern.

That prosody has the upper hand in the distribution of second-position clitics comes from examples such as the following, where the prosodic constituency of the metrical line creates the left edge of an intonational phrase:

(12) *Verse edge is intonational phrase edge*

hót-an d’ híkeːtai, teːnikaũt’ egòː
when-MOD PTCL come.PRES.SBJV.MP.3SG, then 1SG.NOM
kakòs
remiss.M.NOM.SG
(*mèː drõːn*)$_\omega$=***àn*** *éieːn*
NEG do.PTCP.PRES.ACT.M.NOM.SG=MOD be.PRES.OPT.ACT.1SG
pánt^h’ hós’ àn deːloĩ
all.N.ACC.PL so.much.REL.N.ACC.PL MOD indicate.PRES.OPT.ACT.3SG
t^heòs.
god.M.NOM.SG
‘When he gets here, I would be remiss if I didn’t do whatever god indicates.’ (Soph. *OT* 76–77)

The modal particle *án* is hosted by the third prosodic word of the clause. Crucially, it is not possible to analyze the preceding two prosodic words as adjoined phrases with either topic or focus functions. (*mèː drõːn*)$_\omega$ is a licit host prosodic host here because the left edge of the metrical line is the left edge of an intonational phrase. So the prosodic properties of the metrical line satisfy the lexical entry of the clitic.

3 Analysis

3.1 Multiple-context free grammars

Ordinary context-free grammar (CFGs) rules conflate category formation and yield computation. A rule such as DP $\rightarrow$ D NP says both that the category DP is formed of a D and a NP, and that the yield of the resulting DP is formed by concatenating the yields of D and NP. Multiple context free grammar (MCFG) is a generalization of CFG which retains ordinary CFG productions for the expression of categorial structure, but uses explicit *yield functions* to compute the yield of the mother node from the yields of the daughters. In effect, then, a CFG can be seen as an MCFG with concatenation as the only yield function.[5]

To allow for greater expressivity, MCFG allows yields to be *tuples* of strings. For example, we may want to say that the yield of DP is a pair (2-tuple) consisting of the yields of D and NP. This pair will then be the input to further yield functions that apply to productions with DP on the right-hand side. The start symbol of the grammar is required to yield a string.

Symbolically, we will write $\langle x, y\rangle$ to refer to the y'th component in the yield of the x'th category on the right-hand side of a production. We use semicolon (;) for concatenation and square brackets to delimit components in the yield of the left-hand side. (13) gives sample yield functions.

(13) i. $c = [\langle 1, 1; 2, 1\rangle]$
ii. $s_1 = [\langle 1, 1\rangle][\langle 2, 1\rangle]$

c says that the yield of the mother node is formed by concatenation of the first component of the first daughter and the first component of the second daughter. s_1 (mnemonic for "split after first daughter") instead yields a pair of these two components. Notice that both rules are only defined when applied to productions with two categories on the right-hand side (since they only refer to two daughters), both of whose yields is a string (since they only refer to one component in the yield of each daughter). Since the yield of s_1 has two components, we say that it is two-dimensional.

If a daughter node is discontinuous (has dimensionality > 1), that discontinuity may be propagated to the yield of the mother node. (14) gives an example. Here the production references a particular yield function which is independently specified.

(14) PP $\to p_2$(P DP), $p_2 = [\langle 1, 1\rangle; \langle 2, 1\rangle][\langle 2, 2\rangle]$

p_2 (mnemonic for "propagate a discontinuity in the second daughter") forms a two-dimensional yield for the PP. The first component is the yield of P concatenated with the first component in the yield of DP and the second component is the second component in the yield of DP. p_2 is only defined for productions with two daughters, the second of which has a two-dimensional yield.

The split and propagation rules create two-dimensional yields but say nothing about the positioning of the two components, or what elements can intervene between them. Their function is like that of a *head* in Head Grammar (Pollard, 1984), i.e. they create a distinguished position in the yield, after which it can be split.

The idea behind our analysis is that yield functions can exploit this gap by hosting clitics in that position. This is achieved with the yield function in (15).

(15) $h_2 = [\langle 2, 1\rangle][\langle 1, 1\rangle; \langle 2, 2\rangle; \langle 3, 1\rangle]$

[5] Our treatment of MCFG here is necessarily brief, see Clark (2014) for an accessible introduction for linguists.

This corresponds to the wrap operation of Head Grammar: the (discontinuous) second daughter is wrapped around the (continuous) first daughter. The result is a two component yield, where the first component of the second daughter is the first component of the mother, and the second component of the mother is the concatenation of the first daughter, the second part of the second daughter and the third daughter. This yield function, then, "hosts" the first daughter within the second one, but otherwise leaves the gap in the second daughter open so that more elements can be hosted in the gap. We also need a variant of this rule which hosts the first daughter while resolving the gap in the second daughter. This is given in (16):

(16) $r_2 = [\langle 2,1\rangle; \langle 1,1\rangle; \langle 2,2\rangle; \langle 3,1\rangle]$

So far, all the yield functions we have seen deal with only one discontinuous argument. More complex situations can also arise, where there are two discontinuous arguments. In such cases we can intertwine the two to create a continuous yield for the mother node, as in (17), which takes two discontinuous arguments and one continuous one (the third).

(17) $i_1 = [\langle 1,1\rangle; \langle 2,1\rangle; \langle 1,2\rangle; \langle 2,2\rangle; \langle 3,1\rangle]$

As we will see, this intertwining pattern is found in examples like (11ii).

This exhausts the yield functions that we need. Importantly, c, s, h, r and i are *families* of yield functions. As we have seen, yield functions must specify how many daughter nodes they apply to and how many components they have. We leave the specification of the number of daughter nodes implicit as it is retrievable for the categorial structure. In addition, s, h and r require exactly one of the daughter nodes to have a two-component yield, and the subscript on the function designates this node. i requires two discontinuous daughter nodes; by convention, for any given i_n, this will be the n'th and $n+1$'th daughters, so that a single parameter is enough. The number of yield functions in the grammar will therefore remain reasonable even if the branching factor is high. As we will see in section 3.2, their actual application to productions will be controlled by a prosodic *HOST* feature.

The fact that yield functions must explicitly state the number of arguments and discontinuities means that discontinuities created this way do not interact with the recursive mechanism in the categorial structure, so that the maximum number of discontinuities that a grammar permits can be read directly off the most complex yield function in the grammar. In our case, no rule outputs a yield of more than 2 dimensions, so our grammar is a 2-MCFG.[6]

It is instructive to compare this with how LFG can otherwise model syntactic discontinuities through reentrancies, i.e. multiple c-structure nodes corresponding

[6]This assumes that the grammar can be binarized without increasing the dimensionality, an assumption that may in fact not hold in the presence of *ill-nested* yield functions such as i, see, e.g., Kuhlmann (2013). Nevertheless, it is clear that the complexity will remain low especially given the fact that the set of potential intrudes—the admissible clitic sequences—will be finite, see section 4.4.

to the same f-structure. This does interact with the recursive mechanism of the c-structure. The result is that a given LFG may not provide an upper bound on the number of c-structure nodes corresponding to a single f-structure and hence features may in principle be transmitted across unbounded distances in the tree.

This also gives some indications of how the MCFG approach differs from linearization-based HPSG (Reape 1994; Kathol 1995), which is another attempt at dissociating categorial structure and yield computation. In linearization HPSG, the schemata that build the categorial structure do not at the same time build yields of terminals; instead, they build word order domains, over which linear precedence constraints can then be stated. A daughter's word order domain can either be *compacted*, i.e. it enters its mother's word order domain as a contiguous string yield; or it can be *domain unioned* with its mother, i.e. the elements of its word order domain can appear discontinuously in the mother's domain as long as the relative order of the elements is preserved in the mother's domain. Here too, the building up of word order domains interacts with recursion, which means that the size of word order domains may not be bounded by a given grammar.

3.2 Application

We assume an architecture of the prosody-syntax interface along the lines of Dalrymple & Mycock (2011). That is, we assume that the grammar builds syntactic and prosodic trees in tandem and that trees meet in the s- and p-strings, which are associated via their co-occurrence in lexical entries: the string is therefore the sole point of interface between syntax and prosody. Following Mycock & Lowe (2013), we do not use dedicated projections (e- and χ-structure) to pass information through the prosodic and syntactic trees, but rather assume that the terminal elements of the s- and p-strings are AVMs that can store information beyond the form of the relevant string elements.

Concretely, we need the p-string to contain information about prosodic hosting patterns. This is done via annotation on the prosodic structure building rules in the same way as in Mycock & Lowe (2013). As discussed in section 2.1.1 there are two patterns that are special to clitics: recursive prosodic word formation (example 4 above) and hosting of the clitic by a subprosodic word (example 7 above). The latter process is less well understood, but for concreteness we will assume stray adjunction of syllables. The relevant prosodic structure building rules are given in (18)–(19).[7]

(18)
$$\begin{array}{lccc} \omega & \rightarrow & \omega & \sigma \\ & & \omega \in (\swarrow \mathrm{L}) & \omega \in (\searrow \mathrm{R}) \\ & & \mathrm{IntP} \in_c (\swarrow \mathrm{L}) & (\searrow \mathit{HOST}) = \omega \end{array}$$

(19)
$$\begin{array}{lcccc} \omega & \rightarrow & \sigma & \sigma^* & \sigma \\ & & \omega \in (\swarrow \mathrm{L}) & & \omega \in (\searrow \mathrm{R}) \\ & & \mathrm{IntP} \in_c (\swarrow \mathrm{L}) & & (\searrow \mathit{HOST}) = \sigma \end{array}$$

[7]Following Mycock & Lowe (2013), we use italics for p-string features.

The first line in each of these rules simply passes down edge information to the terminal left ($\swarrow$) and right ($\searrow$) daughters.[8] The interesting things happen in the second line: ($\downarrow$ *HOST*) = $\omega|\sigma$ records the type of prosodic host (syllable or prosodic word) in the p-string of the clitic, and IntP $\in_c$ ($\swarrow$ L) ensures that hosting can only apply at the left edge of IntP. Notice that the rules leave open whether the relevant IntP is the core clause or a verse-induced IntP as in (12).

We now come to the lexical entry of clitics. As we saw in section 2.1.1, sentential clitics are typically happy to accept both a PW and a sub-PW as their host, whereas clausal clitics require a PW host. (20) gives sample lexical entries for the sentential clitic *gár* (roughly 'for', signalling a causal connection between the sentence in which it occurs and some piece of preceding discourse) and the clausal clitic *me* ('me').

(20)	*gár*	$\widehat{\text{Adv}}$	($\searrow$ *HOST*)	*me*	$\widehat{\text{D}}$	($\searrow$ *HOST*) $=_c \omega$
			$\vdots$			($\uparrow$ PRED) = 'PRO'
						($\uparrow$ CASE) = ACC
						$\vdots$

The constraints on the *HOST* feature ensure that *me* can only be inserted in the prosodic structure via (18) whereas *gár* can be inserted with (18) or (19). To keep yield functions from overgenerating, we assume that the yield functions h, r and i modify the productions they apply to by inducing a syntactic existential constraint ($\uparrow$ HOST) on the argument hosted by that yield function (i.e., the $n-1$ first daughters for h_n, r_n and the $n+1$'th daughter for i_n), with the effect that non-concatenative yield functions on the syntactic side must be licensed by (18)–(19) in the prosody.[9]

This is in fact all we need to derive clitic behaviour. On the prosodic side, the *HOST* feature requires the clitic to go after an appropriate host at the left edge of its IntP. Since the prosodic and syntactic structures are built over the same string, with no reordering, the syntax must position the clitic appropriately. Notice that no information passing between prosody and syntax is required; prosody influences syntactic positioning because both structures must be simultaneously well-formed.

To see how this works, consider (10). The core clause S in this example is *pánta kakà ék^{h}ein*, which by our assumptions forms an IntP. So the clitic *sp^{h}i* must find a suitable host at the left edge of this IntP while at the same time being a daughter of S. Since in this case the first constituent of S, *pánta*, is exactly one prosodic word and the clitic qua $\widehat{\text{D}}$ can be the second constituent of S, this is straightforward and requires no yield functions beyond concatenation. For the sentential clitic *gàr* things are somewhat more complicated: it must find a prosodic host within the

[8] As noted by Mycock & Lowe (2013), these can probably be stated as more general principles and need not actually be stated on every rule.

[9] This requires a slight change to the principle of Interface Harmony (Dalrymple & Mycock, 2011; Mycock & Lowe, 2013) since an existential constraint on the syntactic side is verified by a constructive constraint on the prosodic side.

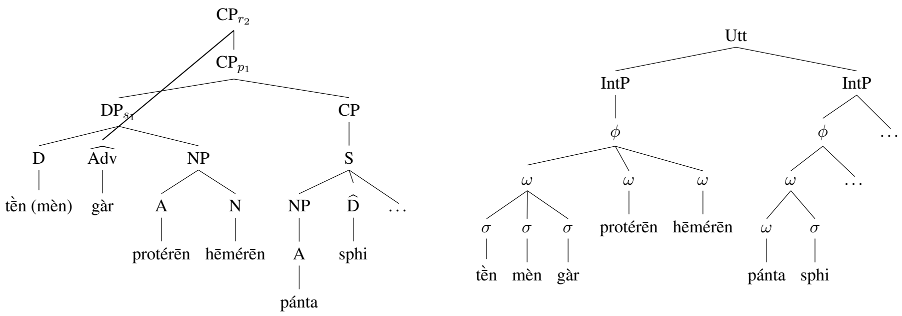


Figure 2: Syntactic and prosodic structure for (10)

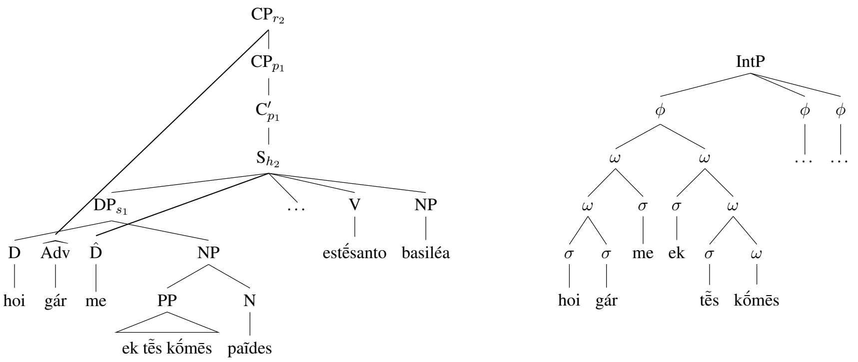


Figure 3: Syntactic and prosodic structure for (8i)

topicalized phrase *tè:n=mèn protére:n he:mére:n* which has been adjoined to CP and forms a separate IntP. At the same time it must adjoin to the highest CP, i.e. above the phrase which must host it prosodically. This can be done by exploiting the yield functions p_1 and r_2, as shown in Figure 2.[10]

We can also derive the correct behavior in sentences like (8i) where there is no S-external material apart from the sentential clitic. Such sentences make up only one IntP and consequently, the sentential clitic must be hosted lower than its own position in the syntactic structure. This is achieved by hosting the clausal clitic in a gap in the first constituent DP in S, and then propagating this gap up to the CP-level, where it is resolved, as shown in Figure 3.

A salient feature here is that the host D *hoi* is not itself a prosodic word, but forms one with the first clitic, which can take a syllable host. The result is a prosodic word, which can therefore host *me*. Finally, we note that we can capture the complex example in (11ii) where the clitic is hosted in Spec,CP even if it belongs in the S domain. This is shown in Figure 4.[11]

[10]Needless to say, given our knowledge of AG, much of the prosodic structure assumed in the following figures is based on conjecture. But all the points crucial to our analysis have been argued for here.

[11]We add that there are also examples where Spec,CP is apparently not in the same IntP as S even

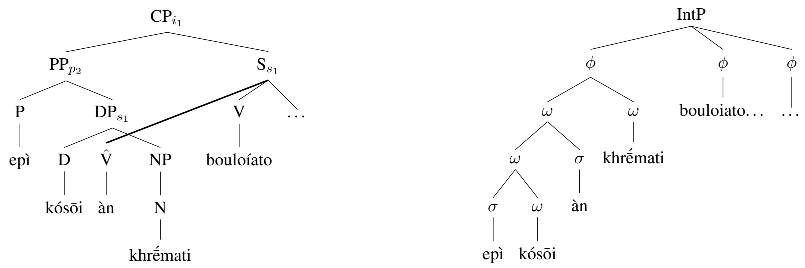


Figure 4: Syntactic and prosodic structure for (11ii)

4 Discussion

The main advantage of our proposal, we contend, is that we can treat clitics as syntactically normal words as far as the categorial structure goes.[12] What is special about them is their need for a prosodic host, which drives their special linearization. MCFG is the perfect tool to capture this, with its distinction between categorial structure and linearization (i.e., yield computation).

By contrast, all other approaches try to put clitics in the "correct" position in an ordinary, context-free c-structure, where they can get the correct grammatical function, and then displace them at some point in LFG's projection architecture. For Bögel et al. (2009, 2010) and Bögel (2015), they are displaced between the syntactic and the prosodic string, i.e. we have a syntax-prosody mismatch. For Lowe (2015), they are displaced between the c-structure and the syntactic string, i.e. we have a syntax-internal mismatch where the yield of the c-structure is not identical to the syntactic string. Architectural differences apart, there are similar conceptual and empirical problems with both approaches: both need to motivate a non-surface position in the c-structure and a means of enforcing this position.

4.1 Non-surface positions

Traditionally, it has been a hard constraint on LFG c-structures that their yield should match the output string. This means that the precedence order on the string can guide our assumptions about the precedence order on the terminals of the c-structure. If we give up this assumption, we need other heuristics to determine c-structure position.

Consider our examples (8i) and (10) with Figures 2–3. On a displacement view, we must motivate non-surface positions for the clitics in these trees. For *gár* this is perhaps not too difficult, since this particle has sentential scope and so we can assume that it adjoins high up and hence its underlying position is to the left of

in the absence of a focalized constituent and where consequently clausal clitics are hosted after the first prosodic word in S, discounting Spec,CP. We have as of yet no account for this variation.

[12]Apart from their non-projecting status, that is, but non-projection (to the extent it is needed) is independently motivated for these words and in fact not essential to derive their clitic behavior.

the rest of the sentence. But consider *me* in Figure 3. On the displacement view, this must be realized to the left of the rest of S in order to be targeted by prosodic inversion. But what is the evidence that *me* actually has this c-structure position? None of the traditional LFG heuristics work: AG does not have a dedicated object position, and the stressed variant *emé* can occur in various positions in the clause, both pre- and post-verbally. The assumption that *me* is S-initial is entirely driven by the need to collect all clitics in a single position where they can be targeted by prosodic inversion. Even worse, the displacement view is forced to assume that *sphi* in Figure 2, which satisfies all the prosodic and syntactic constraints in its surface position, is actually displaced from an underlying S-initial position. Again, there seems to be no motivation for this assumption, except the need for the clitic to undergo inversion. Finally, it is unclear how to derive the position of *an* in Figure 4 on a prosodic inversion view: for functional reasons *an* belongs inside S, but it actually surfaces to the *left* of the edge of S.

4.2 Clitic as a syntactic category

While the assumption that clitics have a dedicated covert position in the c-structure is doubtful, the means of enforcing this position is no less problematic. Theories of this kind standardly assume that all clitics belong to a single lexical category CL. But clisis is a prosodic category, not a functional one. Cross-linguistically and even within single languages such as AG the functional categories of clitics are so diverse (encompassing at least pronominal elements, connectives, discourse particles, and tense and modal auxiliaries) that a single unified category is unappealing (O'Connor, 2002, 316).

Moreover, sequences of clitics are typically hosted under a CCL node (e.g., examples 16–17 in Lowe 2011). There is, however, little to no evidence that clitic clusters form a syntactic or a prosodic constituent. This analysis also assigns clitics with diverse functional profiles (pronouns, auxiliary verbs, discourse particles, modal quantifiers) the same lexical category. The lexical category is in fact defined by the need for a prosodic host. We see little appeal in this move, because lexical categories and prosodic properties are independent dimensions of lexemes. By allowing them to interact, one opens the door to a lexical category for all stressed words, which is an unwelcome possibility. In contrast, our approach lets clitics have exactly the category that we expect from their functional profile.

4.3 Comparison with linearization HPSG

As we saw in section 3.1, yield functions as found in MCFG are a less powerful device than the word order domains found in linearization-based HPSG, which were developed to distinguish between tectogrammatical and phenogrammatical structure, something which is already present in the c-/f-structure distinction in LFG. Since linearization is a more powerful device, designed to do more than what our yield functions do, it must be constrained in order not to overgenerate. For example, the analysis of Penn (1999) requires four word order domains, the pre-clitic, clitic, post-clitic and remainder fields, in order to prevent post-clitic material that belongs syntactically with the pre-clitic field to mix with the remainder field. In our approach, there is never more than a binary split of the yield.

But modulo the formal device involved, the basic idea is the same in our approach and that of Penn (1999) and Crysmann (2006). In the HPSG version, prosody can effect compaction, whereas in our approach non-standard yield functions are applied for essentially prosodic reasons. In HPSG, prosodic and syntactic compaction is modelled in the same structure, which predicts a much closer relationship between syntax and prosody than we assume. That, however, seems to be a result of the underdeveloped role of prosody in HPSG: there is only one level where constraints can be satisified, but as Penn (1999, 6) points out, "it is entirely unsatisfying to represent prosodic structure as a disconnected list of prosodic words to be carried around in an otherwise syntactic derivation. Ultimately, the notion of sign in HPSG must be changed to allow parallel derivations of prosody, syntax and discourse." Our LFG account gets this for free on account of the parallel architecture of the framework.

4.4 Architectural issues

Increasing the generative power of the c-structure from a CFG to a 2-MCFG does not increase the complexity of the combined LFG formalism. As shown by Seki et al. (1993), any MCFG can be translated into an LFG, and any LFG which bounds the number of c-structure nodes corresponding to each f-structure can be translated into an MCFG. Though we omit the proof, the same is true for the MCFG c-structure component of our analyses: the linearization facts that we analyze with the help of yield functions could in principle be captured with reentrancies, i.e. with multiple c-structure nodes mapping to a single f-structure. This would entail treating clitic-induced discontinuities in the same way as scrambling, but it would require a lot of linguistically unmotivated bookkeeping in the f-structure to enforce the significant differences between clitic-induced discontinuities and scrambling, which for example can never separate a determiner from its noun phrase.

Similar problems could arise if one wanted to apply our MCFG-approach to e.g., scrambling discontinuities. At present it is not clear to us what the domain of application for non-concatenative yield functions is. We have assumed here that they are governed by the HOST-feature, which means that only clitics can have non-concatenative syntax, but future research may show that the domain of application is wider. A larger question is whether grammars can vary on this point or whether there are universal principles of non-concatenative syntax.

Finally, one aspect that does not fall out from our account, but which is stressed by Bögel et al. (2010), is the finiteness of the set of admissible clitic sequences in any given language. This is predicted by their account because it models the syntax-prosody relation as a regular transducer, which can only handle a finite set of elements to be displaced at the interface. There is nothing directly comparable in our theory, but it may be possible to prove that finiteness of the set of possible fillers of gaps in two-dimensional yields decreases the parsing complexity of the MCFG grammar, perhaps down to that of an ordinary CFG.

5 Conclusion

We have argued that the distribution of second-position clitics motivates an increase in the power of c-structure to that of a 2-MCFG in analyzing natural language syntax. Our analysis obviates both covert positions in c-structure and the assumption that "clitic" is a syntactic category. Adopting an MCFG-based c-structure does not increase the expressivity of the LFG formalism, but it does mean that c-structure can describe more complex phenomena that would otherwise be left to other projections in the LFG architecture. We leave for future research the question of whether the yield functions of MCFG could be used to insightfully model other linguistic phenomena.

References

Agbayani, Brian & Chris Golston. 2010a. Phonological movement in classical Greek. *Language* 86. 133–167.

Agbayani, Brian & Chris Golston. 2010b. Second-position is first-position. *Indogermanische Forschungen* 115. 1–21.

Anderson, Stephen R. 2005. *Aspects of the theory of clitics*. Oxford: Oxford University Press.

Anderson, Stephen R. 2012. Clitics. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, vol. 4, 2002–2018. Malden, MA: Blackwell.

Bögel, Tina. 2015. *The syntax–prosody interface in Lexical Functional Grammar*: Universität Konstanz dissertation.

Bögel, Tina, Miriam Butt, Ronald M. Kaplan, Tracy Holloway King & John T. Maxwell III. 2009. Prosodic phonology in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG09 conference*, 146–166. Stanford: Center for the Study of Language and Information.

Bögel, Tina, Miriam Butt, Ronald M Kaplan, Tracy Holloway King & John T Maxwell. 2010. Second position and the prosody-syntax interface. In Miriam Butt & Tracy Holloway King (eds.), *Fifteenth international lexical functional grammar conference (LFG10)*, 106–126. Stanford: Center for the Study of Language and Information.

Ćavar, Damir & Melanie Seiss. 2011. Clitic placement, syntactic discontinuity, and information structure. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 conference*, 131–151. Stanford: Center for the Study of Language and Information.

Chandler, Henry W. 1881. *A practical introduction to Greek accentuation*. Oxford: Oxford University Press 2nd edn.

Clark, Alexander. 2014. An introduction to multiple context free grammars for linguists. http://www.cs.rhul.ac.uk/home/alexc/lot2012/mcfgsforlinguists.pdf.

Crysmann, Berthold. 2006. Floating affixes in Polish. In Stefan Müller (ed.), *Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar, Varna*, 123–139. Stanford: Center for the Study of Language and Information.

Dalrymple, Mary & Louise Mycock. 2011. The prosody-semantics interface. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 conference*, 173–193. Stanford: Center for the Study of Language and Information.

Devine, Andrew M. & Laurence D. Stephens. 1994. *The prosody of Greek speech*. Oxford: Oxford University Press.

Devine, Andrew M. & Laurence D. Stephens. 1999. *Discontinuous syntax*. Oxford: Oxford University Press.

Dik, Helma. 1995. *Word order in ancient Greek*. Amsterdam: Gieben.

Fortson, Benjamin W. 2010. *Indo-European language and culture*. Malden, MA: Blackwell 2nd edn.

Fraenkel, Eduard. [1933] 1964. Kolon und Satz. In *Kleine Beiträge zur klassischen Philologie*, vol. 1, 93–130. Roma: Storia e Letteratura.

Goldstein, David. 2016. *Classical Greek syntax: Wackernagel's law in Herodotus*. Leiden: Brill.

Hale, Mark R. 1987a. Notes on Wackernagel's law in the language of the Rigveda. In Calvert Watkins (ed.), *Studies in memory of Warren Cowgill (1929–1985)*, 38–50. Berlin: de Gruyter.

Hale, Mark R. 1987b. *Studies in the comparative syntax of the oldest Indo-Iranian languages*: Harvard University dissertation.

Kathol, Andreas. 1995. *Linearization-based German syntax*: Ohio State University dissertation.

Kuhlmann, Marco. 2013. Mildly non-projective dependency grammar. *Computational Linguistics* 39. 355–387.

Lowe, John J. 2011. Ṛgvedic clitics and 'prosodic movement'. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 conference*, 360–380. Stanford: Center for the Study of Language and Information.

Lowe, John J. 2014. Accented clitics in the *Ṛgveda*. *Transactions of the Philological Society* 112. 5–43.

Lowe, John J. 2015. Clitics. *Journal of Linguistics* 52. 375–419.

Mycock, Louise & John Lowe. 2013. The prosodic marking of discourse functions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG13 conference*, 440–460. Stanford: Center for the Study of Language and Information.

O'Connor, Robert. 2002. Clitics and phrasal affixation in constructive morphology. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG02 conference*, 315–332. Stanford: Center for the Study of Language and Information.

Penn, Gerald. 1999. A generalized-domain-based approach to serbo-croatian second position clitic placement. In Gosse Bouma, Erhard W. Hinrichs, Geert-Jan M. Kruijff & Richard T. Oehrle (eds.), *Constraints and resources in natural language syntax and semantics*, 119–136. Stanford: Center for the Study of Language and Information.

Pollard, Carl. 1984. *Generalized phrase structure grammars, head grammars, and natural language*: Stanford University dissertation.

Probert, Philomen. 2003. *A new short guide to the accentuation of ancient Greek*. London: Bristol Classical Press.

Reape, Mike. 1994. Domain union and word order variation in german. In John Nerbonne, Klaus Netter & Carl Pollard (eds.), *German in Head-Driven Phrase Structure Grammar*, 151–197. Stanford: Center for the Study of Language and Information.

Seki, Hiroyuki, Ryuichi Nakanishi, Yuichi Kaji, Sachiko Ando & Tadao Kasami. 1993. Parallel multiple context-free grammars, finite-state translation systems, and polynomial-time recognizable subclasses of lexical-functional grammars. In *Proceedings of the 31st Annual Meeting on Association for Computational Linguistics (ACL '93)*, 130–139. Morristown, NJ: Association for Computational Linguistics. doi:10.3115/981574.981592.

Taylor, Ann. 1990. *Clitics and configurationality in ancient Greek*. Philadelphia: University of Pennsylvania dissertation.

Toivonen, Ida. 2003. *Non-projecting words*. Dordrecht: Kluwer.

The syntax-prosody interface in Korean: Resolving ambiguity in questions

Stephen Jones
University of Oxford

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 318–338

Keywords: Korean, prosody, syntax, experimental, questions

Jones, Stephen. 2016. The syntax-prosody interface in Korean: Resolving ambiguity in questions. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 318–338. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.17.

Abstract

The paper[1] considers a phenomenon in Korean where ambiguity in the written language is resolved prosodically. An LFG analysis is provided which extends the proposals of Mycock and Lowe (2013) to Korean, based on experimental evidence on the prosodic expression of focus in Korean which challenges the phrase-boundary based account of Jun and Oh (1996), and suggests that considering expanded pitch range may give a more robust account of focus expression.

1 Introduction

This paper aims to give a LFG analysis of the role of prosody in determining the three meanings of the Korean sentence[2] given at (1).

(1) *acwumeni-ka* *nwukwu-lul* *manna-syess-eyo*
auntie-SBJ someone/who-OBJ met-SH.PST-POL

a. 'Auntie met someone.'

b. 'Did auntie meet someone?'

c. 'Who did auntie meet?'

The three meanings result from interactions between three elements of Korean grammar. First, in the polite speech style there is no morphological marking of questions. Instead, mood is specified by a sentence-final tone pattern on the last syllable. Thus the pattern HL (a high tone immediately followed by a low tone) is associated with indicative mood, whereas the pattern LH (a low tone immediately followed by a high tone) is associated with interrogative mood (Jun, 2005). Second, Korean has a set of words that function both as indefinite pronouns and wh-interrogatives, e.g. *edi* 'where/somewhere', *encey* 'when/sometimes', *ettehkey* 'how/somehow', *mwe(s)* 'what/something', *nwukwu* 'who/someone'. Given this dual function, I will refer to them as content pro-forms (CPFs).

Third, Korean allows for scrambling and, as a consequence, does not require a particular position for the content pro-form in open questions (wh-in-situ). These three elements combine to give the possible readings in (1).

The difference between open and polar question readings can be analysed as an alternation in the scope of question focus (Dalrymple, 2001). When the focus includes the predicate together with the CPF, termed *broad focus*, a polar reading is obtained, with the minimal answer 'yes/no' (2), whereas when only the CPF is

[1]I am grateful for support and contributions from Dr. Jieun Kiaer and Dr. Louise Mycock; Korean language advisers Yoolim Kim and Myungsu Kang; the experimental participants; and the anonymous reviewers. I am also grateful for financial support to attend from Kellogg College, Oxford and the International Lexical Functional Grammar Association.

[2]Romanizations of Korean follow the Yale system (Martin, 1992) unless otherwise indicated.

in focus, termed *narrow focus*, an open reading is obtained, with a minimal answer referring to a specific individual (3).

(2) *acwumeni-ka* ***nwukwu-lul*** ***manna-syess-eyo***
auntie-SBJ **someone-**OBJ **met-**SH.PST-POL
'Did auntie meet someone?'

(3) *acwumeni-ka* ***nwukwu-lul*** *manna-syess-eyo*
auntie-SBJ **who-**OBJ met-SH.PST-POL
'Who did auntie meet?'

Native speakers report no ambiguity when hearing a spoken sentence as each reading is associated with a distinctive prosodic pattern (e.g. Jun and Oh, 1996). However, there is no generally agreed-upon account of acoustic cues by which focus is realised. Some accounts consider the acoustic realisation of narrow focus, rather than comparing narrow and broad focus. These include those of Kim (2000), who describes the focused element as "prominent" and Yun (2012), who concludes that phonological dephrasing after the focused element is more important than the characteristics of the focused element itself.

An alternative account is proposed by Jun and Oh (1996) based on the findings of a perception experiment, who propose that the open and polar readings are distinguished by the placement of phrase boundaries in the whole sentence, rather than the acoustic characteristics of any particular element. Figure 1 is reproduced from their paper and shows the two different readings of the question[3] in (4). In the diagram, for each reading, the x-axis gives the syllables of the question, with each syllable annotated with the symbol σ and, where appropriate, with L or H denoting low or high tone respectively. The y-axis is an unscaled representation of pitch. Phrase boundaries are marked with vertical lines. The boxed sections of each question show the syllables that are in focus.

(4) atʃuməninɨn əntʃe ətʃiləwəjo
madam.TOP sometimes/when dizzy.POL

a. "Is madam dizzy at any time?"

b. "When is madam dizzy?"

In Jun and Oh's account, the prosodic pattern seen in (a), where a there is a phrase boundary between the CPF and the final verb, is associated with the polar reading, where the CPF and the final verb are both in focus. However, in the open reading (b) when the CPF alone is in focus, there is no phrase boundary between the CPF and the verb.

[3]I have given Jun and Oh's phonetic rendering of the sentence. The equivalent Yale romanization is *acwumeninun encey ecileweyo*.

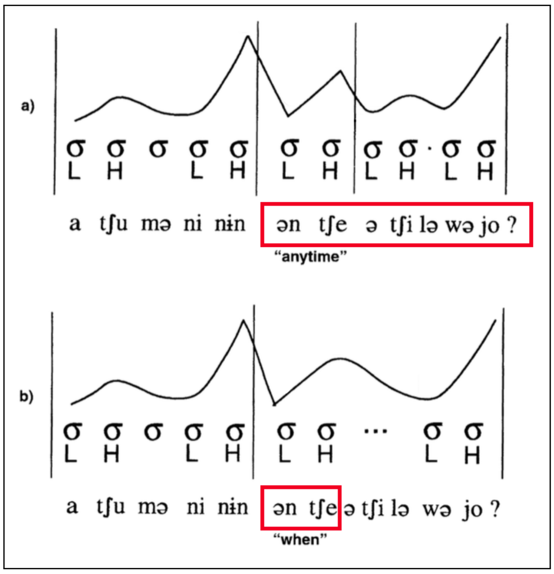


Figure 1: Prosodic disambiguation of focus (from Jun and Oh, 1996, p. 48)

2 Korean prosody

I adopt the account of Korean prosody given by Jun (2005). This differs from English prosody in that there is no evidence of nuclear pitch accent, and that phrasal constituents are not syntactically determined (cf. claims for English by e.g. Selkirk, 1984; Nespor and Vogel, 1986; Hayes, 1989). A prosodic hierarchy exists with nested elements in line with the Strict Layer Hypothesis (Selkirk, 1984). Jun identifies the following elements: syllables (Syll), prosodic words (PW), accentual phrases (AccP), and intonational phrases (IntP). Prosodic word boundaries are not marked, and there is no evidence of lexical stress. However, a prosodic word can be identified as the domain of some sandhi phenomena, such as lenis obstruent tensing. Phrasal constituents AccP and IntP are marked prosodically with distinctive tone patterns and phrase-final lengthening. The hierarchy is shown in Figure 2, reproduced from Jun (2005, p. 205).

Each AccP has the minimal structure shown within the double lines in Figure 2. The left edge of an AccP is marked tonally with a pattern described by Jun as *T*-H, where *T* is underlyingly L, but can be phonologically conditioned to appear as H by an initial tense or aspirated consonant. The right edge of an AccP is marked tonally L-H, with lengthening of the final syllable. The boundary pattern *T*-H. . . L-H is autosegmentally associated (Goldsmith, 1976) with the syllables of the phrase. Accentual phrases can be as short as a single word, and AccPs with a length greater than 7 syllables are seen infrequently. Where a phrase is longer than 4 syllables, the

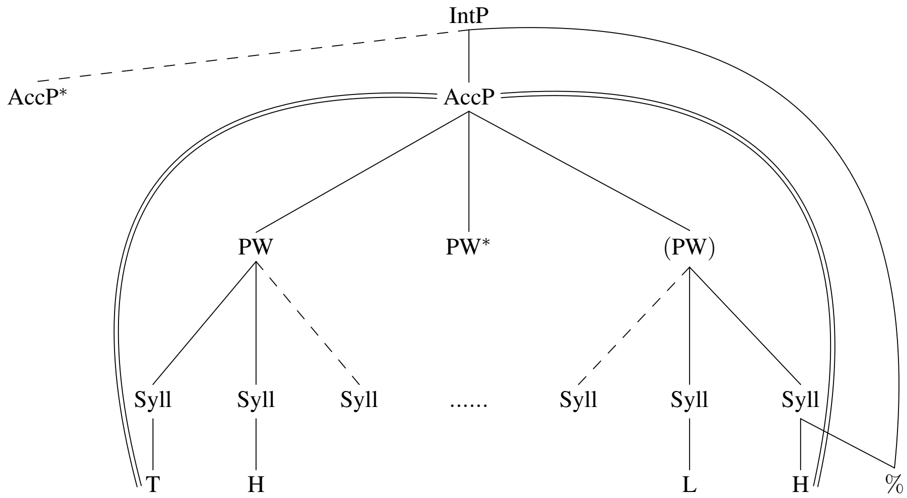


Figure 2: Korean prosodic hierarchy proposed by Jun (2005, p. 205)

left edge H tone may attach to the third rather than the second syllable, and there is a gradual decrease in pitch between the two edge patterns. Where a phrase is shorter than 4 syllables, tone deletion occurs, which results in a variety of possible tone patterns: a full description of this is given in Jun (2005).

Intonational phrases are often a whole sentence or major clause. Their left edge has no particular marking, but the right edge has a boundary tone pattern which replaces the final H tone of the rightmost AccP within the IntP. These boundary tones carry semantic/pragmatic meaning, including the patterns associated with mood mentioned in Section 1: HL ⇒ declarative; LH ⇒ interrogative. In this paper I denote these boundary tones using a variable, %.

Figures 3 and 4 show the respective prosodic patterns for the declarative (a) and interrogative (b) readings respectively of example (5).

(5) *acwumenika nwunalul mannasyesseyo*
auntie.SBJ older.sister.OBJ meet.HON.PST.POL

a. "Auntie met older sister."

b. "Did auntie meet older sister?"

In both examples, AccPs can be distinguished for each of the three words, together with the characteristic IntP final tone pattern. Declination of pitch is seen across the IntP. In Figure 4, the pitch range for the IntP-final LH tone pattern is much greater than that for the HL pattern in Figure 3.

One difficulty for Jun and Oh's account of the prosodic disambiguation is that there is no requirement for the CPF and final verb to be adjacent. Where a con-

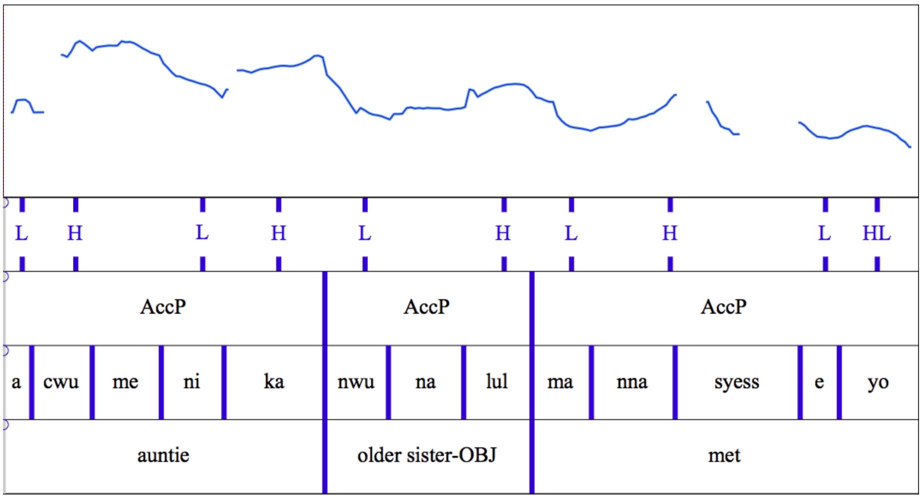


Figure 3: Example prosody for (5a) "Auntie met older sister."

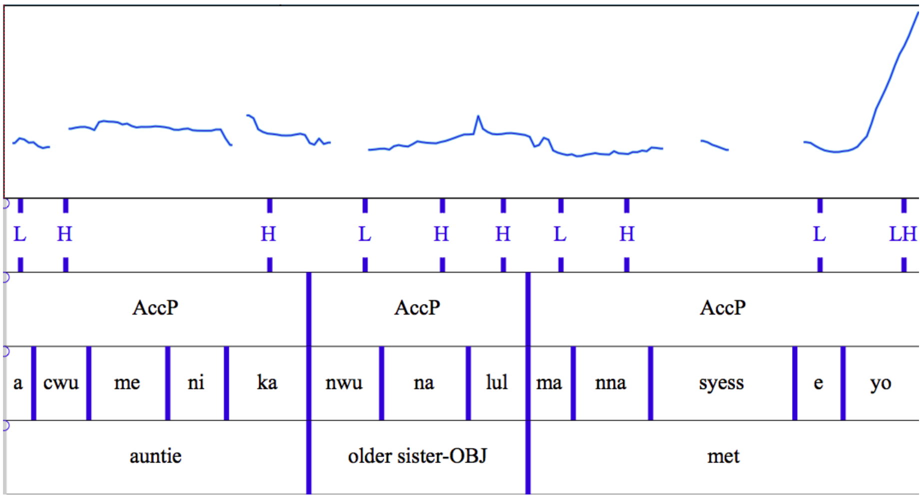


Figure 4: Example prosody for (5b) "Did auntie meet older sister?"

stituent interposes between the CPF and the verb, Jun and Oh's account predicts prosodic patterns that are highly unlikely. Consider example (6).

(6) *acwumeni-nun encey simhakey swusikantongan ecilewe-syess-eyo*
madam-TOP when severely for.several.hours dizzy-HON.PST-POL

a. "Was madam sometimes severely dizzy for several hours?"

b. "When was madam severely dizzy for several hours?"

Jun and Oh frame their account in terms of the presence or absence of an AccP boundary between the CPF and the final verb. For the open reading (6b), Jun and Oh's model predicts the pattern of AccP boundaries shown in Figure 5.

Subject	CPF	Adjunct	Adjunct	Verb
a.cwu.me.ni.nun	*en.cey*	*sim.ha.key*	*swu.si.kan.to.ngan*	*e.ci.le.we.syes.se.yo*
madam-TOP	when	severely	for.several.hours	dizzy.PST.POL
AccP	AccP			
5 syllables	17 syllables			

Figure 5: Predicted prosody: "When was madam severely dizzy for several hours?"

The requirement for no AccP boundary between the CPF *encey* and the final verb *ecilewesyesseyo* gives an AccP of 17 syllables. This is unlikely given that AccPs generally have a maximum length of around 7 syllables, and very long AccPs (e.g. >10 syllables) are infrequently produced. It is questionable whether Jun and Oh's model is robust in these circumstances.

However, Jun and Oh's constraint on AccP boundary placement might also be described as the presence or absence of an AccP boundary either immediately after the CPF, or immediately before the verb, or both. Where the CPF and verb are adjacent, this distinction is moot, but a reframing in terms of the right edge of the CPF and/or the left edge of the verb would avoid the improbably long AccP presented in Figure 5.

3 Experiment

An elicitation experiment was carried out to test the hypothesis that Jun and Oh's account makes the wrong prediction for questions where a constituent intervened between CPF and final verb, in other words, that AccP boundary placement alone would not disambiguate between open and polar readings. The null hypothesis was that Jun and Oh's model was correct.

Stimuli Experimental stimuli were six sets of four context-utterance pairs generated according to the template in Figure 6 with 2×2 variation in length of utterance and target reading (open vs. closed). There was an equal number of similarly-constructed fillers.

An example of one context-utterance pair is given below. The context was given for a target open reading (7) or polar reading (8) and this was matched alternately with the short (9) or long (10) version of the question to generate short open (9a), short polar (9b), long open (10a) and long polar (10b) readings respsectively.

Background information	Target = open:	“You know some, but not all, details of an event.”
	Target = polar:	“You don’t know whether an event happened.”
Structure		Introductory constituent | CPF | (intervening constituent) | verb
Variation		{ open reading / polar reading } × { short: − intervening constituent / long: + intervening constituent }

Figure 6: Template used to generate experimental stimuli

(7) *tangshinun ecey hwanan sarami issessten kesul alko*
you.TOP yesterday angered person.SBJ existed thing.OBJ knowing
issupnita
exists

“You know that someone got angry yesterday.”

(8) *tangshinun ecey hwanan sarami issessnun ci*
you.TOP yesterday angered person.SBJ existed uncertain.thing
moruko isssupnita
unknowing exists

“You don’t know whether anyone got angry yesterday.”

(9) *ecey nwuka hwanasseyo*
yesterday who/someone.SBJ became.angry.POL

a. “Who got angry yesterday?”

b. “Did someone get angry yesterday?”

(10) *ecey nwuka orayn.sikan hyepsang tongan*
yesterday who/someone.SBJ lengthy negotiations during
hwanasseyo
became.angry.POL

a. “Who got angry during the lengthy negotiations yesterday?”

b. “Did someone get angry during the lengthy negotiations yesterday?”

In order to consider the generality of Jun and Oh’s model, the lengths of utterances varied as shown in Table 1. There was also variation in the CPF used in the utterance: *nwuka* ‘who/someone.SBJ’ (Sets D, H); *nwukwulul* ‘who/someone.OBJ’ (Set B); *nwukwuhako kathi* ‘with whom/with someone’ (Set L); *mwelul* ‘what/something.OBJ’ (Sets F, J).

Table 1: Variation in length of experimental utterances

Stimulus	Constituent length				Total length	
Set	Intro	CPF	(Intervening)	Verb	Short	Long
B	5	3	9	5	13	22
D	7	2	7	4	13	20
F	4	2	6	4	10	16
H	2	2	8	4	8	16
J	5	2	8	3	10	18
L	4	6	6	6	16	22

Participants Participants were 9 native speakers of Seoul Korean (7 female, 2 male) aged between 18 and 35, studying at Oxford University, recruited following approval by the university's Research Ethics Committee.

Procedure Context-utterance pairs and fillers were presented visually to participants in a random order: participants first saw a blank slide, then pressed a button to reveal the context. After reading the context, they then pressed the button to reveal the target question, which was shown underneath the context. After uttering the target question, they pressed the button again to clear the screen in preparation for to the next pair. Utterances were recorded digitally and manually analysed into syllables using Praat (Boersma and Weenink, 2016). Each participant produced all of the utterances.

For one participant a technical error resulted in only two of the six utterance sets being recorded completely. The other four incomplete sets from this participant were excluded from the data, giving a total of 200 utterances. Pitch maxima and minima for each syllable were obtained using a Praat script modified from a script published by Mietta Lennes[4]. Following this, Accentual Phrases boundaries were determined by the author in line with Jun (2005) and correspondences between AccP boundaries and the edges of constituents were identified.

Results: Phrase boundary placement in short utterances For the short utterances, three patterns of AccP boundary placement were observed, shown in Table 2. In pattern (i) the CPF and the verb were together in a single AccP, which following Jun and Oh (1996) was predicted to be associated with an open question reading. In patterns (ii) and (iii), the CPF and the final verb were separated by an AccP boundary. Following Jun and Oh, this was predicted to be associated with a polar question reading.

[4]http://www.helsinki.fi/~lennes/praat-scripts/public/collect_pitch_data_from_files.praat, accessed 3 January 2015.

Table 2: Short utterances: Phrase boundary patterns and overall frequency

	AccP pattern	CPF + verb	Prediction (Jun and Oh)	Frequency $n = 100$
(i)	•Intro•CPF verb•	together	open	50
(ii)	•Intro•CPF•verb•	separate	polar	48
(iii)	•Intro CPF•verb•	separate	polar	2

As seen in Table 3, there was no categorical association between the prosodic patterns and open or closed readings. Overall, 66/100 utterances were according to the prediction. However, there was variation between the different question sets, from 15/18 (83%) utterances following prediction for set H to 9/16 (56%) following prediction for sets B and D.

Table 3: Short utterances: Variation in pattern frequency between stimulus sets

Stimulus Set	Open CPF+verb together	Open CPF/verb separate	Polar CPF+verb together	Polar CPF/verb separate	In line with prediction	Contra prediction
B	3	5	2	6	9	7
D	5	3	4	4	9	7
F	5	3	2	6	11	5
H	9	0	3	6	15	3
J	8	0	4	4	12	4
L	3	6	2	7	10	8
Totals	33	17	17	33	66	34

Results: Phrase boundary placement in long utterances The AccP boundary patterns seen for long utterances are given in Table 4. None of the 100 utterances had the CPF and verb in the same AccP. This suggests that Jun and Oh's account as originally framed does have the problem illustrated in Figure 5 in Section 2. Additionally, all of the long utterances had an AccP boundary at the right edge of the CPF. Accordingly, I reframed the account in terms of the absence or presence of an AccP boundary at the left edge of the verb, with the prediction that this would be associated with an open or polar reading respectively.

Table 4: Long utterances: Phrase boundary patterns and overall frequency

	AccP pattern	Vb = AccP	Prediction	Freq. $n = 100$
(i)	•Intro•CPF•intervening•constituent•verb• (1, 2, 3, 4, 5)	yes	polar	48
(ii)	•Intro•CPF•intervening constituent•verb• (1, 2, 3, 4)	yes	polar	35
(iii)	•Intro•CPF•intervening•constituent verb• (1, 2, 3, 4)	no	open	15
(iv)	•Intro•CPF•intervening constituent verb• (1, 2, 3)	no	open	1
(v)	•Intro CPF•intervening•constituent•verb• (1, 2, 3, 4)	yes	polar	1

Overall 84/100 utterances had an AccP boundary at the left edge of the verb, and 16/100 had no AccP boundary at the left edge of the verb. As with the short utterances, there was no categorical distinction between open and polar readings, and considerable variation between the stimulus sets, shown in Table 5. Frequencies are too small for full statistical analysis, but note that for stimulus sets B and L, every utterance for both open and polar readings was produced with an AccP boundary at the left edge of the verb. Set J, which showed the greatest distinction between open and polar utterances, had 75% of utterances in line with the prediction.

Table 5: Long utterances: Variations in pattern frequency between stimulus sets

Stimulus Set	Open		Polar		In line with predicton	Contra prediction
	Verb **not** start AccP	Verb starts AccP	Verb **not** start AccP	Verb starts AccP		
B	0	8	0	8	8	8
D	3	5	1	7	10	6
F	1	7	0	8	9	7
H	0	9	1	8	8	10
J	7	1	3	5	12	4
L	0	9	0	9	9	9
Totals	11	39	5	45	56	44

Results: Pitch variance The utterances were also analysed to identify if there were any differences in pitch that might contribute to the differentiation between the two readings (cf. Kim, 2000; Yun, 2012). These results must be treated with caution as the aim of the experiment was to explore a range of conditions (e.g. constituent length) and so there was considerable variation between stimulus sets and between speakers.

Figure 7 shows the maximum pitch data, averaged across all participants, for the short utterance pair from stimulus set B, where no constituent intervenes between CPF and verb. For the open reading, only the CPF is in focus, whereas for the polar reading, the CPF and the verb are in focus. The mean pitch levels show that there is pitch movement against declination at the rightmost AccP of the site of focus. For the open reading, this is seen in the third syllable *lul*, for the polar reading this is seen at the antepenultimate syllable *syess*. The pitch range of the utterance-final LH tone is also greater for the polar reading, where the verb is in focus, than the open reading, where it is not.

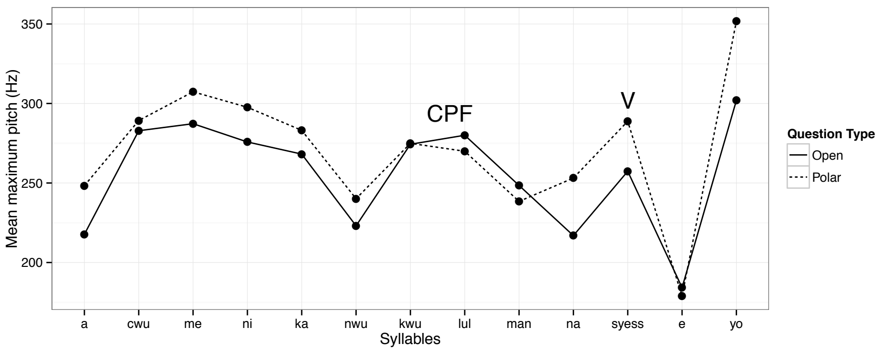


Figure 7: Mean maximum syllable pitch, all participants, set B short utterances

Figure 8 shows the corresponding average maximum pitch data for the longer pair from stimulus set B, where a constituent intervened between CPF and verb. Again, there is an overall pattern of declination through the utterance until the final LH tone. In the open reading, where only the CPF is in focus, there is movement against declination at the CPF. For the polar reading, this is not the case, but the mean pitch levels at syllables *na.syess* in the final verb are higher than in the open reading, and the pitch range of the final LH tone is also greater.

For both readings, there is a further pitch elevation within the constituent that intervenes between the CPF and verb. Discussions with native-speaker language informants suggest that the weight of the intervening constituent might make its position between the object and the verb marked, a phenomenon comparable to heavy-NP shift, but this requires more investigation.

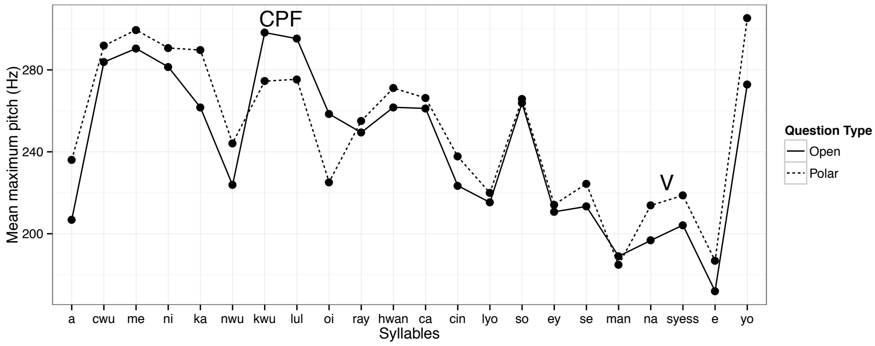


Figure 8: Mean maximum syllable pitch, all participants, set B long utterances

Discussion In summary, the experimental results do not support the categorical account proposed by Jun and Oh (1996), even when reframed to take account of intervening constituents. Without a constituent between CPF and verb, 66/100 utterances followed the prediction. Where a constituent intervened, 56/100 utterances followed the prediction. Instead, the results indicate that this is a gradient phenomenon. The differences between question sets suggest that AccP boundary placement in questions with CPFs is determined by factors other than the open/polar reading, which might include the length of preceding constituents or AccPs, and variation between individual speakers.

Based on this initial exploration, I have drawn the following tentative conclusions about the prosodic disambiguation of open and polar readings of questions with content pro-forms in Korean. First, the account of disambiguation provided by Jun and Oh (1996) is not sufficient to explain the observed data. Second, there appears to be a link between the rightmost AccP in the focused constituent and pitch peaks that move against declination. This may be similar to the expanded pitch range for focus described by Peng et al. (2005) for Mandarin. Finally, although I did not see evidence of post-focal dephrasing, it may be that some degree of pitch compression following a focused constituent may affect the placement of subsequent AccP boundaries. This could explain the AccP boundary patterns reported by Jun and Oh (1996), but also the gradient nature of the phenomenon seen in this experiment.

4 An LFG analysis of the phenomenon

This section provides an LFG analysis of the phenomenon in (1) by formally defining the relationship between prosodic expression of focus and the differing information structures of the open and polar readings. The analysis follows the framework used by Mycock and Lowe (2013) to describe prosodic expression of focus in English. A brief summary of the model is given here: for a more thorough discussion please refer to the original paper.

The analysis assumes the structural elements given in Table 6. It also assumes that there are semantic features of focus and question semantics, which each have c-structure and p-structure reflexes. The syntactic scope of focus is denoted by the feature **DF_Focus**, and question semantics within the c-structure by the feature **Sem_QSem**. The counterparts for prosodic exponence of these features are *DF_Focus* and *Sem_QSem* respectively. Language-specific cascade rules govern the presence of these features in the edge sets of s- and p-string items.

Table 6: Structural elements of the analysis

Element	Derivation
constituent structure (c-structure)	Language-specific phrase structure rules.
prosodic structure (p-structure)	Language-specific prosodic phrase structure rules.
syntactic string (s-string)	Lexical entries.
phonological string (p-string)	Lexical entries, language-specific phonological rules.
edge sets	These are a property of each s- and p-string item. They contain information about the left and right edges of c-structure and p-structure constituents respectively, together with semantic features derived from cascade rules.

The principle of interface harmony (Dalrymple and Mycock, 2011) determines grammaticality and is tested as follows. First, corresponding s-string and p-string units are identified. These are generated from lexical entries, with the p-string subject to language specific phonological processes that may include resyllabification, stress assignment, tone alignment etc. As a result, there may not be a 1-1 correspondence between s-string and p-string units[5], but the lines of correspondence between the s-string and p-string units should not cross. The edge sets of the corresponding s-string and p-string units are then compared, L with L and R with R. Interface harmony holds if the semantic features are coherent, i.e. if **DF_Focus** appears in an s-string unit's R edge set and *DF_Focus* in the R edge set of the corresponding p-string unit.

[5]For example, a prosodic unit may relate to more than one syntactic constituent, such as the syllable [hi:z] in the sentence *He's coming tomorrow*.

4.1 Structural rules for Korean

The following structural elements are used for the analysis. The c-structure rules in (11) are a subset of those proposed by Sohn (1999).

(11) a. $\mathrm{S} \rightarrow \left(\begin{array}{c} \mathrm{NP} \\ \uparrow=\downarrow \end{array} \right) \begin{array}{c} \mathrm{VP} \\ \uparrow=\downarrow \end{array}$

b. $\mathrm{NP} \rightarrow \begin{array}{c} \mathrm{N'} \\ \uparrow=\downarrow \end{array}$

c. $\mathrm{N'} \rightarrow \begin{array}{c} \mathrm{N*} \\ \uparrow=\downarrow \end{array} \begin{array}{c} \mathrm{N} \\ \uparrow=\downarrow \end{array}$

d. $\mathrm{VP} \rightarrow \begin{array}{c} \mathrm{NP*} \\ \uparrow=\downarrow \end{array} \begin{array}{c} \mathrm{V'} \\ \uparrow=\downarrow \end{array}$

e. $\mathrm{V'} \rightarrow \left(\begin{array}{c} \mathrm{NP} \\ \uparrow=\downarrow \end{array} \right) \begin{array}{c} \mathrm{V} \\ \uparrow=\downarrow \end{array}$

The p-structure rules governing the prosodic hierarchy (12) and phonological processes (13)-(15) are derived from Jun (2005). Prosodic words are not included as they play no role in marking phrase boundaries.

(12) Timing tier: p-structure
 a. IntP $\rightarrow$ AccP^{+}
 b. AccP $\rightarrow$ Syll^{+}

(13) Timing tier: final syllable lengthening
 Syll $\rightarrow$ Syll: / ___#

(14) Intonation tier: edge tones
 a. IntP $\rightarrow$ ___%##
 b. AccP $\rightarrow$ #TH_____LH#

(15) Intonation tier: assimilation of IntP final tones
 H $\rightarrow$ Ø/ ___%##

Cascade rules[6] for question semantics (following Dalrymple and Mycock, 2011) are shown below for syntactic scope (16) and prosodic exponence (17). They are based on the link made by Jun (2005) to the semantic and pragmatic function of IntP-final tones, of which question semantics is taken to be an instance.

[6]Operators ⭸R and ↘R denote the right edge set of the rightmost terminal node within the constituent within c- and p-structures respectively (Mycock and Lowe, 2013).

(16) S $\rightarrow$ NP VP

$\left(\textbf{Sem\_Qsem} \in (\Searrow \text{R})\right)$

(17) IntP $\rightarrow$ AccP* AccP

(% = TONE LH) $\Rightarrow$

Sem_QSem $\in$ ($\searrow$R)

Cascade rules for focus (following Mycock and Lowe, 2013) are shown below, again for syntactic scope (18) and prosodic exponence (19). The prosodic exponence rules are derived from the experimental evidence assuming that expanded pitch range is marked by the phonological feature PITCH = EXP. A new operator $\wr$R is proposed for Korean, which shows the correspondence between expanded pitch range and the presence of *DF_Focus* in the p-string R edge set. This was not used in Mycock and Lowe's account of English, where nuclear pitch accent plays a role in exponence of focus.

(18) $\Sigma \rightarrow \Sigma^*$ Σ^* Σ Σ^*

	$(\uparrow_\sigma$ DF) = FOCUS	**DF_Focus** $\in$ ($\Searrow$ R)	
	$(\downarrow_\sigma$ DF) = FOCUS	$(\uparrow_\sigma$ DF) = FOCUS	
		$(\downarrow_\sigma$ DF) = FOCUS	

(19) AccP $\rightarrow$ Syll* Syll^{+} Syll

(PITCH = EXP) $\Rightarrow$

DF_Focus $\in$ ($\wr$R)

5 Analyses of sentences

Analyses using the framework in Section 4.1 are given for example (20), the shorter utterances of experimental stimulus set B. Mean pitch data for these utterances was given in Figure 7.

(20) *acwumeni-ka* *nwukwu-lul* *manna-syess-eyo*
auntie-SBJ who/someone-OBJ meet-PST-POL
"Who did auntie meet? / Did auntie meet someone?"

The full analysis of the open reading 'Who did auntie meet?' is given at the end of the paper in Figure 11. A larger-scale extract from this analysis, showing the right-hand edge of the focused constituent *nwukwu-lul* 'who-OBJ' is given in Figure 9.

In the c-structure, the feature **DF_Focus** cascades according to rule (18) to the R edge set of the s-string unit 'nwukwulul'. Expanded pitch range is seen in the

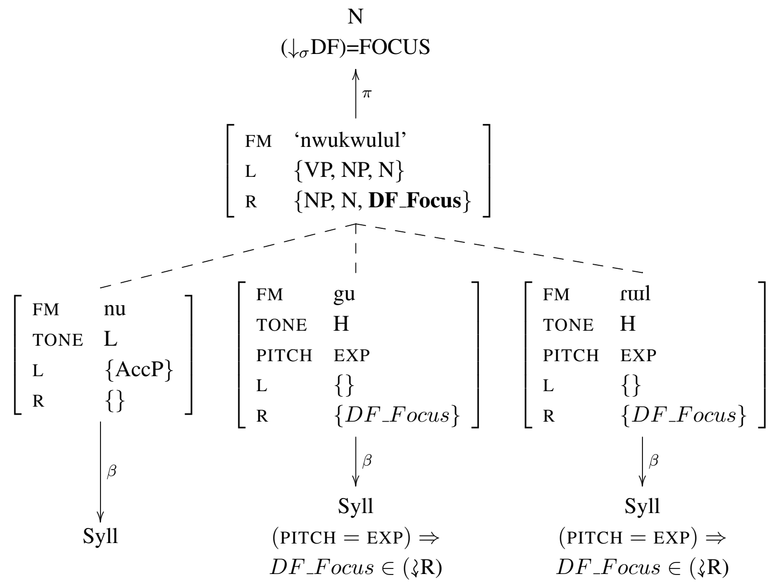


Figure 9: Focused constituent in the open reading (Extract from Figure 11)

syllables [gu] and [ɾɯl] and, following cascade rule (19), the feature *DF_Focus* appears in the R edge sets of these syllables.

Question semantic features are shown in shown in Figure 11. The question semantics s-structure feature **Sem_Qsem** cascades according to rule (16) to the R edge set of the rightmost s-structure element in the utterance, 'mannasyesseyo'. In the p-structure, prosodic expression of question semantics is given by the LH tone on the final syllable of the final AccP. Following rule (17), this places feature *Sem_Qsem* into the R edge set of its rightmost syllable, [joː].

Interface harmony is tested by comparing the L and R edge sets of corresponding s- and p-string units. As there is a 1:3 relationship between the s-string 'nwukwulul' and its syllables [nu.gu.ɾɯl], comparison is made between the L and R edge sets of 'nwukwulul' and the unions of the three syllables' L and R edge sets. The feature **DF_Focus** is found in the s-string R edge set, and its counterpart *DF_Focus* in the p-string union R edge set. A similar process is followed for 'mannasyesseyo' and its syllables [man.na.ʃʌs.sʌ.joː], where features **Sem_Qsem** and *Sem_Qsem* are found in the R edge sets of 'mannasyesseyo' and the union of its syllables respectively. Accordingly, the principle of interface harmony is upheld and the utterance is grammatical.

A similar process is followed for the analysis of the polar reading, 'Did auntie meet someone?' The full analysis is given at the end of the paper in Figure 12 and a larger-scale extract of the final constituent at Figure 10. As was the case for

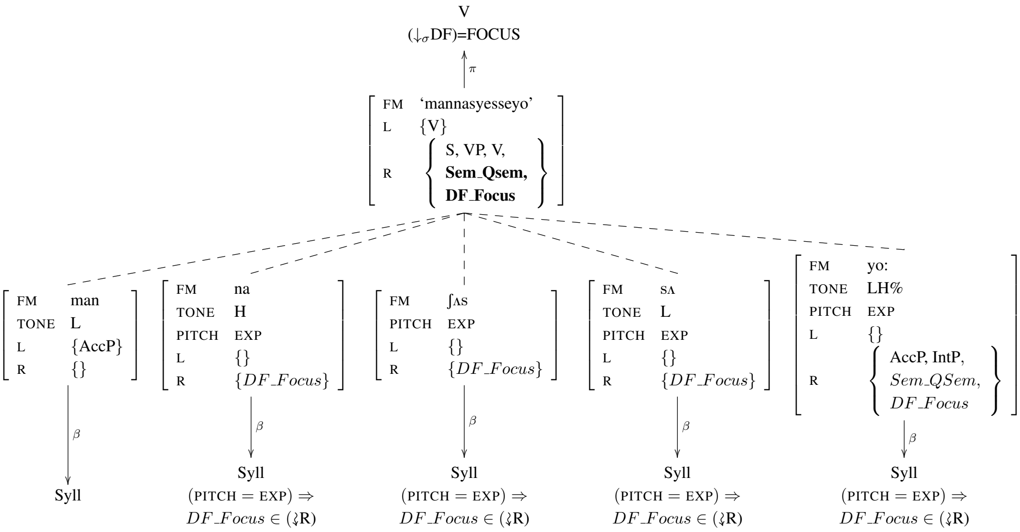


Figure 10: Focused constituent in the polar reading (Extract from Figure 12)

the open reading, rules (16) and (17) place features **Sem_QSem** and *Sem_QSem* into the R edge sets of 'mannasyesseyo' and [joː] respectively. From rule (18), the R edge set of 'mannasyesseyo' also contains feature **DF_Focus**. Prosodically, expanded pitch range is seen on syllables [na.ʃʌs.se.joː] and following rule (19), this places feature *DF_Focus* in their R edge sets. Comparing the corresponding edge sets as for the open reading, the principle of interface harmony is again seen to apply.

6 Conclusions

The analysis gives an account of the prosodic contribution to semantics by analysing the syntax-prosody interface. The method used does not assume that syntax determines prosody, but rather that the two are mutually constraining. It offers a way to unify the various accounts of the phenomenon in Korean and shows that the model of the syntax-prosody interface proposed for English by Mycock and Lowe (2013) can be adapted for Korean, including the introduction of a new operator to describe particular features of focus expression.

The experimental data from which the LFG analysis is derived led to tentative conclusions. Further experiments, designed specifically to gather pitch information, are necessary to explore the exact nature of EXPANDED PITCH RANGE, and the possible presence and nature of post-focal pitch compression. Other areas for exploration include multiple wh-questions, and other Korean prosodic phenomena.

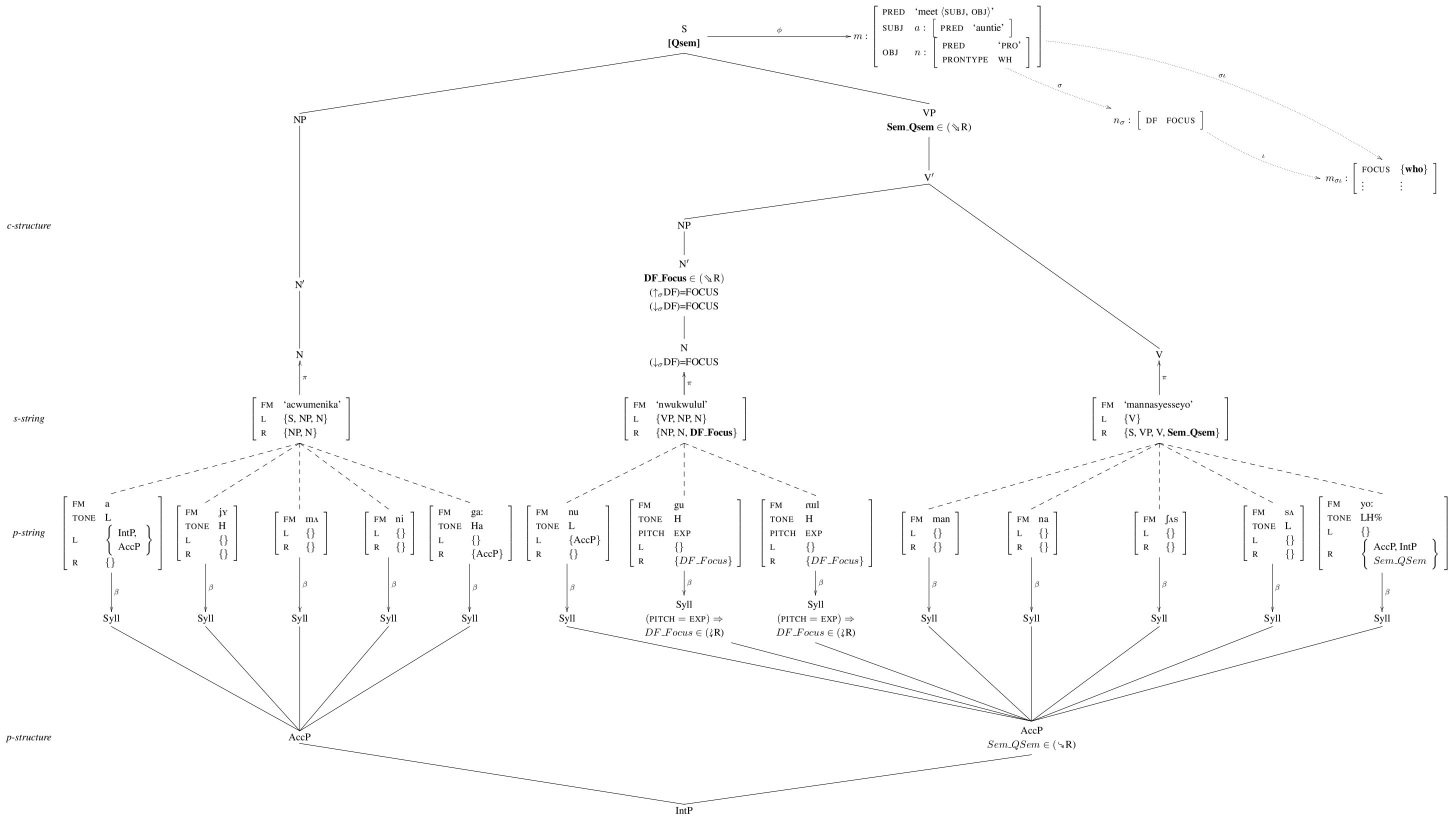


Figure 11: Short open question: ‘Who did auntie meet?’

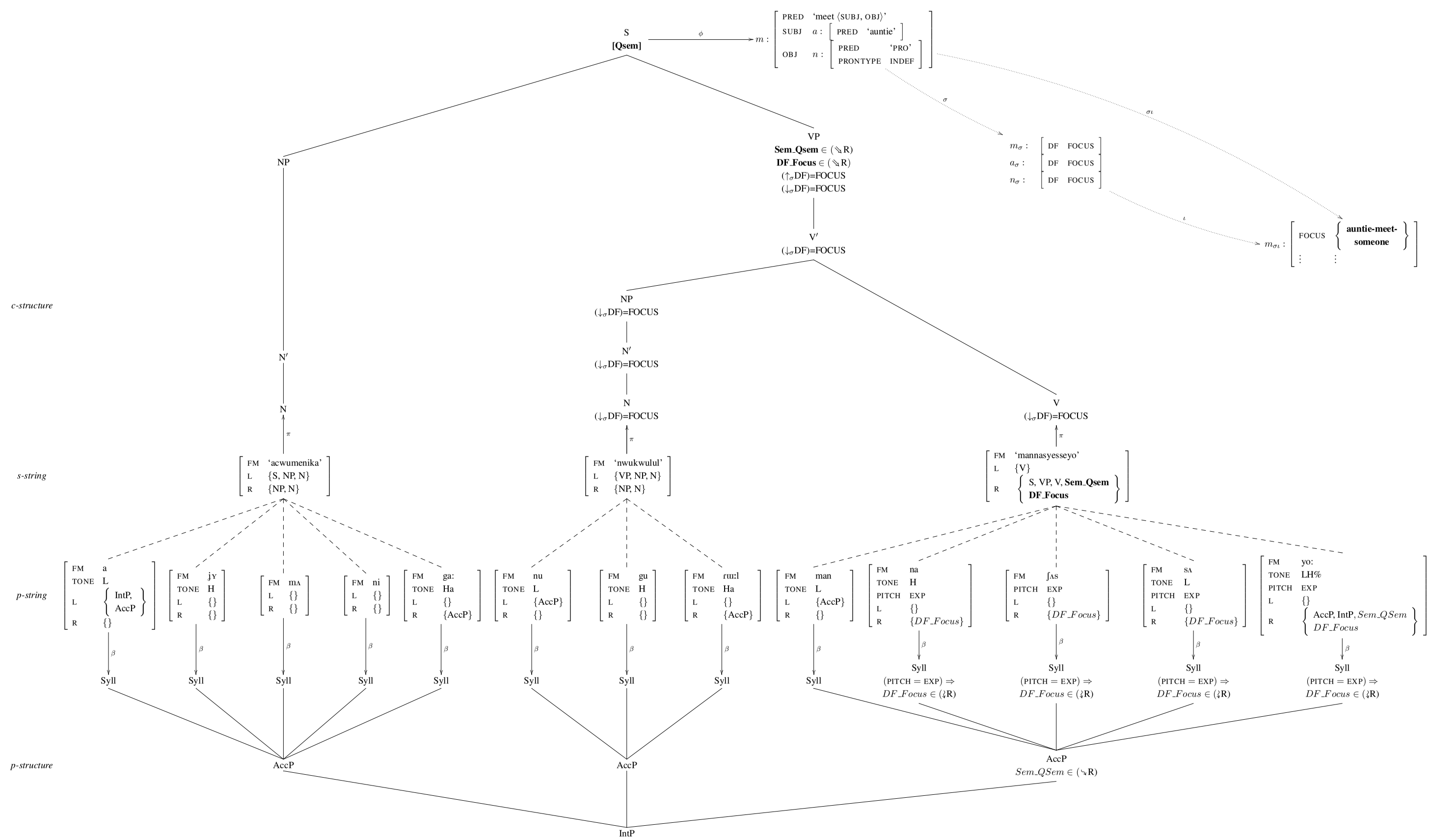


Figure 12: Short polar question: 'Did auntie meet someone?'

References

Boersma, P. and Weenink, D. (2016). Praat: doing phonetics by computer [computer program]. version 6.0.21. http://www.praat.org, retrieved 2016/12/03.

Dalrymple, M. (2001). *Lexical Functional Grammar*. Academic Press, San Diego.

Dalrymple, M. and Mycock, L. (2011). The prosody-semantics interface. In Butt, M. and King, T. H., editors, *Proceedings of LFG11*, Stanford, CA. CSLI Publications.

Goldsmith, J. A. (1976). *Autosegmental phonology*. PhD thesis, Massachusetts Institute of Techonology, Cambridge, MA.

Hayes, B. (1989). The prosodic hierarchy in meter. In Kiparsky, P. and Youmens, G., editors, *Perspectives on Meter*, pages 203–260. Academic Press, New York.

Jun, S.-A. (2005). Korean intonational phonology and prosodic transcription. In *Prosodic typology: the phonology of intonation and phrasing*, chapter 8, pages 202–228. Oxford, UK: Oxford University Press.

Jun, S.-A. and Oh, M. (1996). A prosodic analysis of three types of wh-phrases in Korean. *Language and Speech*, 39(1):37–61.

Kim, A.-R. (2000). *A derivational quantification of "wh-phrases"*. PhD thesis, Indiana University.

Martin, S. E. (1992). *A reference grammar of Korean: a complete guide to the grammar and history of the Korean language*. Tuttle, Rutland, VT.

Mycock, L. and Lowe, J. (2013). The prosodic marking of discourse functions. In Butt, M. and King, T. H., editors, *Proceedings of LFG13*, Stanford, CA. CSLI Publications.

Nespor, M. and Vogel, I. (1986). *Prosodic phonology*. Foris.

Peng, S.-h., Chan, M. K. M., Tseng, C.-y., Huang, T., Lee, O. J., and Beckman, M. E. (2005). Towards a Pan-Mandarin system for prosodic transcription. In Jun, S.-A., editor, *Prosodic typology: the phonology of intonation and phrasing*, chapter 9, pages 230–270. Oxford University Press.

Selkirk, E. O. (1984). *Phonology and syntax: the relation between sound and structure*. MIT Press.

Sohn, H.-M. (1999). *The Korean language*. Cambridge University Press, Cambridge, UK.

Yun, J. (2012). The deterministic prosody of indeterminates. In Choi, J., editor, *Proceedings of the 29th West Coast Conference on Formal Linguistics*, pages 285–293, Somerville, MA. Cascadilla Proceedings Project.

Theoretical linguistics and grammar engineering as mutually constraining disciplines

Tracy Holloway King
A9 Amazon

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 339–359

Keywords: grammar engineering, indeterminacy, restriction, LFG, HPSG

King, Tracy Holloway. 2016. Theoretical linguistics and grammar engineering as mutually constraining disciplines. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 339–359. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.18.

Abstract

This paper desribes four areas in which grammar engineers and theoretical linguists can interact. These include: using grammar engineering to confirm linguistic hypotheses; linguistic issues highlighted by grammar engineering; implementation capabilities guiding theoretical analyses; and insights into architecture issues. It is my hope that we will see more work in these areas in the future and more collaboration among grammar engineers and theoretical linguists. This is an area in which HPSG and LFG have a distinct advantage, given the strong communities and resources available.

1 Introduction

LFG and HPSG are in the privileged position of having not only a community of theoretical linguists but also of grammar engineers, with significant crossover between the theoretical and grammar-engineering communities. Grammar engineering involves the implementation of linguistically-motivated grammars so that natural language utterances and text can be processed to produce deep syntactic, and sometimes semantic, structures. In this paper, I outline four areas in which grammar engineering and theoretical linguistics interact (see also King (2011)). These are areas in which significant contributions have already been made and in which I foresee the possibility of even greater impact in the future.

Both LFG and HPSG have large-scale grammar engineering projects which span across institutions and across typologically-distinct languages. The projects test the underlying tenets of the theories, especially their universality across a broad range of languages. In addition, the projects build resources to support applications requiring the structures provided by the theories such as machine translation, question answering, summarization, and language teaching. The LFG-based ParGram and ParSem projects began with English, French, and German and have been expanded to include Japanese, Norwegian, Hungarian, and Indonesian among others. The ParGram grammars (Butt et al., 1999, 2002) are written within the LFG linguistic framework and with a commonly-agreed-upon set of grammatical features, using XLE (Crouch et al., 2011) as a grammar development platform. ParSem develops semantic structures based on the ParGram syntactic structures; most of the ParSem systems use the XLE XFR (transfer) system (Crouch and King, 2006), although some use a Glue Semantics implementation (Dalrymple, 2001; Asudeh et al., 2002). There are two HPSG-based grammar engineering projects which share these same goals: DELPH-IN and CoreGram. DELPH-IN and the LinGO Grammar Matrix (Bender et al., 2002, 2010) is a framework for the development

[†]I would like to thank the audience of HeadLex 2016 for lively discussion, the two anonymous reviewers for their detailed comments, and especially the HeadLex 2016 organizers for inviting me and for bringing together the theoretical and grammar engineering communities of HPSG and LFG. I would also like to thank the ParGram community who have been such an important part of my life for two decades.

of broad-coverage, precision, implemented grammars for diverse languages. The project builds from experience with the broad-coverage implemented HPSG grammars of English (the LinGO ERG (Flickinger, 2000)), German (DFKI's grammar (Müller and Kasper, 2000)), and Japanese (the JACY grammar (Siegel and Bender, 2002)) to extract components that are common across these grammars and therefore may be useful in the development of new grammars. They facilitate the development of grammars for different languages, producing semantic representations in a common format (MRS (Copestake et al., 2005)). The CoreGram (Müller, 2013, 2015) project is a multilingual grammar engineering project that develops HPSG grammars for typologically diverse languages, including German, Chinese, Danish, Maltese, and Persian. These grammar share a common core and are implemented in TRALE.

In the remainder of this paper I discuss four areas in which grammar engineering interacts with theoretical linguistics, illustrating these with examples from the LFG-based ParGram project. These include: using grammar engineering to confirm linguistic hypotheses (section 2); linguistic issues highlighted by grammar engineering (section 3); implementation capabilities guiding theoretical analyses (section 4); and insights into architecture issues (section 5). It is my hope that we will see more work in these areas and more collaboration among grammar engineers and theoretical linguists. This is a domain in which HPSG and LFG as a distinct advantage compared to many other linguistic theories, given the strong communities and resources available.

2 Grammar Engineering to Confirm Hypotheses: Indeterminacy by Underspecification

Grammar engineering can be used to confirm linguistic hypotheses (Butt et al., 1999; Bender, 2008; Bender et al., 2011; Fokkens, 2014). By implementing a fragment of a grammar that focuses on the hypothesis in question, the linguist can explore the details of the analysis and understand whether key issues or corner cases have been missed in the initial analysis. One caveat for this approach is that limitations of the grammar engineering platform may limit what types of hypotheses can be tested, e.g. in XLE there is no implementation of standard Lexical Mapping Theory (see LMT references in Dalrymple (2001)) and so testing hypotheses about LMT via grammar engineering is difficult.

In general this approach has been taken by linguists who work both in theoretical and computational linguistics and hence are able to straightforwardly test theoretical hypotheses through grammar engineering. However, this is a fruitful area for collaboration between theoretical linguists and grammar engineers. Some examples of LFG and HPSG work which used grammar engineering to confirm theoretical hypotheses include: Bender (2010)'s work on Wambaya which takes Nordlinger (1998a,b)'s detailed LFG analysis of Wambaya (morpho-)syntax and implements it in HPSG; Butt et al. (1997)'s work on extensions of LFG's Linking

Theory; Asudeh (2004)'s work on the analysis of resumptive pronouns using Glue semantics; Crysmann (2015)'s work on Hausa resumption and extraction and Crysmann (2016)'s work on Hausa tone and the phonology-syntax interface in HPSG; Beyaev et al. (2015)'s work on adjective coordination in LFG; Sadler et al. (2006) and Villavicencio et al. (2005)'s work on agreement with coordinated nouns in Brazilian Portuguese; Müller (1999)'s work on German syntax.

In this section, I review Dalrymple et al. (2009)'s proposal for handling indeterminacy by the underspecification of features (see Ingria (1990) on indeterminacy in general, Dalrymple and Kaplan (2000) on indeterminacy in LFG, and Crysmann (2005) and references therein on indeterminacy in HPSG). Dalrymple et al. (2009) examines the formal encoding of feature indeterminacy, focussing on case. Forms that are indeterminately specified for the value of a feature can simultaneously satisfy conflicting requirements on that feature and thus are a challenge to constraint-based formalisms which model the compatibility of information carried by linguistic items by combining or integrating that information. Dalrymple et al. (2009) views the value of an indeterminate feature as a complex and possibly underspecified feature structure. This complex feature structure allows for the incremental, monotonic refinement of case requirements in particular contexts. The proposed structure uses only atomic boolean-valued features. It covers the behaviour of both indeterminate arguments and indeterminate predicates (i.e. predicates placing indeterminate requirements on their arguments).

German has four cases (nominative, accusative, dative, genitive). Many nouns are fully specified for case; that is, they can only be interpreted as being a single case. However, some nouns are indeterminant for case. (1) shows an example of a noun which is indeterminate for all four German cases.

(1) Papageien
parrots
NOM/ACC/DAT/GEN
'parrots' (nominative, accusative, dative, or genitive) [deu]

In (2a) the indeterminate noun is the object of a verb which requires accusative case, while in (2a) the same noun is the object of a verb which requires dative case.

(2) a.

Er	findet	Papageien.
he	finds	parrots
	OBJ=ACC	NOM/ACC/DAT/GEN

'He finds parrots.' [deu]

b.

Er	hilft	Papageien.
he	helps	parrots
	OBJ=DAT	NOM/ACC/DAT/GEN

'He helps parrots.' [deu]

The data in (2) could be indicative of indeterminate case on a noun or of an ambiguously case-marked form. What distinguishes indeterminate forms is that they can simultaneously satisfy more than one case. This can be seen in (3). The indeterminate form can be the object of coordinated verbs, one which requires accusative case on its object (as in (2a)) and one which requires dates case on its object (as in (2b)).

(3)	Er	findet	und	hilft	Papageien.
	he	finds	and	helps	parrots
		OBJ=ACC		OBJ=DAT	NOM/ACC/DAT/GEN

‘He finds and helps parrots’ [deu]

The question is how to analyze indeterminate case so that the shared object can simultaneously satisfy the requirements to be dative and accusative. The proposal is to have case be a feature structure which for some nouns is indeterminate. For nouns like *Papageien* the only information that is available as to the case of the noun in the underlying form is that it must have a case, similar to all nouns heading noun phrases in German. It is only within a specific linguistic construction that the case values are specified. Determinate forms would have case feature values such as in (4), while indeterminate nouns have case features values such as in (5).[1]

(4) Determinate accusative case: Determinate dative case:

$$\left[\text{CASE}\ \begin{bmatrix}\text{NOM} & - \\ \text{ACC} & + \\ \text{GEN} & - \\ \text{DAT} & -\end{bmatrix}\right] \qquad \left[\text{CASE}\ \begin{bmatrix}\text{NOM} & - \\ \text{ACC} & - \\ \text{GEN} & - \\ \text{DAT} & +\end{bmatrix}\right]$$

(5) Indeterminate case:

$$\left[\text{CASE}\ \begin{bmatrix}\text{NOM} \\ \text{ACC} \\ \text{GEN} \\ \text{DAT}\end{bmatrix}\right]$$

Consider the situation with an indeterminate noun occuring with the same dative assigning verb in (6). The noun initially has no case specification and so within the f(unctional)-structure it appears with only CASE DAT=+ due to the case assignment from the verb. This works similarly for a verb taking an accusative object when the noun is indeterminate, only with an ACC=+ specification.

(6)	a.	Er	hilft	Papageien.
		he	helps	parrots
			OBJ=DAT	NOM/ACC/GEN/DAT

‘He helps parrots.’ [deu]

[1]Partially indeterminate forms would have some values of the case features specified, but not all.

b. $$\left[\begin{array}{ll} \textsc{obj} & \left[\begin{array}{ll} \textsc{pred} & \text{'parrots'} \\ \textsc{case} & \left[\begin{array}{ll} \textsc{nom} & \\ \textsc{acc} & \\ \textsc{gen} & \\ \textsc{dat} & + \end{array}\right] \end{array}\right] \end{array}\right]$$

When coordinating a verb taking a dative object with a verb taking an accusative object there is no clash, as shown in (7). If the nouns had been specified as having CASE=ACC or CASE=DAT there would be a clash of feature values. Similarly, a fully determinate noun would be ungrammatical because the DAT=− specification would clash with DAT=+ and vice versa for ACC. (The indices on the f-structures, e.g. the *1* in (7), indicate re-entrancy, i.e. an f-structure shared across two or more parts of the larger f-structure.)

(7) a. Er findet und hilft Papageien.
he finds and helps parrots
OBJ=ACC OBJ=DAT NOM/ACC/DAT/GEN
'He finds and helps parrots' [deu]

b. $$\left\{\begin{array}{l} \left[\begin{array}{ll} \textsc{pred} & \text{'find'} \\ \textsc{obj} & \left[\begin{array}{ll} \textsc{pred} & \text{'parrots'} \\ \textsc{case} & \left[\begin{array}{ll} \textsc{nom} & \\ \textsc{acc} & + \\ \textsc{gen} & \\ \textsc{dat} & + \end{array}\right] \end{array}\right]_{\mathit{1}} \end{array}\right] \\ \left[\begin{array}{ll} \textsc{pred} & \text{'help'} \\ \textsc{obj} & [\ ]_{\mathit{1}} \end{array}\right] \end{array}\right\}$$

Next consider how adjectival modification interacts with indeterminacy. An unambiguously dative modifier like *alten* 'old' imposes a negative specification for ACC in addition to the positive specification for DAT. This ACC − clashes with the ACC + of the accusative-taking verb *findet*, as in (8).

(8) a. *Er findet und hilft alten Papageien.
he finds and helps old parrots
OBJ=ACC OBJ=DAT DAT NOM/ACC/DAT/GEN
'He finds and helps old parrots.' [deu]

b. Ill-formed f-structure:

$$\begin{bmatrix} \textsc{pred} & \text{'parrots'} \\ \textsc{adjunct} & \left\{\begin{bmatrix} \textsc{pred} & \text{'old'} \end{bmatrix}\right\} \\ \textsc{case} & \begin{bmatrix} \textsc{nom} & - \\ \textsc{acc} & +/- \\ \textsc{gen} & - \\ \textsc{dat} & + \end{bmatrix} \end{bmatrix}$$

In contrast to determinate adjectives like *alt* 'old', the adjective *rosa* 'pink' is fully indeterminate, and imposes no case restrictions on the noun it modifies.

(9) rosa: [no case restrictions]

This means that the noun *Papageien* can be modified by *rosa* and still satisfy simultaneous accusative and dative requirements, as in (10).

(10) a.

Er	findet	und	hilft	rosa	Papageien.
he	finds	and	helps	pink	parrots
	OBJ=ACC		OBJ=DAT		NOM/ACC/DAT/GEN

'He finds and helps pink parrots.' [deu]

b.
$$\begin{bmatrix} \textsc{pred} & \text{'parrots'} \\ \textsc{adjunct} & \left\{\begin{bmatrix} \textsc{pred} & \text{'pink'} \end{bmatrix}\right\} \\ \textsc{case} & \begin{bmatrix} \textsc{nom} & \\ \textsc{acc} & + \\ \textsc{gen} & \\ \textsc{dat} & + \end{bmatrix} \end{bmatrix}$$

This appeared to be a plausible analysis, but before proposing it, to confirm the analysis, we implemented a grammar fragment in XLE and developed a testsuite (Chatzichrisafis et al., 2007) with one instance of each adjective, determiner, noun, and verb type. We then ran all the (un)grammatical sentences composed of these lexical items in order to see whether the predictions held. Grammatical sentences should be accepted by the grammar and ungrammatical ones rejected. In this case, the implementation confirmed that our proposed analysis captured the data. This was especially helpful for untangling how adjectives and determiners combinations with indeterminate nouns in different syntactic positions.

3 Linguistic Issues Highlighted by Grammar Engineering: Copulas, Adjectives, and Subjects

Implementing a broad coverage grammar requires, by definition, analyzing a wide range of syntactic phenomena in such a way that they interact correctly with one another. This contrasts with theoretical linguistics which tends to focus on a phenomenon in isolation. Analyzing a wide range of phenomena simultaneously highlights interesting facts about the language. These are often obvious in hindsight, but they fall out from implementing each part of the analyses and from working on corpora where constructions are often more complicated than they originally seemed (Bender et al., 2011). These interactions are indicative of how the formalism and theory need to be structured. As an example of where implementing a broad coverage grammar highlights linguistic issues, consider the interaction of copular constructions, predicate adjectives, and subjecthood. The topics have been debated for decades in the theoretical literature, but implementing them unearthed interactions not captured in the standard LFG analyses.

In many languages, copular constructions show predicate adjective agreement between the adjective and the subject of the copular clause, as in the French example in (11). This leads to the question of whether adjectives have subjects given their predicative nature and the agreement facts. If they do, then the adjective can agree with its subject, which in turn can be identified with the subject of the copular clause, as in (12). If they do not, then the adjective must agree with the subject of the copula, as in (13). (See Dalrymple et al. (2004a), Butt et al. (1999) and references therein for more details on the copular construction in LFG; see van Eynde et al. (2016) on using treebanks to inform theoretical HPSG analyses of copular constructions.)

(11) Elle est petite.
she.F.Sg is small.F.Sg
'She is small.' [fra]

(12) a.
$$\begin{bmatrix} \text{PRED} & \text{'be<XCOMP>SUBJ'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'pro'} \\ \text{NUM} & \text{sg} \\ \text{GEND} & \text{fem} \end{bmatrix} \mathit{1} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'small<SUBJ>'} \\ \text{SUBJ} & [\;]\mathit{1} \end{bmatrix} \end{bmatrix}$$

b. petite (↑ PRED) = 'small<SUBJ>'
(↑ SUBJ NUM) =c sg
(↑ SUBJ GEND) =c fem

c. est (↑ PRED) = 'be<XCOMP>SUBJ'
(↑ SUBJ) = (↑ XCOMP SUBJ)

(13) a. $$\begin{bmatrix} \text{PRED} & \text{'be<SUBJ,PREDLINK>'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'pro'} \\ \text{NUM} & \text{sg} \\ \text{GEND} & \text{fem} \end{bmatrix} \\ \text{PREDLINK} & \begin{bmatrix} \text{PRED} & \text{'small'} \end{bmatrix} \end{bmatrix}$$

b. petite $(\uparrow \text{PRED}) = \text{'small'}$
$((\text{PREDLINK} \uparrow)\ \text{SUBJ NUM}) =_c \text{sg}$
$((\text{PREDLINK} \uparrow)\ \text{SUBJ GEND}) =_c \text{fem}$

The open complement (XCOMP) analysis with the adjectives with subjects shown in (12) makes it easy to capture agreement of predicate adjectives with their subjects, the semantic predication relation between the adjective and the subject, and the control relations for raising adjectives, as in (14) (Dalrymple et al. (2004a), p194). This is easy to encode because the subject information is passed up with standard function application and hence becomes local, as in (14). Because of these facts, this analysis was implemented in the English ParGram grammar.

(14) a. It is likely/bound/certain to rain.

b. They are eager/foolish/loathe to leave.

(15) $$\begin{bmatrix} \text{PRED} & \text{'be<XCOMP>SUBJ'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRON-FORM} & \text{it} \end{bmatrix} 1 \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'likely<XCOMP>SUBJ'} \\ \text{SUBJ} & [\ ]1 \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'rain<>SUBJ'} \\ \text{SUBJ} & [\ ]1 \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

However, as the ParGram grammar was used to parse large corpora and served as the input to further semantic and abstract knowledge representations (Crouch and King, 2006; Bobrow et al., 2007), it was discovered that this analysis fails when the post-copular element already has a subject, as in (16) (Dalrymple et al. (2004a), p193).

(16) a. The problem is that they appear.

b. The problem is their appearing.

c. The problem is (for them) to leave before 6.

Constructions like those in (16) are incompatilbe with analyses where copulas are analyzed as taking an open complements due to the conflict as to the subject of the complement. This is shown in (17) for (16a) where the subject of *appear* should be *they*, but the open complement construction also assigns the subject of the copula *the problem* to be the subject of *appear*.[2]

(17) Open Complement

$$\begin{bmatrix} \textsc{pred} & \text{'be<}\textsc{xcomp}\text{>}\textsc{subj}\text{'} \\ \textsc{subj} & \begin{bmatrix} \textsc{pred} & \text{'problem'} \end{bmatrix} \\ \textsc{xcomp} & \begin{bmatrix} \textsc{pred} & \text{'appear<}\textsc{subj}\text{>'} \\ \textsc{subj} & \begin{bmatrix} \textsc{pred} & \text{*'they/problem'} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

When the object of the copular is a closed complement, there is no conflict for what the subject is, as shown in (18). However, this open complement analysis of this construction requires more machinery for the simple adjectival copular and raising adjective cases.

(18) Closed Complement

$$\begin{bmatrix} \textsc{pred} & \text{'be<}\textsc{predlink}\text{>}\textsc{subj}\text{'} \\ \textsc{subj} & \begin{bmatrix} \textsc{pred} & \text{'problem'} \end{bmatrix} \\ \textsc{predlink} & \begin{bmatrix} \textsc{pred} & \text{'appear<}\textsc{subj}\text{>'} \\ \textsc{subj} & \begin{bmatrix} \textsc{pred} & \text{'they'} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

How best to analyze copular constructions including their interactions with the argument structure of adjectives is still not resolved and continues to be the topic of debate among theoretical linguists and grammar engineers. The issues raised by these competing analyses could have been pursued in a purely theoretical setting, but implementing these constructions in a broad-coverage grammar clarified some of the issues, especially the interaction amongst constructions, even if a satisfactory solution has not yet been found.

4 Implementation Guiding Analysis: Complex Predicates via Restriction

Complex predicates are a major area of study in LFG (Butt, 1995; Mohanan, 1994; Butt et al., 2009) and HPSG (Müller, 2002). The fundamental issue with complex predicates is that it is not possible to exhaustively list them in the lexicon. Instead, they must be formed productively through the combination of main verbs and light verbs. There have been many theoretical linguistic proposals for how to analyze

[2]With an open complement analysis, the grammar has to create a dummy layer with a dummy PRED to protect the subject of the lower clause: an unsatisfying, unintuitive analysis. Details of this analysis are not discussed to space limitations.

complex predications, but most of them are not currently implementable in XLE (for the LFG proposals) because they involve Lexical Mapping Theory and complex operations in argument-structure. One analysis of complex predicates which is not used theoretically is to employ the restriction operator. By implementing complex predicates via restriction, it is possible to determine whether restriction is a theoretically-feasible option.

Consider the complex predicate in (19): (19) is a N(oun)-V(erb) complex predicate in which *kahAnI* 'story' is an argument which is contributed (i.e. licensed) by the noun, yet functions as the direct object of the clause. The finite verb *kI* 'did' has two arguments: Nadya and *yAd* 'memory'. The noun *yAd* 'memory' plays a double role: it is an argument of the finite verb, yet it also contributes to the overall predication of the clause (Mohanan, 1994).

(19) nAdiyah nE kahAnI yAd k-I
Nadya.F.Sg Erg story.F.Sg.Nom memory.F.Sg.Nom do-Perf.F.Sg
'Nadya remembered a/the story.' [urd]

In theoretical analyses, complex predicates comprise a single PRED form as in (20a) constructed from two underlying forms as in (20b) and (20c). This composition is generally argued to take place in argument structure or as a pre-syntactic operation over the lexicon.

(20) a. Standard LFG complex PRED: $(\uparrow \text{PRED}) = {'}\text{memory-do}{<}\text{SUBJ,OBJ}{>}{'}$

b. $(\uparrow \text{PRED}) = {'}\text{memory}{<}\text{OBJ}{>}{'}$

c. $(\uparrow \text{PRED}) = {'}\text{do}{<}\text{SUBJ,OBJ}{>}{'}$

Given the prevalence of complex predicates in Urdu, when implementing the Urdu ParGram grammar (Butt and King, 2002, 2007) it was imperative to analyze complex predicates from the outset. However, the XLE platform has no implementation of argument structure. Instead, the choices were (1) to include all complex predicates in the lexicon, (2) to use the lexical rules standardly used for passive formation, or (3) to explore the restriction operator (Kaplan and Wedekind, 1993). The lexicon and lexical rules options were not viable due to the productive nature of complex predicates and the types of argument changes they require. However, restriction can construct predicates on the fly, forming a new predicate form and altering the argument structure. Thus it was decided to explore this option.

First consider how the restrition operator works (Kaplan and Wedekind, 1993). The restriction operator allows for f-structure features to be "restricted out", i.e., to cause the grammar to function as if these features did not exist. A restricted f-structure is identical to the original f-structure except that it does not contain the restricted attribute. Monotonicity, which is fundamental to LFG, is still preserved since the original, non-restricted f-structure still exists.[3] An example of restricting case from the f-structure of a noun phrase is shown in (21).

(21) Original f-structure: F-structure with case restricted out:

$$\begin{bmatrix} \textsc{pred} & '\text{Nadya}' \\ \textsc{num} & \text{sg} \\ \textsc{pers} & 3 \\ \textsc{case} & \textsc{erg} \end{bmatrix} \qquad\qquad \begin{bmatrix} \textsc{pred} & '\text{Nadya}' \\ \textsc{num} & \text{sg} \\ \textsc{pers} & 3 \end{bmatrix}$$

For complex predicates, this operation can construct complex predicate-argument structures dynamically (Butt et al., 2003a, 2009). The resulting PRED contains the same information as proposed in theoretical analyses, but arranged differently. Contrast the two PREDs in (22). In the theoretical analysis in (22a) there is a single predicate *memory-do* which takes two arguments, a SUBJ and an OBJ. In the restriction analysis in (22b) the PRED is spread across *do* and *memory* but there are again two arguments, a SUBJ and an OBJ. The f-structure for (19) is shown in (23).

(22) a. Standard LFG PRED: $(\uparrow \textsc{pred}) = '\text{memory-do}{<}\textsc{subj},\textsc{obj}{>}'$

b. Restriction-based PRED: $(\uparrow \textsc{pred}) = '\text{do}{<}\textsc{subj},'\text{memory}{<}\textsc{obj}{>}'{>}'$

(23) $$\begin{bmatrix} \textsc{pred} & '\text{do}{<}\textsc{subj},'\text{memory}{<}\textsc{obj}{>}'{>}' \\ \textsc{subj} & [\ \textsc{pred}\ '\text{Nadya}'\ ] \\ \textsc{obj} & [\ \textsc{pred}\ '\text{story}'\ ] \end{bmatrix}$$

This f-structure is achieved by a dynamic composition of the subcategorization frames contributed by *kar* 'do' and *yAd* 'memory'. The restriction operator is invoked as part of the f-structure annotations on the c(onstituent)-structure rules and is represented by a backslash. Grammatical functions and attributes listed after the backslash are restricted out of the f-structure when forming the new f-structure. Any grammatical functions or attributes not mentioned are inherited by the new f-structure. (24) shows the annotated c-structure rule which creates a complex predicate from a noun and a light verb. As is standard with LFG complex predicate analyses, the N and the Vlight are both heads of the Vcp, as indicated by $\uparrow=\downarrow$, since they both contribute to form the single, complex predicate in the f-structure. For the $\uparrow=\downarrow$ on the N, the PRED is restricted out ($\backslash$PRED) and instead its PRED becomes the second argument (ARG2) of the Vcp PRED (($\uparrow$ PRED ARG2)=($\downarrow$ PRED)).

(24) $$\text{Vcp} \rightarrow \begin{array}{c} \text{N} \\ \uparrow\backslash\textsc{pred}=\downarrow\backslash\textsc{pred} \\ (\uparrow \textsc{pred arg}2)=(\downarrow \textsc{pred}) \end{array} \quad \begin{array}{c} \text{Vlight} \\ \uparrow=\downarrow \\ \ \end{array}$$

[3]This multiplicity of f-structures is often considered aesthetically unstatisfying, especially in theoretical linguistics.

In the lexicon the light verb's subcategorization contributes a subject but its second argument is incomplete, as in (25) where %Pred represents a variable to be filled in. This predicate is provided by the N in (24), e.g. the noun *yAd* 'memory'. The annotation ($\uparrow$ PRED ARG2)=($\downarrow$ PRED) in (24) substitutes the PRED value of *yAd* as the second argument of the light verb. The subcategorization frame of *yAd* is lexically specified to contribute an object, as in (26).

(25) $(\uparrow \textsc{pred}) = {'}\text{do}\!< \textsc{subj}\ \%\text{Pred} >{'}$

(26) $(\uparrow \textsc{pred}) = {'}\text{memory}\!<\!\textsc{obj}\!>{'}$

To reiterate, the restriction operator restricts out those pieces of information which are "changed" as part of complex predication, namely the PRED. When the light verb and the noun are combined, they create the PRED in (27), and the annotated c-structure rules create the f-structure in (29) from the f-structures of the N and Vlight in (28).

(27) $(\uparrow \textsc{pred}) = {'}\text{do}\!<\!\textsc{subj},{'}\text{memory}\!<\!\textsc{obj}\!>{'}\!>{'}$

(28)

$$
\begin{array}{cc}
\text{N} & \text{Vlight} \\
\downarrow \backslash \textsc{pred}{=}\uparrow \backslash \textsc{pred} & \uparrow{=}\downarrow \\
(\uparrow \textsc{pred arg2}){=}(\downarrow \textsc{pred}) & \\
\begin{bmatrix}
\textsc{pred} & {'}\text{memory}\!<\!\textsc{obj}\!>{'} \\
\textsc{num} & \text{sg} \\
\textsc{gend} & \text{fem} \\
\textsc{obj} & [\ ]
\end{bmatrix}
&
\begin{bmatrix}
\textsc{pred} & {'}\text{do}\!<\!\textsc{subj},\ \%\text{Pred}\!>{'} \\
\textsc{perf} & + \\
\textsc{num} & \text{sg} \\
\textsc{gend} & \text{fem} \\
\textsc{subj} & [\ ]
\end{bmatrix}
\end{array}
$$

(29)

$$
\begin{array}{c}
\text{Vcp} \\
\begin{bmatrix}
\textsc{pred} & {'}\text{do}\!<\!\textsc{subj},\ {'}\text{memory}\!<\!\textsc{obj}\!>{'}\!>{'} \\
\textsc{perf} & + \\
\textsc{num} & \text{sg} \\
\textsc{gend} & \text{fem} \\
\textsc{subj} & [\ ] \\
\textsc{obj} & [\ ]
\end{bmatrix}
\end{array}
$$

Based on experiences with the Urdu ParGram grammar, experiences which were driven largely out of implementational necessity, Butt et al. (2003a) demonstrated that the restriction operator is indeed able to model different types of complex predicates in the Urdu grammar and can even model cases of stacked complex predicates (Butt et al., 2009). Having a complex predicate analysis for these constructions is necessary to meet the linguistic requirements of Urdu and to allow the

Urdu ParGram analysis of these constructions to be parallel to those of the other ParGram grammars. This implementation via the restriction operator opens a new theoretical approach to complex predicates. The jury is still out as to whether this analysis is superior to existing ones, but the theory is richer for having restriction as a possible formal device for complex predicate formation.

5 Insights Into Architecture Issues: Passive-Causative Interactions

A final area in which grammar engineering informs theoretical linguistics is by providing insights into architectural issues. This arose in ParGram in the interaction of passives and causatives where complex predication via restriction occurred in conjunction with passive sublexical rules. This interaction was observed in the Turkish (Çetinoğlu, 2009; Çetinoğlu and Oflazer, 2009) and Urdu grammars. In this section the focus is on Urdu, but the same issue arises in Turkish and was noticed there first.

Causatives in Urdu are formed morphologically. The causative morpheme *-A* adds an argument, the causer, to the PRED of the verb, as in (30). With a transitive verb, the subject of the transitive is realized as the causee and is marked with the dative/accusative *kO*.

(30) a. yassIn=nE kHAnA kHa-yA
Yassin=Erg food.M.Sg.Nom eat-Perf.M.Sg
'Yassin ate food.' [urd]

b. nAdyA=nE yassIn=kO kHAnA kHil-A-yA
Nadya=Erg Yassin=Dat food.M.Sg.Nom eat-Caus-Perf.M.Sg
'Nadya had Yassin eat (fed Yassin).' [urd]

Causatives can be analyzed as complex predicates: the overall argument structure is co-determined by more than one predicational element (Alsina, 1993). The Urdu grammar treats morphologically formed causatives on a par with syntactically formed complex predicates (Butt et al., 2003b; Butt and King, 2006). The predicate-argument structure is calculated dynamically based on the information contributed by each of the predicational parts. The final subcategorization frame is created by the restriction operator. For causatives, as shown in (32), both the PRED and the SUBJ are restricted out from the verb; the causative morpheme will provide the subject for the causativized verb ((33a)). With morphological causatives, this plays out at the level of sublexical rules (see Frank and Zaenen (2004) for discussion of sublexical rules in XLE and LFG). The morphological analyzer provides the analysis in (31) for the verb *kHilvAyA* 'made to eat'. The tags are terminal nodes of sublexical rules,[4] and the *+Cause* tag provides a phrase-structure locus for the restriction operator.

[4]The _BASE notation indicates a sublexical rule. The only difference between sublexical rules

(31) eat.Causative
kHilvAyA $\Leftrightarrow$ kHA +Verb +Cause +Perf +Masc +Sg

(32) $$\begin{array}{lcc} \text{V} \rightarrow & \text{V\_BASE} & \text{CAUSE\_BASE} \\ & \downarrow\backslash\text{PRED}\backslash\text{SUBJ}{=}\uparrow\backslash\text{PRED}\backslash\text{SUBJ} & \uparrow{=}\downarrow \\ & (\downarrow\text{SUBJ}){=}\ (\uparrow\text{OBJ-GO}) & \\ & (\uparrow\text{PRED ARG2}){=}(\downarrow\text{PRED}) & \end{array}$$

(33) a. +Cause: ($\uparrow$ PRED) = $'$cause$<$ SUBJ %Pred $>'$

b. eat ($\uparrow$ PRED) = $'$eat$<$ SUBJ OBJ $>'$

The causative lexical entry in (33a) is that of a complex predicate light verb. The variable is filled by the PRED value of the main verb *kHA* 'eat' and the original subject of 'eat' is realized as the causative OBJ-GO (i.e. a goal thematic object). (34) shows the f-structures for the main verb and the causative morpheme. The resulting causative verb's f-structure is in (35).

(34) V_BASE CAUSE_BASE

$$\begin{bmatrix} \text{PRED} & '\text{eat}<\text{SUBJ,OBJ}>' \\ \text{SUBJ} & [\ ]1 \\ \text{OBJ-GO} & [\ ]2 \end{bmatrix} \quad \begin{bmatrix} \text{PRED} & '\text{cause}<\ \text{SUBJ}\ \%\text{Pred}\ >' \\ \text{AGR} & \begin{bmatrix} \text{PERF} & + \\ \text{GEND} & \text{masc} \\ \text{NUM} & 3 \end{bmatrix} \\ \text{SUBJ} & [\ ]3 \end{bmatrix}$$

(35) V

$$\begin{bmatrix} \text{PRED} & '\text{cause}<\text{SUBJ},'\text{eat}<\text{OBJ-GO,OBJ}>'> \\ \text{AGR} & \begin{bmatrix} \text{PERF} & + \\ \text{GEND} & \text{masc} \\ \text{NUM} & 3 \end{bmatrix} \\ \text{SUBJ} & [\ ]3 \\ \text{OBJ-GO} & [\ ]1 \\ \text{OBJ} & [\ ]2 \end{bmatrix}$$

Passives in Urdu are formed by combining the verb *jA* 'go' with the perfect form of the main verb. The agent of the verb is realized as an adjunct and is marked with *se* 'with/from', as shown in (36).

and standard c-structure rules is how they are displayed by the XLE system. By default sublexical rules do not show the internal structure, e.g. the V_BASE and CAUSE_BASE in (31), in the visual display produced by XLE.

(36) a. yassIn=nE kHAnA kHa-yA
Yassin=Erg food.M.Sg.Nom eat-Perf.M.Sg
'Yassin ate food.' [urd]

b. kHAnA yassIn=sE kHa-yA ga-yA
food.M.Sg.Nom Yassin=Inst eat-Perf.M.Sg go-Perf.M.Sg
'The food was eaten by Yassin.' [urd]

Now consider the interaction of the causative with passivization. When a causative is passivized, the agent of the causative is realized as an adjunct and is marked with *se*, as in (37). That is, the causative applies first and then the passive.

(37) yassIn=kO nAdyA=sE kHAnA kHil-A-yA ga-yA
Yassin=Dat Nadya=Inst food.M.Sg.Nom eat-Caus-Perf.M.Sg go-Perf.M.Sg
'The food was fed to Yassin by/through Nadya.' [urd]

However, although the rules for causatives and passives worked independently in the grammar, they did not interact properly to provide an analysis of cases where a verb was both causativized and passivized. In addition the grammar could parse the ungrammatical constructions where the indirect object (OBJ-GO) *Yassin*, which was the agent of the main verb but not the agent of the causative form, was realized as an agentive adjunct, as in (38).

(38) *nAdyA=nE yassIn=sE kHAnA kHil-A-yA ga-yA
Nadya=Erg Yassin=Inst food.M.Sg.Nom eat-Caus-Perf.M.Sg go-Perf.M.Sg
'Nadya made the food be eaten by/through Yassin.' [urd]

That is, the implemented grammar of Urdu could not analyze grammatical combinations and incorrectly provided analyses for ungrammatical ones. The underlying problem is an architectural one. Passivization has traditionally been handled by lexical rules in LFG (Bresnan, 1982). These lexical rules apply in the lexicon and hence are applied directly to the specification of subcategorization frames. For example, as shown in (39a), the transitive verb *kHA* 'eat' states that there is a predicate 'P' (eat) which has a subject and an object and which can optionally be subject to passivization. The '@' sign signals a template call to the template PASSIVE, shown in (39b) which effects the passivization via a lexical rule (see Dalrymple et al. (2004b) and Asudeh et al. (2008) on templates in LFG).

(39) a. $\text{TRANS(P)} = @(\text{PASSIVE } (\uparrow\text{PRED}){=}'\text{P}{<}(\uparrow\text{SUBJ})\ (\uparrow\text{OBJ}){>}')$

b. $\text{PASSIVE(P)} = \begin{array}{l} (\uparrow\text{SUBJ}) \longrightarrow \text{NULL} \\ (\uparrow\text{OBJ}) \longrightarrow (\uparrow\text{SUBJ}) \end{array}$

Since the passive lexical rule is specified in the lexical entry of *kHA* 'eat', passivization always applies before causativization. That is, the lexical rule is applied to the V_BASE. This is followed by the application of the causativization restriction operator in the sublexical rules in the syntax. The intuitive order of application in the original implementation is shown in (40) with passivization occuring in the lexicon and causativization in the grammar.

(40) Ungrammatical derivation of passive+causative:

a. Original Predicate: (↑PRED)=′eat<(↑SUBJ) (↑OBJ)>′

b. Lexical Rule Passive: (↑PRED)=′eat<NULL (↑SUBJ)>′

c. Restriction Causative: (↑PRED) = ′cause<SUBJ,′eat<NULL,OBL-GO>′>

However, passivization should operate on the entire complex predicate, i.e. passivization applies to the causativized verb. Once the source of the issue was identified, passivization was moved to be part of the sublexical rules and analyzed via the restriction operator. This allowed for the intuitive order of operations in (41) since the passive sublexical rule (not shown here) applies to the causative sublexical rule ((32)) in the c-structure.

(41) Grammatical derivation of passive+causative:

a. Original Predicate: (↑PRED)=′eat<(↑SUBJ) (↑OBJ)>′

b. Restriction Causative: (↑PRED) = ′cause<SUBJ,′eat<OBL-GO,OBJ>′>

c. Restriction Passive: (↑PRED) = ′cause<NULL,′eat<OBL-GO,SUBJ>′>

While the solution outlined above in which both causative and passive are handled via restriction captures the linguistic generalization, many theoretical linguists do not consider it a satisfactory analysis. In theoretical LFG, argument alternations occur in a(rgument)-structure and are independent of particular morphological or syntactic realizations. However, there is an architectural flaw in how argument alternations are treated within the computational grammar implementation. In the ParGram grammars, passivization continues to be treated via lexical rules, as per classic LFG (but see Wedekind and Ørsnes 2003) and a-structure is not implemented in the grammars. Instead, predicate-arguments are modeled solely via subcategorization frames pertaining to grammatical functions. The interaction between causativization and passivization at the morphology-syntax interface highlights how traditional lexical rules make incorrect predictions when causativization is morphological but passivization is part of the syntax.

6 Conclusion

In this paper, I explored four areas in which grammar engineers and theoretical linguists can interact. These include: using grammar engineering to confirm linguistic hypotheses (section 2); linguistic issues highlighted by grammar engineering (section 3); implementation capabilities guiding theoretical analyses (section 4); and insights into architecture issues (section 5). These are all areas in which significant contributions have been made and in which I foresee the possibility of greater impact in the future. This is an area in which HPSG and LFG as a distinct advantage, given the strong communities and resources available.

References

Alsina, Alex. 1993. *Predicate Composition: A Theory of Syntactic Function Alternations*. Ph. D.thesis, Stanford University.

Asudeh, Ash. 2004. *Resumption as Resource Mangement*. Ph. D.thesis, Stanford University.

Asudeh, Ash, Crouch, Dick and Dalrymple, Mary. 2002. The Syntax-Semantics Interface: Theory and Implementation, NASSLLI-02 course notes.

Asudeh, Ash, Dalrymple, Mary and Toivonen, Ida. 2008. Constructions with Lexical Integrity: Templates as the LexiconSyntax Interface. In *Proceedings of the LFG08 Conference*, CSLI On-Line Publications.

Bender, Emily M. 2008. Grammar Engineering for Linguistic Hypothesis Testing. In Nicholas Gaylord, Alexis Palmer and Elias Ponvert (eds.), *Proceedings of the Texas Linguistics Society X Conference: Computational Linguistics for Less-Studied Languages*, pages 16–36, CSLI Publications.

Bender, Emily M. 2010. Reweaving a Grammar for Wambaya: A Case Study in Grammar Engineering for Linguistic Hypothesis Testing. *Linguistic Issues in Language Technology* 3, 1–34.

Bender, Emily M., Drellishak, Scott, Fokkens, Antske, Poulson, Laurie and Saleem, Safiyyah. 2010. Grammar Customization. *Research on Language & Computation* 8(1), 23–72.

Bender, Emily M., Flickinger, Dan and Oepen, Stephan. 2002. The Grammar Matrix: An Open-Source Starter-Kit for the Rapid Development of Cross-Linguistically Consistent Broad-Coverage Precision Grammars. In John Carroll, Nelleke Oostdijk and Richard Sutcliffe (eds.), *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pages 8–14.

Bender, Emily M., Flickinger, Dan and Oepen, Stephan. 2011. Grammar Engineering and Linguistic Hypothesis Testing: Computational Support for Complexity in Syntactic Analysis. In Emily M. Bender and Jennifer E. Arnold (eds.), *Language from a Cognitive Perspective: Grammar, Usage and Processing*, pages 5–29, CSLI Publications.

Beyaev, Oleg, Dalrymple, Mary and Lowe, John. 2015. Number Mismatches in Coordination: An LFG Analysis. In *Proceedings of the LFG15 Conference*, pages 26–46, CSLI On-line Publications.

Bobrow, Daniel G., Cheslow, Bob, Condoravdi, Cleo, Karttunen, Lauri, King, Tracy Holloway, Nairn, Rowan, de Paiva, Valeria, Price, Charlotte and Zaenen, Annie. 2007. PARC's Bridge and Question Answering System. In *Proceedings of the Grammar Engineering Across Frameworks (GEAF07) Workshop*, pages 46–66, CSLI Publications.

Bresnan, Joan (ed.). 1982. *The Mental Representation of Grammatical Relations*. MIT Press.

Butt, Miram and King, Tracy Holloway. 2007. Urdu in a Parallel Grammar Development Environment. In T. Takenobu and C.-R. Huang (eds.), *Language Resources and Evaluation: Special Issue on Asian Language Processing: State of the Art Resources and Processing*, volume 41, pages 191–207.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. CSLI Publications.

Butt, Miriam, Dalrymple, Mary and Frank, Anette. 1997. An Architecture for Linking Theory in LFG. In *Proceedings of the LFG97 Conference*, CSLI On-line Publications.

Butt, Miriam, Dyvik, Helge, King, Tracy Holloway, Masuichi, Hiroshi and Rohrer, Christian. 2002. The Parallel Grammar Project. In *Proceedings of COLING2002 Workshop on Grammar Engineering and Evaluation*, pages 1–7.

Butt, Miriam and King, Tracy Holloway. 2002. Urdu and the Parallel Grammar Project. In *Proceedings of COLING-2002 Workshop on Asian Language Resources and International Standardization*, pages 39–45.

Butt, Miriam and King, Tracy Holloway. 2006. Restriction for Morphological Valency Alternations: The Urdu Causative. In Miriam Butt, Mary Dalrymple and Tracy Holloway King (eds.), *Intelligent Linguistic Architecturs: Variations on Themes by Ronald M Kaplan*, pages 235–258, CSLI Publications.

Butt, Miriam, King, Tracy Holloway and III, John T. Maxwell. 2003a. Complex Predicates via Restriction. In Miriam Butt and Tracy Holloway King (eds.), *The Proceedings of the LFG '03 Conference*, pages 92–104, CSLI Publications.

Butt, Miriam, King, Tracy Holloway and Maxwell III, John T.2̇003b. Complex Predicates via Restriction. In *Proceedings of the LFG09 Conference*, CSLI Online Publications.

Butt, Miriam, King, Tracy Holloway, Niño, María-Eugenia and Segond, Frédérique. 1999. *A Grammar Writer's Cookbook*. CSLI Publications.

Butt, Miriam, King, Tracy Hollowy and Ramchand, Gillian. 2009. Complex Predication: Who made the child pinch the elephant? In Linda Uyechi and Lian Hee Wee (eds.), *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, pages 231–256, CSLI Publications.

Çetinoğlu, Özlem. 2009. *A Large Scale LFG Grammar for Turkish*. Ph. D.thesis, Sabanci University.

Çetinoğlu, Özlem and Oflazer, Kemal. 2009. Integrating Derivational Morphology into Syntax. In N. Nicolov et al. (ed.), *Recent Advances in Natural Language Processing V*, John Benjamins.

Chatzichrisafis, Nikos, Crouch, Dick, King, Tracy Holloway, Nairn, Rowan, Rayner, Manny and Santaholma, Marianne. 2007. Regression Testing For Grammar-Based Systems. In *Proceedings of the Grammar Engineering Across Frameworks (GEAF07) Workshop*, pages 128–143, CSLI On-line Publications.

Copestake, Ann, Flickinger, Dan, Sag, Ivan A. and Pollard, Carl J. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language and Computation* .

Crouch, Dick, Dalrymple, Mary, Kaplan, Ron, King, Tracy, Maxwell, John and Newman, Paula. 2011. XLE Documentation, on-line documentation from the Palo Alto Research Center.

Crouch, Dick and King, Tracy Holloway. 2006. Semantics via F-Structure Rewriting. In *Proceedings of LFG06*, pages 145–165, CSLI On-line publications.

Crysmann, Berthold. 2005. Syncretism in German: A Unified Approach to Underspecification, Indeterminacy, and Likeness of Case. In Stefan Müller (ed.), *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar*, pages 91–107, CSLI On-line Publications.

Crysmann, Berthold. 2015. Resumption and extraction in an implemented HPSG of Hausa. In *Proceedings of Grammar Engineering Across Frameworks*, ACL-IJCNLP.

Crysmann, Berthold. 2016. Representing Morphological Tone in a Computational Grammar of Hausa. *Journal of Language Modelling* pages 463–512.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. Academic Press, syntax and Semantics Volume 34.

Dalrymple, Mary, Dyvik, Helge and King, Tracy Holloway. 2004a. Copular Complements: Closed or Open? In *Proceedings of the LFG'04 Conference*, pages 188–189, CSLI Publications.

Dalrymple, Mary and Kaplan, Ronald M. 2000. Feature Indeterminacy and Feature Resolution. *Language* 76, 759–798.

Dalrymple, Mary, Kaplan, Ronald M. and King, Tracy Holloway. 2004b. Linguistic Generalizations over Descriptions. In *Proceedings of the LFG04 Conference*, CSLI On-Line Publications.

Dalrymple, Mary, King, Tracy Holloway and Sadler, Louisa. 2009. Indeterminacy by Underspecification. *Journal of Linguistics* 45, 31–68.

Flickinger, Dan. 2000. On Building a More Efficient Grammar by Exploiting Types. *Natural Language Engineering* 6 (1) (Special Issue on Efficient Processing with HPSG), 15 – 28.

Fokkens, Antske Sibelle. 2014. *Enhancing Empirical Research for Linguistically Motivated Precision Grammars*. Ph. D.thesis, Department of Computational Linguistics, Universität des Saarlandes.

Frank, Anette and Zaenen, Annie. 2004. Tense in LFG: Syntax and Morphology. In Louisa Sadler and Andew Spencer (eds.), *Projecting Morphology*, pages 23–66, CSLI Publications.

Ingria, Robert J. P. 1990. The Limits of Unification. In *Proceedings of the 28th Annual Meeting of the ACL*, pages 194–204.

Kaplan, Ronald M. and Wedekind, Jürgen. 1993. Restriction and correspondence-based translation. In *Proceedings of the 6th Conference of the Association for Computational Linguistics European Chapter (EACL)*, pages 193–202.

King, Tracy Holloway. 2011. (Xx*-)Linguistics: Because We Love Language. *Linguistic Issues in Language Technology: Interaction of Linguistics and Computational Linguistics* 6.

Mohanan, Tara. 1994. *Argument Structure in Hindi*. CSLI Publications.

Müller, Stefan. 1999. *Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche*. Max Niemeyer Verlag.

Müller, Stefan. 2002. *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German*. Studies in Constraint"=Based Lexicalism, No. 13, Center for the Study of Language and Information.

Müller, Stefan. 2013. The CoreGram Project: A Brief Overview and Motivation. In Denys Duchier and Yannick Parmentier (eds.), *Proceedings of the Workshop on High-level Methodologies for Grammar Engineering (HMGE 2013), Düsseldorf*, pages 93–104.

Müller, Stefan. 2015. The CoreGram Project: Theoretical Linguistics, Theory Development and Verification. *Journal of Language Modelling* 3(1), 21–86.

Müller, Stefan and Kasper, Walter. 2000. HPSG Analysis of German. In Wolfgang Wahlster (ed.), *Verbmobil: Foundations of Speech-to-Speech Translation*, Artificial Intelligence, pages 238 – 253, Springer.

Nordlinger, Rachel. 1998a. *Constructive Case: Evidence from Australian languages*. CSLI Publications.

Nordlinger, Rachel. 1998b. *A Grammar of Wambaya*. Pacific Linguistics.

Sadler, Louisa, Arnold, Doug and Villavicencio, Aline. 2006. Portuguese: Corpora, Coordination and Agreement. In *Proceedings of Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*.

Siegel, Melanie and Bender, Emily M. 2002. Efficient Deep Processing of Japanese. In *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization at the 19th International Conference on Computational Linguistics*.

van Eynde, Frank, Augustinus, Liesbeth and Vandeghinste, Vincent. 2016. Number Agreement in Copular Constructions: A Treebank-based Investigation. *Lingua* pages 104–126.

Villavicencio, Aline, Sadler, Louisa and Arnold, Doug. 2005. An HPSG Account of Closest Conjunct Agreement in NP Coordination in Portuguese. In *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar*.

Wedekind, Jürgen and Ørsnes, Bjarne. 2003. Restriction and Verbal Complexes in LFG: A Case Study for Danish. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of LFG03*, CSLI On-line Publications.

Refining the semantics of lexical rules in HPSG

David Lahm
University of Frankfurt

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 360–379

Keywords: lexical rules, SRL, semantics of lexical rules

Lahm, David. 2016. Refining the semantics of lexical rules in HPSG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 360–379. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.19.

Abstract

This paper points out certain flaws in the semantics for lexical rule specifications developed by Meurers (2001). Under certain circumstances, certain words may not be licit inputs to a rule according to this semantics while one would expect them to be from inspecting the specification of the rule, i.e. the rule violates the criterion of *Universality*. It is not universally applicable to all the words that its specification seems to characterise as licit inputs. The reasons for this are shown to be that whether properties of paths should be transferred from the input of a rule to its output is decided considering only the respective paths and their properties in isolation, ignoring the 'non-local' effects that transferring their properties can have. Furthermore, the semantics is insensitive to the possible shapes of inputs to the rule, which also makes it possible that inputs of certain shapes are unexpectedly not accepted. An alternative semantics is developed that does not suffer from these deficits.

1 Introduction

In HPSG theorising, *lexical rules* (henceforth *LRs*) play a prominent role. Such rules are employed to derive new words from existing words systematically. What precisely this means depends, of course, on what a *word* is taken to be. Pollard & Sag (1994) take *word* to mean *lexical entry*, envisaging LRs as a means of deriving new lexical entries from given ones (basic or themselves derived). This approach, which Meurers (2001) calls the *meta-level* approach, has however never been worked out formally in a satisfactory manner. An alternative approach is developed by Meurers (2001). Meurers suggests to view LRs as a means of deriving *lexical items* – the objects described by lexical entries – from given ones (basic or themselves derived). This approach can straightforwardly be formalised in model-theoretic HPSG by introducing a sort *lex_rule* with appropriate attributes IN and OUT, both of which take objects of sort *word* as values. The value of IN can then be viewed as the input and the value of OUT as the output of a LR. The content of any rule, i.e. the input-output relation it is supposed to encode, can be stated by an appropriate description that constrains the licenced *lex_rule* objects to those in which this relation holds between the values of IN and OUT. The actual content of a LR is thus given by an ordinary description of the same kind as a lexical entry. Meurers (2001) thus calls this a *description-level* approach.

In this paper, I exclusively assume this latter approach and will not concern myself with the idea of formalising LRs as relating lexical entries to lexical entries. Furthermore, I assume the model-theoretic foundations for HPSG developed by King (1989) and Richter (2004), where only the formalism developed by King (SRL) will be used here. As a consequence, I will not concern myself with approaches that have not been or cannot be satisfactorily formalised within these

†I thank two anonymous reviewers for helpful comments and Frank Richter, Manfred Sailer and Olivier Bonami for valuable discussion. The work reported here originates with Lahm (2012). The current paper is due to funding by the German Research Foundation (DFG) under grant GRK 2016/1.

frameworks, which leaves (Meurers, 2001) as the only serious attempt at specifying a semantics for LRs.

Usually, LRs are stated using abbreviations, which I also call Lexical Rule Specifications (LR descriptions). An LR descriptions is of the form $A \mapsto B$, where A and B are AVMs. It is supposed to express a rule that (i) accepts any input that is described by A, (ii) yields an output described by B and (iii) effects not more than the smallest changes to the input necessary to guarantee (i) and (ii). So the output should still be like the input to the greatest possible extent compatible with its being described by B. The LRs given in the literature employ this abbreviatory notation, without any exceptions I would be aware of. The central aim of Meurers (2001) and this paper is to explicate what exactly the intuitive ideas (i)-(iii) are supposed to mean.

The idea seems simple enough. If some path π is not mentioned in the specification of a rule, its value can be assumed to be the same (token-identical) in the input and output of the rule and is thus transferred from the input to the output. Similarly, the sorts of path values in the input should become those of the values of the corresponding paths in the output if this is possible. Meurers's semantics of LRs operates according to these ideas, but the details turn out to be more involved than one might expect. In particular, it can be shown that Meurers's system does not fulfill (i) above, which will be stated more explicitly below as the criterion of **Universality**: it is possible to construct LRs $A \mapsto B$ which, according to Meurers's semantics, do not accept every word described by A but impose further restrictions on their inputs in a manner that seems unexpected from their specification and therefore undesirable.

One reason for this will turn out to be that Meurers's semantics transfers properties of paths in a local manner, considering only one path at a time and disregarding the effects that such transfers can have on the possible values of other paths. If the properties of two paths cannot both be transferred at the same time, the semantics will thus not detect this, which leads to inconsistencies and thus violations of Universality. The alternative account offered here will instead allow for transferring the properties of all maximal sets of paths whose properties can consistently be transferred together. It thus takes into account the global effects of the transfers that are performed and acknowledges the fact that transferring one property may come at the cost of not being able to transfer another.

This alone however is not sufficient to guarantee Universality. In addition, the semantics needs to be made sensitive to the possible shapes of a rule's input according to its specification in order to guarantee that all of these are actually accepted by the rule. This problem will also be addressed in the approach presented here, which can be shown to no more violate Universality.

2 Preliminaries

2.1 The scope of this paper

The present paper is an investigation into lexical rules, an expressive mechanism often employed in HPSG grammars, and the way they are specified. It discusses the proposal by Meurers (2001) on how the meaning of LRs specifications can be made precise in the context of model-theoretic HPSG based on SRL as developed by King (1989). It points out certain problems with the approach advocated by Meurers. The problems identified are of a conceptual nature. It is shown that, in principle, the semantics that Meurers defines for LR specifications will fail to fulfill certain expectations that a semantics would be required to fulfill by intuitions which I expect everyone who employs LR specifications to share.[1]

The LRs as well as the sort hierarchy this paper uses are artificially constructed for the purpose of discussing the basic ideas behind the semantics outlined and the conceptual shortcomings of Meurers's approach. What the paper hence does not offer is any concrete example of the semantics developed by Meurers failing in the case of any actual LR that has been employed in the literature. It thus remains silent on whether the conceptual problems outlined also are practical problems that have adverse effects on any actual grammars that employ LRs. Even if this were not the case, it is worth keeping in mind that a merely conceptual problem could easily turn into such a practical problem with each new rule that is proposed.[2]

[1] I am not concerned here with the question whether all of Meurers's ideas about the correct meanings of LRs are tenable or not. LRs have been employed long before any precise semantics for them was available and, as a consequence, the intuitions of linguists can with good reason be expected to provide the standard of adequacy for any semantics to be proposed after the fact, not *vice versa*. Unfortunately however, nothing even guarantees that intuitions are stable enough among linguists to make a semantics that suits them all even possible. Hence it is hardly surprising if any concrete attempt at specifying such a semantics does not satisfy everyone. At the very least however, it provides a precise starting point from which a discussion about the benefits or shortcomings of certain decisions can be productively entered and opinions like that of one reviewer, who thinks that Meurers's system and mine alike are 'broken by design' (which, as I understand it, is supposed to be due at least in part to the choice of SRL) can sensibly be put forth, due to the presence of an actual design. Yet what I will be addressing here are not questions of semantic detail but rather something I think would be a conceptual problem for *any* semantics LRs might be given. This is not to say of course that semantic detail does not matter, but the system presented here is not supposed to carve the semantics in stone but is open to modification. For instance, the sceptical reviewer wondered why a certain rule would transfer to the output the species of paths that were equated in its input but not leave said structure sharing intact. It will be seen below that the present system is readily adapted so as to leave the structure sharing intact as well. Where intuitions run counter to the use of SRL however, there is no remedy in the present context. This does not mean that I think that arguments against its use could not sensibly be made or that the status of LRs might not figure prominently in such arguments. But unlike the arguments made here, such arguments cannot, it seems, be made on grounds that are purely formal apart from appealing to a single intuition about the meanings of LRs, namely that they should respect Universality, which seems to be very basic and has not yet been contested. Instead, they actually need to take into account how actual LRs behave in actual grammars when interpreted according to the system proposed here, and this issue is orthogonal to the one addressed here.

[2] The reviews of this paper remarked that offering an example of a problematic rule with genuine

2.2 The presupposed formalism

I shall, like Meurers (2001), assume the formalism in which HPSG theories are expressed to be Speciate Reentrant Logic (SRL) as developed in King (1989). Speciate Reentrant Logic was expressly designed for the formulation of HPSG grammars and incorporates assumptions on the linguistic ontology that are endorsed in Pollard & Sag (1994), total well-typedness and sort-resolvedness. The present section briefly discusses these assumptions, as they should be kept in mind in what follows.

Each linguistic object is assumed to have some sort. While sorts are organised in a partially ordered sort-hierarchy, so that it makes sense to speak of more general sorts subsuming more specific ones, each object needs to be assigned exactly one maximal sort that does not subsume any further sorts. Maximal sorts are also called species. Thus, if an object is said to be of a sort σ that subsumes exactly the species $\{s_1, ..., s_n\}$, this means that the object is of one and only one of the species in this set.

Attributes may be defined on linguistic objects. If defined on some object, they will have a value on that object, which also is an object. The species of an object determines which attributes are defined on the object. Since the objects are *totally well-typed*, each attribute that can be defined on objects of some species also must be defined on each of them. The information which attributes are appropriate to which species is a part of the sort hierarchy, and so is the information which species the value of an attribute that is defined on some object may have, which depends on the species of this object.

A string of attributes is a *path*. Its value on an object, if it has any, is determined by determining the value of its leftmost attribute on this object and then evaluating the remainder of the path on the result, if any. If any value is undefined on the way, so is the path. Two distinct paths evaluated on an object may yield the same object as their value.

I shall not give a full characterisation of the model theory of SRL, whose details are not essential to understanding the paper, but content myself with a highly intuitive sketch. SRL provides for two types of atomic descriptions, *species assignments* and *path equations*. A species assignment $:\pi \sim s$ describes all objects on which the path π has a value of species s[3] and a path equality $\pi =\pi'$ describes all

linguistic content would be helpful, and I agree. Unfortunately, I have not been able to construct such an example in the available time and it may be the case that no such examples will ever come up in linguistic practice. If they could be shown to be irrelevant, this might actually be seen as a welcome result, as the semantics of (Meurers, 2001) is of a considerably lower complexity than the one offered here while both coincide in cases where violations of Universality as discussed below do not actually occur. Hence, if one could be certain that the problems discussed will never attain relevance in actual grammar writing, staying with (Meurers, 2001) would definitely seem the right choice. Still, if it is agreed that the problems pointed out are genuine conceptual problems (and this no one has denied so far), it seems that the burden of proof regarding their harmlessness should be on those who prefer to leave things unchanged.

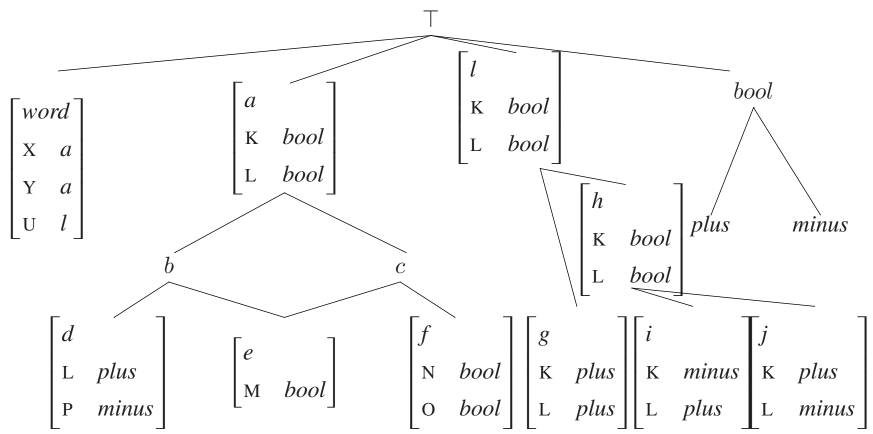
[3]It is convenient to assume that non-maximal sorts can also appear in such assignments, but these can straightforwardly be reduced to disjunctions of species assignments as long as no sort subsumes infinitely many species.

objects on which the value of π is the same object as the value of π'. The set of objects described by a description is called the description's *denotation*. The syntax of SRL is like that of ordinary propositional logic, where e.g. $\wedge$ is interpreted as intersecting the denotations of the conjoined descriptions, but the syntax is hardly important in this paper. If a description cannot have a non-empty denotation, i.e. if no object at all can be described by it, the description is called *unsatisfiable*.

I trust that the reader will see how the familiar AVM notation relates to SRL. Precise specifications of this relation can be found in King (1989) and Richter (2004) (for RSRL, which contains SRL as a fragment). It is important to keep in mind that AVMs, just like SRL expressions, are only descriptions of linguistic objects that are conceived of as total in the sense of being totally well-typed and sort-resolved.

2.3 The sort hierarchy

Throughout the paper, the following sort hierarchy will be presupposed, which is derived from the one in (Meurers, 2001, p. 188). The sorts in the bottom line are species. Lines indicate the subsumtion relation: a sort subsumes another if it is above it and connected to it by a line.



3 Lexical rules and Universality

Lexical Rules are specified by Lexical Rule Specifications, typically given in the shape $A \mapsto B$. These are understood to mean that for every word described by A there is one described by B that has as much in common with A as possible without violating B. The purpose of the semantics discussed in this paper is to make this idea explicit; in order for the output to have 'as much in common with A as possible, properties of the input must be transferred to the output, and how this is supposed to happen needs to be made precise.

The semantics of Meurers (2001) as well as the one proposed here proceed by starting out from the class of objects described by $A \mapsto B$, the LR descriptions as it is before any transfers of properties are specified. $A \mapsto B$ is understood as an abbreviation for the AVM in (1).

(1) $\begin{bmatrix} \textit{lex\_rule} & \\ \textsc{in} & A \\ \textsc{out} & B \end{bmatrix}$

The objects described by an LR descriptions as such are also called *proto-instances*. From the class of proto-instances, those are singled out in which properties are transferred in the way desired. These form the class of the rule's *instances*. The semantics will be spelt out indirectly by specifying a translation function that assigns to each LR descriptions a Lexical Rule Description (LR description) that only describes its licit instances. Section 6 briefly discusses this translation. The discussion of the intended meaning of LR descriptions will however present the contents of the proposals in a more direct manner, talking about objects (proto-instances and instances) directly instead of descriptions of them.

Since the class of instances of a rule is arrived at by preventing an appropriate subclass of the proto-instances from becoming instances, it follows that every instance of a rule, i.e. every actual pairing of an input with an output, is also a proto-instance of that rule. This seems right, as both input and output should at least conform to what is stated by the LR descriptions. Additional requirements, i.e. property transfers, are imposed monotonically without contradicting the specification provided by the LR descriptions.

Furthermore, one would expect an LR descriptions that is not in itself unsatisfiable, i.e. one that has proto-instances, to have instances, too. So merely trying to transfer certain properties should not make it impossible to pair an input with an output. This requirement can be called the criterion of **Preservation**.

Criterion (Preservation). *If a rule has a proto-instance, it also has an instance.*

Preservation is a consequence of the stronger criterion of **Universality**.

Criterion (Universality). *For every PI of an LR, there is an instance of this LR such that the values of the* IN *attribute on the PI and on the instance are congruent.*

Two objects u_1 and u_2 are *congruent* iff there is a bijection f from the components of u_1 to those of u_2 such that for each component v of u_1, v and $f(v)$ are of the same species and for every attribute α and component v of u_1, the value of α on $f(v)$ (if defined) is identical to $f(\text{the value of } \alpha \text{ on } v)$. A *component* of an object u is any object that is the value of some path evaluated on u. In other words, congruent objects are look-alikes and any description that describes one of them also describes the other. Perhaps a bit more intuitively, the criterion can hence be thought of as demanding that an object that is the IN-value of a proto-instance must also be the IN-value of some instance.[4] Universality clearly implies that any rule that has proto-instances has instances, hence Preservation. But it goes beyond that in demanding, in effect, that the rule also must have something to say about every word that can be the IN-value of a proto-instance. This means that inspecting an LR descriptions should be sufficient to determine whether the specified rule will relate a given word to an output. Universality in effect limits the scope of property transfers, requiring that such transfers may not be carried out at the cost of reducing the applicability of the rule to any given input that conforms to its specification. So if, for instance, the input specification of a rule is simply *word* (as will be the case in some of the examples below), the rule should be expected to be applicable to any word whatsoever rather than to nothing at all or only to words that satisfy certain path equations.

In the following section, I shall discuss the semantics proposed by Meurers (2001) and show that it does not fulfill the criterion of Universality and not even that of Preservation. In section 5, an alternative semantics will be developed that can be shown to fulfill Universality (and thus Preservation).

4 Meurers's Semantics of Lexical Rule Specifications

The following exposition of the semantics for LR descriptions given in Meurers (2001) is in part my interpretation of what is intended in this paper. The formal definition of the semantics in the paper comes in the form of an algorithm that is supposed to translate LR descriptions into SRL descriptions that capture their intended semantics. Unfortunately however, the algorithm is flawed in that the rule responsible for transferring species can never apply, which renders it useless and leaves all path species untransferred. This is definitely not what was intended and so remaining faithful to (Meurers, 2001) in this respect would render about half of Meurers's proposal entirely uninteresting. What I present here is thus my understanding of what the algorithm was in fact supposed to achieve, based on the informal discussion in (Meurers, 2001).

It should be noted that the differences between my reading and the actual statement do not affect the way in which values are transferred. In this respect, the discussion below can be regarded as entirely faithful to (Meurers, 2001).

[4] Note that the formulation of Universality differs from demanding that, given an LR descriptions $A \mapsto B$, there must be an instance of the rule for every object described by A such that the instance's IN-value is congruent to that object. This requirement would be to strong since an LR descriptions may specify certain components of the input and output to be token-identical. But if a token-identity between an input path π and an output-path π' is specified, it may be the case, for instance, that π can have values of sorts that are not possible as sorts of values of π'. So there could be objects described by A for which it would simply not be possible to apply the rule in a way that completely conforms to the specification, and in such a case, it seems, the rule should thus not pair the object with any output at all. But for such objects there also will not be any proto-instances with congruent IN-values to begin with. So under the formulation of Universality chosen here, these objects will not be considered at all.

4.1 Sort Transfer

The purpose of having a semantics for LR descriptions is to guarantee the transfer of properties of words which the LR descriptions does not mention from the rule's input to its output. There are two kinds of properties which need to be considered: the sorts of path values and path values themselves. The latter, which will be dealt with in the next section, is perhaps the more salient kind of transfer: (at least) if a path π is not mentioned in an LR descriptions at all, the token identity $\text{IN}\,\pi = \text{OUT}\,\pi$ should hold if this is possible (in a sense to be made more precise below).

But even if a path π is mentioned in the LR descriptions and the LR descriptions prohibits transferring its value because the rule effects some change on it, one might still expect IN π and OUT π to be of the same species whenever this is possible. So even if the value of a path cannot be transferred (which would of course imply the transfer of its species), the species possibly can and, in view of the goal of transferring as much as possible, it should.

The intuition behind the transfer of path value sorts is best conveyed by an example. Consider the simple rule in (2), taken from (Meurers, 2001, p. 188).

(2) $word \mapsto \begin{bmatrix} \text{X} & c \end{bmatrix}$

This rule takes any word as its input and its output should be a word that is like the input except for having an X-value of sort c, which subsumes the species e and f. On *word*s, X is allowed to have values of sort a, which subsumes d, e and f. The rule should then allow for the following configurations:

(3) a. $\begin{bmatrix} \text{IN X} & d \\ \text{OUT X} & e \end{bmatrix}$

b. $\begin{bmatrix} \text{IN X} & d \\ \text{OUT X} & f \end{bmatrix}$

c. $\begin{bmatrix} \text{IN X} & e \\ \text{OUT X} & e \end{bmatrix}$

d. $\begin{bmatrix} \text{IN X} & f \\ \text{OUT X} & f \end{bmatrix}$

In (3c) and (3d), the species of X in the input is subsumed by c and thus compatible with what the output demands. In such a case, the species is transferred: if X is of species, say, e in the input, it also needs to be in the output, ruling out f as a possible output species in such cases. But if X in the input is of species d, which is not subsumed by c, the species of X in the output must differ from that in the input if the LR descriptions is to be obeyed. In this case, where no transfer is possible, any species subsumed by c is allowed.

$$\begin{bmatrix} \text{IN} & \begin{bmatrix} \text{U} & \begin{bmatrix} g & \\ \text{K} & plus \\ \text{L} & plus \end{bmatrix} \end{bmatrix} \\ \text{OUT} & \begin{bmatrix} \text{U} & \begin{bmatrix} i & \\ \text{K} & minus \\ \text{L} & plus \end{bmatrix} \end{bmatrix} \end{bmatrix} \qquad \begin{bmatrix} \text{IN} & \begin{bmatrix} \text{U} & \begin{bmatrix} g & \\ \text{K} & plus \\ \text{L} & plus \end{bmatrix} \end{bmatrix} \\ \text{OUT} & \begin{bmatrix} \text{U} & \begin{bmatrix} j & \\ \text{K} & plus \\ \text{L} & minus \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(a) U-value of species i (b) U-value of species j

Figure 1: Proto-instances cancel each other out.

Meurers (2001) implements this intuition about sort transfer as follows: the basic idea is to allow no proto-instance x of a rule to become an instance if another proto-instance y can be found such that, for some path π, IN π and OUT π have values of the same species s on y, IN π also has a value of species s on x but OUT π has a value of a species distinct from s on this proto-instance. In such a case, x is said to be *cancelled out by* y. In the case of rule (2), proto-instances configured as in (4) are cancelled out due to the existence of those configured as in (3c).

(4) $\begin{bmatrix} \text{IN X} & e \\ \text{OUT X} & f \end{bmatrix}$

4.2 Problems

Meurers's way of transferring sorts leads to a violation of the criterion of Universality. This can be seen by inspecting rule (5).

(5) $\begin{bmatrix} \text{U} & g \end{bmatrix} \mapsto \begin{bmatrix} \text{U} & \begin{bmatrix} h & \\ \text{K} & bool \\ \text{L} & bool \end{bmatrix} \end{bmatrix}$

The input is required to have a U-value of species g. Objects of species g only allow for objects of species *plus* as values of the attributes K and L. In the output, the value of U is supposed to be of sort h, i.e. of species i or j. Objects of species i only allow K-values of species *minus* and L-values of species *plus*. The same in reverse holds for j. This is illustrated in Fig. 1.

It is easily seen that using Meurers's way of transferring sorts will result in the rule having no instances at all, i.e. a violation of the criteria of both Preservation and, *a fortiori*, Universality. Proto-instances configured as in Fig. 1(a) cancel out those configured as in Fig. 1(b): the U L-value of the proto-instances described by Fig. 1(a) is of species *plus* in both the input and output. Regarding the proto-instances described by Fig. 1(b), it is of species *plus* in the input, too, but of species *minus* in the output. According to Meurers's approach to sort transfers, the proto-instances described by Fig. 1(b) are thus cancelled out and may not become instances. The same clearly holds in reverse if the path U K is considered instead of U L.

So while the rule should accept any word as input with a U-value of species g, it in fact will not accept any input at all, thus violating both Preservation and Universality.

4.3 Value Transfer

Meurers suggests to transfer path values from a rule's input to its output along the following lines: (i) only the values of paths which are not mentioned in the rule's output specification are transferred. (ii) the value of each such path has to be transferred whenever possible.

(i) is meant to keep the semantics from becoming overly complex. Because enforcing token-identity between input and output for paths that also get further specified in the output may of course lead to certain conflicts if the input and the specification of the output are incompatible, Meurers suggests that complex strategies will be needed to prevent inconsistencies. In this paper, I follow Meurers in adopting (i), but without strongly endorsing it.

(ii) can be made precise as follows: Let $\mathsf{Approp}(s, \alpha)$ denote the set of species the value of α can have on an object of species s. Now let π be a path such that OUT π is mentioned in the LR descriptions in question and let α be an attribute such that OUT $\pi\alpha$ is not mentioned in the LR descriptions, in accord with (i). Then IN $\pi\alpha$ = OUT $\pi\alpha$ must hold of every instance of the rule on which IN π has a species s and OUT π has a species s' such that $\mathsf{Approp}(s, \alpha) \cap \mathsf{Approp}(s', \alpha) \neq \emptyset$. This means that if it is possible for objects of the species of IN π and OUT π to have identical α-values, then they must have identical α-values.

In terms of cancelling out of proto-instances, this approach amounts to cancelling out any proto-instance for which $\mathsf{Approp}(s, \alpha) \cap \mathsf{Approp}(s', \alpha)$ is not empty but IN $\pi\alpha$ = OUT $\pi\alpha$ does not hold.

This approach to transferring values also leads to violations of the requirement of Universality. For one thing, this is due to the fact that the transfer of an attribute $\pi\alpha$ is required whenever $\mathsf{Approp}(s, \alpha) \cap \mathsf{Approp}(s', \alpha) \neq \emptyset$ for s, s' the species of IN π and OUT π, respectively. Universality is then not generally respected since, if e.g. $\mathsf{Approp}(s, \alpha) = \{t, q\}$ and $\mathsf{Approp}(s', \alpha) = \{q\}$, the condition is fulfilled and the corresponding paths will be identified, but this clearly restricts the possible species for the IN-path to q, ruling out t even if there are proto-instances on which the IN-path has this species.

Additionally, the following rules all lead to violations of Universality.

(6) a. $\begin{bmatrix}\text{X K} & \textit{plus}\end{bmatrix} \mapsto \begin{bmatrix}\text{X} & \boxed{1} \\ \text{Y} & \boxed{1}\begin{bmatrix}\text{K} & \textit{minus}\end{bmatrix}\end{bmatrix}$

b. $\begin{bmatrix} \text{X K} & \textit{plus} \\ \text{Y K} & \textit{minus} \end{bmatrix} \mapsto \begin{bmatrix} \text{X} & \boxed{1} \\ \text{Y} & \boxed{1} \end{bmatrix}$

c. $word \mapsto \begin{bmatrix} \text{X} & \boxed{1} \\ \text{Y} & \boxed{1} \end{bmatrix}$

Regarding (6a), the value of X K will be required to be transferred: according to the signature, all species allowed as values of X (d, e and f) allow for values of K of species *plus* and *minus* alike. So the set of commonly accepted species is non-empty and X K will hence be required to be transferred and consequently of species *plus*. But this contradicts the output specification which requires OUT X K $=$ OUT Y K and that OUT Y K be of species *minus*, which of course implies that OUT X K also is of this species. The description that Meurers's approach derives from the LR descriptions in (6a) is thus contradictory and cannot licence any lex_rule objects at all. The rule thus has no effect.

(6b) leads to the same kind of contradiction, but in this case it is not due to the output specification but to the simultaneously required transfer of the values of X K and Y K, which are required to be of different species in the input but need to be identical in the output. So this rule, again, has no effect at all.

Unlike rules (6a) and (6b), rule (6c) is not contradictory and hence respects Preservation, but it still violates Universality. To see this, consider again the attribute K. Since rule (6c) should accept any word, it should accept words with an X K-value of species *plus* and a Y K-value of species *minus* in particular. But in fact it does not: according to Meurers's approach, the values of both X K and Y K will be transferred from the input to the output, i.e. OUT X K $=$ IN X K and OUT Y K $=$ IN Y K in any instance. Likewise, OUT X and OUT Y need to be the same object in any instance according to the LR descriptions itself. This implies that OUT X K $=$ OUT Y K also holds and hence IN X K $=$ IN Y K. Contrary to expectations, thus, rule (6c) will only accept words as its inputs on which IN X K and IN Y K are the same object. Since there are proto-instances of the rule for which this does not hold, this is again a violation of Universality.

5 The Alternative

To see which issues precisely an alternative proposal needs to address, first note that at the heart of the violations of Universality in the system of (Meurers, 2001) observed in the preceding sections are the following two properties of this semantics:

- Whether a proto-instance is cancelled is decided by inspecting paths in isolation.
- Comparison of proto-instances is not sensitive to whether they do have congruent IN-values or not.

The first point is illustrated by all the problematic rules presented so far. Regarding example (5), a proto-instance can be cancelled because it does not transfer the species of U K while another does. This happens without regard to the fact that such a proto-instance will transfer the species of U L while the one it is cancelled by will not. Similarly rule (6a), for instance, will always 'try' to transfer the value of X K. Since this is impossible without violating the output specification, the rule cannot apply to any word at all. This merely local consideration of paths is misguided and a more global approach called for. Such an approach is developed in this section. It will not rely on the inspection of paths in isolation in order to effect cancellation of proto-instances but instead take into consideration the set of paths whose properties are transferred by a proto-instance. A proto-instance will be cancelled if the set of paths that it transfers the properties of is a proper subset of the corresponding set for some other proto-instance, i.e. if the latter transfers the properties of more paths than the former.

This alone would however not suffice to guarantee Universality, due to the second of the two problems named above. It would still be possible for all proto-instances with an input object of a certain shape to be cancelled, leaving no instance with an input object congruent to those of the cancelled instances. This problem is solved here by allowing cancellation only inside of classes of proto-instances with congruent input objects. Thus, since congruence clearly is an equivalence relation, the class of proto-instances is partitioned into 'input congruence classes' and only inside of these classes proto-instances are cancelled that do not transfer the properties of a maximal set of paths.

5.1 Spelling out the Alternative

As stated above, the alternative approach to transferring path properties proposed here rests on the idea of maximising the set of paths whose properties are transferred. For each proto-instance, call its *SFrame* the set of paths with species transferred and its *VFrame* the set of paths with values transferred, defined as in (7). For any LR descriptions λ, *Men*(λ) is the set of paths π such that OUT π is mentioned in λ. *Edge* is the set of all paths with values that might be transferred by a given LRs, i.e. the set of mentioned paths that extend an unmentioned one by one attribute: *Edge*$(\lambda) = \{\pi\alpha \mid \pi \in Men(\lambda) \,\&\, \pi\alpha \notin Men(\lambda)\}$.

(7) a. *SFrame*$(x) =$
$\{\pi \in \textit{Men}(\lambda) \mid$ IN π and OUT π have the same species on $x\}$

b. *VFrame*$(x) =$
$\{\pi \in \textit{Edge}(\lambda) \mid$ IN π, OUT π are defined and OUT $\pi =$ IN π on $x\}$

As a first approach, one might let a proto-instance x cancel another proto-instance y if *SFrame*$(x) \supset$ *SFrame*(y) or *VFrame*$(x) \supset$ *VFrame*(y), i.e. if x transfers the sorts or values of all paths of which y transfers them and also of some additional paths for which y does not. But, as has already been mentioned above, it could very well happen then that some proto-instance cancels another whose IN-value is not congruent to its own. E.g., in the case of rule (8), only proto-instances with a IN U-value of species g allow for a *VFrame* containing both U K and U L.

(8) $word \mapsto \begin{bmatrix} \text{U} & g \end{bmatrix}$

The existence of such proto-instances would effect the cancellation of all those with IN U-values of species i or j, leading to a violation of Universality again. The solution to this problem, likewise mentioned above, is to allow cancellation only within the equivalence classes of proto-instances according to the equivalence relation of having congruent IN-values.

Letting the meaning of $\mathit{IVC}(x, y)$ be that the *lex_rule* objects x and y have congruent IN-values, the intended way of performing sort and value transfer can now be expressed as in definitions 1 and 2.

Definition 1 (Species Transfers)**.**
$\mathit{STrans}(PI) =$
$\{x \in PI \mid \mathit{For\ no}\ y \in PI : \mathit{IVC}(x, y)\ \mathit{and}\ \mathit{SFrame}(y) \supset \mathit{SFrame}(x)\}$

Definition 2 (Value Transfers)**.**
$\mathit{VTrans}(PI) =$
$\{x \in PI \mid \mathit{For\ no}\ y \in PI : \mathit{IVC}(x, y)\ \mathit{and}\ \mathit{VFrame}(y) \supset \mathit{VFrame}(x)\}$

Let $PI(\lambda)$ denote the set of proto-instances of a LR specification λ.[5] Denote by $\mathit{STrans}(PI(\lambda))$ the set of all proto-instances x of λ such that no proto-instance with an IN-value congruent with that of x exists that transfers the species of a proper superset of the paths whose species are transferred by x. $\mathit{VTrans}(PI(\lambda))$ is the analogous notion for value transfers. Since there will clearly exist maximal elements wrt $\supset$ (note that both *SFrame* and *VFrame* are finite), these sets are guaranteed to be non-empty and to contain, for every proto-instance of λ, some element with a congruent IN-value. So Universality is clearly respected by *STrans* and *VTrans*.

The set of instances of the rule is given by

(9) $\mathit{Transfers}(PI(\lambda)) = \mathit{VTrans}(PI(\lambda)) \cap \mathit{STrans}(PI(\lambda))$

Does *Transfers* still respect Universality? One can show that

$$\begin{aligned} & \mathit{STrans}(\mathit{VTrans}(PI(\lambda))) \\ = \ & \mathit{VTrans}(\mathit{STrans}(PI(\lambda))) \\ = \ & \mathit{VTrans}(PI(\lambda)) \cap \mathit{STrans}(PI(\lambda)) \end{aligned}$$

So the order in which the transfers are performed (values before sorts, sorts before values or in parallel) is immaterial and, since each of the transfer operations respects Universality, so does *Transfers* itself.

[5] Strictly speaking, this 'set' is of course a proper class. It is possible however to find sets which contain, for any PI, a congruent object. Since the actual formalization of the ideas laid out here proceeds indirectly, as sketched in section 6, thus operating on descriptions instead of the objects themselves, there is no reason to worry here.

5.2 Application

Let us consider the problematic rules given above to see the results of the proposed semantics. Since *Transfers* is known to respect Universality, what remains to be seen is only which Transfers the proposed semantics will actually licence in these cases.

Regarding rule (5) and the two interesting classes of proto-instances which this rule has, shown in Figures 1(a) and 1(b), the former of these have *SFrame*s that contain U L but not U K while those of the latter contain U K but not U L. Consequently, neither *SFrame*$(x) \subset$ *SFrame*(y) nor *SFrame*$(y) \subset$ *SFrame*(x) can ever hold if x is a proto-instance described by Figure 1(a) and y is one described by Figure 1(b). Of course, inside of the classes of proto-instances described by each of 1(a) and 1(b), cancellation will take place: e.g. some proto-instances will have *SFrame*s that contain X while others will not, and the same is true regarding the *VFrames*. Thus those proto-instances from whose frames X is missing will be cancelled due to those in whose frames it is contained. But 'across' 1(a) and 1(b), no cancellation can happen and thus proto-instances described by each of these will be among the instances. Transfer of U K and U L thus happens, but only on one of the paths U K and U L at a time, as it is impossible for both together.

Consider next rule (6a). No proto-instance of this rule can fulfill OUT X $=$ IN X, i.e. have a *VFrame* that contains X, which would require X K to have the species *plus* and *minus* at once. But under the present account, this does not have the effect of cancelling all proto-instances of the rule. In Meurers's semantics, this is the result because it uncompromisingly demands that X be transferred in the case of rule (6a) and thus cancels each proto-instance that does not transfer it. Under the approach presented here, there just is no proto-instance whose *VFrame* contains X, but this does of course not mean that there are none with maximal *VFrame*s and *SFrame*s. These will in fact exist for every satisfiable LR descriptions, and these will be instances of the rule.

Regarding (6b), the situation is slightly different and reminiscent of rule 5: there are *VFrame*s that contain X, and these cannot contain Y, which would again require X K to have as its species both *plus* and *minus*. Conversely and analogously, there are also *VFrame*s that contain Y but not X. A proto-instance whose *VFrame* contains X can thus never cancel one whose *VFrame* contains Y and *vice versa*. As a result, analogously to the case of rule (5), one of X and Y will be transferred in each instance but never both.

Basically the same is true regarding rule (6c), but in this case it is important that cancellation can only take place within a congruence class. While proto-instances of this rule can be found whose *VFrame*s contain both X and Y, none of these can have an input with X K and Y K-values of distinct species. But proto-instances with such IN-values exist, and inside of the congruence classes of such proto-instances, transferring the values of both X K and Y K is as impossible as it is in general in the case of rule (6b). For inputs of this shape, thus, it also holds that one and only one of these paths will have its value transferred.

6 Indirect formulation

In this section, I shall briefly discuss how the semantics described above is actually implemented under Meurers's and my approach. As remarked above, the semantics is given in an indirect manner. Thus LR descriptions are not given an interpretation themselves under which they denote their instances but are translated into SRL descriptions (LR descriptions), and these in turn denote the instances according to the semantics of SRL. That the semantics can be realised in this way shows that the expressive means used do not actually go beyond SRL. Specifying LRs is thus just a more convenient way of writing SRL descriptions.

Meurers's translation function builds on the notion of what I call the *Kepser Normal Form* (KNF) of SRL descriptions. This kind of normal form was introduced by Kepser (1994) as a tool for deciding satisfiability of SRL descriptions. A description in KNF is a description in Disjunctive Normal Form (DNF) that has certain closure properties.

For the sake of convenience, a description in DNF can be represented by a set of sets of literals, where a literal is an atomic description or a negated atomic description. Adopting the terminology used by Meurers (2001), I call such a set a *matrix* and its elements *clauses*. The meaning of a DNF δ will then be that of the SRL description

$$\bigvee_{\alpha\in\delta}\bigwedge_{\beta\in\alpha}\beta$$

i.e. the disjunction of the conjunctions of the literals contained in each of its clauses.

A matrix is in KNF iff each of its clauses is in KNF. Discussing the exact definition of a clause in KNF is beyond the scope of this paper, but what matters here is that such clauses are highly explicit about the objects they describe as far as the paths are regarded that are mentioned in them: for each path π that occurs at all in the literals of a clause in KNF and is defined on any of the objects it describes, there is some species s for which the sort assignment $:\pi \sim s$ is contained in the clause. The same holds for the equality $:\pi =:\pi$ (which is not completely trivial but expresses the definedness of π.) Furthermore, each clause in a matrix in KNF is satisfiable.[6]

These properties of normal clauses make them well suited to represent the proto-instances of LRs in the translation process. Hence, as the first step of the translation, the LR descriptions is normalised. This results in a matrix in KNF. Sort transfers can then be formulated as above, but instead of cancelling proto-instances directly, the clauses that denote them are dropped from the matrix. In Meurers's system, if the matrix $\mathcal{M}$ contains clauses $\mathcal{C}$ and $\mathcal{C}'$ such that $:\textsc{in}\,\pi\ \sim$

[6]Hence normalisation, for which Kepser (1994) provides an algorithm, also provides the decision procedure that he is after: a description is satisfiable iff its KNF is not empty.

$s, {:}\textsc{out}\,\pi \sim s \in \mathcal{C}$ and ${:}\textsc{in}\,\pi \sim s, {:}\textsc{out}\,\pi \sim s' \in \mathcal{C}'$ for some path π and (distinct) species s and s', then $\mathcal{C}'$ is dropped from the matrix. This means to drop all proto-instances described by $\mathcal{C}'$, which is what is wanted.

Meurers (2001) realises value transfers by adding ${:}\textsc{in}\,\pi\alpha = {:}\textsc{out}\,\pi\alpha$ to every clause $\mathcal{C}$ that fulfills the following:

- ${:}\textsc{out}\,\pi = {:}\textsc{out}\,\pi \in \mathcal{C}$
- ${:}\textsc{out}\,\pi\alpha = {:}\textsc{out}\,\pi\alpha \notin \mathcal{C}$
- ${:}\textsc{in}\,\pi \sim s, {:}\textsc{out}\,\pi \sim s' \in \mathcal{C}$
- $\mathsf{Approp}(s, \alpha) \cap \mathsf{Approp}(s', \alpha) \neq \emptyset$

Since $\mathcal{C}$ is in KNF, the first condition ensures that $\textsc{out}\,\pi$ is mentioned in the rule and defined on each of the proto-instances licenced by $\mathcal{C}$. The second condition ensures that the path resulting from this one by appending α is not defined and mentioned in the rule. The last two conditions ensure that $\pi\alpha$ can in principle have the same value on the input and output of the proto-instances the clause describes. Adding ${:}\textsc{in}\,\pi\alpha = {:}\textsc{out}\,\pi\alpha$ then requires these values to actually be the same, i.e. it cancels all proto-instances described by the original clause on which they are not, just as required by the semantics stated above.

This highlights an asymmetry between the way in which sorts and values are transferred under Meurers's approach: while sorts are transferred by dropping certain clauses from the matrix, values are transferred by adding literals to clauses. In contrast, the present approach treats sort and value transfers in an exactly parallel fashion, but this requires an additional step. The reason is that normal clauses are not yet explicit enough: while a normal clause assigns a species to the value of every path that occurs in it and is defined on the objects it describes, the clauses are not in general explicit about which path equalities the objects they describe satisfy. So while the notion of an *SFrame* is straightforwardly adapted to the indirect semantic account as in (10), the notion of a *VFrame* is not.

(10) $\mathit{SFrame}(\mathcal{C}) = \{\pi \mid \text{For some species } s, {:}\textsc{in}\,\pi \sim s, {:}\textsc{out}\,\pi \sim s \in \mathcal{C}\}$

While every path that is defined on the objects a normal clause describes is also explicitly assigned a species by the clause, the absence of a literal ${:}\textsc{out}\,\pi = {:}\textsc{in}\,\pi$ from a clause does not mean that this equality may not describe some proto-instances the clause licences; it just does not, in general, need to describe them. It is thus not possible to compare the clauses with regard to the value transfers they enforce because proto-instances in which some such transfer is performed and those in which it is not performed are not in general described by distinct clauses.

The missing explicitness about transferred values can be supplied in the following way. Let the set of value transfer specifications be the set of all possible equations between corresponding paths such that the $\textsc{out}$ path is mentioned in $\mathcal{C}$: $VTr(\mathcal{C}) := \{{:}\,\textsc{in}\,\pi = {:}\,\textsc{out}\,\pi \mid \ {:}\,\textsc{out}\,\pi \in \mathcal{C}\}$. Now consider the KNF of an LR descriptions λ. Under reasonable assumptions about the form of LR descriptions,[7] this will be a matrix $\mathcal{M}$ such that for each $\mathcal{C} \in \mathcal{M}$ and $\mathcal{C}' \in \mathcal{M}$, $VTr(\mathcal{C}) = VTr(\mathcal{C}')$, which can thus be referred to as $VTr(\mathcal{M})$. Now a new matrix can be defined as in (11). This matrix consists of all clauses that are the union of some clause in $\mathcal{M}$ with some subset of $VTr(\mathcal{M})$. So for every set of mentioned paths, there is a clause in this matrix that specifies all elements of this set as transferred.

(11) $\{\mathcal{C} \cup E \,|\, \langle \mathcal{C}, E \rangle \in \mathcal{M} \times \mathcal{P}(VTr(\mathcal{M}))\}$

Obviously, not all of these clauses can be expected to be satifsiable, i.e. to describe anything at all. But normalising the new matrix again will remove all the inconsistent clauses.

The matrix that results from this second normalisation step thus contains, for each set of paths whose values can jointly be transferred, a clause that explicitly states the values to be transferred on the objects it describes. On the basis of this matrix it is now possible to define the notion of a *VFrame* appropriately, which is done in (12)

(12) $\mathit{VFrame}(\mathcal{C}) = \{\pi \in Edge(\mathcal{C}) \,|\, \text{:IN}\,\pi = \text{:OUT}\,\pi \in \mathcal{C}\}$

$Edge(\mathcal{C})$ corresponds to $Edge(\lambda)$ as used in section 5 and can be defined as in definition 3.

Definition 3. *Edge of a clause*
$Edge(\mathcal{C}) =$
$\{\pi\alpha \,|\, \text{:OUT}\,\pi = \text{:OUT}\,\pi \in \mathcal{C}\ \&\ \text{:OUT}\,\pi\alpha = \text{:OUT}\,\pi\alpha \notin \mathcal{C}, \textit{for } \alpha \textit{ an attribute.}\}$.

The last missing ingredient is a way to ensure that cancellation of clauses only takes place within what I called a "congruence class" above. The details of how this can be done are rather involved, but the basic idea is straightforward: from any clause $\mathcal{C}$, a description $\mathcal{C}^{:\text{IN}}$ can be derived that describes all and only objects which are congruent with the :IN-value of some object described by $\mathcal{C}$.[8] Call $\mathcal{C}^{:\text{IN}}$ the IN-value description for $\mathcal{C}$. The notion of proto-instances having congruent IN-values can then be replaced by the notion of clauses having equivalent IN-value descriptions, which I notate as $EQV(\mathcal{C}, \mathcal{C}')$.[9]

The indirect semantics can now be stated in a fashion parallel to that used for the direct semantics above, making use of the following definitions.

[7]To be precise, it is assumed that an LR descriptions is stated by giving the input AVM, the output AVM and a (possibly empty) set of path inequalities. About the AVMs it is assumed that they do not contain any logical symbols. This means that each AVM is equivalent to a conjunction of SRL literals, but where it is allowed to use sort assignments that assign non-maximal sorts. Rule specifications found in the literature typically obey these constraints and a rule whose specification does not can equivalently be expressed by a set of rules that do.

[8]A detailed account of how to do this is given in Lahm (2012).

[9]$EQV(\mathcal{C}, \mathcal{C}')$ is decidable, e.g. by using the fact that δ and δ' are equivalent iff bot $\delta \wedge \neg\delta'$ and $\neg\delta \wedge \delta'$ are unsatisfiable.

Definition 4 (Species Transfers)**.**
$STrans(\mathcal{M}) =$
$\{\mathcal{C} \in \mathcal{M} \mid \textit{For no } \mathcal{C}' \in \mathcal{M} : EQV(\mathcal{C},\mathcal{C}') \textit{ and } SFrame(\mathcal{C}') \supset SFrame(\mathcal{C})\}$

Definition 5 (Value Transfers)**.**
$VTrans(\mathcal{M}) =$
$\{\mathcal{C} \in \mathcal{M} \mid \textit{For no } \mathcal{C}' \in \mathcal{M} : EQV(\mathcal{C},\mathcal{C}') \textit{ and } VFrame(\mathcal{C}') \supset VFrame(\mathcal{C})\}$

Each of the clauses in each of these two matrices has as its denotation a subclass of the proto-instances that the semantics specified in section 5 admits into the respective *STrans* and *VTrans* sets as defined in that section. Since the matrices are interpreted disjunctively, their denotation is the union of all these classes and thus the same as that of the transfer functions that were specified directly.

7 Conclusion

In this paper, I have shown that the semantics of LRs specified in Meurers (2001) disrespects what I have called the criterion of Universality. Words that one would expect to be licit inputs to a rule may not in fact be that under the semantics suggested by Meurers, i.e. there will be no output for them. In extreme cases, rules may become unsatisfiable although their specifications are not. I have shown the reason for that to be that in Meurers (2001) whether properties of a path are transferred from the input to the output is decided by inspection of that path in isolation, without regard for the non-local effects that transferring the properties may have, and that the system developed performs transfers without regard to the possible shapes of input objects to the rule. I further introduced an alternative semantics that solves both of these problems.

An interesting question that I must leave open is whether the problems pointed out actually affect realistic grammar writing. To argue that they do not would require showing that sort hierarchies and LRs of the kind that lead to the problems are by their very nature pathological and can be expected not to occur in actual grammar writing. This would actually be a welcome result since the complexity of the criticised account by Meurers (2001) seems to be considerably lower than that of mine.[10] I do not consider it unlikely that this might be the case, but determining whether it is is not the goal of this paper, but rather to show that the question matters.

References

Kepser, Stephan. 1994. A satisfiability algorithm for a logic for typed feature structures. Arbeitspapiere des SFB 340 60 Universität Tübingen, Germany.

[10]Note that the size of the set specified in (11) grows exponentially with the number of $Edge$ paths.

King, Paul. 1989. *A logical formalism for head-driven phrase structure grammar*: University of Manchester dissertation.

Lahm, David. 2012. Revisiting the description-level approach to lexical rules in HPSG. Unpublished master's thesis, University of Tübingen.

Meurers, Walt Detmar. 2001. On Expressing Lexical Generalizations in HPSG. *Nordic Journal of Linguistics* 24(2). 161–217.

Pollard, Carl J. & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Richter, Frank. 2004. *A mathematical formalism for linguistic theories with an application in Head-Driven Phrase Structure Grammar*: Eberhard-Karls-Universität Tübingen Phil. dissertation (2000).

Verbal present participles in Norwegian: Controlled complements or parts of complex predicates

Helge Lødrup
University of Oslo

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 380–400

Keywords: complex predicates, present participles, presentational focus, Norwegian syntax, Scandinavian syntax

Lødrup, Helge. 2016. Verbal present participles in Norwegian: Controlled complements or parts of complex predicates. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 380–400. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.20.

Abstract

Norwegian has a limited option for verbal present participles. These participles only exist for a limited number of verbs, and they are selected by a handful of predicates. The analysis of sentences with these participles raises some challenges. Taking the analysis of Thurén (2008) as my point of departure, I argue that verbal present participles are in some cases controlled complements, and in other cases parts of complex predicates. The presentational focus construction gives important evidence for this analysis.

1. Introduction[1]

This paper is about a limited construction that has been considered problematic in Scandinavian syntax. Even if present participles are usually adjectival in Norwegian, some verbs seem to take a verbal present participle. An example is (1).

(1) Han kom styrtende ned trappen
he came rushing down stairs.DEF
'He came rushing down the stairs'

These present participles have been analyzed as main verbs, with their selecting verbs as auxiliaries. There are several reasons that this cannot be correct, as will be shown below. A more interesting analysis is given by Thurén (2008) (on Swedish), who proposed that sentences such as (1) are restructuring (also called reanalysis) sentences in which the two verbs together constitute one complex predicate. The selecting verb is then a "light verb". I will partly delimit this analysis by arguing that the restructuring is optional, and partly extend it by applying it to sentences with the verb *ha* 'have'. I will also present new data, and show how the presentational focus construction gives important evidence for the optionality of restructuring.

[1] I have received valuable input from audiences in Oslo (September 2015), Gothenburg (Gramino, May 2016), and Warsaw (HeadLex16, July 2016). I am grateful to the two anonymous reviewers for their thorough and constructive comments, and to the proceedings editors.

2. The basic facts

2.1 Present participles in Norwegian

Present participles represent a more limited phenomenon in Norwegian — and Swedish and Danish — than in e.g. English and French (see e.g. Egerland 2002). Many verbs lack a present participle, including frequent verbs. Norwegian present participles are primarily adjectival (see e.g. Faarlund et al. 1997:119). To what extent there are also verbal present participles is a difficult question. One problem is that varieties of Norwegian differ in their use of present participles. Written Norwegian uses some present participles that must be considered verbal because they show syntactic options that are typical for verbs, such as taking an object (Western 1921:368-76, Kinn 2014). An example is (2).

(2) Fændrik sitter bakerst i sin fantebåt (..) syngende en munter vise
Fændrik sits hindmost in his hobo.boat singing a merry song
'Fændrik sits hindmost in his hobo-boat, singing a merry song'
filmklubb.no/filmer/fant/ 11/02/16

Sentences such as (2) are, however, not acceptable in colloquial Norwegian. My focus here is upon options that are intuitively acceptable in the colloquial language. I follow Western (1921:368-71) in assuming that the colloquial language allows verbal present participles with four predicates: the verbs *komme* 'come', *bli* 'remain',[2] *ha* 'have', and the preposition *med* 'with'. Examples are (3)-(6).

(3) Han kom styrtende ned trappen
he came rushing down stairs.DEF
'He came rushing down the stairs'
(4) Han ble liggende i gresset
he remained lying in grass.DEF
'He remained lying in the grass'
(5) Jeg har en fin gammel portvin stående (Faarlund et al 1997:752)
I have a fine old port.wine standing
'I have a fine old port wine standing'
(6) Vi kan ikke fortsette med John liggende under bordet
we can not continue with John lying under table.DEF
'We cannot continue with John lying under the table'

[2] The Norwegian verb *bli* also has other uses, which are not directly relevant here. It can mean 'become', and it is used as a passive auxiliary.

2.2 Selection of present participles

What verbal participles are acceptable with the four predicates mentioned is not easy to delimit exactly. Even so, it is clear that these predicates restrict what participles they take as complements (Kinn 2014).

With the verb *komme* 'come', the central participles denote movement, with a focus on manner (e.g. *gå* 'walk', *snike* 'sneak', *kjøre* 'drive').[3]

With the verbs *bli* 'remain' and *ha* 'have' and the preposition *med* 'with', the central participles are forms of the posture verbs (*ligge* 'lie', *sitte* 'sit', *stå* 'stand'), and some other verbs for moving (e.g. *gå* 'walk') or being at a place (e.g. *bo* 'live'). The verb *bli* 'remain' in addition allows the participles of the verbs *hete* 'be.called' and *være* 'be'.

The latter two aside, the selected participles are all forms of one-place verbs (e.g. *snike* 'sneak'), or verbs that take an oblique (e.g. *bo* 'live'). A fact that will be of interest later is that they all allow the presentational focus construction with an expletive subject, as in (7).

(7) Det ligger / går / bor en mann her
there lies / walks / lives a man here
'A man lies / walks / lives here'

2.3 Are they really verbal?

Three arguments for considering the selected present participles inflectional verbal forms will be presented briefly, others will follow later.[4]

Argument 1): The verbs *komme* 'come' and *bli* 'remain' do not select an AP, but they select a present participle. The verb *komme* can precede an adjective,

---

3 Present participles can be used as adjectival adjuncts in sentences with *komme* 'come'. For example, the adjunct participle *triumferende* 'triumphant' in (i) is in an adjunct position, which is not available for a complement such as *styrtende* 'rushing'. Other differences between complement and adjunct participles are mentioned in section 4.

(i) At de triumferende kommer inn her nå ..
that they triumphant come in here now
'That they come here triumphant now ..'
forum.bataljonen.no/index.php?topic=9247.190;wap 11/04/16

4 Swedish present participles sometimes end in an *-s*. According to Thurén (2008:56), these forms are unambiguously verbal. This effect of *-s* is unknown in Norwegian. Present participles in *-s* occur in some dialects, e.g. *spisendes* 'eating.S'. They often have a "passive" interpretation, but this interpretation can also be found without the *-s* (Western 1921:372, Faarlund et al 1997:119).

as in (8), but they are adjuncts. The verb *bli* 'remain' can take a locative, but not an AP, as shown in (9).[5]

(8) Han kom full / sur hjem
he came drunk / grumpy home
'He came home drunk / grumpy'

(9) Han ble hjemme / i byen / * full / *sur
he remained home / in town.DEF / drunk / grumpy
'He stayed home / in town / *drunk / *grumpy'

When the verbs *komme* 'come' and *bli* 'remain' select a present participle, this form must be inflectional according to the classical distinction between inflection and derivation: Syntax can "see" inflectional morphology, but not derivational morphology, and thus not distinguish an adjectival participle from another adjective.

Argument 2): Kinn (2014:94) mentions that verbal present participles keep the meaning of the stem, while this is not necessarily the case with adjectival participles — as expected from general properties of inflection and derivation. For example, posture verbs have a rather wide meaning, allowing abstract and metaphorical uses (Holm 2013). These uses can also be found with verbal present participles, as in (10), but not necessarily with adjectival participles, as in (11).

(10) Konklusjonen blir stående i kontrast til innholdet
conclusion.DEF remains standing in contrast to content.DEF
'The conclusion is in contrast to the content'

(11) *Stående i kontrast til innholdet er konklusjonen uheldig
standing in contrast to content.DEF is conclusion.DEF unfortunate
'The conclusion is unfortunate, being in contrast to the content' [intended]

Argument 3): Verbs and adjectives have different options for compounding and derivation. As expected, adjectival present participles have the potential of adjectives, and not of verbs. For example, the compound *traktorkjørende* 'tractor.driving' can be adjectival, as in (12), but not verbal, as in (13). This is expected, when there is no verb **traktorkjøre* 'tractor.drive'.

---

[5] The verb *komme* can take an AP in a lexicalized expression such as *komme løs* 'come loose'. The verb *bli* 'remain' can take an AP in archaic language, as in (i).

(i) mennesket blir sig dog altid ligt (Ibsen)
man.DEF remains himself still always alike
'Man always resembles himself'

It could be mentioned that Thurén (2008:62-63) rejects argument 1) for Swedish, because the corresponding Swedish verbs can select AP.

(12) traktorkjørende menn
tractor.driving men
'men driving tractors'

(13) *Han kom traktorkjørende
he came tractor.driving

Verbal present participles raise several challenges for morphological theory which cannot be pursued here — including the question of how a language can have an inflectional form that occurs with a limited set of verbs only.

3. Complex predicates?

The analysis of sentences with selected verbal present participles is considered problematic by those who have discussed it. One possibility is that the verbs that take verbal present participles are auxiliaries, which take the participles as main verbs. Teleman et al. (1999:618-19) (on Swedish) say that *komma* 'come' and *bli* 'remain' are close to having auxiliary status, but they also say that the present participle is a predicate complement with these verbs. Faarlund (1997:472, 532) and Ebeling (2003:154-177) assume that *bli* 'remain' is an auxiliary when it takes a present participle. Kinn (2014:77-78) also considers *bli* 'remain' an auxiliary. This is also his view of *komme* 'come' — with some reservations (Kinn 2014:83).

The traditional concept of auxiliary covers a rather heterogeneous group. Even so, it is clear that the verbs that take verbal present participles have very different properties (more later). A striking difference is that verbs that take verbal present participles only combine with a small number of verbs, while most auxiliaries can take any verb.

Thurén (2008) (on Swedish) has an interesting approach to sentences with selected present participles. She proposes that they are restructuring sentences with complex predicates. The selecting verbs are then light verbs.[6] (Lundquist 2009 also assumes this analysis, without discussing it.) There is, however, more to be said. Thurén (2008) does not discuss the predicates *ha* 'have' and *med* 'with', and she does not take presentational focus sentences into account. There is also a question if restructuring can give a full account of these sentences. I will argue that there are sentences in which selected verbal present participles are parts of complex predicates, as well as sentences in which they are not. The question then arises how the sentences in question should be analyzed when they do not show restructuring.

[6] Thurén (2008) seems to be more "liberal" than me concerning what verbs select verbal present participles, and what participles should be considered selected (as opposed to adjuncts). It is not clear to me to what extent differences between Swedish and Norwegian are relevant to differences between our analyses.

It is a standard assumption that two verbs can be restructured (or reanalyzed) to a complex predicate in a monoclausal structure which takes one single set of syntactic functions (see e.g. Butt 1995, 2010, Alsina 1996, Cinque 2004, Wiklund 2007, Wurmbrand 2001, 2004). For example, the Norwegian verb *prøve* 'try' is a verb that allows restructuring with an infinitive. The resulting complex predicate can passivize, as in (14) (Lødrup 2014a). It also allows the second verb to take on verbal features of the first verb. This "feature agreement" has been established as a restructuring phenomenon (Niño 1997, Sells 2004, Wiklund 2007). Example (14) shows feature agreement with the preceding passive verb, (15) with the preceding imperative form, and (16) with the preceding participle form (Lødrup 2014a, Havnelid 2015, Aagaard 2016).

(14) Dette må prøves å gjøre(s)
this must try.PASS to do(.PASS)
'One must try to do this'

(15) Prøv å gjør det!
try.IMPERATIVE to do.IMPERATIVE it
'Try doing it!'

(16) Han har prøvd å gjort det
he has try.PART to do.PART it
'He has tried doing it'

Restructuring is usually an optional process. Verbs that can be light verbs in complex predicates also appear as full verbs in e.g. Italian (Monachesi 1998), German (Wurmbrand 2004), and Urdu (Butt 2010). The verb *prøve* 'try', which shows restructuring in (14)-(16), also occurs in sentences with properties that are incompatible with restructuring. An example (17), where the infinitive is realized as a passive subject.

(17) Å gjøre dette er aldri blitt prøvd før
to do this is never been tried before
'Doing this has never been tried before'

4. The verb *komme* 'come'

4.1 *komme* 'come' without restructuring

I first discuss sentences with the verb *komme* 'come' thoroughly, before showing how the other verbs may throw light upon the analysis.

When a sentence with *komme* 'come' such as (18) does not have restructuring, its analysis is rather straight forward in LFG. In the f-structure in (19), the verbal present participle is an XCOMP — a complement with an unrealized subject which is obligatorily controlled by an argument in the main clause.

(18) En mann kom styrtende
 a man came rushing
 'A man came rushing'

(19)
$$\left[\begin{array}{ll} \text{PRED 'come <(}\uparrow\text{SUBJ) (}\uparrow\text{XCOMP)>'} & \\ \text{SUBJ [PRED 'man']} & \\ \text{XCOMP} & \left[\begin{array}{l} \text{SUBJ} \\ \text{PRED 'rush <(}\uparrow\text{SUBJ)>'} \end{array}\right] \\ \text{VFORM PAST} & \end{array}\right]$$

The XCOMP with *komme* does not alternate with a DP/NP, as shown in (20). It cannot topicalize or enter into other unbounded dependencies, as shown in (21), contrasting with adjectival adjuncts, as shown in (22).

(20) *Han kom den store bilen / denne / det
 he came the big car.DEF / this / it
(21) *Styrtende kom de ut av kirken
 rushing came they out of church.DEF
 'They came rushing out of the church' [intended]
(22) Syngende kom de ut av kirken
 singing came they out of church.DEF
 'They came singing out of the church'

These are properties that are shared by verbal XCOMPS in general — except the verbal XCOMPs of auxiliaries, if they are assumed to take XCOMPs.[7] These properties recur with other XCOMPs that are realized by verbs — infinitives or participles — in sentences with or without subject raising (see Lødrup 2004). Examples are (23)-(26).

(23) Hun sies å vinne *Hun sies det *Å vinne sies hun
 she say.PASS to win - she say.PASS that - to win say.PASS she
 'She is said to win'
(24) Hun akter å vinne *Hun akter det *Å vinne akter hun
 she intends to win - she intends that - to win intends she
 'She intends to win'
(25) Bilen bes flyttet *Den bes det *Flyttet bes den
 car.DEF ask.PASS moved - it ask.PASS that - moved ask.PASS it
 'They ask somebody to move the car'

[7] This question has been discussed a number of times, see Butt et al. (1996), Sells (2004), Wedekind and Ørsnes (2004), Falk (2008).

(26) Vi så ham stupe *Vi så ham det *Stupe så vi ham
we saw him dive - we saw him that - dive saw we him
'We saw him dive'

Even if most constituents can topicalize, this kind of verbal complement usually cannot. The reason is not clear. One possibility is that it could be connected to the classical Higgins' generalization (see e.g. Higgins 1973, Dalrymple and Lødrup 2000, Lødrup 2012), which can be paraphrased to say that a clausal argument can only topicalize if it is in a position in which a DP/NP is an alternative. The traditional auxiliaries are different. They can take some cases of a DP/NP in Norwegian, and their complement can topicalize, as shown in (27).

(27) Hun ville ikke tape Hun ville ikke det Tape ville hun ikke
she would not loose - she would not that - loose would she not
'She didn’t want to loose'

4.2 The verb *komme* 'come' with restructuring

The distinction between c-structure and f-structure is important to account for restructuring. The c-structure does not reflect restructuring directly. I assume that the basic c-structure for a sentence such as (28) is as in (29) — with or without restructuring.

(28) Han kom styrtende
he came rushing
'He came rushing'

(29)
```
        IP
       /  \
     DP    I'
     he  /  \
        I    VP
     came    |
             VP
           rushing
```

The crucial level of representation is f-structure, as in (30), where the two verbs constitute one predicate which takes a single set of syntactic functions.

(30)
$$\left[\begin{array}{l} \text{PRED 'come-rush <(}\uparrow\text{SUBJ)>'} \\ \text{SUBJ [PRED 'man']} \\ \text{VFORM PAST} \end{array}\right]$$

The verb *komme* 'come' has a separate lexical entry for its use as a restructuring verb. It is then an incomplete verb with an argument structure in which the internal argument is not a thematic role, but an open position, as in (31) (Alsina 1996:201-3). When the f-structure is built, a process of predicate composition combines the restructuring verb and the verb below it, creating a complex predicate. The argument structure of the complex predicate is the result of combining the argument structures of the two verbs, as in (32). The lines indicate that the external argument of the second verb is identified with the external argument of the first verb. (Some technical questions involved are not in focus here, see e.g. Andrews and Manning 1999, Sells 2004, Lowe 2015.)

(31) *komme* < theme < . . > >
|_____|
(32) *komme styrtende* < theme < agent > >
|_______|

Sentences with *komme* 'come' and a verbal present participle share certain properties with and without restructuring. Sentences with restructuring also have a second part that cannot be replaced by a DP/NP, and not topicalize (see (20)-(21) above). These properties can also be found with other cases of complex predicates, such as long passives, as shown in (33)-(35).

(33) Dette må prøves å gjøres
this must try.PASS to do.PASS
'One must try to do this'
(34) *Dette må prøves det
this must try.PASS that
(35) *Å gjøres må dette prøves
to do.PASS must this try.PASS

The fact that the second part of a complex predicate cannot topicalize follows from the treatment of unbounded dependencies in LFG. They are accounted for on the level of f-structure, and what is topicalized must have a syntactic function.[8] With restructuring, the second verb and its complements, if any, are not a unit with a syntactic function, and thus not expected to topicalize (Lødrup 2011:166-67).

[8] Complements of auxiliaries are potentially problematic in this respect, if auxiliaries are assumed to be functional heads. See Wedekind and Ørsnes (2004) for a proposal.

4.3 The verb *komme* 'come' in presentational focus sentences

As far as I know, the syntax of presentational focus sentences with selected verbal present participles has never been discussed. Scandinavian presentational focus sentences are usually assumed to have an expletive subject and an argument that is realized as an object (see e.g. Lødrup 1999 and references there, for an alternative analysis, see Börjars and Vincent 2005). Most one-place verbs take this construction, including many unergative verbs.

There are two possible word orders in the relevant presentational focus sentences; the object can precede or follow the present participle, as shown in (36)-(37). (This fact is mentioned, but not discussed for Danish in Hansen and Heltoft 2011:1603.)

(36) Det hadde kommet en mann styrtende (*object - participle*)
there had come a man rushing
'A man had come rushing'

(37) Det hadde kommet styrtende en mann (*participle - object*)
there had come rushing a man
'A man had come rushing'

With present participles that are adjuncts, the participle cannot precede the object, and it is of course not expected that an adjunct should be positioned between the (non-finite) main verb and its object as in the ungrammatical (38).

(38) *Det hadde kommet syngende en mann
there had come singing a man

I will argue that the difference in word order reflects a deeper difference between the sentences — (37) has restructuring, while (36) does not.

In a presentational focus sentence without restructuring such as (36), the tripartite c-structure (39) is assumed for *main verb - object - present participle*. The f-structure assumed is (40).

(39)
```
          VP
        /  |  \
       V  DP    VP
  come   a man rushing
```

(40)
```
PRED 'come <(↑OBJ) (↑XCOMP)> (↑SUBJ)'
SUBJ [FORM  'there']
OBJ [PRED  'man']
XCOMP      [SUBJ
            PRED 'rush <(↑SUBJ)>']
VFORM PLUPERFECT
```

In (40), the verb *komme* 'come' takes expletive *det* 'there' as its subject. The DP *en mann* 'a man' is assumed to be its object, while the present participle takes *en mann* 'a man' as its subject (via functional control of its subject position). Given this analysis, there cannot be a complex predicate in (40). A complex predicate takes one single set of syntactic functions — it cannot be the case that the first verb takes one subject, while the second verb takes a different subject. Sentence (36) would thus represent a problem if *komme* + a participle were assumed to have obligatory restructuring (which seems to be the position in Thurén 2008).

In a presentational focus sentence with restructuring, such as (37) above, the word order is *present participle* - *object*. I assume that the object is a part of the present participle VP, as in (41). The f-structure assumed is (42).

(41)
```
            VP
           /  \
          V    VP
       come   /  \
             V    DP
       rushing     |
                 a man
```

(42)
```
PRED 'come-rush <(↑OBJ)> (↑SUBJ)'
SUBJ [FORM 'there']
OBJ [PRED 'man']
VFORM PLUPERFECT
```

There is a complex predicate *komme-styrtende* 'come rushing' which takes an expletive subject, and *en mann* 'a man' as its object. In the argument structure, there is an "empty" role that is realized as the expletive subject, visualized as underlining in (43).

(43) *komme styrtende* ___ < theme < agent > >

The thematic argument is realized as an object of the complex predicate. The rule for the presentational focus construction has applied to the complex predicate as a whole. Independent evidence that the presentational focus rule can apply to a complex predicate comes from sentences such as (14) above, in which it has applied to a complex predicate with an infinitive as its second part.

5. The verb *bli* 'remain'

The verb *bli* 'remain' is not among the verbs that allow the presentational focus construction when it is a matrix verb, as shown in (44). It also does not allow this construction when a present participle follows the object — the word order for sentences without restructuring — as shown in (45). However, it is allowed when the participle precedes the object — the word order for sentences with restructuring — as in (46).

(44) *Det ble en mann her lenge
there remained a man here long
(45) *Det ble en mann liggende på bakken
there remained a man lying on ground.DEF
(46) Det ble liggende en mann på bakken
there remained lying a man on ground.DEF
'A man remained lying on the ground'

The ungrammaticality of (45) and the grammaticality of (46) follow from the analysis given here. The point is that *bli* 'remain' constitutes the matrix predicate alone in (45), while it is a part of a complex predicate in (46). This is another argument that the difference between the word orders reflects the deeper analysis. The contrast (45) - (46) also shows that the presentational focus construction with a complex predicate requires that the verb occurring as a present participle allows this construction. When the verb does not allow this construction, the complex predicate as a whole does not. This restriction can also be seen in sentences with the verb *hete* 'be.called'. This verb never takes the presentational focus construction, neither as a matrix verb, as in (47), nor as the second part of a complex predicate, as in (48).

(47) *Det het en hund Troll
there was.called a dog Troll
'A dog was called Troll' [intended]
(48) *Det ble hetende en hund Troll
there was be.called.ing a dog Troll

Another argument for my analysis concerns the position of an object relative to an oblique. In (46) above, the oblique must follow the object; it cannot precede the object as in (49).

(49) *Det ble liggende på bakken en mann
there remained lying on ground.DEF a man

The oblique in (46) must be selected by the present participle, but it cannot immediately follow it. The word order of the embedded VP is simply the general VP word order *verb - object - oblique* (which is also found when the posture verb is the main verb in a presentational focus sentence). If *en mann* 'a man' were an object of *bli* 'remain', these facts would be difficult to account for.[9]

6. Against an auxiliary analysis

It was mentioned above that *komme* and especially *bli* have been considered auxiliaries — with some reservations (Teleman et al. 1999:618-19, Faarlund 1997:472, Ebeling 2003:154-177, Kinn 2014:77-78, 83). Auxiliary is a difficult concept, which is used of verbs with rather different properties. Even so, there are some general properties that are assumed to distinguish auxiliaries from light verbs (Butt 2010, Butt and Lahiri 2013, Seiss 2009). Properties relevant to the case at hand include the following:

- Light verbs such as *komme* and *bli* are used in all forms and periphrases, while auxiliaries are often used in some forms only.
- Light verbs such as *komme* and *bli* often have limited combinatorial options, while auxiliaries usually occur with all kinds of verbs.

There are also language specific syntactic differences: *komme* og *bli* differ from auxiliaries in not topicalizing their complement. Furthermore, *komme* takes the presentational focus construction, while auxiliaries do not.

---

[9] The verb *bli* 'remain' has one property that could give an argument for auxiliary status. Kinn (2014:77) points out that its present participle VP can be pronominalized.

(i) (Ble det liggende en mann på bakken?) Ja, det ble det
remained there lying a man on ground.DEF yes there remained that
'Did a man remain lying on the ground? Yes, he did'

In my view, this is not a real argument. The verb *bli* can pronominalize a complement independently of its category. This is true of all its uses as a main or auxiliary verb (as an alternative to pronominalizing the larger VP with *gjøre det* 'do it). An example is (ii).

(ii) (Ble han hjemme?) Ja, han ble det
remained he home Yes he remained that
'Did he stay home? Yes, he did.'

7. The verb *ha* 'have' and the preposition *med* 'with'

Sentences with *ha* 'have' and *med* 'with' are not mentioned in Thurén (2008). The syntax of *ha* 'have' offers many challenges. I assume that the object of *ha* 'have' can be non-thematic (see e.g. Sæbø 2009), and that a sentence such as (50) with *ha* 'have' and a verbal participle takes an XCOMP and a raised object. Its f-structure is as in (51).

(50) Jeg har en fin gammel portvin stående (Faarlund et al 1997:752)
I have a fine old port.wine standing
'I have a fine old port wine standing'

(51)
```
[PRED 'have <(↑SUBJ) (↑XCOMP)> (↑OBJ)'               ]
[SUBJ [PRED PRO]                                     ]
[OBJ [PRED 'port wine']                              ]
[XCOMP        [SUBJ                       ]          ]
[             [PRED 'stand <(↑SUBJ)>'     ]          ]
[VFORM PRES                                          ]
```

As in the presentational focus sentences discussed above, word order gives an argument for optional restructuring. Faarlund et al (1997:752-53) discuss the fact that an object can precede the participle, as in (50), or follow it, as in (52). An adjectival present participle cannot follow the object, cf. (53).

(52) Jeg har stående en fin gammel portvin (Faarlund et al 1997:753)
I have standing a fine old port.wine
'I have a fine old port wine standing'

(53) *Vi kan ikke ha snokende en skatteinspektør her
we can not have snooping a tax.inspector here
'We cannot have a tax inspector snooping here' [intended]

Faarlund et al. claim that the object can follow the participle if it is indefinite. This is reminiscent of the definiteness restriction in presentational focus sentences (not mentioned by Faarlund et al.). The definiteness restriction applies, as expected, to the object of a complex predicate with an expletive subject, as shown in (54).

(54) Det hadde kommet styrtende en mann / *mannen
there had come rushing a man / man.DEF
'A / *The man had come rushing'

In a sentence such as (52), however, the subject of the complex predicate is not expletive, and there is no reason there should be a definiteness restriction.

Text searches give acceptable examples with a definite object, such as (55),[10] so the restriction in Faarlund et al. does not seem to be empirically correct.

(55) Noen som også har liggende den siste oppdaterte versjonen?
anybody that also has lying the last updated version.DEF
'Anybody who has the last updated version as well?'
mac1.no/forum/viewtopic.php?p=1903139 11/03/16

We see, then, that there is optional restructuring with the verb *ha*.[11] This fact gives another argument against the idea that verbal present participles are main verbs with auxiliary verbs selecting them, since it would be impossible to analyze two-place *ha* as an auxiliary.

With restructuring, the simplified f-structure of (52) above is as in (56).

(56)
$$\begin{bmatrix} \text{PRED 'have-stand <(}\uparrow\text{SUBJ) (}\uparrow\text{OBJ)>'} \\ \text{SUBJ [PRED PRO]} \\ \text{OBJ [PRED 'port wine']} \\ \text{VFORM PRES} \end{bmatrix}$$

The preposition *med* 'with' also gives interesting evidence concerning the relation between word order and restructuring. It has often been pointed out that this preposition shares aspects of its syntax with the verb *ha*. One of these is the option of selecting a verbal present participle. There is one

---

[10] A reviewer points out that an object following the participle cannot be pronominal. A pronominal object must follow *ha* 'have', cf. (i).
(i) Jeg har den stående her / *Jeg har stående den her
I have it standing here / I have standing it here
'I have it standing here'
This might be interpreted as a case of clitic climbing. For another possible case of clitic climbing in Norwegian, see Lødrup (1996:84).

[11] There is a meaning difference between sentences with and without restructuring, which follows from the account given here. In sentences without restructuring, I assume subject-to-object raising, which means that there is no thematic relation between *ha* 'have' and the object. In sentences with restructuring, on the other hand, there is a thematic relation here; the object is assumed to realize both the internal role of *ha* 'have' and the role of the present participle. Consider (i)-(ii).
(i) Han har tungen hengende ut av munnen
he has tongue.DEF hanging out of mouth.DEF
'He has his tongue hanging out of his mouth'
(ii) ??Han har hengende tungen ut av munnen
he has hanging tongue.DEF out of mouth.DEF
It is strange to say that a person 'has' his tongue. Even so, there is nothing strange about (i), because the object is non-thematic relative to *ha* 'have'. On the other hand, (ii) preserves this strangeness, because restructuring does not sever the object's thematic relation to *ha* 'have'.

important difference, however: The preposition *med* only allows one word order: the object preceding the present participle, cf. (57)-(58). This follows when one assumes that a preposition cannot take restructuring.

(57) med en fin gammel portvin stående ..
with a fine old port.wine standing
'with a fine old port wine standing ..'
(58) *med stående en fin gammel portvin ..
with standing a fine old port.wine

Some Swedish and Danish dialects have a construction in which 'have' takes a second verb with the same inflectional form as 'have'. Examples are (59)-(60). Swedish and Danish differ in that Danish has the conjunction *og* 'and' preceding the second verb.

(59) Jag har cykeln står på gården (Swedish, Nordberg 1977:117)
I have bike.DEF stands in backyard.DEF
'I have my bike standing in the backyard'
(60) Vi havde en lang bænk og stod i køkkenet (Danish, Pedersen 2014:223)
we had a long bench and stood in kitchen.DEF
'We had a long bench standing in the kitchen'

The group of possible second verbs seems to be the same as in the construction with a present participle (Nordberg 1977:118, Pedersen 2014:229). A complex predicate analysis is proposed informally in Pedersen (2014) (see also Larsson 2014). The morphological form of the second verb must be seen as a case of verbal feature agreement, which has been established as a restructuring phenomenon (Niño 1997, Sells 2004, Wiklund 2007, see also (14)-(16) above). The word order is not expected from a Norwegian point of view, but Scandinavian languages and dialects do not necessarily have the same word order in these and related constructions (Larsson 2014).

There is a use of the verb *få* 'get' that gives an interesting parallel to the restructuring sentences with *ha* 'have'. In (61), *få* 'get' takes a participle with a following object.

(61) Han fikk reparert bilen
he got repaired car.DEF
'He got the car repaired'

Example (61) is ambiguous. It has an "active" interpretation, where the subject of *få* is the agent of the repairing, as well as a "passive" interpretation, where the subject of *få* is a benefactive, and the agent of the repairing is not specified. Lødrup (1996) argues that the latter interpretation is syntactically distinct from the first one. In his analysis, (61) with the "passive" interpretation is a complex predicate construction with a passive second verb. The parallel to the sentences with *ha* 'have' is striking. In both cases, there is an embedded VP, with a verb realizing its argument as an object. This embedded VP has no subject of its own, and combines with the verb above it to form a complex predicate.

The parallels go even further. Example (61) above has the word order *participle - object*. As with *ha*, there are also sentences with the word order *object - participle*, such as (62).

(62) Han fikk bilen reparert
he got car.DEF repaired
'He got the car repaired'

Lødrup (1996) shows that there are grammatical differences between sentences with different word orders, and suggests that sentences like (63) are not restructuring sentences — the participle is an XCOMP.

8. Conclusion

Verbal present participles do not show uniform behavior in syntax. There is no evidence that they can be main verbs in sentences with auxiliaries. A verbal present participle can be a verbal XCOMP[12], or it can take part in restructuring with its selecting verb. In these constructions, they have the properties expected. The LFG theories of controlled complements and complex predicates make a simple account possible.

---

[12] There is one problem with my analysis that cannot be discussed due to lack of space: The analysis is not compatible with the analysis of pseudocoordinations given in Lødrup (2002, 2014b, 2014c). The problem concerns sentences such as (i).

(i) Det kom en mann styrtende og brølte
there came a man rushing and roared
'A man came rushing and roared'

This is a presentational focus sentence without restructuring, in which the present participle is an XCOMP. Following the participle is the constituent *og brølte* 'and roared'. Example (i) is not a coordinate structure, however, but a so-called pseudocoordination. Lødrup (2002, 2014b, 2014c) assumes that pseudocoordinations are subordination constructions, in which the second part is (usually) an XCOMP. Combining my analyses of pseudocoordinations and verbal present participles would make both *styrtende* 'rushing' and *og brølte* 'and roared' XCOMPs. This would be an impossible situation, a violation of the uniqeness condition. I have no solution to this problem — maybe the real question is the analysis of pseudocoordinations?

References

Aagaard, Teodor Ekblad. 2016. *Doble partisipper i norsk: Verbal trekkongruens og restrukturering*. MA thesis. University of Oslo.

Alsina, Alex. 1996. *The Role of Argument Structure in Grammar: Evidence from Romance*. Stanford, CA: CSLI Publications.

Andrews, Avery D. and Christopher D. Manning. 1999. *Complex Predicates and Information Spreading in LFG*. Stanford, CA: CSLI Publications.

Börjars, Kersti and Nigel Vincent. 2005. Position versus function in Scandinavian presentational construction. In Miriam Butt and Tracy Holloway King (eds.) *Proceedings of the LFG05 Conference*, 54-72. Stanford CA: CSLI Publications.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford, CA: CSLI Publications.

Butt, Miriam. 2010. The light verb jungle: Still hacking away. In Mengistu Amberber, Brett Baker, and Mark Harvey (eds.) *Complex Predicates in Cross-Linguistic Perspective*, 48-78. Cambridge University Press.

Butt, Miriam, and Aditi Lahiri. 2013 Diachronic pertinacity of light verbs. *Lingua* 135, 7-29.

Butt, Miriam, Maria-Eugenia Niño and Frédérique Segond. 1996. Multilingual processing of auxiliaries in LFG. In Dafydd Gibbon (ed.) *Natural Languages Processing and Speech Technology*, 111-122. Berlin: Mouton de Gruyter.

Cinque, Guglielmo. 2004. Restructuring and functional structure. In Adriana Belletti (ed.) *Structures and Beyond. The Cartography of Syntactic Structures* 3, 132-191. Oxford: Oxford University Press.

Dalrymple, Mary and Helge Lødrup 2000. The grammatical functions of complement clauses. In Miriam Butt and Tracy Holloway King (eds.) *Proceedings of the LFG00 Conference*. CSLI Publications.

Ebeling, Signe Oksefjell. 2003. *The Norwegian Verbs bli and få and their Correspondences in English: A Corpus-Based Contrastive Study*. Dissertation, Faculty of Arts, University of Oslo.

Egerland, Verner. 2002. On absolute constructions and the aquisition of tense. *Working Papers in Scandinavian Syntax* 69, 77–120.

Faarlund, Jan Terje, Svein Lie and Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.

Falk, Yehuda. 2008. Functional relations in the English auxiliary system. *Linguistics* 46, 5, 861-889.

Hansen, Erik and Lars Heltoft. 2011. *Grammatik over det danske sprog*. København: Det Danske Sprog- og Litteraturselskab.

Havnelid, Inga Margrethe. 2015. *"Prøv å fakk skurken som gjemmer seg på Manhattan": Om dobbel imperativ i norsk*. MA thesis. University of Oslo.

Higgins, F. R. 1973. On J. Emonds's analysis of extraposition. In John P. Kimball (ed.), *Syntax and Semantics 2*, 149–195. Academic Press.

Holm, Elisabeth. 2013. *Det sto aldri på noe papir, men det lå i luften og satt i veggene: Kroppspositurverb brukt om eksistens og lokasjon i norsk*. MA Thesis. University of Oslo.

Kinn, Torodd. 2014. Verbalt presens partisipp. *Norsk Lingvistisk Tidsskrift* 32, 1, 62-99.

Larsson, Ida. 2014. Serial verb constructions. *Nordic Atlas of Language Structures (NALS) Journal*, Vol. 1, 370-80. http://tekstlab.uio.no/nals#/chapter/65

Lødrup, Helge. 1996. The theory of complex predicates and the Norwegian verb *få* 'get'. *Working Papers in Scandinavian Syntax* 57, 76-91.

Lødrup, Helge. 1999. Linking and optimality in the Norwegian presentational focus construction. *Nordic Journal of Linguistics* 22, 2, 205-229.

Lødrup, Helge. 2002. The syntactic structures of Norwegian pseudocoordinations. *Studia Linguistica* 56, 2, 121-143.

Lødrup, Helge. 2004. Clausal complementation in Norwegian. *Nordic Journal of Linguistics* 27, 1, 61-95.

Lødrup, Helge. 2011. Lexical Functional Grammar: Functional Structure. In Robert D. Borsley and Kersti Börjars (eds.) *Non-Transformational Syntax: Formal and Explicit Models of Grammar*, 141 - 180. Chichester: Wiley-Blackwell.

Lødrup, Helge. 2012. In search of a nominal COMP. In Miriam Butt and Tracy Holloway King (eds.) *Proceedings of the LFG12 Conference*, 383-403. Stanford, CA: CSLI Publications.

Lødrup, Helge. 2014a. Long passives in Norwegian: Evidence for complex predicates. *Nordic Journal of Linguistics* 37, 3, 367-391.

Lødrup, Helge. 2014b. There is no reanalysis in Norwegian pseudocoordinations (except when there is). In Hans Petter Helland and Christine Meklenborg Salvesen (eds.), *Affaire(s) de grammaire: Mélanges offerts à Marianne Hobæk Haff à l'occasion de ses soixante-cinq ans*, 43-65. Oslo: Novus.

Lødrup, Helge. 2014c. How can a verb agree with a verb? Reanalysis and pseudocoordination in Norwegian. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG14 Conference*, 367-86. Stanford, CA: CSLI Publications.

Lowe, John J. 2015. Complex predicates: an LFG + glue analysis. *Journal of Language Modelling* 3, 2, 413-462.

Lundquist, Björn. 2009. *Nominalizations and Participles in Swedish*. Ph.D. Dissertation. University of Tromsø.

Monachesi, Paola. 1998. Italian restructuring verbs: A lexical analysis. In Erhard Hinrichs, Andreas Kathol and Tsuneko Nakazawa (eds.) *Complex Predicates in Nonderivational Syntax*. *Syntax and Semantics* 30, 313-368. San Diego: Academic Press.

Niño, Maria-Eugenia. 1997. The multiple expression of inflectional information and grammatical architecture. In Francis Corblin, Danièle Godard and Jean-Marie Marandin (eds.), *Empirical Issues in Formal Syntax and Semantics*, 127-47. Bern: Peter Lang.

Nordberg, Bengt. 1977. Har du Dahlstedt ligger på nattduksbordet? En syntaktisk egendomlighet i mellansvenskt talspråk. In Claes-Christian Elert, Stig Eliasson, Sigurd Fries and Sture Ureland (eds) *Dialectology and Sociolinguistics*. Umeå Studies in the Humanities 12, 113-136.

Pedersen, Karen Margrethe. 2007. Nogle danske dialektale pseudosideordninger. In Torben Arboe (ed.) *Nordisk dialektologi och sociolingvistik*, 314-322. Århus: University of Århus.

Sæbø, Kjell Johan. 2009. Possession and pertinence: the meaning of *have*. *Natural Language Semantics* 17, 4, 369–397.

Seiss, Melanie. 2009. On the difference between auxiliaries, serial verbs and light verbs. *Proceedings of the LFG09 Conference*, 501-19. CSLI Publications.

Sells, Peter. 2004. Syntactic information and its morphological expression. In Louisa Sadler and Andrew Spencer (eds.) *Projecting Morphology*, 187–225. Stanford, CA: CSLI Publications.

Teleman, Ulf, Staffan Hellberg and Erik Andersson. 1999. *Svenska akademiens grammatik. Volume 2*. Stockholm: Svenska Akademien.

Thurén, Camilla. 2008. *The Syntax of the Swedish Present Participle*. Ph.D. Dissertation. Lund University.

Wedekind, Jürgen, and Bjarne Ørsnes. 2004. An LFG account of the Danish verbal complex and its topicalization. *Acta Linguistica Hafniensia* 36, 1, 35-64.

Western, August. 1921. *Norsk riksmåls-grammatikk*. Kristiania: Aschehoug.

Wiklund, Anna-Lena. 2007. *The Syntax of Tenselessness*. Berlin: Mouton de Gruyter.

Wurmbrand, Susanne. 2001. *Infinitives: Restructuring and Clause Structure*. Berlin: Mouton de Gruyter.

Wurmbrand, Susi. 2004. Two types of restructuring - Lexical vs. functional. *Lingua* 114, 991-1014.

Participles, gerunds and syntactic categories

John Lowe
University of Oxford

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 401–421

Keywords: mixed categories, English gerund, participles, morphology

Lowe, John. 2016. Participles, gerunds and syntactic categories. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 401–421. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.21.

Abstract

The phenomenon of so-called 'mixed' categories, whereby a word heads a phrase which appears to display some features of one lexical category, and some features of another, raises questions regarding the criteria used for distinguishing syntactic categories. In this paper I critically assess some recent work in LFG which provides 'mixed category' analyses. I show that three types of evidence are typically utilized in analyses of supposed mixed category phenomena, and I argue that two of these are not, in fact, crucial for determining category status. I show that two distinct phenomena have become conflated under the 'mixed category' heading, and argue that the term 'mixed category' should be reserved for only one of these.

1 Mixed categories

So-called 'mixed' categories have been the subject of a number of works within LFG, in particular by Bresnan (1997, 2001, 289–296) and Bresnan et al. (2016, 309–319), Spencer (2004), Bresnan & Mugane (2006), Seiss (2008), Nikitina (2008), and Alsharif (2014). Most recently, Nikitina & Haug (2016), Spencer (2015) and Börjars et al. (2015) have discussed phenomena which they analyse in terms of mixed categories.

The most commonly discussed mixed category is undoubtedly the English gerund in *-ing*. In fact, the English gerund is a particularly complicated example, because there are three different phrasal configurations in which the gerund may appear, that is with entirely nominal, entirely verbal, or 'mixed' phrasal structure:

(1) a. Type A: His stupid missing of the penalty lost us the game.
b. Type B: Him stupidly missing the penalty lost us the game.
c. Type C: His stupidly missing the penalty lost us the game.

I refer to these as English gerund types A, B and C respectively. In all these examples, the gerund heads a phrase which functions as subject of the sentence. In type A (1a), the syntax of the phrase headed by *missing* is entirely nominal: *missing* is premodified by an adjective and a possessor phrase, and the logical object of *missing* appears as a prepositional complement. The exclusively nominal syntax of the phrase means that *missing* here is unambiguously a noun, of category N, heading an unremarkable NP. The type A gerund type is therefore not a mixed category, and will not be considered further.

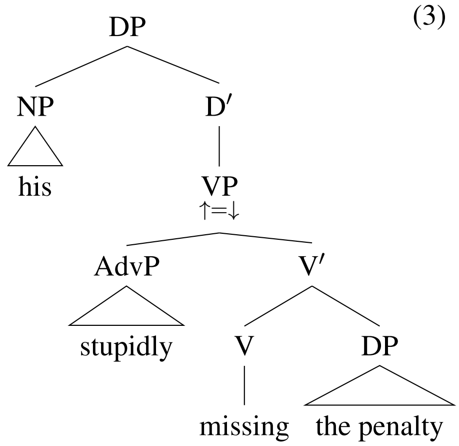
[†]I am very grateful for insightful comments and criticisms to Andrew Spencer, to the audience at SE-LFG 20, 21 May 2016, in particular Louisa Sadler, John Payne, Miriam Butt and Jamie Findlay, and to the audience at HeadLex16, 25 July 2016, in particular Bob Borsley and Dag Haug. This work was supported by a grant from the Jill Hart Fund for Indo-Iranian Philology at the University of Oxford, and parts of the work were undertaken while I was in receipt of an Early Career Research Fellowship from the Leverhulme Trust. All errors are of course my own.

In contrast, the syntax of the phrase headed by *missing* in type B (1b) is entirely verbal: the logical object appears in the same 'bare' form as an object of a finite verb (i.e. not embedded under a preposition); the logical subject likewise appears in the 'bare' form, with oblique/accusative case; and the modifier is an adverb. This type will be discussed below.

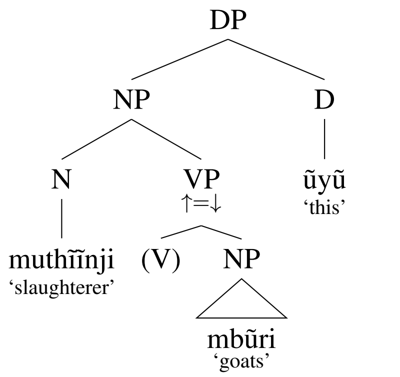
The unambiguously mixed construction is type C (1c): the logical object and the modifier are of the verbal type (as in (1b)), but the logical subject appears as a possessive phrase (as in (1a)). The analysis proposed by Bresnan et al. (2016, 311–319) involves a 'head-sharing' construction with a VP as (co-)head of a DP (the DP is the 'extended head' of VP); this is shown in (2).

The construction in (1c) involves a lexical category serving as (co-)head of a functional category ('lexical-functional head-sharing'). It is also possible, however, for a lexical category to have a different lexical category as its extended head ('lexical-lexical head-sharing'). Following Bresnan & Mugane (2006), Gĩkũyũ nominalizations involve a VP with NP as extended head (3).[1]

(2)



(3)



Other types of head-sharing construction have also been proposed, which do not correspond precisely to either of the types discussed above. For example, Nikitina (2008) proposes that IP may take DP as an extended head, i.e. that 'functional-functional' head sharing is possible, while Bresnan et al. (2016) allow the exocentric category S to take DP as extended head, and similarly Nikitina & Haug (2016) permit S to take NP as an extended head.

Bresnan (1997) contrasts the LFG approach to mixed categories with a number of alternative possibilities, most importantly 'the indeterminate category projection theory' and 'projection-switching' approaches. The 'indeterminate' approach assumes that the head of a mixed category is lexically underspecified for category, and projects a phrase which is likewise underspecified, such that it may contain e.g. both nominal and verbal structure. The 'projection-switching' approach is similar,

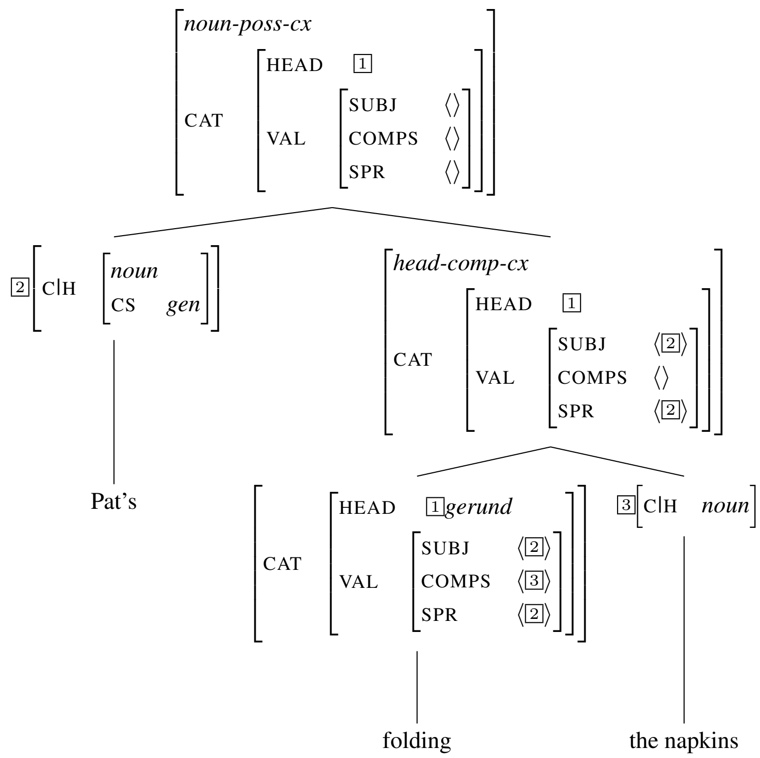
[1]In the trees below I show only the crucial (co-)head annotation ($\uparrow=\downarrow$) which establishes the mixed category. Other annotations are as expected: heads are annotated $\uparrow=\downarrow$, non-heads have other annotations.

only slightly more constrained, assuming that the head is assigned some kind of intermediate, or dual category, status, which enables it to project e.g. as a verb up to a certain level, and as a noun above.

Indeterminate or intermediate approaches to mixed categories have been widely proposed, in particular by Lapointe (1993), Hudson (2003), and Malouf (2000a) within HPSG. For example, Lapointe (1993) treats category labels as bipartite, and argues for an intermediate category (a “dual lexical category”) ⟨N|V⟩ to account for the English gerund. This category projects verbally, except for its top layer (ensuring it has the distribution of a noun phrase, for example), and the *-ing* suffix has a morphological function in deriving a category ⟨N|V⟩ from a plain category ⟨V|V⟩.

In the HPSG analysis by Malouf (2000a), the mixed properties of the English gerund are accounted for by means of the multiple inheritance hierarchy for HEAD values. The HEAD value *gerund* is a subtype of both *noun* and *verbal*. As a subtype of *noun*, its external distribution is that of an NP (it can “occur anywhere an NP is selected for”), and does not necessarily correspond to the distribution of verbs (which are a separate subtype of *verbal*). Adverbial modification applies to elements of category *verbal* (which includes *adjective*). Since adjectives modify only *c(ommon)-nouns*, gerund phrases cannot be modified by adjectives.

(4) English gerund in HPSG (Malouf, 2000b, 22)



This is a more sophisticated intermediate category analysis, and the use of a multiple inheritance hierarchy permits cross-classification across more than one grammatical category: it is possible to define a category that shares some features with nouns and some with verbs, without requiring it to be fully one or the other.

Within LFG, there is no concept of intermediate categories: there is a finite inventory of distinct lexical categories, and every lexical word must belong to exactly one category.[2] The different lexical categories are not unrelated; for example, Bresnan et al. (2016, 103) propose that the major lexical categories N, V, P and A can be analyzed according to a categorial feature matrix, whereby V and A are $+$PREDICATIVE and V and P are $+$TRANSITIVE, but these features are used purely to cross-classify pairs of categories, and not to license underspecified or intermediate categories. LFG also admits the possibility of complex categories (Butt et al., 1999, 192): within any one category it is possible to distinguish an unrestricted number of subcategories. However, categorization remains absolute, in the sense that every word is necessarily a member of one and only one syntactic (sub)category: it is not possible for a word or class of words to be intermediate between one category and another, or to be underspecified with respect to category membership.

The LFG approach to mixed categories violates a strict approach to endocentricity, both in the fact that words may head phrases of different categories (albeit restricted to cases of morphological derivation from the category concerned), and in the fact that lexical phrases are permitted to lack an explicit head internal to the phrase. This is not to say that the approach is therefore not viable, given that non-X′ theoretic structures are admitted in LFG, but assuming that non-endocentric structures are a marked feature of grammar, it does suggest that mixed categories should be admitted only where an alternative, X′ theoretic, analysis cannot adequately account for the linguistic data.

2 'Mismatched' categories

Whatever approach one adopts to deal with mixed categories, it is clearly necessary to distinguish phrases that are mixed from those that are not. The English type C gerund (1c) is uncontroversially mixed: the phrase concerned contains both a nominal element, a possessor, and verbal elements, such an object and adverb. However, many authors also treat as mixed phrases which are consistently verbal in terms of their internal structure. Analyses of mixed categories proposed within LFG reveal three major properties implicated in the categorization of a word:

(5)
a. Internal syntax: the internal structure of the phrase, for example whether it contains determiners, adjectives, objects, adverbs.
b. Distribution: the distribution of the phrase at a clausal level, for example whether it can appear in the same structural positions, and fill the

[2] Some words may belong to one lexical and one functional category, and/or more than one functional category.

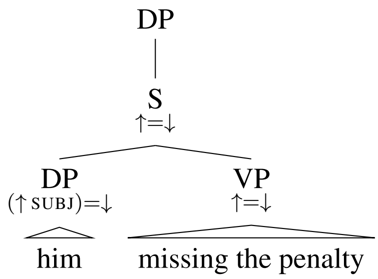
same grammatical functions, as noun phrases that function as subjects and objects.

c. Morphosyntax: the morphosyntactic properties of the head of the phrase, for example whether it shows the agreement features typical of a verb or an adjective.

Consider the type C gerund in (1c) and (2): the internal syntax of the phrase headed by the gerund is mixed, in that the phrase contains elements which are specific to DPs in English (possessive modifier) and elements which are specific to VPs in English (object and adverb modifier); the distribution of the phrase is nominal, since it can function as a subject (as in the example provided), object, or other grammatical function, or indeed can appear in any of the positions in a clause that an ordinary noun phrase can; given the lack of morphology in English, the gerund in *-ing* is morphologically unclear, but it shows the same kind of tense/aspect distinctions as finite verb sequences (cf. *his having missed the penalty...*), and so must be morphologically verbal on some level, at least.

In many discussions of mixed categories in LFG, mixed internal syntax is not considered necessary for the categorization of a phrase as mixed. So, Bresnan et al. (2016, 318) treat the type B English gerund (1b) via the same kind of head-sharing analysis as the type C construction (1c), the only difference being the category of the (co-)head phrase:

(6)



As with (2) above, the gerund appears in the head of a VP, with an object complement like any transitive V, but in this case the VP and accusative case subject phrase *him* together constitute a clausal phrase S, which serves as (co-)head of the higher DP. Thus the phrase as a whole is a DP, headed by a V in an embedded VP in an embedded S. This is very similar to (2). But the internal syntax of the phrase is entirely verbal (or clausal): there can be no adjectival modifier, or determiner or possessor phrase. Given that there are no explicit morphological properties of the gerund that require a DP node, it is clear that the DP node is assumed purely to account for the distribution of the phrase, i.e. distribution is taken as a sufficient criterion for mixed category status. Similarly, Haug & Nikitina (2016, 15) assume a head-sharing construction for a participle construction in Latin (the 'dominant' participle construction) which shows the external syntactic distribution of a noun, but whose internal syntactic structure is that of an S. Since Latin lacks a DP, they assume an S as a (co-)head of a lexical category. Again, the only justification for the NP projection above S is the distribution of the phrase.

3 Attributive participles

The importance of both distribution and morphosyntax to categorization is evident in some LFG approaches to participles. Spencer (2015) discusses participles in Sanskrit, as treated by Lowe (2015), and also in Lithuanian, proposing an analysis that is in many respects the same as that proposed for German participles by Bresnan (1997, 2–3). All these participles are diachronically related, and are of the same 'subject-oriented' type; that is, to the extent that an attributive participial clause can be treated as a reduced relative clause, it is necessarily the subject that is relativized on. The participle agrees with and attributively modifies a noun, which is identified with the gapped subject of the participial clause. Ex. (7) shows an attributive participle in Sanskrit, (8) one from Lithuanian, and (9) one from German.

(7)	sárasvatī	sādháyantī	dhíyaṃ	naḥ
	S.NOM.SG	perfect.PTC.PRS.N.SG.F	thought.ACC.SG	us.GEN

'Sarasvatī who perfects our thought' (Sanskrit, RV 2.3.8a)

(8)	ateisiančios	žiemos	ilgumo
	come.PTC.FUT.ACT.GEN.SG.F	winter.GEN.SG	length

'the length of the coming winter' (Lithuanian, from Spencer, 2015, ex. 3)

(9)	ein	mehrere	Sprachen	sprechender	Mann
	a	several	language.PL	speak.PTC.NOM.SG.M	man.NOM.SG

'A man who speaks several languages' (German, cited by Bresnan, 1997, 2)

The attributive use is the most canonically adjectival use of participles, but adjectives can also, to a slightly more limited extent, be used as clausal adjuncts, so in both uses participles appear to display adjectival distribution. These participles can also have other uses. For example, as discussed by Lowe (2015), Sanskrit participles can also be used predicatively, that is in 'converbal' or clausal adjunct use (10).

(10)	víṣūco	áśvān	yuyujāná	īyata	ékaḥ
	separated.ACC.PL	horses.A	yoke.PF.PTC.MED.NOM.SG.M	speeds	alone

'Having yoked the separated horses, he speeds (off) alone.' (RV 6.59.5cd)

I focus on attributive participles here, following Spencer (2015) and Bresnan (1997), but the analysis advanced could easily be extended to deal with predicative participles.

My concern in this paper is the phrasal structure of participle phrases, in particular their syntactic category, and not their functional representation. However, Spencer (2015) draws a connection between the categorization and the f-structure representation, so it is necessary to briefly address the latter question. There are two proposals for the f-structural representation of attributive participle clauses in LFG. Lowe (2015, 87–94) proposes that attributive ('adnominal') participles be analysed as heading an ADJ in the f-structure, with a null subject anaphorically identified with the element which the participle modifies:

(11)
$$\left[\begin{array}{ll} \mathrm{PRED} & \text{'Sarasvatī'} \\ \mathrm{ADJ} & \left\{ \left[\begin{array}{ll} \mathrm{PRED} & \text{'make\_perfect}\langle \mathrm{SUBJ,OBJ} \rangle\text{'} \\ \mathrm{REL\text{-}TOPIC} & \left[\mathrm{PRED}\ \ \text{'pro'}\right] \\ \mathrm{SUBJ} & \left[\ \right] \\ \mathrm{OBJ} & \left[\mathrm{PRED}\ \ \text{'thought'}\right] \end{array}\right] \right\} \end{array}\right]$$

This analysis captures the similarity between attributive participial clauses and relative clauses, treating participles as essentially a reduced form of the latter, lacking an explicit relative pronoun. The alternative, assumed by e.g. Haug & Nikitina (2012, 2016) and Spencer (2015), is that the subject of the participle is not anaphorically but functionally controlled by the element the participle modifies, the participial phrase thus being an open adjunct XADJ at f-structure:

(12)
$$\left[\begin{array}{ll} \mathrm{PRED} & \text{'Sarasvatī'} \\ \mathrm{XADJ} & \left\{ \left[\begin{array}{ll} \mathrm{PRED} & \text{'make\_perfect}\langle \mathrm{SUBJ,OBJ} \rangle\text{'} \\ \mathrm{SUBJ} & \left[\ \right] \\ \mathrm{OBJ} & \left[\mathrm{PRED}\ \ \text{'thought'}\right] \end{array}\right] \right\} \end{array}\right]$$

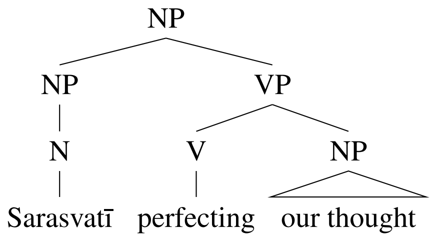
For participles of the Indo-European type considered here there is no evidence against the functional control analysis; it involves a simpler f-structure, and also means that both attributive and predicative uses of participles can be analysed in the same way at f-structure (since predicative participles are uncontroversially XADJs). On the other hand, the attributive XADJ analysis requires a cyclical f-structure, where the f-structure that serves as the SUBJ inside the f-structure of the participle is identical with the f-structure that contains the participle's f-structure (and therefore also contains the SUBJ of the participle, i.e. itself). At a certain degree of cyclicity, cyclic f-structures are problematic: Wedekind (2014) shows that the universal generation problem for unification grammars can be undecidable for cyclic f-structures, whereas it is decidable for acyclic f-structures, as shown by Wedekind & Kaplan (2012). The degree of cyclicity required to render the universal general problem undecidable is considerably greater than in the simple case seen here, but the very fact that cyclicity can cause problems for unification grammars suggests that it is perhaps better avoided.[3]

In any case, neither f-structure possibility is necessarily more adjectival or verbal than the other. The ADJ analysis is closer to the standard analysis of attributive adjectives in LFG as ADJ, but if adjectives were assumed to select for subjects, a far from unlikely possibility, then attributive adjectives would be XADJs, just like the participle in (12).[4] Furthermore, neither f-structure analysis depends to any degree on the c-structure representation: either can be obtained without difficulty regardless of whether an A(dj) node is assumed in the projection of the participle phrase. The f-structure question is therefore tangential to the question of category.

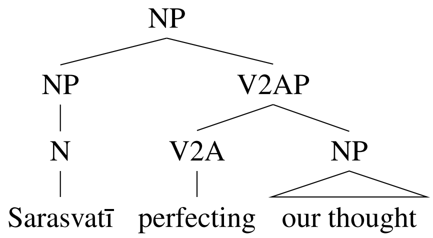
Regarding the structure of participle phrases, Lowe (2015) takes internal syntax to be the primary criterion for category status. The internal syntax of participle phrases is exclusively verbal: e.g. participles of transitive verbal stems may take objects (7, 9), and participles may be modified by adverbs, but never by adjectives. Therefore Lowe (2015) proposes that participle phrases in Sanskrit are VPs headed by participle Vs. This is show in (13), using English words for the Sanskrit of (7).

In contrast, Spencer (2015) makes a somewhat different proposal regarding the c-structure of attributive participles of the Sanskrit/Lithuanian/German type. Spencer's interest is in the interface between morphology and syntax, specifically how the apparently mixed morphological and mixed syntactic properties of participles can be linked. Building on Spencer (2013), a semantic argument structure representation in the morphology determines the category of a word. The morphological process which derives (or inflects) a participle from a verb creates a composite Semantic Function (SF) role, which projects to c- and f-structure in such a way that participles display both adjectival and verbal morphosyntax. In terms of syntactic category, Spencer (2015) proposes that the composite SF role maps to an intermediate category, which he labels 'V2A' (for 'verb-to-adjective transposition'); this is shown in (14).

(13) Ex. (7), fllg. Lowe (2015):



(14) Ex. (7), fllg. Spencer (2015):



Spencer's intermediate category analysis is, as he notes, somewhat similar to other intermediate category proposals, such as that of Malouf (2000a). However, within LFG there is no concept of intermediate categories, meaning that the various features of the category must be stipulated. That is, it must be stipulated that V2A

[3]Thanks to Ron Kaplan (p.c.) for discussion of this point.

[4]On the selection of subjects by adjectives, see Dalrymple et al. (2004a, 197–198).

has the internal syntax of V and the distribution of Adj, since it does not fall out from any principle of intermediate category formation.[5]

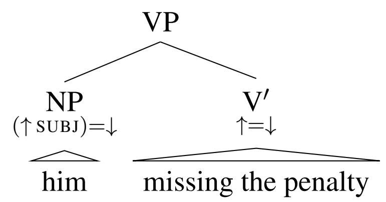
This problem only arises for Spencer (2015), however, because he assumes that attributive participles are a mixed category, partly verbal and partly adjectival, and that there must therefore be an adjectival element to the participial phrase. But participles are not a mixed category in the same sense as the type C English gerund in (1c), since the internal syntax of participial phrases is exclusively verbal. Rather, participles are more like the type B English gerund or the Latin 'dominant' participle construction: it is only the distribution and the morphosyntactic properties of participles which suggest adjectival categorization.

4 Distribution

As discussed in detail in the preceding sections, distribution is taken as key evidence for the category of a phrase within LFG work on mixed categories. In the case of the type B English gerund (1b) and the Latin 'dominant' participle construction, distribution alone has been taken as sufficient evidence to justify a nominal projection, by Bresnan et al. (2016) and Haug & Nikitina (2016).

A detailed analysis of the category of the English gerund is undertaken by Seiss (2008). Seiss (2008) discusses a number of tests which have been proposed to show that English gerunds, including the type B gerund, have the distribution of nouns: functioning as subject or object, complementing prepositions, coordination with NPs, *it*-replacement, tough movement, topicalization, clefting, and pseudo-clefting (see e.g. Bresnan et al., 2016, 309–311). Seiss shows that all these tests also apply to some clausal phrase types, namely *that*-clauses and *to*-infinitive clauses; she argues that since all these distributional tests apply not only to nominal phrases but also to a subset of clausal phrase types, there is no reason to assume a nominal projection in the structure of the type B gerund. In place of Bresnan's analysis of the English type B gerund as involving a DP dominating an S (6), Seiss (2008) proposes the following:

(15)



Under this analysis, the type B gerund is not a mixed category. The gerund itself is of category V, as in Bresnan et al.'s analysis, but it heads only a VP. The

[5] Spencer (2015) notes that an extended head approach would work for this type of participle, but he prefers the intermediate category analysis because the other requires the participle to be of category Adj, and in Spencer's (2013) model, a participle of category Adj would necessarily be a different lexeme from the base verb (which would have category V).

only unusual features of this VP are the gerundal form of the verb, and the fact that the subject appears in SpecVP, not a standard position for subjects in English.

However, some of the tests discussed by Seiss (2008) are somewhat more ambiguous than they appear at first sight. While *that*-CPs and *to*-infinitives can appear it what looks like the same subject and object positions as ordinary DPs, Bresnan et al. (2016, 15–19) argue that these positions are subtly different. On the apparent appearance of CPs in subject position in English, Bresnan et al. (2016) note that auxiliary inversion is impossible with 'subject' *that*-CPs, while it is unproblematic with subject DPs:

(16) a. *That/that he fell over* was quite unexpected.
b. Is *that/*that he fell over* unexpected?
c. How unexpected was *that/*that he fell over*.

On this basis, Bresnan et al. (2016) argue that 'subject' CPs do not appear in the standard structural position for subjects in English, SpecIP, but appear in a higher, adjoined, topic position, and can be identified as the subject by a default principle that subjects are topics. Bresnan et al. propose that SpecIP as the structural subject position in English can only be filled by noun phrases. The data for complement clauses in object position is more complicated, and is related to the ongoing uncertainty over the functional status of complement clauses in LFG. Bresnan et al. (2016, 18) note that complement clauses in English cannot generally enter into the passive alternation:

(17) a. I don't care that languages are learnable.
b. *That languages are learnable isn't cared.

Bresnan et al. (2016, 18) argue that complement clauses are not objects, since only noun phrases can be objects, and this explains why they cannot be passivized. However, Alsina et al. (2005) argue that complement clauses may be OBJ, showing that some complement clauses in Catalan can become passive subjects. In fact some English verbs take clausal complements which may become passive subjects, including some *that*-CP complements:

(18) a. We did not debate whether this was a good thing. / Whether this was a good thing was not debated.
b. They did not consider why he had come. / Why he had come was not considered.
c. They soon forgot that he had previously criticized them. / That he had previously criticized them was soon forgotten.

Clauses introduced by wh-words can relatively freely serve as passive subjects. In addition, subject wh-clauses can even undergo auxiliary inversion, suggesting that they can, in fact, appear in SpecIP, despite being clausal:

(19) a. Whether he knows is not important.

b. Is whether he knows important?

c. How important is whether he knows?

The data for embedded clauses serving as subjects and objects is therefore relatively complicated, but it does not seem possible to restrict the canonical subject and object positions (and SUBJ and OBJ roles at f-structure) to noun phrases. In fact, other lexical categories may also appear in SpecIP:

(20) a. Is slowly the best way to do it?

b. Is under the chair a good place to hide it?

Whether or not a particular embedded clause can serve as a subject or object depends partly on the matrix verb, and partly on the type of embedded clause concerned. Most importantly for the present topic, these restrictions do not seem to be enforcable by purely structural means, e.g. by restricting structural subject and core object positions in English to NP/DP. So distribution cannot be used as a *sufficient* criterion for a particular syntactic categorization.

It would alternatively be possible to consider distribution a *necessary* criterion for categorization: one could argue that two phrases must differ in distributional terms in at least one respect in order for them to be considered members of different categories, at least at the top level. That is, identity of distribution could be taken as evidence for identity of category. But it is not clear that embedded wh-clauses, in particular those introduced by *whether*, are any different from noun phrases in terms of their distribution in English. Given the difficulties discussed above in determining whether embedded clauses may occur in subject or object positions, two of the most easily identifiable positions, it seems reasonable to claim that if identity of distribution necessarily means identity of category, the burden of proof is on establishing absolute identity of category, rather than the other way around.

Consider also the relative benefits of treating distributional differences a necessary criterion for categorization. Given that identity of distribution may be accompanied by differences in internal syntax (as appears to be the case with the type B gerund and ordinary noun phrases), it would be necessary to admit mixed categories in such cases, e.g. where a phrase with the internal structure of a clause or VP could appear in all and only the positions that DPs could appear in. One could of course suppose that even *whether*-clauses and other embedded clause types are mixed, i.e. analysed as CP co-heads of DP. The obvious benefit of treating say the English type B gerund, or even a *whether*-clause, as a DP would be that it simplifies the phrase structure rules. For example, as assumed by Bresnan et al. (2016), one could argue that only DPs can appear in subject and object position in English, and it is neither necessary nor desirable to add an additional rule to the grammar licensing CP, S, or VP in those positions. However, such an apparent simplification of the grammar requires complication elsewhere. So, if only DP can appear in subject and object position, but DPs can also serve as extended heads to clausal categories, some additional constraint will be required to ensure that only clauses of the correct type can serve as (co-)head in a DP. But such a constraint will be effectively equivalent, in terms of load on the grammar, to licensing clausal phrases of the appropriate type in subject and object position. For example, assume that English admits only DPs as the object complement of V:

(21)
$$\begin{array}{lccc} \mathrm{V}' & \rightarrow & \mathrm{V} & \mathrm{DP} \\ & & \uparrow=\downarrow & (\uparrow \text{OBJ}) =\downarrow \end{array}$$

Now it becomes necessary to license type B gerunds as (co-)heads of DP, in order to account for sentences such as:

(22) I resented him giving my book away.

This can be achieved via the machinery for mixed categories discussed above. Embedding S under DP is not formalized by Bresnan et al. (2016), but it would require an additional stipulation, because it does not fall out from any of the principles discussed above. Alternatively, if we assume that the English type B gerund is a VP, its (co-)heading DP will be licensed by the principles discussed above, but we will require an additional rule to license the subject in SpecVP only in this construction. Either way, it will also be necessary to restrict the construction to cases where the V heading the phrase is a gerund, i.e. finite and infinitival Vs must be ruled out.

In contrast, this machinery can be considerably simplified by simply admitting S or VP as a possible object complement of V, when headed by a gerund. The restriction of the construction to a gerund is necessary under any analysis, and under this analysis can be easily achieved e.g. by a f-structure feature or a complex category. The following rule uses an f-structure feature VFORM to constrain the form of the verb via f-structure.

(23)
$$\begin{array}{lccccc} \mathrm{V}' & \rightarrow & \mathrm{V} & \{\mathrm{DP} & \mid & \mathrm{S}\} \\ & & \uparrow=\downarrow & (\uparrow \text{OBJ}) =\downarrow & & (\uparrow \text{OBJ}) =\downarrow \\ & & & & & (\downarrow \text{VFORM}) =_c \text{GER} \end{array}$$

Thus it is by no means clear that a mixed analysis is any simpler than the alternative. Furthermore, given that mixed categories necessarily violate X′ principles of endocentricity, it may be preferable to avoid an analysis involving a mixed construction whenever an equally empirically adequate alternative exists.

The distributional differences and similarities between participles and adjectives in languages like Sanskrit are similarly problematic. Attributive participle phrases do have the same distribution as attributive adjective phrases, in English as well as Sanskrit. But restrictive relative clauses also show the same distribution, meaning that attributive modification does not necessarily involve an AdjP.[6] At the same time, there may be differences in distribution between participle phrases and adjective phrases. In R̥gvedic Sanskrit, for example, participles and adjectives share much of their distribution, both being able to function as attributive modifiers and as clausal adjuncts, but they differ fundamentally in that adjectives, but not participles, can freely occur as primary clausal predicates, with explicit or null copula (Lowe, 2015, 116–127). For example, it is not possible to form a sentence from the noun and participle phrase in (7) by making the latter the main predication, whereas this is unproblematic if the participle is replaced by a derived adjective with roughly the same meaning:

(24) *sárasvatī sādháyantī dhíyam̩ naḥ
S.NOM.SG perfect.PTC.PRS.ACT.NOM.SG.F thought.ACC.SG us.GEN/DAT
Intended meaning: 'Sarasvatī (is the one who) perfects our thought.'

(25) sárasvatī sā́dhanā dhíyo naḥ
S.NOM.SG perfecting.NOM.SG.F thought.GEN.SG us.GEN/DAT
'Sarasvatī (is the one who) perfects our thought.'

Thus it does not seem possible to use distribution as evidence for (even partial) categorial identity between adjectives and participles: participle phrases share much of their distribution and functionality with adjective phrases, but they are not identical, at least in English, Sanskrit and Lithuanian, and most likely in most or all languages with Indo-European type participles.

Given that participle phrases of the Indo-European type have the internal syntax of verb phrases, and that their distribution, while similar to adjectives, does not necessitate even partial adjectival categoriality, I propose that, just as assumed in Lowe (2015), participle phrases are VPs, headed by participial Vs, and do not involve any kind of category mixing. Of course, as with the English gerund discussed above, if participial phrases are VPs, then it is necessary to constrain their distribution, so that for example they can function as attributive modifers and clausal adjuncts, but cannot function as primary clausal predicates. Once again, this is a trivial matter: for example a complex category V[ptc] (26), or reference to an f-structure feature VFORM PARTICIPLE (as in Lowe, 2015), can be used to effectively subdivide the category V in such a way that participle Vs can distribute differently from finite and other Vs. An advantage of assuming VP rather than an intermediate category like Spencer's (2015) V2A, is that the exclusively verbal internal syntax of participle phrases falls out naturally if participle phrases are, in fact, VPs, whereas under an intermediate category analysis it must be stipulated.

[6] In English, attributive modification is restricted by the fact that right branching structures are prohibited in prenominal position, which rules out relative clauses, but not all adjectives and participle phrases, in prenominal position. But this constraint is purely to do with right branching, and not the category of the phrases involved.

(26) $\text{VP[ptc]} \rightarrow \ldots \underset{(\text{ADJ} \in \uparrow)}{\text{V}'\text{[ptc]}} \ldots$

However, Spencer's (2015) argument that participle phrases should be treated as at least partially categorially adjectival is not based on distribution alone, but also on the third criterion used for categorial status in LFG: morphosyntactic agreement features.

5 Morphology and agreement

At first sight, it might seem intuitively obvious that agreement and similar morphosyntactic properties provide evidence of categoriality. In most languages with rich morphological systems, some features are typically expressed on verbs, and others on nouns and adjectives. Many Indo-European languages, including those discussed in the previous section, typically mark person and number on verbs, and case, number and gender on nouns and adjectives: thus person is a feature of verbal morphosyntax, and case and gender features of nominal and adjectival morphosyntax. However, even cross-linguistically regular tendencies may have exceptions; for example, Nordlinger & Sadler (2004a,b) discuss languages in which tense features can be marked on nominals. Within languages too, there are often exceptions to the typical division of morphosyntactic labour between nouns and verbs.

In particular, it is not uncommon for unequivocal finite verb categories to show 'adjectival' or 'nominal' agreement in some languages, in cases where non-finite verbal categories such as participles or agent nominalizations have become grammaticalized and integrated into the finite verbal paradigm. For example, finite present and future verb forms in Russian (and other Slavic languages) mark person and number, but past tense verbs, marked with the formant *-l-*, agree rather in gender and number. The gender/number morphology of past tense verbs in Russian is essentially identical to that of nouns and adjectives in Russian. Diachronically, the past tense in *-l-* derives from a construction involving a verbal adjective in **-lo-*; at some point in the pre-history of Slavic, the forms in *-l-* were categorial adjectives, but now there is no justification for analysing the past tense formation in *-l-* as anything other than fully verbal. Many languages attest equivalent developments; e.g. in many Indo-Aryan languages the (often ergative) perfective aspect derives from what is in Sanskrit a verbal adjective.

So diachronic developments may lead to the recategorization of a particular formation, without necessitating any corresponding morphosyntactic reformation. Thus from a diachronic perspective, it is problematic to assume that morphosyntax necessarily tells us anything about the synchronic status of a formation. If a verb form happens to show adjectival agreement, this does not mean it is synchronically an adjective.

On a purely synchronic level, however, Spencer (2015) argues that it should not be an accident that participles show exactly the same agreement features, and even paradigms, as lexical adjectives. For example, on Lowe's (2015) analysis of Sanskrit participles as Vs, there is no account of the fact that participles and adjectives show the same morphosyntactic properties: it appears to be a synchronic accident. If a synchronic account can be given which could account for the shared morphological properties of adjectives and participles, all the better.

Any such account must be fundamentally morphological. Spencer's (2015) argument that participles are categorially adjectival (at least in part) due to their adjectival morphosyntax depends on the assumption that the morphological feature(s) shared between adjectives and participles necessarily influence the syntactic category of a word. In Spencer's model this is true, since it is the adjectival semantic function in the argument structure representation of a lexeme which determines both agreement properties and syntactic category.

In a different morphological model, however, there need not be a one-to-one correspondence between morphological category and syntactic category. Dalrymple (2015) proposes a model for the morphology-syntax interface as a way of representing the contribution of morphology to the syntactic properties of words. The model presupposes a realizational theory of morphology, which Spencer (2013, 2015) likewise assumes.

Dalrymple (2015) proposes that lexical entries, the building blocks of syntax, are constructed on the basis of three sub-syntactic components: lexemic entries LE, the realization relation R, and the functional description function D. A lexemic entry is a pairing between a lexeme and grammatical information common to all word forms of that lexeme. Formally, LE is a three-place relation defining: a. the form(s) of the root; b. f-descriptions common to all forms of the lexeme; and c. the Lexemic Index (LI), a unique identifier of the lexeme. For example, the LE for the Sanskrit lexeme underlying the adjective *ugra-* 'fierce' is:

(27) LE <{ROOT: ugra}, {(↑ PRED)='fierce'}, FIERCE1>

The first part of the relation defines the form of the root; there are no suppletive or irregular forms, so only the single form appears. The second part of the relation defines syntactic information common to all forms of the lexeme; here, only the f-structure PRED value is represented. This information is represented as part of the lexemic entry itself, since it cannot be changed by morphological processes. The third part of the relation is the LI, which is simply a unique label identifying this lexeme, which could be used, for example, to constrain the application of a particular morphological rule if it were specific to this lexeme.

The realization relation R is a set of four-place relations, *m-entries*, which associate a LI, an s-form, and a p-form with a set of m(orphological)-features.[7] Ex. (28) shows the m-entry for the word form *ugraḥ*, nominative singular masculine of the lexeme FIERCE1. The first part of the relation specifies the lexeme by its LI: this m-entry applies only to this particular lexeme (other lexemes will have equivalent m-entries). The second two parts of the relation define the s-form and p-form for

[7]On p-forms and s-forms, see Dalrymple & Mycock (2011) and Mycock & Lowe (2013).

the word. The fourth part of the relation specifies certain m-features which are associated with this morphological form of the lexeme.

(28) R <FIERCE1, ugraḥ, /ugrah/, {M-CAT:ADJ, M-CASE:NOM, M-NUM:SG, M-GEND:MASC}>

The set of m-features specified in (28) can be understood as the set of m-features associated with the adjectival suffix in a given case/number/gender combination. Abstracting over sets of m-features using templates (Dalrymple et al., 2004b), we can rewrite (28) as in (29), where the template is defined as in (30)

(29) R <FIERCE1, ugraḥ, /ugrah/, {@M-ADJ(NOM,SG,MASC)}>

(30) M-ADJ(_CASE,_NUM,_GEND) ≡ M-CAT:ADJ, M-CASE:_CASE, M-NUM:_NUM, M-GEND:_GEND

The use of a template simply permits us to generalize over the morphological contribution of adjectival formants. Thus M-ADJ(_CASE,_NUM,_GEND) represents morphological 'adjective-hood' (including the specification M-CAT:ADJ), and is instantiated to different case/number/gender combinations in particular instances by means of the argument variables.

The description function D maps a set of m-features to the relevant c-structure category and f-descriptions, given a particular LI. That is, D converts m-features introduced by the m-entry of a particular word form into syntactic features. So, in (31), the lexeme labelled FIERCE1 and the m-features M-CAT:ADJ, M-CASE:NOM, M-NUM:SG, M-GEND:MASC are associated with the syntactic category Adj and the f-structure features $(\uparrow$ CASE$) =$ NOM, $(\uparrow$ NUM$) =$ SG, and $(\uparrow$ GEND$) =$ MASC.

(31) Description function D for *ugraḥ*:
D <FIERCE1, {M-CAT:ADJ, M-CASE:NOM, M-NUM:SG, M-GEND:MASC}, Adj, {$(\uparrow$ CASE$) =$ NOM, $(\uparrow$ NUM$) =$ SG, $(\uparrow$ GEND$) =$ MASC}>

For the present purposes, the description function D is of central importance, as it relates morphological features to syntactic features, including syntactic category. D is a composite function, elements of which can be specified separately; in particular Dalrymple (2015) defines the description function D_{cat} which specifies the c-structure category alone, based on the LI and m-features.

Turning now to participles, the lexemic entry for the Sanskrit verbal root *gam* 'go' will be as in (32); the m-entry for the nom. pl. masc. present participle *gácchantaḥ* 'going' will be as in (33).

(32) LE <{ROOT: gam; STEM1: gaccha}, {$(\uparrow$ PRED$)$='go'}, GO1>

(33) R <GO1, gacchantaḥ, /gátʃantah/, {M-TENSE:PRES, M-VOICE:ACT, @M-ADJ(NOM,PL,MASC)}>

For our purposes the important point is that both lexical adjectives and participles, which are forms of verbal lexemes, make use of the template M-ADJ and thus share the same adjectival m-features. Thus morphological 'adjective-hood' has a unified and coherent morphological contribution in the lexicon, addressing Spencer's (2015) concern.

In order to produce a lexical entry for both *ugraḥ* and *gacchantaḥ*, we require description functions, which map the set of m-features to the appropriate c-structure category and f-description. The present concern is the c-structure category. Dalrymple (2015) proposes the following basic structure for D_{cat}:

(34) D_{cat} <LI, m-features, N> iff M-CAT:N $\in$ m-features.
D_{cat} <LI, m-features, V> iff M-CAT:V $\in$ m-features.
D_{cat} <LI, m-features, Adj> iff M-CAT:ADJ $\in$ m-features. etc.

These functions relate lexical indices and sets of m-features with a particular syntactic category, and the function is constrained to apply if and only if a particular m-feature appears in the set of m-features associated with the word form (i.e. specified in the word form's m-entry). So, in the case of *ugraḥ*, we can assume that the third line of (34) will apply: M-CAT:ADJ appears in the m-features for *ugraḥ*, so it will map to the syntactic category Adj.

These examples represent the expected mappings, and are likely to be the defaults cross-linguistically. However, the ability to specify these mapping by the lexemic index LI means that in principle the mappings between M:CAT features and grammatical category need not be uniform, even within a single language. So we can assume that the following applies for all lexical adjectives in Sanskrit, including *ugra-* (FIERCE1):

(35) D_{cat} <LI$_{adj}$, m-features, Adj> iff M-CAT:ADJ $\in$ m-features.

where LI$_{adj}$ is the set of lexemic indices associated with lexemic entries which are fundamentally adjectival, i.e. {HAPPY1, SAD1, TALL1, FIERCE1...}.

However, it is equally possible to assume that in the case of morphological adjectives based on verbal lexemes, there is a different specification:

(36) D_{cat} <LI$_{vb}$, m-features, V[ptc]> iff M-CAT:ADJ $\in$ m-features.

where LI$_{vb}$ is the set of LIs associated with a verbal lexeme, i.e. {GO1, FIND1, LOVE1...}. Crucially, the specification in (36) associates the m-feature M-CAT:ADJ with the syntactic category V, for a certain set of lexemes in the language. Thus, although the morphological adjective-hood of a participle supplies the same m-features as for any lexemic adjective, the one specifying grammatical category is treated differently due to the verbal nature of the base lexeme, resulting in a mapping from M-CAT:ADJ to the grammatical category V in the case of participles.

Thus the model for representing the morphology-syntax interface proposed by Dalrymple (2015), which presupposes a realizational theory of morphology just as Spencer (2015) does, permits both a unified, coherent morphological representation of adjective-hood, which can include not only lexical adjectives but also participles, and permits participles, as instantiations of verbal lexemes, to map to the category V, while lexical adjectives map to the category Adj.

Thus both conceptually and formally, there is no support for using morphosyntactic features such as agreement properties as evidence for categorial status. This means that of the three criteria used in discussions of mixed categories in LFG, two do not necessarily reveal anything about the syntactic category of a word. We are left with internal syntax as the only remaining criterion for distinguishing syntactic categories. On some level, this seems intuitively reasonable. Granted that English DPs and CPs are not necessarily distinguished by distribution, as discussed above, and given that English is relatively lacking in morphology, why do we have no problem in identifying a particular phrase as a DP rather than a CP, or vice versa? The obvious difference between them is their internal syntax: one has one type of internal syntax, including the possibility of adjectival modification and determiners, while the other has a different type of internal syntax, including complementizers, subject positions, and verb phrases.

Whether an approach to syntactic categories based purely on internal syntax is viable is a larger topic that cannot be fully addressed here. But the evidence of mixed categories, at least, seems to bear it out. I conclude therefore that mixed categories are necessary only when the internal syntax of a phrase is mixed, as in the case of the English type C gerund, but that a mixed category analysis is not necessary when the internal structure of a phrase is uniformly of one category, where we are dealing merely with a mismatch between internal syntax and distribution and/or morphosyntax.

6 Conclusion

I have shown that recent work on mixed categories in LFG depends on three main criteria for category membership: internal syntax, distribution, and morphosyntax. I have also shown that two distinct phenomena have received mixed category analyses within LFG. Truly mixed phrases are those where the internal syntax of the phrase is itself mixed. A number of mixed category proposals have been made, however, for phenomena where there is no evidence for mixed internal syntax, but where there is merely a mismatch between the internal syntax and the distribution and/or morphosyntax of the phrase. I have shown that distribution and morphosyntax are not reliable as evidence for categoriality, and thus argue that such 'mismatched' categories (e.g. participles) should not be analysed as mixed categories.

References

Alsharif, Ahmad. 2014. *The syntax of negation in Arabic*: University of Essex dissertation.

Alsina, Alex, K. P. Mohanan & Tara Mohanan. 2005. How to get rid of the COMP. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG05 Conference*, Stanford, CA: CSLI Publications.

Börjars, Kersti, Safiah Madkhali & John Payne. 2015. Masdars and mixed category constructions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG15 Conference*, 26–42. Stanford, CA: CSLI Publications.

Bresnan, Joan. 1997. Mixed categories as head sharing constructions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG97 Conference*, Stanford, CA: CSLI Publications.

Bresnan, Joan. 2001. *Lexical-functional syntax*. Oxford: Blackwell.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-functional syntax*. Oxford: Wiley-Blackwell. Second edition. First edition by Joan Bresnan, 2001, Blackwell.

Bresnan, Joan & John Mugane. 2006. Agentive nominalizations in Gĩkũyũ and the theory of mixed categories. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*, 201–234. Stanford, CA: CSLI Publications.

Butt, Miriam, Tracy Holloway King, María-Eugenia Niño & Frédérique Segond. 1999. *A grammar writer's cookbook*. Stanford, CA: CSLI Publications.

Dalrymple, Mary. 2015. Morphology in the LFG architecture. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG15 Conference*, 43–62. Stanford, CA: CSLI Publications.

Dalrymple, Mary, Helge Dyvik & Tracy Holloway King. 2004a. Copular complements: Closed or open? In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG04 Conference*, 188–198. Stanford, CA: CSLI Publications.

Dalrymple, Mary, Ronald M. Kaplan & Tracy Holloway King. 2004b. Linguistic Generalizations over Descriptions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG04 Conference*, 199–208. Stanford, CA: CSLI Publications.

Dalrymple, Mary & Louise Mycock. 2011. The prosody-semantics interface. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 Conference*, 173–193. Stanford, CA: CSLI Publications.

Haug, Dag T. T. & Tanya Nikitina. 2012. The many cases of non-finite subjects: The challenge of "dominant" participles. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference*, 292–311. Stanford, CA: CSLI Publications.

Haug, Dag T. T. & Tanya Nikitina. 2016. Feature sharing in agreement. *Natural Language & Linguistic Theory* 34. 865–910.

Hudson, Richard A. 2003. Gerunds without phrase structure. *Natural Language & Linguistic Theory* 21. 579–615.

Lapointe, Steven G. 1993. Dual lexical categories and the syntax of mixed category phrases. In A. Kathol & M. Bernstein (eds.), *Proceedings of the Eastern States Conference on Linguistics*, 199–210. Ithaca, NY: Cornell University.

Lowe, John J. 2015. *Participles in Rigvedic Sanskrit: The syntax and semantics of adjectival verb forms*. Oxford: Oxford University Press.

Malouf, Robert. 2000a. *Mixed categories in the hierarchical lexicon*. Stanford, CA: CSLI Publications.

Malouf, Robert. 2000b. Verbal gerunds as mixed categories in HPSG. In R. Borsley (ed.), *The nature and function of syntactic categories*, 133–166. New York: Academic Press.

Mycock, Louise & John J. Lowe. 2013. The prosodic marking of discourse functions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG13 conference*, 440–460. Stanford, CA: CSLI Publications.

Nikitina, Tatiana V. 2008. *The mixing of syntactic properties and language change*: Stanford University dissertation.

Nikitina, Tatiana V. & Dag T. T. Haug. 2016. Syntactic nominalization in Latin: a case of non-canonical subject agreement. *Transactions of the Philological Society* 114(1). 25–50.

Nordlinger, Rachel & Louisa Sadler. 2004a. Tense beyond the verb: Encoding clausal tense/aspect/mood on nominal dependents. *Natural Language & Linguistic Theory* 22(3). 597–641.

Nordlinger, Rachel & Louisa Sadler. 2004b. Nominal tense in crosslinguistic perspective. *Language* 80(4). 776–806.

Seiss, Melanie. 2008. The English *-ing* form. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG08 Conference*, 454–472. Stanford, CA: CSLI Publications.

Spencer, Andrew. 2004. Towards a typology of mixed categories. In C. Orhan Orgun & Peter Sells (eds.), *Morphology and the web of grammar: Essays in memory of Steven G. Lapointe*, Stanford, CA: CSLI Publications.

Spencer, Andrew. 2013. *Lexical relatedness: A paradigm-based approach*. Oxford: Oxford University Press.

Spencer, Andrew. 2015. Participial relatives in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG15 Conference*, 378–398. Stanford, CA: CSLI Publications.

Wedekind, Jürgen. 2014. On the universal generation problem for unification grammars. *Computational Linguistics* 40(3). 533–538.

Wedekind, Jürgen & Ronald M. Kaplan. 2012. LFG generation by grammar specialization. *Computational Linguistics* 38(4). 867–915.

Seminumerals, determiners and nouns in English

Takafumi Maekawa
Ryukoku University

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 422–441

Keywords: seminumerals, singular countable nouns, determiners, agreement, noun phrase syntax

Maekawa, Takafumi. 2016. Seminumerals, determiners and nouns in English. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 422–441. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.22.

Abstract

A singular countable noun in English normally requires a determiner and they should agree in number. However, there is a type of noun phrase, such as *those thousand teachers*, which does not conform to this generalisation. As a subtype of singular countable noun *thousand* requires a determiner, but the determiner has number agreement with the head noun *teachers*. The standard HPSG treatment, in which the determiner requirement and the determiner-noun agreement are both represented in the SPR speci cations of the head noun, cannot capture this special agreement pattern. Our analysis, in which the determiner requirement and the determiner-noun agreement are dissociated from each other, can provide a straightforward account of the data.

1 Introduction

This paper addresses the issue of the syntactic relation between the determiner and the noun in a noun phrase. Especially we will focus on the syntax of singular countable nouns in English. The discussion will raise some fundamental questions about noun phrase syntax.

A singular countable noun in English is di erent from a plural countable noun and an uncountable noun in that it normally requires a determiner.[1]

(1) a. I haven't got *(a) *pen*.
   b. There were *cats* in every room.
   c. Her coat is made of pure *wool*. (Swan 2005:65-66)

The noun *pen* in (1a), which is a singular countable noun, requires a determiner to combine with, and the determiner *a* satis es this requirement. A plural countable noun (1b) and an uncountable noun (1c), on the other hand, can stand on their own without a determiner.

When a determiner combines with a countable noun, they should agree in number.

(2) a. this book/*this books
   b. these books/*these book (Huddleston & Pullum 2002:352)

[†]I would like to thank the participants at HEADLEX 2016, especially Mary Dalrymple, Frank Van Eynde, John Payne and Nigel Vincent, for their feedback and discussions. I am grateful to anonymous reviewers for their constructive and valuable comments. Thanks are also due to Bob Borsley for his valuable comments on the earlier version of this paper. Any shortcomings are my responsibility. This research was supported by the Japan Society for the Promotion of Science (Grant-in-Aid for Scienti c Research (C) 26370466).

[1]We follow Huddleston & Pullum (2002:355) in assuming that the term 'determiner' refers to the following expressions: determinatives (*the tie*), determiner phrases (*almost every tie*), genitive NPs (*my tie*), plain NPs (*what colour tie*), PPs (*over thirty ties*).

In the ungrammatical examples in (2) the noun and the determiner do not agree in number.

The above observations lead to something like the following generalisation:

(3) A singular countable noun in English requires a determiner and they should agree in number.

It is often assumed in HPSG that a determiner is a speci er of a head noun (Pollard & Sag 1994, Sag et al. 2003, Kim 2004, Kim & Sells 2008). In this assumption a singular count noun in English can be described as in (4).

(4)
$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \boxed{1}\begin{bmatrix}\text{NUM} & \textit{sg}\end{bmatrix}\end{bmatrix} \\ \text{SPR} & \left\langle \begin{bmatrix}\text{AGR} & \boxed{1}\end{bmatrix} \right\rangle \end{bmatrix}$$

(cf. Sag et al. (2003:107), Kim (2004:1114), Kim & Sells (2008:108))

(4) is a partial lexical description of a singular countable noun, stating that the latter has a speci er which agrees with it in number. The feature HEAD encodes such information as is propagated from a head to a phrase, including information about parts of speech. In (4) the value of this feature identi es this expression as a noun. The value of the HEAD feature includes the AGR (AGREEMENT) feature, whose value represents information about morpho-syntactic properties of the expression. The NUM (NUMBER) value represents the information about the grammatical number, and the *sg* value indicates that the word is morpho-syntactically singular. The SPR (SPECIFIER) feature shows that this expression has a speci er and indicates what kind of speci er it is.

Thus, the determiner requirement of a countable singular noun is encoded as a matter of valency. The boxed tag $\boxed{1}$ means that the speci er has the same AGR value as the head noun, representing determiner-noun agreement. (4) captures generalisation (3) and accounts for the unacceptability of (1a) *(*a*) *pen* and (2) **this books*/**these book*: the former lacks a speci er and the latter NPs do not show determiner-noun agreement.

It is possible to say that in the standard HPSG treatment of a singular countable noun in (4), the determiner requirement and the determiner-noun agreement are both represented in the SPR speci cations of the head noun.

In this paper we will rst argue that the numeral quanti ers in (5), which we argue to be subtypes of count nouns, cannot be captured by the standard treatment in (4).

(5) a. those *thousand* teachers

(BYU-BNC[2]: GUR W_ac_polit_law_edu)

[2] Davies (2004–)

b. these *hundred* women (BYU-BNC: FL5 S_brdcast_discussn)

Then we will argue that a satisfactory account of the data is possible within the framework in which prenominal elements such as determiners and numerals are non-heads selecting a head. In this approach the determiner requirement and the determiner-noun agreement are dissociated from each other (Van Eynde (2006), Allegranza (1998)).

Following Jackendoff (1977:126), we will assume that there are two varieties of numeral quantifiers, which will be called 'cardinals' and 'seminumerals', respectively. Cardinals are numerals such as *two* and *ten*, which do not require a determiner before them.

(6) (the/these/those) two (cats)

Seminumerals are words such as *hundred*, *thousand*, *million*, *billion*, *trillion* and *dozen*, which should be preceded by a determiner.

(7) *(the/these/those/a) hundred (cats)

In this paper we will mainly focus on seminumerals, but cardinals are also mentioned in comparison with seminumerals.

The organisation of this paper is as follows. In section 2 we will introduce the data which are problematic for the standard HPSG treatment of noun phrase syntax. In section 3 we will look at some possible analyses, and we will argue that they include important weaknesses. Section 4 and 5 present our claim that the seminumerals are functors, and we will see how it is able to account for the facts. In section 6 we will also look at some data to which we can provide much the same explanation as seminumerals and other data which are no problem to our analysis. In section 7 we will briefly discuss some remaining issues. Section 8 is the conclusion.

2 Problems

Before pointing out the problems that the numerically quantified NPs in (5) pose, we will show that seminumerals are subtypes of countable nouns. First, like a singular countable noun as in (1a), the singular form of seminumerals needs a determiner in order to be grammatical.

(8) a. Today there are a dozen. (COCA[3]: 1992 MAG Ebony)
b. *Today there are dozen.

Second, like a plural countable noun as in (1b), the plural form of seminumerals does not require a determiner.

(9) There are dozens. (COCA: 1993 FIC Mov:Arcade)

[3]Corpus of Contemporary American English (Davies 2008–)

These observations lead to a conclusion that seminumerals are subtypes of countable nouns. We can thus expect that seminumerals in singular have much the same lexical properties as (4).[4]

Given that seminumerals are subtypes of countable nouns with the basic structure in (4), they might be analysed to have the following properties.

(10)
$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \boxed{1}\begin{bmatrix}\text{NUM} & \textit{sg}\end{bmatrix} \\ \text{MOD} & \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\textit{noun} & \\ \text{AGR}\,|\,\text{NUM} & \textit{pl}\end{bmatrix}\end{bmatrix}\end{bmatrix} \\ \text{SPR} & \left\langle \begin{bmatrix}\text{AGR} & \boxed{1}\end{bmatrix} \right\rangle \end{bmatrix}$$

Here it is assumed that a seminumeral has a specifier which agrees with it and that it modifies a plural noun via the MOD feature. In *thousand teachers*, for example, *a thousand* is treated as an adjunct of *teachers*.

The NPs in (5), repeated below, pose challenges for (10).

(11) a. those *thousand* teachers
b. these *hundred* women [= (5)]

In the NPs in (11) the only possible determiner that can satisfy the determiner requirement of the seminumeral is the one just before it: *those* in (11a) and *these* in (11b).

(12) a. *(those) thousand teachers
b. *(these) hundred women

Since a seminumeral is a type of countable common noun, the singular form requires a determiner. (12a) and (12b) show that the determiner is obligatory. Since *teachers* and *women* are plural nouns, they do not require their own determiner. We can conclude that the determiner is required by the seminumeral (See also Hudson (2004:36)).

[4] However, seminumerals behave differently from typical countable nouns in the following respects. First, they combine with a noun without any intermediate element such as a preposition.

(i) a dozen (*of) cats

(ii) a pair *(of) cats

Second, seminumerals should be singular even when they combine with a numeral denoting a number larger than one.

(iii) two hundred / *two hundreds

(iv) two cats / *two cat

However, note that there is no number agreement between the determiner and the seminumeral: the seminumeral is singular but the determiner is plural. Instead, the determiner agrees with the plural noun after the seminumeral.

(13) a. *those thousand teacher

b. *these hundred woman

It is clear, then, that the NPs in (5) do not have the properties in (10).[5]

3 Possible Analyses

One possible analysis would be to propose that the specifier agrees with the plural noun that the seminumeral modifies, giving (14) in place of (10).

(14)
$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \begin{bmatrix}\text{NUM} & \textit{sg}\end{bmatrix} \\ \text{MOD} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix}\textit{noun} & \\ \text{AGR} & \boxed{1}\begin{bmatrix}\text{NUM} & \textit{pl}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix} \\ \text{SPR} & \left\langle \begin{bmatrix}\text{AGR} & \boxed{1}\end{bmatrix} \right\rangle \end{bmatrix}$$

However, examples like those in (15), where the determiners are singular, pose a problem.

(15) a. that thousand pounds (BYU-BNC: KCX S_conv)

b. this hundred houses (BYU-BNC: J8G S_interview_oral_history)

The examples in (15) have singular determiners but have plural heads. Thus, there is no number agreement between the specifier (*that*/*this*) and the head noun (*pounds*/*houses*). Accordingly, we will not pursue this analysis.

Another possibility would be to propose that seminumerals are 'weak heads' (Tseng 2002, Abeillé et al. 2006, Przepiórkowski 2013) or 'transparent heads' (Flickinger 2008). With these mechanisms, it is possible to preserve

[5]Jackendoff (1977:133) assumes that the underlying structure of *those dozen weeks* is *those a dozen of weeks*, where the plural determiner is in the specifier position of *weeks* and the seminumeral has its own determiner *a*. (This underlying form undergoes a couple of transformations (i.e. Pseudopartitive *a*-Deletion and Numeral *of*-Deletion) to obtain the surface form.) Under these assumptions it is possible to avoid the problems discussed above: the plural determiner agrees with the head noun because they are in the determiner-head relationship; the determiner requirement of the seminumeral is satisfied by its own underlying determiner *a*. However, this analysis cannot accommodate the fact observed in (12a). (12a) shows that the determiner is obligatory although *teachers* is a plural countable noun, which normally does not require its own determiner. This casts doubt on the assumption that *those* is the specifier of *weeks*.

some important properties of their complement on the phrase. This propagation of information from non-heads to phrases can account for the data in (5) if we assume that seminumerals preserve the grammatical number of the complement on the phrase.

However, examples like those in (15) pose a problem for analyses along these lines too. In (15) the seminumeral has a singular determiner. The seminumeral inherits the plural number from its complement and passes it to the phrase. The plurality of the phrase does not match the singular determiner. We conclude, then, that the weak/transparent head approach is unsatisfactory.

4 Prenominal Elements as Functors

We will turn to an analysis which we think provides a satisfactory account of the data. In this analysis determiners and seminumerals are functors: non-heads which select the head.

In recent HPSG, a notion of functor has been introduced as an alternative to the dichotomy between modi ers and determiners (Van Eynde 2006, Allegranza 1998). In languages such as Italian, the determiners and the adjectives have the same morphological variation and the same patterns of agreement.

(16) questa bella bambina [Italian]
this-SG.FM beautiful-SG.FM child-SG.FM
'this beautiful child' (Van Eynde 2007:419)

The singular feminine determiner *questa* 'this' selects a singular feminine nominal as does the singular feminine adjective *bella* 'beautiful'. The dichotomy between modi ers and determiners does not capture such shared properties. For this reason, prenominal elements, such as adjectives and determiners, are uniformly treated as functors.

We will adopt the functor analysis of determiners and assume that singular determiners such as *this* have the following syntactic properties (Van Eynde 2006).

(17) *this*:

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{determiner} & \\ \text{AGR} & \boxed{1}\begin{bmatrix}\text{NUM} & \textit{sg}\end{bmatrix} \\ \text{SEL} & \left\langle \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} \\ \text{AGR} & \boxed{1} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{MRK} & \textit{marked} \end{bmatrix}$$

The information about selection is indicated by the SEL (SELECT) feature of a non-head, which represents the constraints which a non-head daughter imposes on the head daughter. The SEL value of (17) shows that this word selects a singular noun.

MARKING (MRK) indicates whether the expression involves a determiner or it can stand alone without these elements. The MRK value is *marked* if the expression contains a determiner or it itself is a determiner, and *unmarked* otherwise. Plural nouns and abstract nouns are [MRK *bare*] because they can stand alone without a determiner. Singular countable nouns such as *pen* and *week* have an *incomplete* value because they require a determiner

The AGR value $\boxed{1}$ [NUM *sg*] shared between *this* and its head noun means determiner-noun agreement between them. The partial description of *these* is the same as (17) except for the AGR value: as a plural determiner, the value is [NUM *pl*].

Singular countable nouns such as *pen* have the following syntactic properties.

(18) *pen*:

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR}\,|\,\text{NUM} & \textit{sg} \end{bmatrix} \\ \text{MRK} & \textit{incomplete} \end{bmatrix}$$

The NUM value of a singular noun is *sg*. The *incomplete* value of the MRK feature indicates that a singular countable noun requires a determiner to combine with.

A determiner and a nominal combine to form a head-functor phrase, which is subject to the following constraint on *head-functor-phrase* (*hd-funct-ph*) (Van Eynde 2006, Allegranza 1998).

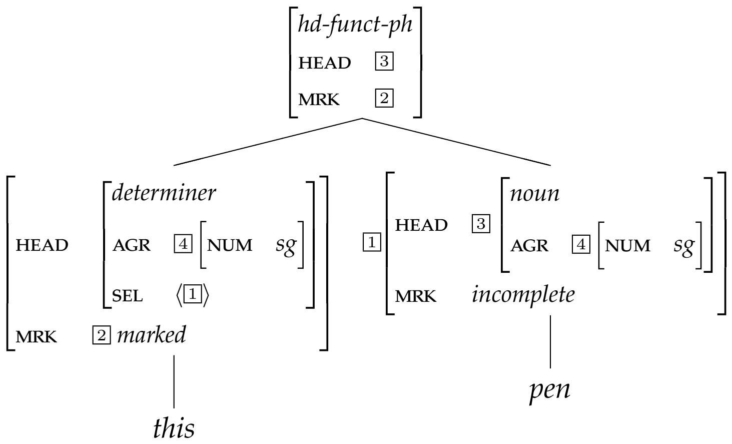
(19) Constraint for head-functor phrase (Van Eynde 2006:164,166)

$$\textit{hd-funct-ph} \rightarrow \begin{bmatrix} \text{MRK} & \boxed{1} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{MRK} & \boxed{1} \\ \text{SEL} & \langle\boxed{2}\rangle \end{bmatrix}, \boxed{3}\begin{bmatrix} \text{SYNSEM} & \boxed{2} \end{bmatrix} \right\rangle \\ \text{H-DTR} & \boxed{3} \end{bmatrix}$$

Constraint (19) states that in a head-functor phrase the non-head daughter selects a head daughter, and the MRK value of the mother is identical to that of the non-head daughter.

Let us consider how generalisation (3) is captured in this approach. (20) shows how functor *this* combines with a singular countable noun.

(20)



The combination of *this* and *pen* is an instance of a head-functor phrase, in which *this* selects the head noun and the MRK value *marked* is inherited to the mother node. The AGR | N value of *pen* is *sg*, indicating that it is a singular nominal. The AGR value $\boxed{4}$ shared between *this* and its head noun means determiner-noun agreement between them. The MRK feature of *pen* has a value whose type is *incomplete*, which means that the word is incomplete on its own, requiring some sort of determiner. The HEAD value is propagated from the head daughter to the mother node. This propagation is due to the constraint on phrases of type *headed-phrase* (*hd-ph*), which is a supertype of *hd-funct-ph*.

In this approach generalisation (3) is captured in terms of two separate specifications. First, the determiner requirement of a singular countable noun is represented by the *incomplete* value of the MRK feature of the head nominal. Second, the determiner-noun agreement is represented by the shared value of the AGR | N feature between the determiner and the head noun.

The difference between the head-functor analysis in (20) and the standard HPSG treatment in (4) can be summarised as follows. In standard HPSG the determiner requirement and the determiner-noun agreement are both represented in the SPR specifications of the head noun. In the head-functor analysis, on the other hand, the determiner-noun agreement and the determiner requirement are dissociated from each other.

5 Seminumerals as Functors

Based on the earlier observations that seminumerals are subtypes of countable nouns (section 2), we propose that the singular form of a seminumeral has the following syntactic properties.

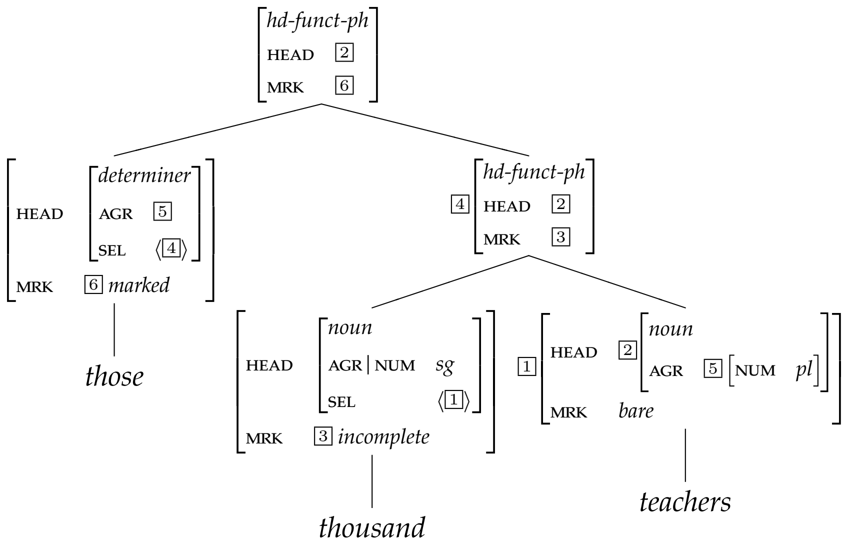
(21) *hundred*:

$$\left[\begin{array}{ll} \text{HEAD} & \left[\begin{array}{ll} \textit{noun} & \\ \text{AGR}\,|\,\text{NUM} & \textit{sg} \\ \text{SEL} & \left\langle \left[\begin{array}{ll}\text{HEAD} & \left[\begin{array}{ll}\textit{noun} & \\ \text{AGR}\,|\,\text{NUM} & \textit{pl}\end{array}\right]\end{array}\right]\right\rangle \end{array}\right] \\ \text{MRK} & \textit{incomplete}\end{array}\right]$$

This lexical entry of a seminumeral is the same as that of a singular countable noun given in (18), except that it has a SEL speci cation. The SEL value in (21) indicates that the seminumeral selects a plural noun. The MRK feature of seminumerals has a value whose type is *incomplete*, which means that the word is incomplete on its own, requiring some sort of determiner.

Our syntactic analysis of *those thousand teachers* is given in (22).

(22)



The combination of *thousand* and *teachers* is an instance of a head-functor phrase. In (22) *thousand* as a functor daughter selects the head daughter *teachers*, and the MRK value of the functor daughter is propagated to the mother node. The MRK value is of type *incomplete*, which means that the expression is incomplete on its own, requiring some determiner. The *pl* value of AGR | N, which is propagated from *teachers* via the HEAD feature, enables this phrase to combine with the plural determiner *those*.

It should be noted that the determiner requirement of the seminumeral is fully satis ed by the plural determiner. Agreement mismatch does not occur, however, because the determiner and the seminumeral are not in the determiner-head relationship. The determiner agrees with the plural noun *teachers* via the AGR | N feature.

Our approach can account for the agreement pattern with the verb when the construction in question is a subject.

(23) a. (...), and when he's drunk **those hundred things** *become* a thousand. (COCA: 2011 FIC Bk:AftertasteNovel)

b. (...) **these thousand pages** *have* been gathered, (...) (COCA: 2011 FIC Bk:GreatCircleMayfield)

c. **These dozen men** *have* been close enough to hear them. (COCA: 1995 MAG SportingNews)

As (22) shows, the head of the whole construction is the plural noun after the seminumeral. The plurality of the head noun accounts for the plural agreement with verb, illustrated by the examples in (23).

We can argue that examples like (24) also have structures like (22).

(24) a. all thousand stones (BNC: CAM W_fict_prose)

b. all hundred modifications (COCA: 2003 SPOK NPR_ATCW)

All requires a plural noun and in the examples in (24) the head nouns *stones* and *modifications* satisfy this requirement, respectively.

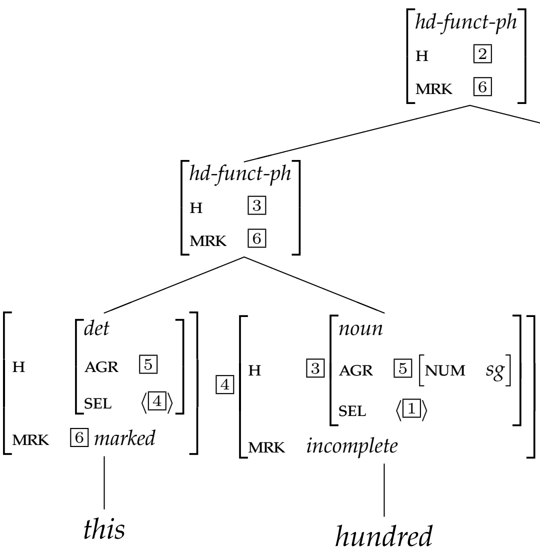
Our functor analysis of seminumerals can give an account of the data in (15), in which the seminumeral takes a singular determiner.

(25) a. that thousand pounds

b. this hundred houses [= (15)]

(26) is a structure we propose for (25b) *this hundred houses*.

(26)



In this construction the seminumeral first combines with the determiner as a head-functor phrase. The determiner should be singular because its head, *hundred*, is [AGR | N *sg*]. The SEL value of *hundred* is inherited to the mother node via the HEAD feature. The phrase *this hundred* combines with the head noun *houses* to form another head-functor phrase.

The subject-verb agreement shown in (27) can be accounted for in terms of the number of the head noun.[6]

(27) Let's see what **this hundred women** *make* of the question, do men hate women? (BYU-BNC: FL7 S_brdcast_discussn)

The head is the plural noun *women* in (27). This accounts for the plural agreement with the verb.

The following NPs have much the same structure as (26).

(28) a. (...) **a dozen men** *move* back and forth, (...) (BYU-BNC: G0F W_ ct_prose)

b. In Bombay, for instance, **every thousand people** *have* only 0.1 hectares of open space – and this includes tra c islands. (BYU-BNC: B7E W_non_ac_nat_science)

Determiners *a* and *every* only combine with a singular nominal. In (28) they combine with the seminumeral which is a subtype of a singular common noun. The resulting phrase combines with the head noun.

We have illustrated that our analysis can accommodate the data which are problematic to the other analyses we discussed in section 3.

6 Further Data

We have argued that a seminumeral is a functor selecting a head and accounted for the agreement between the determiner and the head noun. In this section we will present some pieces of data which are closely related to seminumerals.

6.1 *Sort*-nouns

The syntactic behaviour of seminumerals which we have seen so far is very similar to the type of NP shown in (29).

(29) a. these *sort* of skills

b. those *kind* of pitch changes

c. these *type* of races (Keizer 2007:170)

[6] If the head noun of the subject is a measure noun, the plural subject combines with a singular verb.

(i) **Five pounds** *is*/**are* a lot of money. (Hudson 1999:174)

The singular verb in the following example can be accounted for along the same lines.

(ii) ..., but **that thousand pounds** *is* not a sum that the rm can a ord to lose. (BYU-BNC: EV1 W_ ct_prose)

We call the nouns *sort*, *kind* and *type* collectively as *sort*-nouns. The *sort*-nouns are countable nouns, and those in (29) are singular. Note that they are preceded by a plural determiner. As in the case of seminumerals, the determiner requirement of the singular *sort*-noun is satis ed by the plural determiner. The determiner agrees with the noun following *of*.

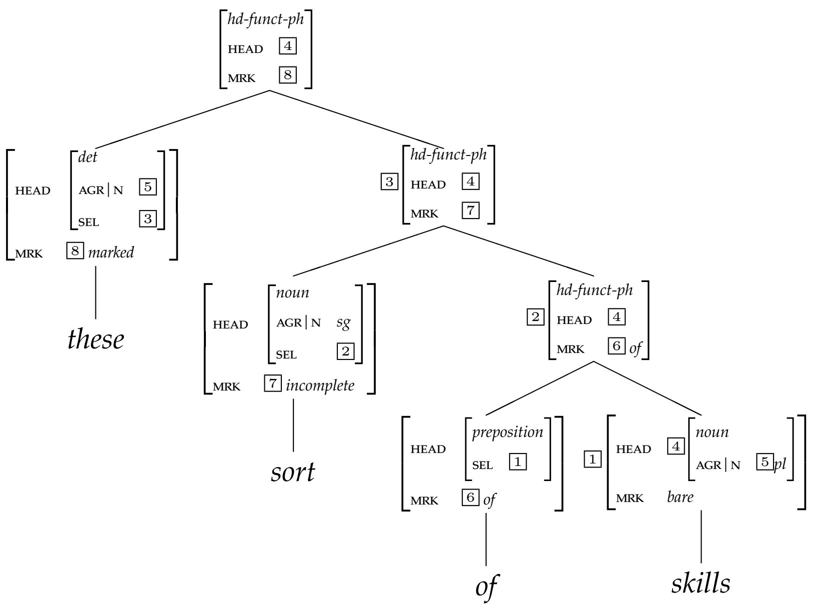
(30) a. *these sort of skill
b. *those kind of pitch change
c. *these type of race

Another similarity is that this construction causes plural agreement with the verb.

(31) Well I'd actually expect that **those sort of courses** *are/*is* very uh heavily subscribed uh, heavy just like **these sort of problems** *are/*is* very hard to solve. (Keizer (2007:175); adapted from ICE-GB)

These similarities lead us to expect that the constructions in (29) have much the same structure as (22). Maekawa (2015) argues that the NPs in (29) have structures like the following .

(32)



Assuming that preposition *of* is a functor (Van Eynde 2005), the combination of the preposition and *skills* is a head-functor phrase, in which the former selects the latter. Preposition *of* has the MRK feature whose value is *of*. This value is inherited from *of* to *of skills*. The *sort*-noun in this construction is a functor with much the same properties as (21). As a functor, it selects the *of*-marked phrase via the SEL value $\boxed{2}$. In this head-functor phrase the *sort*-noun is a non-head daughter, and the head-daughter is *of skills*. The HEAD value of the resulting phrase comes from the head daughter. The *pl* value of AGR | N, which is propagated from *skills* via the HEAD feature, enables this phrase to combine with the plural determiner *these*. The combination of the determiner with the head nominal is also an instance of a head-functor phrase, and the MRK value *marked* is inherited from *these* to *these sort of skills*. The head of the whole construction is the plural noun after *of*, and this explains plural agreement with the verb.

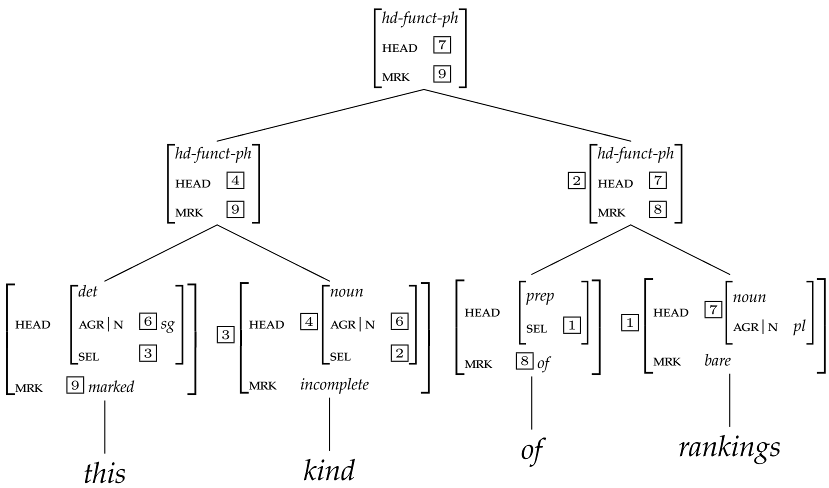
Moreover, the determiner can also be singular but have plural agreement with the verb.

(33) a. **This kind of rankings** *have* given ammunition to conservatives (...) (COCA: 2001 NEWS CSMonitor)

b. (...) **this type of women** *like* to be around rich and powerful men. (COCA: 2008 SPOK Fox_Gibson)

This fact can be captured along the same lines as (25), where the seminumeral takes a singular determiner but have a plural verb. The subject noun phrases in (33) have structures like the following.

(34)



In (34) the *sort*-noun *kind* rst combines with the determiner to form a head-functor phrase. They have singular agreement because the head (*sort*) is singular. The SEL value of *sort* is inherited to the mother node because it is a HEAD feature. The phrase *this sort* combines with the *of*-marked phrase to form another head-functor phrase. Like (32), the head of the whole phrase is the head-daughter of the *of* phrase. This accounts for plural agreement with the verb.

For more details of a head-functor analysis of these constructions, see Maekawa (2015).

6.2 Cardinals

In this subsection we will turn to cardinals, which are another type of numeral exempli ed by words like *three* and *ten*. We will argue that the difference between seminumerals and cardinals is small and most part of the analysis given to seminumerals can be applied to cardinals as well.

Seminumerals and cardinals di er only in that the former require a determiner but the latter need not have one.

(35) a. *(the) hundred weeks
b. (the) three weeks

Cardinals can have much the same range of determiners as seminumerals.

(36) a. *those* three weeks
b. *that* three weeks
c. *every* three weeks
d. *all* three weeks (Jackendo 1977:132)

(37) a. *those* thousand teachers [= (5a)]
b. *that* thousand pounds [= (25a)]
c. *every* thousand people [= (28b)]
d. *all* thousand stones [= (24a)]

Based on the above observations we can propose something like the following partial lexical description of a cardinal.

(38) *three*:

$$\left[\begin{array}{ll} \text{HEAD} & \left[\begin{array}{ll} \textit{noun} & \\ \text{AGR}\,|\,\text{NUM} & \textit{sg} \\ \text{SEL} & \left\langle \left[\text{HEAD} \left[\begin{array}{ll}\textit{noun} & \\ \text{AGR}\,|\,\text{NUM} & \textit{pl}\end{array}\right]\right]\right\rangle \end{array}\right] \\ \text{MRK} & \textit{bare}\end{array}\right]$$

The SEL value in (38) indicates that the cardinal selects a plural noun. The MRK feature has a value whose type is *bare*, which means that a cardinal does not require a determiner to be used in NP positions. This description of a cardinal di ers from that of seminumerals only in the speci cation of the MRK value, which is *bare* for cardinals but *incomplete* for seminumerals. That captures the di erences shown in (35).

6.3 Plural Seminumerals

In this subsection we will have a look at plural seminumerals and compare them with seminumerals in singular.

It is possible to say that plural seminumerals have quite di erent syntactic properties from their singular counterparts. First, a plural seminumeral cannot select a plural noun, but instead they are followed by a prepositional phrase headed by *of*.[7]

(39) a. hundreds *(of) cats
b. a hundred (*of) cats

Second, singular and plural seminumerals di er in the possibility of extraction, according to Kayne (2005:160).

(40) a. (?) What (else) does he have hundreds of?
b. *What (else) does he have a hundred? (Kayne 2005:160)

The above data show that a plural seminumeral can allow extraction of the noun after *of*, but it is impossible to extract a noun after the singular seminumeral.

Based on these observations, we can argue that the partial lexical description of a plural seminumeral is something like the following.

(41) *hundreds*:

$$\left[\begin{array}{ll} \textsc{head} & \left[\begin{array}{ll} \textit{noun} & \\ \textsc{agr}\,|\,\textsc{num} & \textit{pl} \\ \textsc{comp} & \left\langle \left(\left[\textsc{head} \; \left[\begin{array}{ll}\textit{preposition} & \\ \textsc{form} & \textit{of}\end{array}\right]\right]\right)\right\rangle \end{array}\right] \\ \textsc{mrk} & \textit{bare}\end{array}\right]$$

The COMP value in (41) indicates that a plural seminumeral optionally takes a prepositional phrase as its complement. The optionality of the PP is illustrated by the following example.

(42) Hailes-Valentine's study required only 12 patients; others involve hundreds. (COCA: 1990 NEWS USAToday)

The PP complement is constrained to have *of* as its head ([FORM *of*]) (Sag et al. 2003:316), which captures the (un)grammaticality of (39a).

The claims that *of* in this construction is a complement-taking preposition and that the *of*-phrase is a complement of the seminumeral account for the extraction of the noun seen in (40a). We assume that the SLASH value of argument-taking words is determined in terms of the SLASH values of their arguments (Ginzburg & Sag 2001:168). This allows extraction of complements. Extraction is impossible in (40b) because the extracted element is not an argument of the seminumeral: it is the head which the seminumeral selects.

The MRK value of a plural seminumeral is *bare*, which means that the word can occur in a sentence with or without a determiner.

[7] We postulate two lexical entries for *of*: a complement-taking preposition as in (39a) and a functor preposition discussed in 6.1.

(43) (these) hundreds / thousands / dozens of patients

7 Remaining Issues

We have given an analysis of the syntactic relation of seminumerals, determiners and nouns in English, but there are some remaining issues.

One such issue is the spurious ambiguity that structures like (44) have.

(44) the hundred people

The determiner *the* is ambiguous and can be either singular or plural. The NP in (44) has two analyses, which corresponds to (22) and (26), respectively. In the former structure the seminumeral combines with the head noun and the resulting phrase then combines with *the*. In the latter structure the seminumeral combines with *the* rst. The determiner *the* causes such an ambiguity, but it is not clear whether this results in any di erence in interpretation.

Another issue is how this analysis can be applied to similar constructions in other languages. In Polish, for example, numerals show a complicated behaviour concerning case and number. In the subject NP in (45) the numeral is accusative and the noun is genitive. The determiner can be either accusative or genitive. The verb is third person, singular and neutral.

(45)

Te	/ Tych	tysiąc	osób	już
these-PL,ACC	/ these-PL,GEN	thousand-ACC	people-GEN	already

przyszło.	[Polish]
came-3RD,SG,NEUT	

'The thousand people already came' (Przepiórkowski 1999:195)

These issues should be left for future research.

8 Conclusion

Let us summarise the discussion. In English a singular countable noun normally requires a determiner and they should agree in number. It looks, however, as if a seminumeral in examples like *those thousand teachers* does not conform to this generalisation: it is a singular countable noun that requires a determiner, but the determiner satisfying this requirement does not agree with it. Instead, the determiner agrees with the noun following the seminumeral. We argued that the functor treatment of seminumerals can provide a satisfactory account of the data. We also suggested that examples such as *that thousand pounds*, in which the singular determiner agrees with the seminumeral, can be accommodated in our analysis.

In HPSG it has been normally assumed that the SPR value represents the constraints which the head imposes on the determiner, including both the determiner requirement of a singular countable noun and determiner-noun agreement. In the head-functor analysis of determiner-noun relation, however, the determiner requirement and determiner-noun agreement are represented separately: the former is represented as the *incomplete* value of the MRK feature of the head daughter, and the latter is encoded as part of constraints which the determiner imposes on the head daughter. These mechanisms interact to allow the plural determiner to satisfy the determiner requirement of a singular seminumeral and to agree with the plural head noun in examples like *those thousand teachers*.

References

Abeillé, Anne, Olivier Bonami, Danièle Godard & Jesse Tseng. 2006. The syntax of French *à* and *de*: An HPSG analysis. In Patrick Saint-Dizier (ed.), *Syntax and Semantics of Prepositions* (Text, Speech and Language Technology 29), 147–162. Springer Verlag.

Allegranza, Valerio. 1998. Determiners as functors: NP structure in Italian. In Sergio Balari & Luca Dini (eds.), *Romance in HPSG*, 55–108. Stanford: CSLI Publications.

Davies, Mark. 2004–. BYU-BNC (Based on the British National Corpus from Oxford University Press). `http://corpus.byu.edu/bnc/`.

Davies, Mark. 2008–. The Corpus of Contemporary American English: 450 Million Words, 1990-Present. `http://corpus.byu.edu/coca/`.

Flickinger, Dan. 2008. Transparent heads. In Stefan Müller (ed.), *The Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar*, 87–94. Stanford: CSLI Publications. `http://cslipublications.stanford.edu/HPSG/9/`.

Ginzburg, Jonathan & Ivan A. Sag. 2001. *Interrogative investigations: the form, meaning, and use of english interrogatives*. Stanford, California: CSLI Publications. `http://cslipublications.stanford.edu/site/1575862786.html`.

Huddleston, Rodney & Geo rey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Hudson, Richard. 1999. Subject-verb agreement in English. *English Language and Linguistics* 3(2). 173–207.

Hudson, Richard. 2004. Are determiners heads? *Functions of Language* 11(1). 7–24.

Jackendo , Ray. 1977. *X′ Syntax: A Study of Phrase Structure*. Cambridge, Mass.: MIT Press.

Kayne, Richard S. 2005. On some prepositions that look DP-internal. In *Movement and Silence*, chap. 7, 136–175. Oxford: Oxford University Press.

Keizer, Evelien. 2007. *The English Noun Phrase*. Cambridge: Cambridge University Press.

Kim, Jong-Bok. 2004. Hybrid agreement in English. *Linguistics* 42(6). 1105–1128.

Kim, Jong-Bok & Peter Sells. 2008. *English Syntax: An Introduction*. Stanford: CSLI Publications.

Maekawa, Takafumi. 2015. 'Agreement mismatch' between sort/kind/type and the determiner. In Stefan Müller (ed.), *Proceedings of the 22nd International Conference on Head-Driven Phrase Structure Grammar, Nanyang Technological University (NTU), Singapore*, 136–156. Stanford, CA: CSLI Publications. http://cslipublications.stanford.edu/HPSG/2015/maekawa.pdf.

Pollard, Carl J. & Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Przepiórkowski, Adam. 1999. *Case Assignment and the Complement-Adjunct Dichotomy: A Non-con gurational Constraint-based Approach*. Germany: Universität Tübingen dissertation.

Przepiórkowski, Adam. 2013. The syntax of distance distributivity in Polish: Preserving generalisations with weak heads. In Stefan Müller (ed.), *Proceedings of the 20th International Conference on Head-driven Phrase Structure Grammar, Freie Universität Berlin*, 161–181. http://cslipublications.stanford.edu/HPSG/2013/przepiorkowski.pdf.

Sag, Ivan A., Thomas Wasow & Emily M. Bender. 2003. *Syntactic Theory: A Formal Introduction: Second Edition*. Stanford: CSLI Publications.

Swan, Michael. 2005. *Practical English Usage: Third Edition*. Oxford: Oxford University Press.

Tseng, Jesse L. 2002. Remarks on marking. In Frank Van Eynde, Lars Hellan & Dorothee Beermann (eds.), *Proceedings of the 8th International Conference on Head-driven Phrase Structure Grammar*, 267–283. Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/2/.

Van Eynde, Frank. 2005. A head-driven treatment of asymmetric coordination and apposition. In Stefan Müller (ed.), *The proceedings of the 12th* *international conference on head-driven phrase structure grammar, department of informatics, university of lisbon*, 396–409. Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/6/.

Van Eynde, Frank. 2006. NP-internal agreement and the structure of the noun phrase. *Journal of Linguistics* 42(1). 139–186.

Van Eynde, Frank. 2007. The big mess construction. In Stefan Müller (ed.), *The Proceedings of the 14th International Conference on Head-driven Phrase Structure Grammar*, 415–433. Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/8/.

Basic copula clauses in Indonesian

David Moeljadi
Nanyang Technological University

Francis Bond
Nanyang Technological University

Luís Morgado da Costa
Nanyang Technological University

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 442–456

Keywords: copula, Indonesian, HPSG

Moeljadi, David, Francis Bond & Luís Morgado da Costa. 2016. Basic copula clauses in Indonesian. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 442–456. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.23.

Abstract

We want to show how basic copula clauses in Indonesian can be dealt with within the framework of Head Driven Phrase Structure Grammar (HPSG) (Pollard & Sag, 1994). We analyzed three types of basic copula clauses in Indonesian: copula clauses with noun phrase complements (NP) expressing the notions of 'proper inclusion' and 'equation', adjective phrases (AP) expressing 'attribution', and prepositional phrases (PP) expressing relationships such as 'location'. Our analysis is implemented in the Indonesian Resource Grammar (INDRA), a computational grammar for Indonesian (Moeljadi et al., 2015).

1 Introduction

Every language has a copula clause type, which may take a copula verb (Dryer, 2007). Some languages lack a copula verb; the copula slot is left blank and we have 'verbless clauses'. In addition, some languages have more than one kind of copula verb. Most commonly, one will just refer to 'a state' and the other to 'coming into a state', similar to *be* and *become* in English (Dixon, 2009, p. 175). In this paper, we limit our discussion to the stative 'be' clause. Indonesian, a Western Malayo-Polynesian language of the Austronesian language family,[1] has multiple copula verbs, distributed over different semantic relations in addition to 'verbless clauses'. We give an analysis that covers both multiple copula verbs and verbless clauses.

Analyses of Indonesian copulas can be found in reference grammars, such as Alwi et al. (2014), Mintz (2002), and Sneddon et al. (2010). A syntactic analysis in Lexical Functional Grammar (LFG) (Kaplan & Bresnan, 1982; Dalrymple, 2001) was done by Arka (2013). However, to the best of our knowledge, no work has been done on modeling Indonesian copula clauses in HPSG (Sag et al., 2003) and Minimal Recursion Semantics (MRS) (Copestake et al., 2005). This paper aims to fill in this gap, referring to existing HPSG literature on copulas, such as Bender (2001) and Van Eynde (2009). Our analysis is implemented in the Indonesian Resource Grammar (INDRA), a computational grammar for Indonesian (Moeljadi et al., 2015).[2]

Basic copula clauses in Indonesian can be divided roughly into three types, depending on the part-of-speech of the predicate: noun phrase (NP), adjective phrase (AP), or prepositional phrase (PP). Copula clauses taking an NP predicate typi-

[1] Indonesian (ISO 639-3: ind) belongs to the Malayic branch with Standard Malay spoken in Malaysia, Brunei Malay in Brunei, Local Malay in Singapore and other Malay varieties spoken at various places in Indonesia (Lewis, 2009). The Indonesian language is spoken mainly in the Republic of Indonesia as the sole official and national language and as the common language for hundreds of ethnic groups living there (Alwi et al., 2014, p. 1-2). In Indonesia it is spoken by around 22.8 million people as their first language and by more than 140 million people as their second language. It is over 80% cognate with Standard Malay (Lewis, 2009).

[2] http://moin.delph-in.net/IndraTop

cally express the notions of ‘proper inclusion’ and ‘equation’,[3] those taking an AP predicate express ‘attribution’, and the ones taking a PP predicate typically express ‘location’ (Payne, 2008, p. 111-123). Table 1 shows an outline of the three types of basic copula clauses in Indonesian.

Relation	Subject	Predicate	
Proper inclusion, Equation	*Budi*	(*adalah*)	*guru* (NP)
	Budi	is	a teacher
Attribution	*Budi*	ø	*pandai* (AP)
	Budi	is	clever
Location	*Budi*	(*ada*)	*di rumah* (PP)
	Budi	is	at home

Table 1: Three types of basic copula clauses in Indonesian

All three types of basic copula clauses in Table 1 can appear without a copula verb. In fact, ‘attribution’ is typically expressed without a copula verb. The copula verbs shown in Table 1 are *adalah*[4] for ‘proper inclusion’ and ‘equation’, and *ada* for ‘location’. However, as mentioned before, there are more than one copula for some semantic relations. These other types will be discussed in the following section.

2 Basic Data

2.1 Copula clauses with Noun Phrase Predicates

Copula clauses with an NP as predicate may or may not have a copula verb *adalah*, *ialah*,[5] or *merupakan*[6] (Alwi et al., 2014, p. 358-359). These clauses express the notions of ‘proper inclusion’ and ‘equation’. Indonesian does not distinguish these notions syntactically, as shown in Example [1a] and [1b]. The three copula verbs behave the same way.

Since *ialah* is historically derived from 3SG *ia*, it only occurs with a third person subject (Sneddon et al., 2010; Mintz, 2002). Example [1c] shows that *saya* “1SG” cannot be the subject of a copula clause with *ialah*.

The copula verb *merupakan* is a verb which is in the process of becoming a copula (see Footnote 6). At its present stage it cannot appear if the NP predicate

[3] ‘Proper inclusion’ is when a specific entity is asserted to be among the class of items specified in the nominal predicate, as in English sentence “He is a teacher”. Usually the subject is specific (“he”) and the nominal predicate is non-specific (“a teacher”). ‘Equation’ is when a particular entity is identical to the entity specified in the predicate nominal, e.g. “He is my father” (Payne, 2008, p. 114).

[4] *adalah* is derived from the existential verb *ada* and a focus particle *-lah*.

[5] *ialah* is derived from 3SG *ia* “s/he” and a focus particle *-lah*.

[6] *merupakan* is derived from a noun *rupa* “form, figure, appearance, sort”, an agent-trigger prefix *me-*, and an applicative suffix *-kan*. The original meaning is “to form, to shape, to constitute”.

is a specific referent, such as a proper name, demonstrative, or pronoun, as shown in Example [1d]. However, it can precede a unique referent NP with a definite marker or a possessive marker as shown in Example [1b]. In addition, it can take an aspect or tense marker, while *adalah* and *ialah* cannot, as shown in Example [1e]. These have been confirmed in the Indonesian section of the Nanyang Technological University — Multilingual Corpus (NTU-MC) (Tan & Bond, 2012), a parallel corpus containing 2,975 sentences from three sources: Singapore Tourism Board website,[7] a Sherlock Holmes short story *The Adventure of the Speckled Band*, and a Japanese short story written by Akutagawa Ryunosuke: *The Spider's Thread*.

To the best of our knowledge, there is no meaning difference among the three copula verbs. Sneddon et al. (2010), Alwi et al. (2014), Macdonald (1976), and Mintz (2002) note that *adalah* and *ialah* are interchangable and most common in noun clauses where either the subject or predicate is long or structurally complex in formal, written language. Alwi et al. (2014) mention that *adalah* can also be changed with *merupakan*.

(1) a. *Budi (adalah/ialah/merupakan) guru.*
Budi COP teacher
"Budi is a teacher."

b. *Budi (adalah/ialah/merupakan) guruku.*
Budi COP teacher=1SG
"Budi is my teacher."

c. *Saya (adalah/*ialah/merupakan) guru*
1SG COP teacher
"I am a teacher."

d. *Orang itu (adalah/ialah/*merupakan) Budi.*
person that COP Budi
"That person is Budi."

e. *Ini sudah/akan *adalah/*ialah/merupakan hal yang luar biasa.*
this PERF/FUT COP case REL beyond ordinary
"This has been/will be an extraordinary case." (based on NTU-MC sentence ID 11938)

2.2 Copula clauses with Adjective Phrase Predicates

Copula clauses which express the notion of 'attribution' are the ones which have an AP as the main semantic content and are called 'predicate adjectives' (Payne, 2008, p. 120-121). A copula is usually absent in predicate adjectives, as shown in

[7] www.yoursingapore.com

Example [2a]. As Sneddon et al. (2010, p. 246-247) note, a copula *adalah* may be used by some speakers in adjective clauses, as illustrated in Example [2b]. According to the first author's intuition, a copula *ialah* may be less commonly used than *adalah*, but *merupakan* cannot occur with predicate adjectives. Not all speakers agree with the grammaticality of this and we did not find any occurrence of predicate adjectives with copulas in the NTU-MC; further Arka (2013, p. 31, 33) states that a copula cannot precede an adjective. Even so, we do provide an analysis for copula + AP in this paper (and INDRA).

(2) a. *Budi pandai.*
Budi clever
"Budi is clever."

b. *Pernyataan itu (?adalah/??ialah/*merupakan) benar.*
statement that COP true
"That statement is true." (based on Sneddon et al., 2010, p. 247)

2.3 Copula clauses with Prepositional Phrase Predicates

Copula clauses which express the notion of 'location' are the ones which have a PP as the main semantic content and are called 'predicate locatives' (Payne, 2008, p. 121-123). An existential verb *ada* or *berada* may be used optionally in predicate locatives, as illustrated in Example [3a]. The copulas *adalah* or *ialah* may appear, too, as shown in Example [3b]. Both in Example [3a] and [3b], the PP is a complement, not an adjunct.

(3) a. *Budi (ada/berada) di rumah.*
Budi EXIST at home
"Budi is at home."

b. *Satu-satunya air yang ada (adalah/ialah/*merupakan) dari telaga.*
one-RED=DEF water REL EXIST COP from lake
"The only water there is is from the lake." (based on Sneddon et al., 2010, p. 247)

There is another 'benefactive' clause in which the main semantic content of the predication is realized in a PP, marked by a preposition *untuk* "for", and its syntactic pattern usually follows the one of predicate locatives (Payne, 2008, p. 122). In Indonesian, an optional copula verb *adalah* or *ialah* may appear in this 'benefactive' clause, as shown in Example [4a].

Regarding *ialah*, for the same reason mentioned in Section 2.1, it can only appear with a third person subject. Example [4b] shows that *engkau* "2SG" cannot be the subject of *ialah*, while Example [4c] shows that 3SG subject *presiden* "the president" can be the subject.

(4) a. *Ini* (adalah/ialah/*merupakan) *untuk* *Budi.*
this COP for Budi
"This is for Budi."

b. **Engkau* ialah *untukku.*
2SG COP for=1SG
Intended meaning: You are for me.

c. *Presiden* ialah *untuk* *rakyat.*
president COP for people
The president is for the people.

3 Analysis

3.1 Copula clauses with Noun Phrase Predicate

The copula verbs *adalah*, *ialah*, and *merupakan* take two arguments, syntactically similar to simple transitive verbs. Our analysis follows the Montagovian treatment as presented in Van Eynde (2009, p. 368), in the sense that this analysis treats the copula as a transitive verb, covering both the predicating and identifying uses.

In order to model the shared meaning of the various copulas, we use a simple type hierarchy, as illustrated in Figure 1. The supertype of all the NP copulas *cop-verb-lex* inherits from *transitive-verb-lex* with an obligatory complement. This then has two children. The copula *adalah* is an instance of *v_np_cop_noasp_le* which inherits from *cop-verb-lex* with an additional constraint: it cannot occur with any aspect or tense marker (see Example [1e]). The copula *merupakan* also inherits from *cop-verb-lex*, but with a different constraint: the head of the complement should be a common noun, not a proper noun, pronoun, or a demonstrative. We divided `noun` into `commonnoun`, `propername`, and `pronoun`. The copula *ialah* (*v_np_cop_3_le*) inherits from *v_np_cop-_noasp_le* with another constraint: the subject should be third person.

We use MRS as our semantic framework (Copestake et al., 2005). The MRS representation is the same as the one for transitive sentences (see Figure 2 where we show the dependency MRS representation: DMRS.[8]) *_cop_v_ialah_rel* is an event,[9] its `ARG1` has a constraint: the value of the `PNG.PERNUM` is *3sg*. So *named_rel* "*budi*" (`ARG1`) must be third person, while there is no constraint on the `ARG2`, *_guru_n_rel*. Figure 3 shows the parse tree of Example [1a] with a copula *adalah*.

[8]In the simplified version of the graph shown in this paper, properties of the predicates such as semantic type, aspect, tense and number are not shown. If they are important to the analysis they will be discussed in the text. Referential individuals will be in the restriction of a quantifier (shown with the link RSTR/H). All other predicates are events.

[9]Currently we do not distinguish between dynamic and stative meanings, referring to both as events.

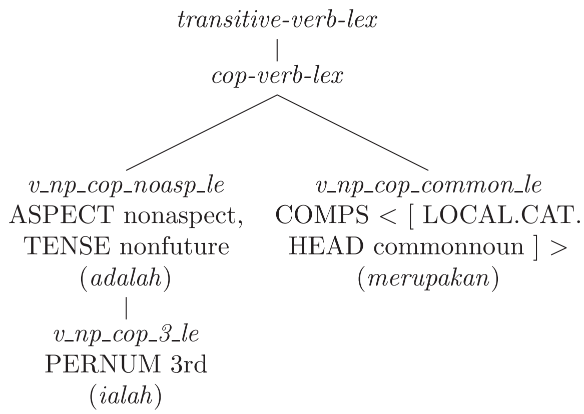


Figure 1: Type hierarchy of Indonesian copula verbs

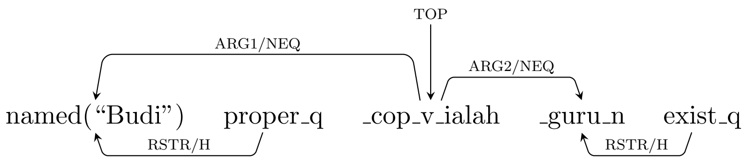


Figure 2: DMRS representation of *Budi ialah guru* "Budi is a teacher"

For zero copula clauses, we made a pumping rule[10] which pumps (or converts) an NP to a VP as shown in Figure 4. It also adds a copula predicate to the semantics; and links its daughter to `ARG2` and the subject to `ARG1`. This pumping rule introduces a predicate *cop_v_zero_rel* with the subject as the first argument and the NP predicate as the second argument, denoting a relation of coreference between them, covering both equational (identificational) and proper inclusion (predicational) relations. The MRS is similar to that produced by the copula verbs *adalah*, *ialah*, or *merupakan*.

This syntactic structure is similar to the one in Arka (2013, p. 38) where any

[10] A unary rule that changes the type (Copestake, 2002, p. 120).

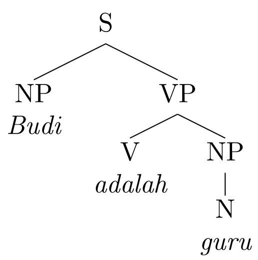


Figure 3: Parse tree of *Budi adalah guru* "Budi is a teacher"

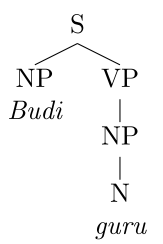


Figure 4: Parse tree of *Budi guru* "Budi is a teacher"

lexical category (VP, NP, AP, and PP) can be a predicate XP; the NP subject takes this XP to make an Indonesian clause. Our analysis corresponds to 'Constructional analysis II' in Bender (2001, p. 101-118). There are three kinds of facts which make such an analysis unsuccessful to deal with African American Vernacular English (AAVE) copula absence: the possibility of copulaless existentials, a curious interaction of negation and ellipsis, and the possibility of complement extraction (Bender, 2001, p. 107). These three things do not exist in Indonesian: Indonesian has an obligatory existential verb *ada*, compared with AAVE which has *there* and a zero copula in existential sentences, as shown in Example `[5a]`; AAVE has the possibility of copula ellipsis in case it strands *not*, while Indonesian uses a negation marker *tidak* or *bukan*[11] which does not occur with any copula, as shown in Example `[5b]`; finally, AAVE has a long distance dependency in which the complement of the silent copula can be extracted, while in Indonesian in Wh-question the complement or the question word must appear without a copula, as shown in Example `[5c]` and `[5d]`.[12] In short, because of differences in syntactic structure, the constructional analysis which does not work for AAVE can be implemented for Indonesian.

(5) a. *Ada mobil yang menghalangi jalanku.*
EXIST car REL block way=1SG
"There a car blocking my way." (based on Bender, 2001, p. 107)

b. *Mereka berkata mereka sahabat, tetapi sebenarnya bukan.*
3PL say 3PL best.friend but actually NEG
"They say they're best friends, but they not." (based on Bender, 2001, p. 115)

c. *Di mana (*adalah/*ialah/*merupakan) mobilmu (ada/berada)?*
at where COP car=2SG EXIST
"Where your car?" (based on Bender, 2001, p. 117)

d. *Mobilmu (*adalah/*ialah/*merupakan/ada/berada) di mana?*
"Where your car?" (based on Bender, 2001, p. 117)

3.2 Copula clauses with Adjective Phrase Predicate

As mentioned in Section 2.2, the predicate and the main semantic content of copula clauses with AP predicates is the AP. Predicative APs take one argument (NP as the subject), similar to intransitive predicates.[13] Figure 5 shows the parse tree of Example `[2a]`.

As mentioned in Section 2.2, a copula may or may not precede AP. In this paper, we provide an analysis for copula + AP, too. The copula *adalah* is treated as

[11] see Section 4 for negation.

[12] Sneddon et al. (2010, p. 324-328) note that question words may occur first in the clause, as in Example `[5c]`, or in the normal position, as in Example `[5d]`. If the question word is predicate of a non-verbal clause, it often precedes the subject, as in Example `[5c]`.

[13] There is a lexical rule that converts these to attributive adjectives for the modifier use.

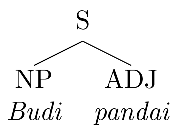


Figure 5: Parse tree of *Budi pandai* "Budi is clever"

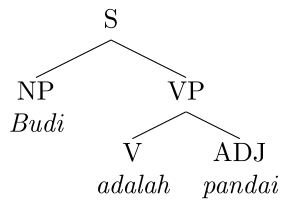


Figure 6: Parse tree of *Budi adalah pandai* "Budi is clever"

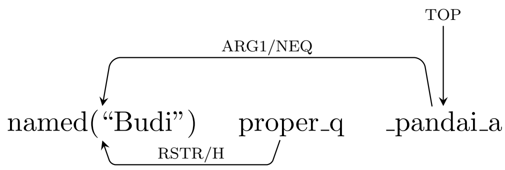


Figure 7: DMRS representation of *Budi (adalah) pandai* "Budi is clever"

a raising auxiliary which does not introduce a predicate and links its subject to the subject of its complement (the adjective). Figure 6 shows the parse tree of Example `[2a]` with *adalah*.

The MRS representation is the same as the one for intransitive sentences (see Figure 7 where we show the dependency MRS representation). The MRS of the clauses with and without *adalah* are the same. The event, *_pandai_a_rel* is the semantic head and hook for composition. Its `ARG1` is linked to the subject: *named_rel*("*Budi*").

3.3 Copula clauses with Prepositional Phrase Predicate

Predicate locatives have a PP as the main semantic content and an optional verb *ada* or *berada*, or a copula *adalah* or *ialah*. Predicative prepositions, such as *di* "in/on/at", take two arguments, similar to transitive predicates, as shown in Figure 8. When appearing with PPs, we also treat *ada*, *berada*, *adalah*, and *ialah* as auxiliaries which do not introduce a predicate of their own. The head of the subject is a noun and the head of the complement is a preposition. Figure 9 shows the parse tree of Example `[3a]` with an existential verb *ada*. The MRS of predicate locatives with *ada*, *berada*, *adalah*, and *ialah* is exactly the same as the one without, as shown in the dependency MRS representation in Figure 10.

In the MRS representation, the semantic head daughter and hook for composition is the event *_di_p_rel*. Its `ARG1` and `ARG2` are linked to *named_rel("Budi")* and *_rumah_n_rel* respectively.

Regarding 'benefactive' clauses, our analysis is the same as the one for predicate locatives. We treat *adalah* and *ialah* in these clauses as auxiliaries which do not introduce a predicate. The MRS (and DMRS) representation is similar to the one in Figure 10.

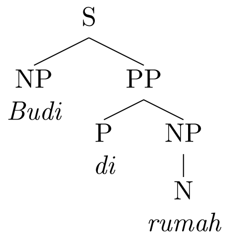


Figure 8: Parse tree of *Budi di rumah* “Budi is at home”

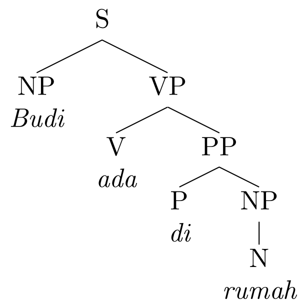


Figure 9: Parse tree of *Budi ada di rumah* “Budi is at home”

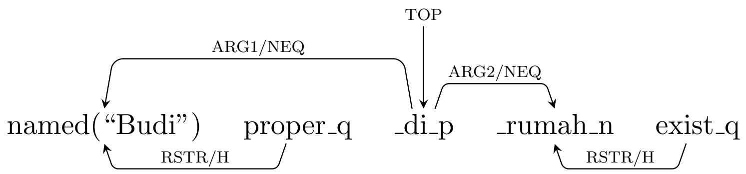


Figure 10: DMRS representation of *Budi (ada) di rumah* “Budi is at home”

4 Negation

Indonesian has two main negation markers for clauses, placed before the negated element. Examples 6, 7, and 8 summarize the interaction of negation with copula verbs in Indonesian, for NP, AP, and PP, respectively. The standard negation marker *tidak* is used when the predicate is verbal, including the copula verb *merupakan* and existential verbs *ada* and *berada*, as shown in Example `[6b]`, `[7a]`, and `[8c]`, or adjectival, as in Example `[7b]`, and with PP predicates, as shown in Example `[8]`. It cannot negate copula *adalah* or *ialah*, as illustrated in Example `[6b]`, `[6d]`, `[7c]`, and `[8b]`. In Example `[6d]`, *tidak* is not compatible with *adalah* and *ialah* and *merupakan* is ruled out because the NP predicate is a proper name (see also Example `[1d]`).

The special negation marker *bukan* “be not” is used when the predicate is nominal, as in Example `[6]` (Kroeger, 2014, p. 137),[14] or prepositional,[15] as shown in Example `[8]`. However, it cannot negate copula *adalah* or *ialah*, as illustrated in Example `[8b]`, or existential verbs *ada* and *berada*, as in Example `[8c]`.

(6) a. *Budi* *bukan/*tidak* *guru.*
Budi NEG teacher
“Budi is not a teacher.”

[14]Kroeger (2014, p. 137) notes that in certain kinds of contexts, *bukan* can be used to negate verbal clauses and argues that it is a marker of ‘external’ (sentential) negation. We will not discuss it because this is beyond the scope of this paper.

[15]Sneddon et al. (2010, p. 202) mention that a number of prepositions can be negated by either *bukan* or *tidak*.

b. *Budi tidak *adalah/*ialah/merupakan guru.*
Budi NEG COP teacher
"Budi is not a teacher."

c. *Guru itu bukan/*tidak Ali.*
teacher that NEG Ali
"That teacher is not Ali." (Arka, 2011, p. 85)

d. **Guru itu tidak adalah/ialah/merupakan Ali.*
teacher that NEG COP Ali
Intended meaning: That teacher is not Ali.

(7) a. *Mereka tidak/*bukan menolong kami.*
3PL NEG help 1PL.EXCL
"They didn't help us." (Sneddon et al., 2010, p. 202)

b. *Budi tidak/*bukan pandai.*
Budi NEG clever
"Budi is not clever."

c. **Budi tidak adalah/ialah/merupakan pandai.*
Budi NEG COP clever
Intended meaning: Budi is not clever.

(8) a. *Tempatnya tidak/bukan di sini.*
place=DEF NEG at here
"The place is not here."

b. *Ini tidak/bukan *adalah/*ialah/*merupakan untuk Budi.*
this NEG COP for Budi
"This is not for Budi."

c. *Budi tidak/*bukan ada/berada di rumah.*
Budi NEG EXIST at home
"Budi is not at home."

We treat *tidak* as an adverb modifying VP, AP, or PP, as shown in Figure 11. It is represented as *neg_rel* in the MRS. The value of its `ARG1` is equated with the `LBL` of the VP, AP, or PP predicate, as illustrated in Figure 12. We treat *bukan* as a non-modifier verb, a combination of *copula_v_rel* as the head and *neg_rel* as the daughter, which takes an NP subject and an NP or PP complement, as shown in Figure 13 and 14.

In order to block *tidak adalah* and *tidak ialah* from parsing, we added a restriction in *tidak*: the value of the `ASPECT` of the VP which it modifies should be `perf-and-prog`, which means it modifies verbs that can take a perfect or progressive aspect marker. Because *adalah* and *ialah*'s `ASPECT` is `nonaspect`, which means they cannot take aspect markers, it is not compatible.

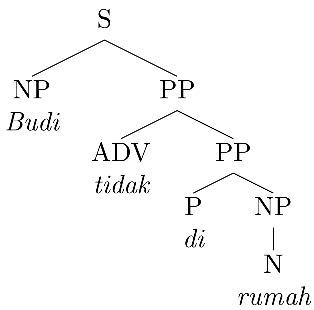


Figure 11: Parse tree of *Budi tidak di rumah* “Budi is not at home”

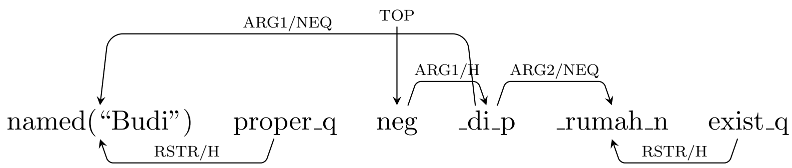


Figure 12: DMRS representation of *Budi tidak di rumah* “Budi is not at home”

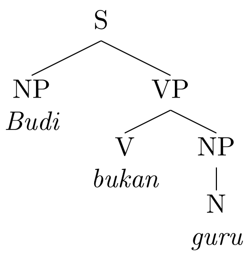


Figure 13: Parse tree of *Budi bukan guru* “Budi is not a teacher”

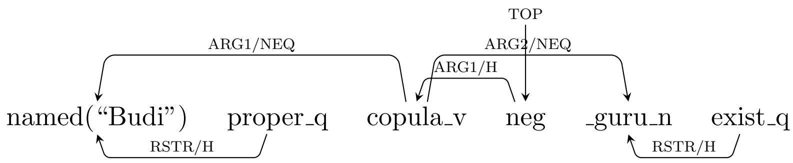


Figure 14: DMRS representation of *Budi bukan guru* “Budi is not a teacher”

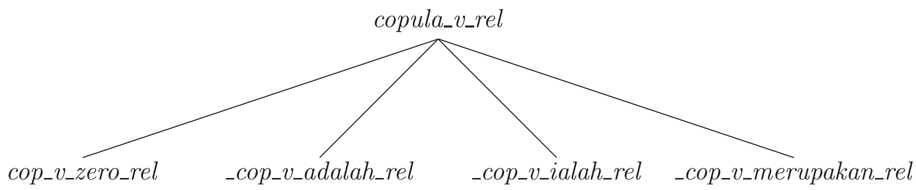


Figure 15: Semantic hierarchy of copulas

5 Generation

Again, we model similarities, in this case of meaning, using a type hierarchy, as illustrated in Figure 15. We can use this to underspecify the input to the generator. For example, for *copula_v_rel* it will then try to generate all predicates that are subsumed by it, i.e. all copula constructions, and only succeed for the grammatical ones.

Input:	Input:
`Budi merupakan guru`	`Budi ada di rumah`
Output:	Output:
`Budi guru`	`Budi di rumah`
`Budi ialah guru`	`Budi ada di rumah`
`Budi adalah guru`	`Budi berada di rumah`
`Budi merupakan guru`	`Budi adalah di rumah`
`...`	`...`

6 Conclusion

Our analyses of Indonesian copula clauses are similar to Arka (2013)'s LFG analysis but cover more copula verbs with a refined type hierarchy. Because of differences in syntactic structure between AAVE and Indonesian, the analysis that builds a VP out of a predicative NP, which does not work for AAVE, can be successfully implemented for Indonesian.

References

Alwi, Hasan, Soenjono Dardjowidjojo, Hans Lapoliwa & Anton M. Moeliono. 2014. *Tata Bahasa Baku Bahasa Indonesia*. Jakarta: Balai Pustaka 3rd edn.

Arka, I Wayan. 2011. On modality and finiteness in Indonesian: complexities of =nya nominalisation. In *Workshop on TAM markers and evidentiality in Indonesian languages*, 17–18. Tokyo University of Foreign Studies.

Arka, I Wayan. 2013. Nonverbal predicates in Austronesian and Papuan languages: an LFG perspective. In Ida Bagus Putra Yadnya, Aron Meko Mbete, Ni Luh Ketu Mas Indrawati, I Gusti Ayu Gde Sosiowati & I Wayan Suardiana (eds.), *Proceedings of the 6th International Seminar on Austronesian and non-Austronesian Languages and Literatures*, 30–46. Unud, Bali.

Bender, Emily M. 2001. *Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence*: Stanford University dissertation.

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. Stanford: CSLI Publications.

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan A. Sag. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language and Computation* 3(4). 281–332.

Dalrymple, Mary. 2001. Lexical-Functional Grammar. *Syntax and Semantics* 34. Academic Press.

Dixon, R. M. W. 2009. *Basic Linguistic Theory 2: Grammatical Topics*. Oxford: Oxford University Press.

Dryer, Matthew S. 2007. Clause types. *Language Typology and Syntactic Description* 1. 224–275. Cambridge University Press.

Kaplan, Ronald & Joan Bresnan. 1982. Lexical Functional Grammar: A formal system for grammatical representation. In *The Mental Representation of Grammatical Relations*, 173–281. Cambridge: the MIT Press.

Kroeger, Paul. 2014. External negation in Malay/Indonesian. *Language* 90(1). 137–184.

Lewis, M. Paul. 2009. *Ethnologue: Languages of the World*. Dallas, Texas: SIL International 16th edn. http://www.ethnologue.com.

Macdonald, R. Ross. 1976. *Indonesian Reference Grammar*. Washington: Georgetown University Press 2nd edn.

Mintz, Malcolm W. 2002. *An Indonesian and Malay Grammar for Students*. Perth: Indonesian / Malay Texts and Resources 2nd edn.

Moeljadi, David, Francis Bond & Sanghoun Song. 2015. Building an HPSG-based Indonesian Resource Grammar (INDRA). In *Proceedings of the GEAF Workshop, ACL 2015*, 9–16. http://aclweb.org/anthology/W/W15/W15-3302.pdf.

Payne, Thomas E. 2008. *Describing Morphosyntax: A Guide for Field Linguists*. New York: Cambridge University Press.

Pollard, Carl & Ivan A. Sag. 1994. *Head Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Sag, Ivan A., Thomas Wasow & Emily M. Bender. 2003. *Syntactic Theory: A Formal Introduction*. Stanford: CSLI Publications 2nd edn.

Sneddon, James Neil, Alexander Adelaar, Dwi Noverini Djenar & Michael C. Ewing. 2010. *Indonesian Reference Grammar*. New South Wales: Allen & Unwin 2nd edn.

Tan, Liling & Francis Bond. 2012. Building and Annotating the Linguistically Diverse NTU-MC (NTU-Multilingual Corpus). *International Journal of Asian Language Processing* 22(4). 161–174.

Van Eynde, Frank. 2009. On the copula: from a Fregean to a Montagovian treatment. In Stefan Müller (ed.), *Proceedings of the 16th International Conference on HPSG*, 359–375. http://cslipublications.stanford.edu/HPSG/2009/vaneynde.pdf.

Flexible phrasal constructions, constituent structure and (cross-linguistic) generalizations: A discussion of template-based phrasal LFG approaches

Stefan Müller

Humboldt Universität zu Berlin

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 457–477

Keywords: LFG, template, lexicalism, Construction Grammar

Müller, Stefan. 2016. Flexible phrasal constructions, constituent structure and (cross-linguistic) generalizations: A discussion of template-based phrasal LFG approaches. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 457–477. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.24.

Abstract

This paper discusses recent LFG proposals on resultative and benefactive constructions. I show that neither resultative nor benefactive constructions are fully fixed and that this flexibility requires traces or a stipulation of constructional templates at several unrelated places in the grammar, something that is not necessary in lexical approaches. A second part of the paper deals with the active/passive alternation and shows that language-internal generalizations are missed if constraints are assumed to be contributed by phrase structure rules. A third part examines the parallel constructions in German and shows that cross-linguistic generalizations are not captured by phrasal approaches.

1 Introduction

Goldberg (1995, 2006), Tomasello (2003) and others argue for a phrasal view on argument structure constructions: lexical entries for verbs come with minimal specifications as to which arguments are required by a verb but they come with a specification of argument roles. Verbs can be inserted into phrasal constructions and these constructions may express the arguments that belong to a verb semantically or even add further arguments. A frequently discussed example is the one in (1):

(1) He runs his sneakers threadbare.

run is an intransitive verb, but in (1) it enters the resultative construction, which licenses an additional argument (*his sneakers*) and a result predicate (*threadbare*). The resultative semantics is said to be contributed by the whole phrasal pattern rather than by one of its elements. The lexical approach assumes that there are several lexical items for verbs like *run*. There is the lexical item that is needed to analyze simple sentences with the intransitive verb and its subject and there is a further lexical item that is used in the analysis of sentences like (1). The latter lexical item selects for a subject, an object and a result predicate and contributes the resultative semantics. Both lexical items are related by a lexical rule. See Simpson, 1983, Verspoor, 1997, Wechsler, 1997, Wechsler & Noh, 2001, Wunderlich 1997, 120–126, Kaufmann & Wunderlich, 1998, Müller, 2002, Chapter 5, and Kay, 2005 for lexical analyses in several frameworks).

[2]I thank Ash Asudeh for extensive personal discussion and Mary Dalrymple, and Ida Toivonen for a long discussion via email. I thank the reviewers of HeadLex 2016 for their comments in the pre-conferecne and post-conference reviewing. I thank Steve Wechsler, Martin Haspelmath, and Dick Hudson for discussion of an earlier version of this paper. Thanks to Jonas Kuhn for discussion of the attachment of constraints to c-structures and Economy of Expression. Thanks to Tom Wasow and Philip Miller for answering my request to the HPSG mailing list regarding extraction of primary objects. I also want to thank the participants of HeadLex 2016 for (intense) discussion. Miriam Butt, Mary Dalrymple, Ron Kaplan, and Anna Kibort deserve special mention. This paper was also presented at the Computational Linguistics Colloquium in Düsseldorf. I thank the working group for the invitation and the audience for discussion.

The question whether constructions like (1) should be treated as lexical or as phrasal constructions has been discussed in the literature in several papers (Goldberg & Jackendoff, 2004; Müller, 2006; Müller & Wechsler, 2014; Goldberg, 2013) but since most Construction Grammar publications (intentionally, see Goldberg, 2006) are not formalized the discussion of aspects not treated in the original proposal (e.g., interaction with morphology, application of the approach to non-configurational languages like German, partial verb phrase fronting) was rather hypothetical. There have been Construction Grammar-inspired proposals in HPSG (Haugereid, 2007, 2009) and Simpler Syntax (Culicover & Jackendoff, 2005) and these were shown to have empirical problems, to make wrong predictions or to be not extendable to other languages (Müller, 2013, 2016b). Formal CxG proposals (Bergen & Chang, 2005; van Trijp, 2011) are discussed in Müller, 2016b, Chapter 10.6.3 and Müller, To appear. Recently, several articles have been published suggesting a template-based phrasal approach in LFG that makes use of glue-semantics, a resource-driven semantic theory (Christie, 2010; Asudeh, Giorgolo & Toivonen, 2014). While these proposals seem to avoid many of the challenges that earlier proposals faced, they in fact have many of the problems that were discussed with respect to hypothetical proposals in Construction Grammar. Fortunately, the LFG proposals are worked out in detail and are embedded in a formal theory that provides formalized analyses of the languages and phenomena under discussion. It is therefore possible to show what the new template-based theories predict and to pin down exactly the phenomena where they fail.

The traditional analysis of the resultative construction in the framework of LFG is a lexical one (Simpson, 1983) but, more recently, several researchers have suggested a different view on argument structure constructions in the framework of LFG. For instance, Alsina (1996) and Christie (2010) suggest analyzing resultative constructions as phrasal constructions and Asudeh, Dalrymple & Toivonen (2008, 2013) argue for a phrasal analysis of the (Swedish) caused motion construction. Toivonen (2013) discusses benefactive constructions of the type in (2b).

(2) a. The performer sang a song.
    b. The performer sang the children a song.

Toivonen notices that the benefactive NP cannot be fronted in questions (3) and that passivization is excluded for some speakers of English (4).[1]

(3) a. I baked Linda cookies.
    b. * Who did I bake cookies?
    c. The kids drew their teacher a picture.
    d. * Which teacher did the kids draw a picture?

(4) * My sister was carved a soap statue of Bugs Bunny (by a famous sculptor).

[1] See Hudson (1992, 257) for references to several papers with varying judgments of question formation involving the fronting of the primary object. See Langendoen et al. (1973) for an experimental study.

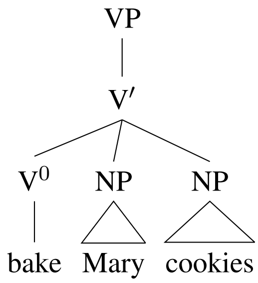


Figure 1: Phrasal configuration for benefactives according to Toivonen (2013, 505)

While she provides a lexical rule-based analysis of benefactives in her 2013 paper, she states in the discussion section:

> The manipulations that involve the word order consistently render the examples ungrammatical; see section 2.3 for the relative ordering test, section 2.4 and examples (47–48) for wh-extraction, section 2.5 for VP anaphora, and section 2.6 for pseudo-clefts. The distribution of benefactive NPs is thus very limited: it can only occur in the frame given in (5). This does not directly follow from the analysis given in section 3, and I will not attempt to offer an explanation for these intriguing facts here. However, it is perhaps possible to adopt an analysis simlar to the one Asudeh et al. (2013) propose for the Swedish directed motion construction (Toivonen 2002). Asudeh et al. (2013) posit a template that is directly associated with a construction-specific phrase structure rule. (Toivonen, 2013, 416)

The configuration that she provides in (5) is given in Figure 1 here. Asudeh, Giorgolo & Toivonen (2014) develop the respective phrasal analysis of the benefactive construction.

Note that Asudeh, Dalrymple, and Toivonen do not argue for a phrasal treatment of argument structure constructions in general. They do not assume that there is a phrasal transitive construction that licenses arguments for normal sentences like *Kim likes Sandy.* or a phrasal ditransitive construction that licenses the objects of normal ditransitive verbs like *give*. They just treat certain specific constructions phrasally, namely those that have a fixed conventionalized form or special idiosyncratic constraints on order that are difficult to capture lexically.

I think the 2014 paper does not reflect the intuition behind the statement in Toivonen, 2013 since Asudeh et al. (2014) are dealing with the grammar of speakers that permit passivization (and as I show below also extraction of the secondary object) and hence the structure of the benefactive construction is not fixed. What I am criticizing here is an approach relying on configurations for phenomena that interact with valence change and extraction and other phenomena that distort phrasal configurations.

The approach of Asudeh et al., 2014 could be seen as a way to formalize phrasal constructional approaches like those by Goldberg (1995, 2004) and Culicover & Jackendoff (2005). What I want to show in this paper is that the phrasal LFG approach has too many drawbacks in comparison to the lexical approaches. Since the phrasal approach is rejected for two specific argument structure constructions (benefactives and resultatives), it follows that it cannot be a viable approach for all argument structure constructions.

The remainder of the paper is structured as follows: I first discuss interactions of the resultative and benefactive construction with extraction and passivization (Section 2), then go on to discuss possible treatments of passivization and point out that generalizations are missed language internally (Section 3). Section 4 examines how the analyses could be adapted to German and I argue that cross-linguistic generalizations are not captured in phrasal analyses. Section 5 shows how restrictions on extraction and passivization can be captured in a lexical analyses. The paper concludes in Section 6.

2 The flexibility of the constructions

Christie (2010) and Toivonen (2013) and Asudeh et al. (2014) suggest phrasal constructions for resultative and benefactive constructions with a fixed number of daughters on the right-hand side of the c-structure rule. Christie (2010) suggested the following c-structure rule for the introduction of the result predicate and its subject:

(5)
$$\begin{array}{llccc} \mathrm{V}' & \rightarrow & \mathrm{V} & \mathrm{DP} & \{\ \mathrm{DP}|\mathrm{AP}|\mathrm{PP}\ \} \\ & & \uparrow=\downarrow & (\uparrow \mathrm{OBJ})=\downarrow & (\uparrow \textsc{xcomp})=\downarrow \\ & & & & (\downarrow\mathrm{SUBJ})=(\uparrow\mathrm{OBJ}) \\ & & & \multicolumn{2}{c}{@\mathrm{RESULT\text{-}T}((\uparrow\mathrm{PRED\ FN}))} \end{array}$$

In Christie's analysis, verbs are assumed to only optionally provide semantic and f-structure constraints. If they enter the resultative construction in (5), the construction takes over and provides a PRED value and specifications for grammatical functions.

The rule for the benefactive construction in (6) is provided in (7).

(6) The performer sang the children a song.

(7)
$$\begin{array}{llccc} \mathrm{V}' & \rightarrow & \mathrm{V} & \mathrm{DP} & \mathrm{DP} \\ & & \uparrow=\downarrow & (\uparrow \mathrm{OBJ})=\downarrow & (\uparrow \mathrm{OBJ}_{\theta})=\downarrow \\ & & (\ @\textsc{Benefactive}\ ) & & \end{array}$$

According to the authors, the noun phrase *the children* is not an argument of *sing* but contributed by the c-structure rule that optionally licenses a benefactive (Asudeh et al., 2014, 81).

As will be shown in the following, neither the resultative construction nor the benefactive construction is fixed in this form. Let us look at resultatives first. Carrier & Randall (1992, 185) discuss extraction data like those in (8):

(8) a. ? How shiny do you wonder which gems to polish?
b. ? Which colors do you wonder which shirts to dye?

These examples show that both the result phrase and the object can be extracted. The examples in (9) show that the objects can be extracted with the result predicate remaining in the V′:

(9) a. I wonder which gems to polish shiny?
b. I wonder which shirts to dye that color?

It is also possible to extract the result predicate and leave the object in place:

(10) a. I wonder how shiny to polish the gems?
b. I wonder which color to dye the shirts?

Apart from extraction, passivization is possible as well:

(11) a. The shoes were polished shiny.
b. The shirts were dyed a different color.

This means that the object, the result predicate, or both the object and the result predicate may be missing from the resultative construction in (5). The same is true for the benefactive construction. Asudeh et al. (2014) deal with grammars of speakers of English that allow for passivization of benefactive constructions. For those speakers all examples in (12) are fine:

(12) a. Her husband prepared her divine and elaborate meals.
b. She had been prepared divine and elaborate meals.
c. Such divine and elaborate meals, she had never been prepared before, not even by her ex-husband who was a professional chef.

The examples show that some speakers permit the promotion of the benefactive to subject as in (12b,c) and the remaining object can be extracted as in (12c).

While the extraction of the benefactive is out (3d), (13) shows that the secondary object in a benefactive construction can be extracted.

(13) a. What kind of picture did the kids draw the teacher?
b. the picture that the kids drew the teacher

The benefactives seem to pattern with normal ditransitives here. For an overview citing several other sources see Hudson, 1992, 258. Hudson reports that the extraction of the primary object of normal ditransitives is also judged as marked or even ungrammatical by many authors and subjects:

(14) a. We give children sweets.
b. Which sweets do you give children _?
c. % Which children do you give _ sweets?

Some variants of LFG account for extraction by assuming that the extracted element is not realized locally. The respective daughter in a rule is optional and the place in the f-structure is filled via functional uncertainty (Kaplan & Zaenen, 1989; Dalrymple 2001, 415; Dalrymple, Kaplan & King, 2001; Zaenen & Kaplan 2002). This means that in (8) and (12c), we have a situation in which it is just the verb that remains in the VP. All other elements are either promoted to grammatical functions that are realized outside of the VP or they are extracted. This means that nothing of the original configuration is left, it is just the verb. Christie's analysis of the resultative would be in deep trouble since she assumed that the resultative template is optionally introduced at the result predicate and overwrites optional information coming from the verb. As is clear from looking at the examples in (8), attaching the constraint to the extracted result predicate would be inappropriate since the result predicate can be fronted and would appear in another local domain (the one of *wonder* rather than *dye*, compare also the discussion of (27)). The constraints would apply to the wrong f-structure. The phrasal approach could be saved by assuming traces (as Berman (2003, Chapter 6) does for extraction crossing clause boundaries). This would be compatible with Christie's proposal since the structure would remain the same with some arguments being realized by empty elements.[2]

The situation with the benefactive construction is similar: in (12c) we have a bare verb and all other items are promoted or extracted. The template is associated with the verb. One could either insist on the phrasal pattern in (7) and assume an additional rule for the passive (see Section 3) and a trace for extraction or assume that constituents are optional and that rules like (7) can be used to account for all examples in (12). If one follows the latter proposal, the c-structure is not really restrictive. In the analysis of (12c) only the verb is present and one therefore could assume a lexical approach in which the benefactive template is associated with the verb right away. (see the discussion of (19), which suggests that there is an advantage for the lexical proposal)

Asudeh et al. (2014, 81) state that *The call to* BENEFACTIVE *is optional, such*

[2]Mary Dalrymple and Miriam Butt (p. c. 2016) pointed out another solution to me: one can annotate the c-structure rule for the CP that combines an extracted phrase and a C′. Extracted phrases find the place in the f-structure that belongs to the place from which they are extracted by functional uncertainty. The resultative template could be associated with the respective place in the f-structure by functional uncertainty as well. However, we would then have a grammar that introduces resultative constructions in at least two places: SpecCP and in a special resultative V′. A generalization about English (and German) is that constituents can be extracted out of their local contexts and be fronted. Although technically possible, I consider it inappropriate to state at the SpecCP node any information about the internal structure of subconstituents from which the extraction took place. For certain types of resultative constructions, a resultative template in fronted position would license an additional object and result predicate in an embedded V′. Note also that authors who assume a phrasal resultative construction probably would also want to assume other phrasal constructions as well. If these allow extraction of crucial parts the respective annotations at SpecCP would be necessary. The generalization about extraction would be missed. (See also the discussion of Figure 4 below.)

In addition the lexical approach assumes one place where the resultative predicate is licensed: the lexical rule. The phrasal approach would assume at least two (unrelated) places. On Occamian grounds the lexical analysis has to be preferred.

that the double-object rule is general and can also apply to non-benefactive cases. If passivization and extraction are treated by declaring arguments to be optional the phrase structure rule in (7) has to be formulated to account for normal ditransitive verbs. If the rule in (7) is supposed to rule out passives like (4) the benefactive NP has to be obligatory. This would also rule out passives of normal ditransitives.

(15) My sister was given a soap statue of Bugs Bunny (by a famous sculptor).

So, if the rule were responsible for normal ditransitives as well as for benefactives, all constraints regarding the obligatory presence of daughters would have to reside in the template since this is the only part that is different between benefactives and normal ditransitives. The templates defined by Asudeh et al. (2014) contain semantic constraints and constraints relevant for argument structure mappings. Nothing syntactic is encoded there. So, either the authors assume that benefactives pattern like normal ditransitives syntactically in the speaker group that they examine and then there would be no need to introduce the benefactive argument phrasally or there is a difference and then a special benefactive c-structure rule should be assumed that is incompatible with normal ditransitive verbs.

3 Phrasal introduction of arguments and missing generalizations about the passive

Asudeh, Giorgolo & Toivonen (2014) discuss the phrasal introduction of benefactives. (16) provides examples of the benefactive construction in an active and a passive variant.

(16) a. The performer sang the children a song.
b. The children were sung a song.

According to the authors, the noun phrase *the children* is not an argument of *sing* but contributed by the c-structure rule that optionally licenses a benefactive (Asudeh et al., 2014, 81). The c-structure rule in (7) is the rule that licenses fixed configurations like the one in Figure 1. Whenever this rule is called, the template BENEFACTIVE can add a benefactive role and the respective semantics, if this is compatible with the verb that is inserted into the structure. One of Toivonen's observations that motivated the phrasal approach was that passivization of benefactive constructions is excluded for some speakers (see example (4)).

Asudeh et al. (2014) deal with those variants of English that allow for a passive and discuss examples like (16b). They show how the mappings for the passive example in (16b) work, but they do not provide the c-structure rule that licenses such examples. Some authors assume that all nodes in c-structures are optional (Bresnan, 2001) but this would contradict the original intention of Toivonen, 2013 since if all constituents on the right-hand side are optional the c-structure rule in (7) would not rule out the ungrammatical instances of question formation in (3). Asudeh (p. c. 25.11.2016) informed me that their intention was to see the arguments in rule (7) as obligatory. Optional arguments are marked by including them in parentheses, which is not the case in (7). So, if one wanted to stick to the c-structure rule with a fixed number of obligatory daughters, one would need a special c-structure rule for passive VPs and this rule has to license a benefactive as well.[3] So it would be:

(17) $$\mathrm{V}' \rightarrow \begin{array}{cc} \mathrm{V[pass]} & \mathrm{DP} \\ \uparrow = \downarrow & (\uparrow \mathrm{OBJ}_\theta) = \downarrow \\ (\ @\mathrm{B{\small ENEFACTIVE}}\ ) & \end{array}$$

The problem is that there is no relation between the rules in (7) and (17). They are independent statements saying that there can be a benefactive in the active and that there can be one in the passive. This is what Chomsky (1957, 43) criticized in 1957 with respect to simple phrase structure grammars and this was the reason for the introduction of transformations. Bresnan-style LFG captured the generalizations by lexical rules (Bresnan, 1982) and later by lexical rules in combination with Lexical Mapping Theory (Bresnan & Kanerva, 1989). But if elements are added outside the lexical representations, the representations where these elements are added have to be related too. One could say that our knowledge about formal tools has changed since 1957. We now can use inheritance hierarchies to capture generalizations. So one can assume a type (or a template) that is the supertype of all those c-structure rules that introduce a benefactive. But since not all rules allow for the introduction of a benefactive element, this basically amounts to saying: c-structure rule A, B, and C allow for the introduction of a benefactive. In comparison lexical rule-based approaches have one statement introducing the benefactive. The lexical rule states what verbs are appropriate for adding a benefactive and syntactic rules are not affected.

As was already mentioned above, (7) and (17) can be generalized over if the daughters in (7) are regarded as optional. With optional daughters, (7) is equivalent to a specification of nine rules. If we ignore the cases in which the verb is omitted, we are left with four rules namely (7) and the three versions of the rule in (18):

(18) a. $$\mathrm{V}' \rightarrow \begin{array}{cc} \mathrm{V} & \mathrm{DP} \\ \uparrow = \downarrow & (\uparrow \mathrm{OBJ}_\theta) = \downarrow \\ (\ @\mathrm{B{\small ENEFACTIVE}}\ ) & \end{array}$$

b. $$\mathrm{V}' \rightarrow \begin{array}{cc} \mathrm{V} & \mathrm{DP} \\ \uparrow = \downarrow & (\uparrow \mathrm{OBJ}) = \downarrow \\ (\ @\mathrm{B{\small ENEFACTIVE}}\ ) & \end{array}$$

c. $$\mathrm{V}' \rightarrow \begin{array}{c} \mathrm{V} \\ \uparrow = \downarrow \\ (\ @\mathrm{B{\small ENEFACTIVE}}\ ) \end{array}$$

[3] See for instance Bergen & Chang, 2005 and van Trijp, 2011 for Construction Grammar analyses that assume active and passive variants of phrasal construction. See Cappelle (2006) on allostructions in general.

(18a) is the variant of (7) in which the OBJ is omitted, (18b) is the variant in which the OBJ_θ is omitted and in (18c) both DPs are omitted. Hence, (7) can be used for V′s containing two objects and for V′s in the passive containing just one object. The template-based approach does not overgenerate since the benefactive template is specified such that it requires the verb it applies to to select for an ARG2. Since intransitives like *laugh* do not select for an ARG2 a benefactive cannot be added. So, in fact the actual configuration in the c-structure rule does not play any role at all: the account entirely relies on semantics and resource sensitivity.[4] This means that it is not the case that an argument is added by a certain configuration the verb enters in. Since any verb may enter (18) and since the only important thing is the interaction between the lexical specification of the verb and the benefactive template, the same structures would be licensed if the benefactive template were added to the lexical items of verbs directly.

Since the actual configuration does not constrain anything, all (alleged) arguments from language acquisition and psycholinguistics for phrasal analyses would not apply to such a phrasal account.[5]

Concluding this section it can be said that the difference between the lexical use of the benefactive template or the phrasal introduction as executed in (7) is really minimal. However, there is one area in grammar where there is a difference: coordination. As Müller & Wechsler (2014, Section 6.1) pointed out it is possible to coordinate ditransitive verbs with verbs that appear together with a benefactive. (19a) is one of their examples and (19b) is an additional example:

(19) a. She then offered and made me a wonderful espresso — nice.[6]
b. My sisters just baked and gave me a nutella cupcake with mint chocolate chip ice-cream in the middle and milk chocolate frosting on

[4]The account would permit (i.b,c) since *give* with prepositional object has an ARG_2 (Kibort, 2008, 317).

(i) a. He gave it to Mary.
b. * He gave Peter it to Mary.
c. * Peter was given it to Mary.

give could combine with the *to* PP semantically and would then be equivalent to a transitive verb as far as resources are concerned (looking for an ARG_1 and an ARG_2). The benefactive template would map the ARG_2 to ARG_3 and hence (i.b) would be licensed. Since there are verbs that take a benefactive and a PP object as shown by (ii), (i.b) cannot be ruled out with reference to non-existing c-structure rules.

(ii) I buy him a coat for hundred dollar.

I assume that (i.b,c) are ruled out on semantic grounds by constraints that forbid two recipients for one verbs. See Toivonen (2013) on the observation that benefactive NPs are recipients.

[5]Note again that Asudeh et al. (2013) and Asudeh et al. (2014) do not argue for a general phrasal account for all argument structure constructions. They did not argue for such a general approach on the basis of language acquisition or psycholinguistic data. I just point out here that their approach should not be mistaken as a formalization of such a general approach.

[6]http://www.thespinroom.com.au/?p=102 07.07.2012

top.[7]

If the benefactive information is introduced at the lexical level the coordinated verbs basically have the same selectional requirements. If the benefactive information is introduced at the phrasal level *baked* and *gave* are coordinated and then the benefactive constraints are imposed on the result of the coordination by the c-structure rule. While it is clear that the lexical items that would be assumed in a lexical approach can be coordinated as symmetric coordination, problems seem to arise for the phrasal approach. It is unclear how the asymmetric coordination of the mono- and ditransitive verbs can be accounted for and how the constraints of the benefactive template are distributed over the two conjuncts. The fact that the benefactive template is optional does not help here since the optionality means that the template is either called or it is not. The optionality does not allow for a distribution to one of the daughters in a coordination.

Mary Dalrymple (p. c. 2016) pointed out that the coordination rule that coordinates two verbs can be annotated with two optional calls of the benefactive template.

(20) V $\rightarrow$ V (@BENEFACTIVE) Conj V (@BENEFACTIVE)

In an analysis of the examples in (19), the template in rule (7) would not be called but the respective templates in (20) would be called instead. While this does work technically, similar coordination rules would be needed for all other constructions that introduce arguments in c-structures. Furthermore, the benefactive would have to be introduced in several unrelated places in the grammar and finally the benefactive is introduced at nodes consisting of a single verb without any additional arguments being licensed, which means that one could have gone for the lexical approach right away. Timm Lichte (p. c. 2016) pointed out an important consequence of a treatment of coordination via (20): since the result of the coordination behaves like a normal ditransitive verb it would enter the normal ditransitive construction and hence it would be predicted that none of the constraints on passive and extraction that are formulated at the phrasal level would hold if an item is coordinated with either another benefactive verb or a normal ditransitive verb like *give*.

4 Missing cross-linguistic generalizations

In Müller & Wechsler (2014) we argued that the approach to Swedish caused motion constructions by Asudeh, Dalrymple & Toivonen (2008, 2013) would not carry over to German since the German construction interacts with derivational morphology. Asudeh & Toivonen (2014) argued that Swedish is different from German and hence there would not be a problem. However, the situation is different with the

[7] http://bambambambii.tumblr.com/post/809470379. 05.06.2012.

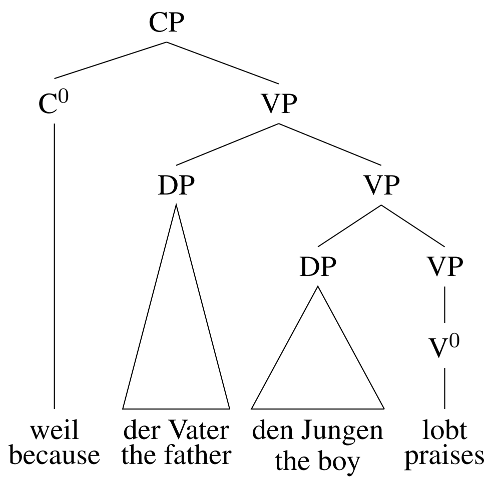


Figure 2: Analysis of German embedded clauses according to Berman (2003, 37)

benefactive construction. Although English and German do differ in many respects, both languages have similar benefactive constructions:

(21) a. He baked her a cake.
  b. Er buk ihr einen Kuchen.
   he baked her.DAT a.ACC cake

Now, the analysis of the free constituent order in German was explained by assuming binary branching structures in which a VP node is combined with one of its arguments or adjuncts (see Berman 1996, Section 2.1.3.1; 2003 and also Choi, 1999). For instance, Berman (2003, 37) assumes the analysis depicted in Figure 2. The c-structure rule is provided in (22):

(22) $$\mathrm{VP} \rightarrow \quad \underset{(\uparrow \mathrm{SUBJ} \,|\mathrm{OBJ}\, |\mathrm{OBJ}_{\theta}) = \downarrow}{\mathrm{DP}} \quad \underset{\uparrow = \downarrow}{\mathrm{VP}}$$

The dependent elements contribute to the f-structure of the verb and coherence/ completeness ensure that all arguments of the verb are present. One could add the introduction of the benefactive argument to the VP node of the right-hand side of the rule as in (23):

(23) $$\mathrm{VP} \rightarrow \quad \underset{(\uparrow \mathrm{SUBJ} \,|\mathrm{OBJ}\, |\mathrm{OBJ}_{\theta}) = \downarrow}{\mathrm{DP}} \quad \begin{array}{c}\mathrm{VP} \\ \uparrow = \downarrow \\ (\ @\textsc{Benefactive}\ )\end{array}$$

However, since the verb-final variant of (21b) would have the structure in (24), one would get spurious ambiguities, since the benefactive could be introduced at several VP nodes:

(24) weil [$_{\mathrm{VP}}$ er [$_{\mathrm{VP}}$ ihr [$_{\mathrm{VP}}$ einen Kuchen [$_{\mathrm{VP}}$ [$_{\mathrm{V}}$ buk]]]]]
  because he her a cake baked

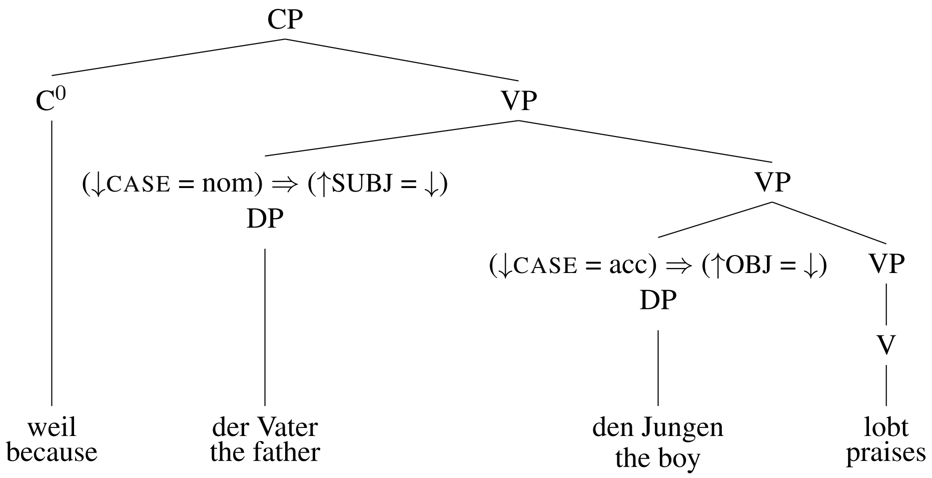


Figure 3: Correspondence between case and grammatical function according to Berman (2003, 37)

So the only option seems to be to introduce the benefactive at the rule that got the recursion going, namely the rule in (25), that projects the lexical verb to the VP level.

(25) VP $\rightarrow$ $\underset{\uparrow = \downarrow}{\text{(V)}}$

Introducing the benefactive at a rule that projects a lexical item to the VP to get some recursion going is almost a lexical approach (for differences see the discussion of (19) above). Note also that the argument above would apply to other constructions as well. So templates for several constructions may be added disjunctively to this projection. Again not much of the original constructional proposal would be left.

Berman (2003) develops an analysis in which the grammatical functions are assigned via implicational constraints that infer the grammatical function from the case of an NP/DP. Figure 3, which is a simplified version of the figure she discusses on p. 37, shows the implicational constraints and that they are attached to certain phrase structure positions. See Bresnan et al., 2015, 113 for the general mechanism. In the case at hand the presence of a dative could be used to infer the grammatical function of a benefactive argument. However, the situation is not as simple as it first may appear. In examples like (26) we have a so-called dative passive. The dative object is promoted to subject and hence gets nominative.

(26) Der Mann bekam einen Kuchen gebacken.
the.NOM man got a.ACC cake baked

This can be accounted for straightforwardly in a lexical approach in which the dative is a dependent of *backen*. Either a lexical rule or the auxiliary verb takes care of the fact that the dative argument has to be realized as nominative in dative-passive

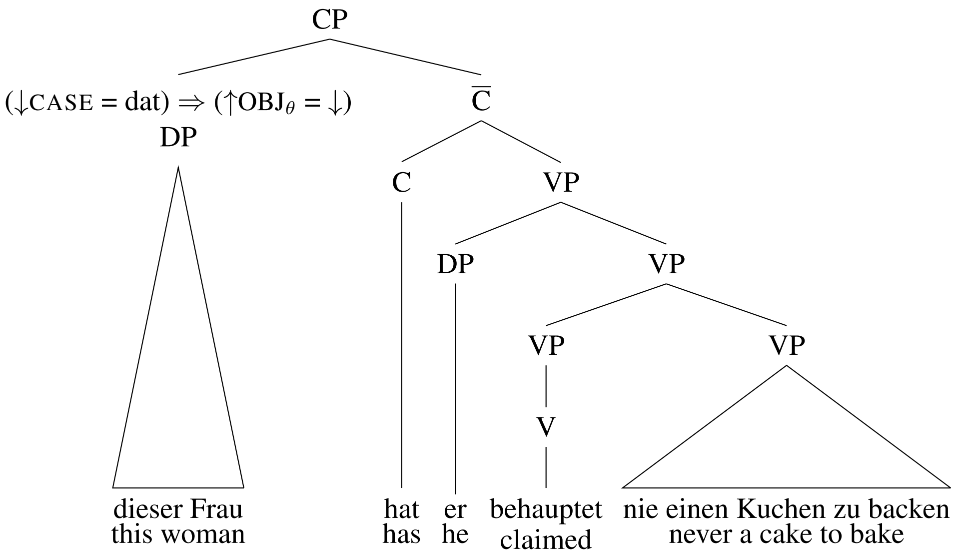


Figure 4: Benefactive construction with fronted dative. Assigment of grammatical functions based on case would exclude such structures

constructions like (26) (see Müller, 2002, Section 3.2.3 for details of an auxiliary-based approach in HPSG). A phrasal approach that wants to assign grammatical functions based on dative case is lost though.

Note also that the dative can be fronted over clause boundaries:

(27) Dieser Frau hat er behauptet, nie einen Kuchen zu backen.
this.DAT woman has he.NOM claimed never a.ACC cake to bake
'He claimed that he never bakes this woman a cake.'

A simple model that adds an OBJ$_\theta$ to the f-structure in which a dative appears would fail here since the OBJ$_\theta$ belongs into the f-structure of *backen* 'bake' rather than into the f-structure of *behauptet* 'claimed' (see also Müller, 2016b, 228). This is due to the fact that the benefactive is extracted and not realized within the VP with the appropriate f-structure (*nie einen Kuchen zu backen* 'never a cake to bake'). The situation is depicted in Figure 4. There seem to be two solutions to the problem: Firstly, one could assume a dative trace in the *backen* VP as is suggested by Berman (2003) for long-distance movement. The assumption of empty elements is usually avoided in LFG (Kaplan & Zaenen, 1989, Dalrymple, 2001, Chapter 14.3, Dalrymple et al., 2001) and in any case empty elements would not be compatible with the view that the phrasal approach restricts extraction since it specifies which daughters have to be present. Secondly one could assume functional uncertainty (Kaplan & Zaenen, 1989) to find the right f-structure. For instance one could say that a dative can be an OBJ$_\theta$ of the local f-structure or an f-structure somewhere on the path of COMPS or XCOMP:

(28) (↓CASE = dat) ⇒ (↑(COMP|XCOMP)* OBJ$_\theta$ = ↓)

This means that benefactive arguments have to "know" where they could come from. This is an unwanted consequence since the treatment of nonlocal dependencies should be independent of the benefactive construction.

Furthermore, if all datives could be associated with deeply embedded f-structures, we would predict that (29) is ambiguous:

(29) Dieser Frau hat er dem Mann versprochen, nie einen Kuchen zu backen.
this.DAT woman has he the.DAT man promised never a.ACC cake to bake
'He promised the man to never bake this woman a cake.'
Predicted: 'He promised the woman to never bake the man a cake.'

The dative *dem Mann* could reach down into the VP f-structure in the same way as the dative NP *dieser Frau*, but *dem Mann* is unambiguously an object of *versprochen*. *dieser Frau* is in initial position and it is the c-structure position (SpecCP) that is connected via functional uncertainty to the deeply embedded VP. If both dative NPs had the potential to fill a grammatical function in embedded f-structures we would expect the ambiguity. Assuming Inside-Out functional uncertainty as suggested by Nordlinger (1998) would not make a difference here.

Note also that benefactive datives appear in adjectival environments as in (30):

(30) a. der seiner Frau einen Kuchen backende Mann
the his.DAT wife a.ACC cake bakeing man
'the man who is baking a cake for his wife'

b. der einen Kuchen seiner Frau backende Mann
the a.ACC cake his.DAT wife bakeing man
'the man who is baking a cake for his wife'

The examples in (30) show that the arguments of *backende* may be scrambled, as is common in verbal environments. Like German verbal projections, adjectival projections with adjectival participles can contain adjuncts at various places. (31) provides two examples:

(31) a. der jetzt seiner Frau einen Kuchen backende Mann
the now his.DAT wife a.ACC cake bakeing man
'the man who is baking a cake for his wife now'

b. der seiner Frau jetzt einen Kuchen backende Mann
the his.DAT wife now a.ACC cake bakeing man
'the man who is baking a cake for his wife now'

In order to account for these datives one would have to assume that the adjective to AP rule that would be parallel to (25) introduces the benefactive. Hence the benefactive template would be introduced in several c-structure rules. In comparison the lexical approach assumes that the benefactive argument is introduced as an argument of the verb. The derivation of the adjectival form just takes over the arguments of the verb (Müller, 2002, 160).

A reviewer of HeadLex 2016 suggested that one could assume VP structures including a benefactive for German as well. While many researchers working in GB, LFG and HPSG assume binary branching structures for German (Haider, 1993; Fanselow, 2001; Berman, 2003; Kiss, 1995), there are indeed LFG accounts that assumes a flat VP for German (Zaenen & Kaplan, 2002). Zaenen & Kaplan's rule has the form in (32):

(32)
$$\text{S|VP} \rightarrow \begin{array}{ccc} \text{NP*} & (\text{V}') & (\text{S|VP}) \\ (\uparrow \text{COMP* NGF}) = \downarrow & \uparrow = \downarrow & (\uparrow \text{XCOMP * COMP}) = \downarrow \end{array}$$

As such the rule looks quite different from the benefactive rule in (7). Note also that this rule could account for benefactives but it does not account for the fact that adjuncts can appear anywhere between arguments in German. This is something that is accounted for by approaches that assume binary branching structures. If one augmented (32) by adjuncts the rule would be even more different from what was assumed for the benefactive rule in English.

Furthermore, Zaenen & Kaplan develop a theory that assumes partial VPs.The partial VPs in (33) are parallel to the VPs in approaches with binary branching. Any LFG of German would have to admit such partial VPs since German allows for partial VP fronting:

(33) a. Backen würde er seiner Frau solchen Kuchen niemals.
bake would he his wife such cake never
'He would never bake such cakes for his wife.'

b. [Seiner Frau backen] würde er solche Kuchen niemals.
his.DAT wife bake would he.NOM such cakes never

c. [Solche Kuchen backen] würde er seiner Frau niemals.
such cakes bake would he.NOM his.DAT wife never

In (33a) the verb is fronted without any argument, in (33b) the verb is realized together with the benefactive but the accusative object is realized outside the verbal projection and in (33c) the accusative is realized together with *backen* but the benefactive stays behind. Hence the idea that the benefactive is introduced in a special phrase structural configuration together with a verb and all other objects would not work for German. See Nerbonne (1986) and Johnson (1986), who introduced lexical valence representations in a Categorial Grammar style into GPSG since there was no way to make the phrasal GPSG approach compatible with German PVP data. See also Müller & Wechsler, 2014, Section 4.3.

Note that I do not claim here that LFG has any problems with partial verb phrase fronting. Zaenen & Kaplan (2002) show that partial verb phrase fronting can be modeled in LFG. What I hope to have shown is that approaches that assume that benefactives are solely licensed in structures like the one in Figure 1 are inappropriate for German and hence do not capture cross-linguistic generalizations.

Concluding this section I must say that the proposal for English in its final form in Asudeh et al., 2014 and its extension to German do not have anything to do with the original constructional proposal envisaged by Toivonen (2013) in which a VP consisting of a verb, the benefactive NP and a further NP is assumed. If one wants to use a similar approach to German one would have to weaken the constructional position and admit that the benefactive may be introduced at several places in the syntax (e.g., at verbs and adjectives). Thirdly, under the assumption of binary branching structures, a unary branching syntactic rule is applying to a lexical item and hence is very similar to a lexical rule and fourthly the analysis does not capture cross-linguistic commonalities of the construction. In a lexical rule-based approach as the one that was suggested by Briscoe & Copestake (1999, Section 5) in the framework of HPSG, and Toivonen (2013) in LFG a benefactive argument is added to certain verbs and the lexical rule is parallel in all languages that have this phenomenon. The respective languages just differ in the way the arguments are realized with respect to their heads. In languages that have adjectival participles, these are derived from the respective verbal stems. The morphological rule is the same independent of benefactive arguments and the syntactic rules for adjectival phrases do not have to mention benefactive arguments.

5 Capturing the constraints on benefactives for speakers with restrictions

Toivonen (2013) stated that a construction-specific phrase structure rule may be the best way to capture the constraints in restrictive idiolects of English. As I showed the restrictions are too strong even for speakers with restrictions on the benefactive construction since extraction of the secondary objects is possible. However, the passive is excluded for some speakers. One easy way to rule out passivization is to explicitly state the case of the benefactive element in the lexical rule. If passive is seen as a promotion of an element with structural case to a position (in a tree or an underlying structure like HPSG's ARG-ST) and subsequent assignment of nominative, passive would be excluded.

An alternative would be to assume that the passive lexical rule for English requires the input to be of type *transitive-verb-lexeme* (Kay, Sag & Flickinger, 2015) and that the lexical rule that licenses benefactive arguments licenses pseudo-transitive lexemes. Pseudo-transitives do not qualify as input to the passive lexical rule and hence passives would be excluded. Both approaches would be just stipulations (as is the phrasal approach) but I prefer the case-based approach since the approach to passive that is developed in Müller & Ørsnes, 2013; Müller, 2016a works for both English and German and does not make any reference to transitivity. (German allows for impersonal passives)

The extraction of primary objects is marked for all verbs that take two objects irrespective of the semantic role. For some speakers the extraction of benefactives is worse than the extraction of other primary objects. If one wanted to block extraction via a hard constraint rather than assuming that performance factors play a role here (Langendoen et al., 1973), one could state that the SLASH value of the primary object is the empty list (Müller, 1999, 98) or – if extraction out of the primary object is to be permitted – different from the LOCAL value of the primary object. Because of this specification a trace would be incompatible with this object. The same applies to an appropriately specified lexical rule for argument extraction or a process like SLASH amalgamation as suggested by (Bouma et al., 2001).

Note that this approach also predicts that constraints on extraction and passivization in coordinated structures affect the result of coordination. The reason is that the constraints on the selected arguments are identified in symmetric coordinations. Hence the SLASH constraints and the case constraints on the benefactive argument are effective on the mother node of verb coordinations as well.

6 Conclusion

This paper argues that fixed-arity rules for the resultative construction and the benefactive construction are empirically inadequate if one does not want to assume traces. Furthermore it is shown that the introduction of arguments at the c-structure results in missing generalizations in the grammar of a single language and that cross-linguistic generalizations are missed in general since c-structures may differ wildly and in some languages they may be less suited for the attachment of templates that introduce arguments.

Because of all these problems I suggest returning to the lexicalist approaches, that is, to analyses that assume that arguments are introduced by lexical means like lexical rules. Examples of such analyses are the lexical analyses of resultatives by Simpson (1983) in LFG and of Wechsler (1997); Wechsler & Noh (2001) and Müller (2002) in HPSG and the lexical analysis of benefactives by Briscoe & Copestake (1999) in HPSG and by Cook (2006) and Toivonen (2013) in LFG.

It may be the case that phrasal constructions are needed in other areas of grammars in other languages (see for instance Butt, 1995 on complex predicates in Urdu), but nothing follows from this for grammars for German and English. In general one should aim for assuming the same descriptive tools if they are appropriate for a given set of languages and supported by language-internal considerations. So, if both German and English allow for a lexical analysis of resultatives and benefactives, an analysis that covers the facts in both languages is to be preferred. If Urdu differs from German and English, this does not necessarily mean that these differences are reflected in the grammars of English and German. See Müller, 2015 on cross-linguistic generalizations without the assumption of Universal Grammar.

The full paper (Müller, 2016c) also contains a discussion of the syntax of resultative constructions, which are argued to form a predicate complex in German but not in English. This is a difference in syntactic structure, which is unproblematic for lexical accounts but results in missing generalizations in phrasal accounts.

The full paper also develops a lexical account of German and English resultatives and benefactives in the framework of HPSG and shows how this account captures the commonalities between German and English despite the superficial dissimilarities between the two languages.

References

Alsina, Alex. 1996. Resultatives: A joint operation of semantic and syntactic structures. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '96 conference, Rank Xerox, Grenoble*, Stanford, CA: CSLI Publications.

Asudeh, Ash, Mary Dalrymple & Ida Toivonen. 2008. Constructions with lexical integrity: Templates as the lexicon-syntax interface. In Butt & King (2008).

Asudeh, Ash, Mary Dalrymple & Ida Toivonen. 2013. Constructions with lexical integrity. *Journal of Language Modelling* 1(1). 1–54.

Asudeh, Ash, Gianluca Giorgolo & Ida Toivonen. 2014. Meaning and valency. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG 2014 conference*, 68–88. Stanford, CA: CSLI Publications.

Asudeh, Ash & Ida Toivonen. 2014. *With* lexical integrity. *Theoretical Linguistics* 40(1–2). 175–186.

Bergen, Benjamin K. & Nancy Chang. 2005. Embodied Construction Grammar in simulation-based language understanding. In Jan-Ola Östman & Mirjam Fried (eds.), *Construction Grammars*, 147–190. Amsterdam: Benjamins.

Berman, Judith. 1996. Eine LFG-Grammatik des Deutschen. In *Deutsche und französische Syntax im Formalismus der LFG*, 11–96. Tübingen: Max Niemeyer Verlag.

Berman, Judith. 2003. *Clausal syntax of German*. Stanford, CA: CSLI Publications.

Bouma, Gosse, Robert Malouf & Ivan A. Sag. 2001. Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory* 19(1). 1–65.

Bresnan, Joan. 1982. The passive in lexical theory. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 3–86. Cambridge, MA/London: MIT Press.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2015. *Lexical-functional syntax*. Oxford, UK/Cambridge, USA: Blackwell Publishing Ltd 2nd edn.

Bresnan, Joan & Jonni M. Kanerva. 1989. Locative inversion in Chichewa: A case study of factorization in grammar. *Linguistic Inquiry* 20(1). 1–50.

Briscoe, Ted J. & Ann Copestake. 1999. Lexical rules in constraint-based grammar. *Computational Linguistics* 25(4). 487–526.

Butt, Miriam. 1995. *The structure of complex predicates in Urdu*. Stanford, CA: CSLI Publications.

Butt, Miriam & Tracy Holloway King (eds.). 2008. *Proceedings of the LFG 2008 conference*. Stanford, CA: CSLI Publications.

Cappelle, Bert. 2006. Particle placement and the case for "allostructions". *Constructions online* 1(7). 1–28.

Carrier, Jill & Janet H. Randall. 1992. The argument structure and syntactic structure of resultatives. *Linguistic Inquiry* 23(2). 173–234.

Choi, Hye-Won. 1999. *Optimizing structure in scrambling: Scrambling and information structure*. Stanford, CA: CSLI Publications.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague/Paris: Mouton.

Christie, Elizabeth. 2010. Using templates to account for English resultatives. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG 2010 conference*, 155–164. Stanford, CA: CSLI Publications.

Cook, Philippa Helen. 2006. The datives that aren't born equal: Beneficiaries and the dative passive. In Hole, Meinunger & Abraham (eds.), *Datives and similar cases: Between argument structure and event structure*, 141–184. Amsterdam: Benjamins.

Culicover, Peter W. & Ray S. Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford Univer-

sity Press.
Dalrymple, Mary. 2001. *Lexical Functional Grammar*. New York: Academic Press.
Dalrymple, Mary, Ronald M. Kaplan & Tracy Holloway King. 2001. Weak crossover and the absence of traces. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG 2001 conference*, 66–82. Stanford, CA: CSLI Publications.
Fanselow, Gisbert. 2001. Features, θ-roles, and free constituent order. *Linguistic Inquiry* 32(3). 405–437.
Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago/London: The University of Chicago Press.
Goldberg, Adele E. 2004. But do we need Universal Grammar? Comment on Lidz et al. (2003). *Cognition* 94. 77–84.
Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford, New York: Oxford University Press.
Goldberg, Adele E. 2013. Explanation and Constructions: Response to Adger. *Mind and Language* 28(4). 479–491.
Goldberg, Adele E. & Ray S. Jackendoff. 2004. The English resultative as a family of Constructions. *Language* 80(3). 532–568.
Haider, Hubert. 1993. *Deutsche Syntax – generativ: Vorstudien zur Theorie einer projektiven Grammatik*. Tübingen: Gunter Narr Verlag.
Haugereid, Petter. 2007. Decomposed phrasal constructions. In Stefan Müller (ed.), *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*, 120–129. Stanford, CA: CSLI Publications.
Haugereid, Petter. 2009. *Phrasal subconstructions: A Constructionalist grammar design, exemplified with Norwegian and English*: Norwegian University of Science and Technology dissertation.
Hudson, Richard. 1992. So-called 'double objects' and grammatical relations. *Language* 68(2). 251–276.
Johnson, Mark. 1986. A GPSG account of VP structure in German. *Linguistics* 24(5). 871–882.
Kaplan, Ronald M. & Annie Zaenen. 1989. Long-distance dependencies, constituent structure and functional uncertainty. In Mark R. Baltin & Anthony S. Kroch (eds.), *Alternative conceptions of phrase structure*, 17–42. Chicago/London: The University of Chicago Press.
Kaufmann, Ingrid & Dieter Wunderlich. 1998. Cross-linguistic patterns of resultatives. SFB 282: Theorie des Lexikons 109 Heinrich Heine Uni/BUGH Düsseldorf.
Kay, Paul. 2005. Argument structure constructions and the argument-adjunct distinction. In Mirjam Fried & Hans C. Boas (eds.), *Grammatical constructions: Back to the roots*, 71–98. Amsterdam: John Benjamins Publishing Co.
Kay, Paul, Ivan A. Sag & Daniel P. Flickinger. 2015. A lexical theory of phrasal idioms. Ms. CSLI Stanford.
Kibort, Anna. 2008. On the syntax of ditransitive constructions. In Butt & King (2008) 312–332.
Kiss, Tibor. 1995. *Infinite Komplementation: Neue Studien zum deutschen Verbum infinitum*. Tübingen: Max Niemeyer Verlag.
Langendoen, D. Terence, Nancy Kalish-Landon & John Dore. 1973. Dative questions: A study in the relation of acceptability to grammaticality of an English sentence type. *Cognition* 4(2). 451–478.
Müller, Stefan. 1999. *Deutsche Syntax deklarativ: Head-Driven Phrase Structure Grammar für das Deutsche*. Tübingen: Max Niemeyer Verlag.
Müller, Stefan. 2002. *Complex predicates: Verbal complexes, resultative constructions,*

and particle verbs in German. Stanford, CA: CSLI Publications.
Müller, Stefan. 2006. Phrasal or lexical Constructions? *Language* 82(4). 850–883.
Müller, Stefan. 2013. Unifying everything: Some remarks on Simpler Syntax, Construction Grammar, Minimalism and HPSG. *Language* 89(4). 920–950.
Müller, Stefan. 2015. The CoreGram project: Theoretical linguistics, theory development and verification. *Journal of Language Modelling* 3(1). 21–86.
Müller, Stefan. 2016a. Germanic syntax. Ms. Humboldt Universität zu Berlin, to be submitted to Language Science Press. http://hpsg.fu-berlin.de/~stefan/Pub/germanic.html.
Müller, Stefan. 2016b. *Grammatical theory: From Transformational Grammar to constraint-based approaches*. Berlin: Language Science Press.
Müller, Stefan. 2016c. Phrasal constructions, derivational morphology, constituent structure and (cross-linguistic) generalizations: A discussion of template-based phrasal LFG approaches. Ms. Humboldt Universität zu Berlin. http://hpsg.fu-berlin.de/~stefan/Pub/lexical-integrity.html.
Müller, Stefan. To appear. HPSG, SBCG, and FCG: Commonalities and differences. *Constructions and Frames* .
Müller, Stefan & Bjarne Ørsnes. 2013. Passive in Danish, English, and German. In Stefan Müller (ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar*, 140–160. Stanford, CA: CSLI Publications.
Müller, Stefan & Stephen Mark Wechsler. 2014. Lexical approaches to argument structure. *Theoretical Linguistics* 40(1–2). 1–76.
Nerbonne, John. 1986. 'Phantoms' and German fronting: Poltergeist constituents? *Linguistics* 24(5). 857–870.
Nordlinger, Rachel. 1998. *Constructive case: Evidence from Australia*. Stanford, CA: CSLI Publications.
Simpson, Jane. 1983. Resultatives. In Lori S. Levin, Malka Rappaport & Annie Zaenen (eds.), *Papers in Lexical Functional Grammar*, 143–157. Indiana: Indiana University Linguistics Club Reprint: Simpson, 2005.
Simpson, Jane. 2005. Resultatives. In Miriam Butt & Tracy Holloway King (eds.), *Lexical semantics in LFG*, 149–161. Stanford, CA: CSLI Publications.
Toivonen, Ida. 2013. English benefactive NPs. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG 2013 conference*, 503–523. Stanford, CA: CSLI Publications.
Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
van Trijp, Remi. 2011. A design pattern for argument structure constructions. In Luc Steels (ed.), *Design patterns in Fluid Construction Grammar*, 115–145. Amsterdam: John Benjamins Publishing Co.
Verspoor, Cornelia Maria. 1997. *Contextually-dependent lexical semantics*: University of Edinburgh dissertation.
Wechsler, Stephen Mark. 1997. Resultative predicates and control. In *Texas Linguistic Forum 38: The syntax and semantics of predication: Proceedings of the 1997 Texas Linguistics Society Conference*, 307–321.
Wechsler, Stephen Mark & Bokyung Noh. 2001. On resultative predicates and clauses: Parallels between Korean and English. *Language Sciences* 23(4). 391–423.
Wunderlich, Dieter. 1997. Argument extension by lexical adjunction. *Journal of Semantics* 14(2). 95–142.
Zaenen, Annie & Ron Kaplan. 2002. Subsumption and equality: German partial fronting in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG 2002 conference*, Stanford, CA: CSLI Publications.

Syntactic, semantic and information structures of floating quantifiers

Ryo Otoguro
Waseda University

Liselotte Snijders
Waseda University

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 478–498

Keywords: quantifier floating, information structure, English, Dutch, Japanese

Otoguro, Ryo & Liselotte Snijders. 2016. Syntactic, semantic and information structures of floating quantifiers. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 478–498. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.25.

Abstract

Quantifiers canonically attach to nouns or noun phrases as modifiers to specify the amount or number of the entity expressed by the noun. However, it has been observed that quantifiers can be positioned outside of the noun phrase. These so-called floating quantifiers (FQs) exhibit intriguing syntactic and semantic characteristics. On the one hand, they appear to have a close relationship with a noun; semantically they quantify a noun in the same way as non-floating quantifiers, and quite often they exhibit agreement with the noun. On the other hand, their phrase structure distribution is very similar to that of VP-adverbs. In this paper, we argue that the distribution of FQs is constrained not purely by syntax, but also by information structure. We show that FQs play a focus role whereas modified nouns are reference-oriented topic expressions. Building upon Dalrymple & Nikolaeva's (2011) recent proposal, we formulate the interaction between syntactic, semantic and information structure features of FQs within LFG's projection architecture.

1 Introduction

As observed in the generative literature since its early days, quantifiers that modify a noun can appear not only inside of the noun phrase (NP) but also outside of it. Since nominal modifiers are canonically located at NP-internal positions, those quantifiers appearing outside of the NP are often called "floating" quantifiers (FQs). FQs can be observed in a wide variety of languages, although there are restrictions on phrase structure positions available for FQs and lexical items that can float. (1) exemplifies FQs in English, French, German and Japanese respectively.

(1) a. The students have *all* finished the assignment.

b. Elles sont *toutes* allées à la plage.
they.F are all.F.PL gone.F.PL to the beach
'They all went to the beach.' (French)

c. Diesen Studenten habe ich gestern *allen* geschmeichelt.
these.DAT.PL students have I yesterday all.DAT.PL flattered
'I flattered all of these students yesterday.' (German)
(Bobaljik, 2003, 107–9)

d. kodomo-tati wa *minna* eiga o tanosinda.
children-PL TOPIC all movie ACC enjoy.PAST
'The children all enjoyed the movie.' (Japanese)

In English and French examples (1a) and (1b), *all* and *toutes* appear between the finite auxiliary and the non-finite lexical verb although they semantically modify

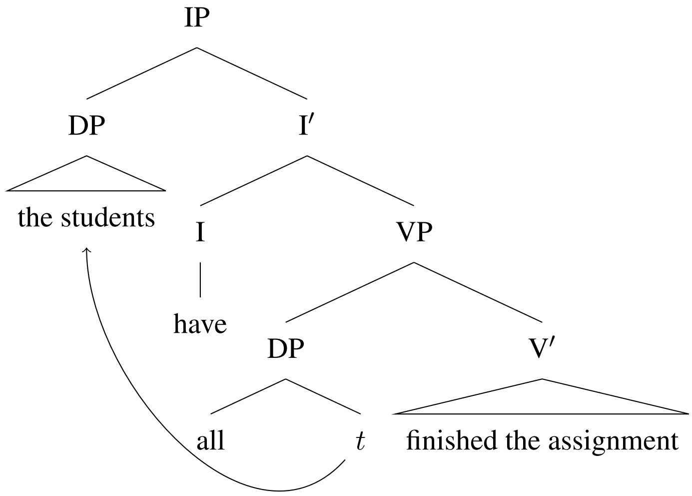
†We thank Anne Abeillé and Mary Dalrymple for their comments and suggestions. We are also indebted to two reviewers for their helpful comments. This work is supported by Grant-in-Aid for Encouragement of Scientists, Japan Society for the Promotion of Science (Grant No. 15F15737).

the subject nouns. In the German example (1c), the quantifier *allen* is located immediately before the clause final non-finite verb and modifies the object noun *Studenten*. Finally in Japanese, the quantifier *minna* appears between the clause-initial topic and the object NP as shown in (1d).

2 Previous analyses

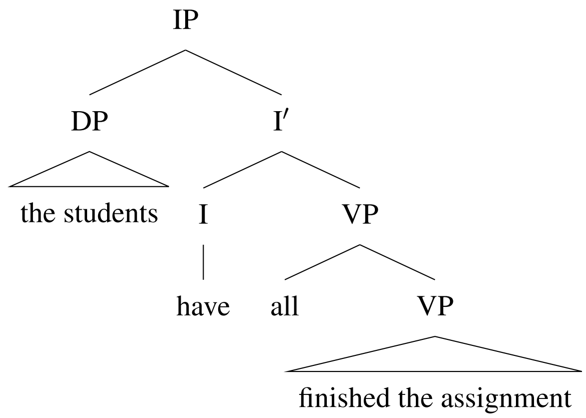
One proposal often found in a derivational approach to syntax is the so-called stranding analysis (Sportiche, 1988; Shlonsky, 1991). In this type of analysis, a quantifier is originally located in an NP (or DP) and the modified NP moves to a higher projection leaving the quantifier behind. For instance, (2) shows that *all the students* originally appears in Spec,VP and *the students* moves to Spec,IP.

(2)



Another approach is often referred to as the VP-modifier analysis, in which a quantifier is essentially treated as a VP adverb and adjoined to VP as illustrated in (3) (Dowty & Brodie, 1984; Baltin, 1982; Bobaljik, 2003; Kim & Kim, 2009). In this approach, the quantifier and the noun do not form a syntactic constituent from the beginning.

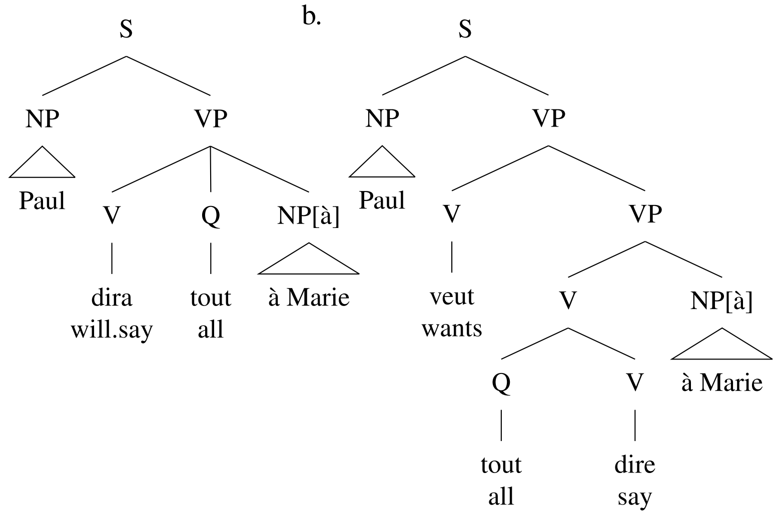
(3)



These two types of approaches differ with respect to the syntactic association between a quantifier and a noun. The stranding analysis tries to capture their relation directly in syntax, while the VP-modifier analysis regards FQs as one type of VP-adverb and their relation to a quantified noun is not formulated in the syntax.

Abeillé & Godard (1998) take a different view and propose a complement and adjunct analysis in Head-driven Phrase Structure Grammar (HPSG) for French quantifier floating. In their approach, a post-verbal quantifier is treated as a complement of the verb as in (4a), while a pre-verbal quantifier is treated as a lexical-level adjunction, i.e. V^0-adjunction, as in (4b).

(4) a. b.



2.1 Against a stranding analysis

One of the counter arguments against the stranding analysis is that an FQ does not always form a constituent with an NP at the NP-internal position. In French, for example, *chacun* 'each' can appear outside of the modified NP as in (5a), while it cannot form a constituent with a head noun as in (5b). Similarly, in English, although *all* cannot appear inside of an NP consisting of co-ordinated nouns as shown in (6b), it is perfectly acceptable as an FQ as in (6a).

(5) a. Ces enfants ont *chacun* lu un livre différent.
these children have each read a book different
'These children have each read a different book.'

b. **Chacun* ces enfants a lu un livre différent.
each these children has read a book different
'Each of these children has read a different book.' (French)
(Bobaljik, 2003, 123–4)

(6) a. John, Bill and Tom *all* came to the class.

b. **All* John, Bill and Tom came to the class.

In addition, some languages have different lexical items for NP-internal and NP-external quantifiers. In Dutch, for instance, *alle* is used in an NP-internal position whereas *allemaal* is used as an FQ as shown in (7). In the same vein, Mandarin Chinese has *suo you* as an NP-internal quantifier and *dou* as an NP-external one as shown in (8).

(7) a. *Alle* toeristen zullen Boston bezoeken.
all tourists will Boston visit
'All tourists will visit Boston.'

b. De toeristen zullen *allemaal* Boston bezoeken.
the tourists will all Boston visit
'The tourists will all visit Boston.' (Dutch)

(8) a. *suo you* de ren zou le
all PRT people left ASP
'All the people have left.'

b. ren *dou* zou le
people all left ASP
'The people have all left.' (Mandarin Chinese)
(Dowty & Brodie, 1984, 82)

2.2 Issues

The above data strongly suggests that there is a dissociation between floating and non-floating quantifiers, namely an FQ is unlikely to be formed by moving a modified NP. The VP-modifier analysis, on the other hand, gives a straightforward account for them. Since an FQ is treated as a VP-adjunct in that approach, it would not be surprising if FQs are distinct from NP-internal quantifiers and are exclusively used for NP-external positions.

However, there are some issues to be resolved even if we assume that FQs are VP-adverbs. Firstly, FQs can appear at non-VP modifier positions in specific constructions, such as ditransitive clauses and secondary predicate constructions. In English, for instance, an FQ that quantifies the object NP cannot appear after that NP as shown in (9). However, in ditransitive constructions like (10a, b) and secondary predicate constructions like (10c), an FQ can occur after the object NP position. A similar pattern is found in Dutch as shown in (11) and many other languages.

(9) a. *The teacher scolded the students *all*.
b. *Tom met the girls *all* in his office.

(10) a. I gave the kids *each* a quarter.
b. Mary put the books *all/both/each* (back) on the proper shelf.
c. We consider the Joneses *both* unbearably pompous.
(Maling, 1976, 712, 715)

(11) a. Marie sloeg de mannen *allebei* op het gezicht.
M. hit the men both in the face
'Marie hit the men both in the face.'
b. Ik vind de talen *allemaal* mooi.
I find the languages all beautiful
'I find the languages all beautiful.' (Dutch)

Another issue is agreement between an FQ and a noun. As (1b, c) illustrates, languages like French and German exhibit long distance agreement. In (1b), for example, *toutes* 'all.F.PL' agrees with the third person feminine plural pronoun *elles*. Hence, an FQ and a quantified noun hold a relationship at some level, so that the modified noun controls the agreement and triggers the inflection of the FQ.

3 Proposal

3.1 Topic–comment structure

To resolve the issues summarised in the previous section, we propose an alternative analysis based on LFG's projection architecture. The essence of our proposal is that an FQ is licensed by information structure partitioning. More specifically, the NP quantified by an FQ is a 'reference-oriented topic expression' (Lambrecht, 1994; Neeleman & van de Koot, 2008; Neeleman & Vermeulen, 2012) and the FQ functions as a focus (cf. Kuno & Takami, 2003; Rochman, 2010). It has been pointed out that the default position of a reference-oriented topic expression is sentence-initial, and the following part functions as a comment that consists of a focus and a background as schematically represented in (12). The outermost structure is construed by the topic–comment partitioning where the Kleene star represents multiple occurrences of topic. In the comment, the structure is divided into a focus and a background.

(12) topic* [$_{\text{comment}}$ focus [$_{\text{background}}$ ...]]
(Neeleman & van de Koot, 2008, 146)

Considering the relation between information structure and phrase structure, one of the most isomorphic syntactic configurations corresponding to the information structure (12) would be (13), in which a quantifier phrase (QP) is adjoined to VP and an NP is in the clause-initial position. Examples of this information structure partitioning and the corresponding phrase structure configuration are given in (14) for English and in (15) for Dutch.

(13) NP [$_{\text{VP}}$ QP [$_{\text{VP}}$ ...]]
topic focus background

(14) a. [$_{\text{topic}}$ The students] [$_{\text{comment}}$ have [$_{\text{focus}}$ *all*] [$_{\text{background}}$ finished the assignment]]
b. [$_{\text{NP}}$ The students] have [$_{\text{VP}}$ [$_{\text{QP}}$ *all*] [$_{\text{VP}}$ finished the assignment]]

(15) a. [$_{\text{topic}}$ De toeristen] [$_{\text{comment}}$ zullen [$_{\text{focus}}$ *allemaal*] [$_{\text{background}}$ Boston bezoeken]]
b. [$_{\text{NP}}$ De toeristen] zullen [$_{\text{VP}}$ [$_{\text{QP}}$ *allemaal*] [$_{\text{VP}}$ Boston bezoeken]]
the tourists will all Boston visit
'The tourists will all visit Boston.' (Dutch)

Another isomorphic syntactic configuration can be found VP-internally as represented in (16). The examples following this pattern are ditransitive constructions and secondary predicate constructions as shown in (17) and (18).

(16) ... [$_{\text{VP}}$ V NP QP XP]
topic focus background

(17) a. I gave [$_{\text{topic}}$ the kids] [$_{\text{comment}}$ [$_{\text{focus}}$ *each*] [$_{\text{background}}$ a quarter]].
b. I [$_{\text{VP}}$ gave [$_{\text{NP}}$ the kids] [$_{\text{QP}}$ each] [$_{\text{NP}}$ a quarter]]

(18) a. Ik vind [$_{\text{topic}}$ de talen] [$_{\text{comment}}$ [$_{\text{focus}}$ *allemaal*] [$_{\text{background}}$ mooi]]
b. Ik vind [$_{\text{VP}}$ [$_{\text{NP}}$ de talen] [$_{\text{QP}}$ *allemaal*] [$_{\text{AP}}$ mooi]]
I find the languages all beautiful
'I find the languages all beautiful.' (Dutch)

According to our proposal, an FQ is not licensed purely by syntax as a VP-adjunct, but by information structure as a focus. We argue that the reason why an FQ frequently appears as a VP-adjunct is that the topic–comment information partitioning can be transparently encoded by such a syntactic configuration as illustrated in (13). Similarly, the reason why a quantifier can float in constructions involving ditransitive verbs and secondary predicates is that the topic–comment structure can be construed VP-internally in those constructions as shown in (16).

3.2 Topic and focus properties in quantifier floating

One piece of evidence for the reference-oriented topic status of a quantified noun involved in an FQ construction comes from the fact that an indefinite noun cannot appear with an FQ as shown in (19).

(19) a. The children *all* visited London.
b. #Children *all* visited London.

If we assume that the subject NP in (19) which is modified by the quantifier must be a reference-oriented topic expression, the unacceptability of (19b) can be straightforwardly accounted for. A reference-oriented topic expression by default refers to an entity that is salient in the discourse. An indefinite noun normally lacks those properties, so *children* in (19b) cannot be a reference-oriented expression and accordingly cannot occur with an FQ under our assumptions.

There are some cases where an indefinite noun can occur with an FQ as shown in Dutch example (20a). In that case, however, the sentence must be a statement of generic properties about the indefinite noun, so in (20a), *Kinderen* 'children' is a reference-oriented topic expression by referring to children in general, and the following comment is a statement about the generic property of children. Thus, if we replace *allemaal* 'all' with *allebei* 'both', the sentence will be unacceptable as shown in (20b).[1] This is because the statement cannot be interpreted as a generic property of children.

(20) a. Kinderen genieten *allemaal* van de film.
children enjoy all of the film
'Children all enjoy the film.'
b. *Kinderen genieten *allebei* van de film.
children enjoy both of the film
'Children both enjoy the film.' (Dutch)

The topic status of quantified nouns is also confirmed by Japanese data. As (21a, b) show, an FQ cannot quantify an NP with the dative particle *ni* or the ablative particle *kara* in the pre-verbal position. In Japanese, like many other verb final languages, elements appearing in the immediately pre-verbal position are given a focus role in the information structure unless the pre-verbal elements themselves are marked as not being focus. In our proposal, the noun quantified by an FQ must be a reference-oriented topic expression, so the unacceptability of (21a, b) is explained by incompatibility of their topic status with a default focus interpretation in the pre-verbal position.

(21) a.??Taroo ga Hanako o sinseki ni *minna* syookai sita.
T. NOM H. ACC relatives DAT all introduce do.PAST
'Taro introduced Hanako to all of his relatives.'

[1] We thank Frank Van Eynde for pointing out the unacceptability of this example.

b.??sono seizika ga kihukin o siensya kara *50-mei*
that politician NOM donation ACC supporter from 50-CL
atumeta.
collect.PAST
'That politician collected donations from 50 supporters.' (Japanese)

Interestingly, if those casemarked NPs are marked by the contrastive topic particle *wa*, the sentences will be acceptable as shown in (22a, b). Since the overt morphological marking of contrastive topic defocuses the casemarked NPs in the pre-verbal position, they can function as a topic and can be modified by a focus FQ.

(22) a. Taroo ga Hanako o sinseki ni wa *minna* syookai
T. NOM H. ACC relatives DAT TOPIC all introduce
sita.
do.PAST
'As for his relatives, Taro introduced Hanako to all of them.'

b. sono seizika ga kihukin o siensya kara wa *50-mei*
that politician NOM donation ACC supporter from TOPIC 50-CL
atumeta.
collect.PAST
'As for supporters, that politician collected donations from 50 of them.'

Finally, it is well-known that certain types of adverbs including manner adverbs that convey new information are by default given a focus interpretation (Kuno & Takami, 2003). Our proposal predicts that they cannot precede an FQ because the positioning of adverbs blocks topic–comment structure involving an FQ. This prediction is borne out. Kuno & Takami (2003) point out that Japanese manner adverbs like *geragera-to* 'loudly' cannot precede an FQ as shown in (23a). Note that the reverse order is acceptable as shown in (23b). A focus-bearing adverb is assigned a focus role when it is adjoined to VP, and any following elements are thereby given a background status. In (23a), the manner adverb is assigned a focus role and forces the FQ to be part of the background information structure role. However, since an FQ is required to be a focus, this information structure role assignment is not permitted. This violation does not happen in (23b) because manner adverbs, despite their default information focus role, can also play other roles including background. This ordering restriction does not arise with non-focus bearing locative adverbials as in (23b) or sentential adverbs as in (23c).

(23) a. *kodomo ga geragera-to *hutari* waratta.
child NOM loudly two.CL laughed
'Two children laughed loudly.'

b. kodomo ga *hutari* geragera-to waratta.
child NOM two.CL loudly laughed

'Two children laughed loudly.'

c. gakusei ga office ni *hutari* kita.
student NOM office to two.CL came
'Two students came to the office.' (Kuno & Takami, 2003, 283–4)

d. gakusei ga kinoo *hutari* kita.
student NOM yesterday two.CL came
'Two students came yesterday.' (Japanese)

Similarly in English and Spanish, an FQ must precede a manner adverb or a completive adverb to mark the VP-edge which forms a topic–comment structure as in (24a) and (25a). The intervention of focus-bearing adverbs prevents the FQ from forming this information structure partitioning, so (24b) and (25b) are unacceptable. The same order restriction is not observed between a sentential adverb and an FQ as shown in (24c, d).

(24) a. These thieves could *all* completely crack this safe in 5 minutes flat.
b. *These thieves could completely *all* crack this safe in 5 minutes flat.
c. The thieves have *all* certainly been apprehended.
d. The thieves have certainly *all* been apprehended.
(Bobaljik, 1995, 231–2)

(25) a. ?Los estudiantes entenderán *todos* completamente (ese problema).
the students will.understand all completely that problem
'The students will understand all completely that problem.'
b. *Los estudiantes entenderán completamente *todos* (ese problema).
the students will.understand completely all that problem
(Spanish; Bošković 2004, 686)

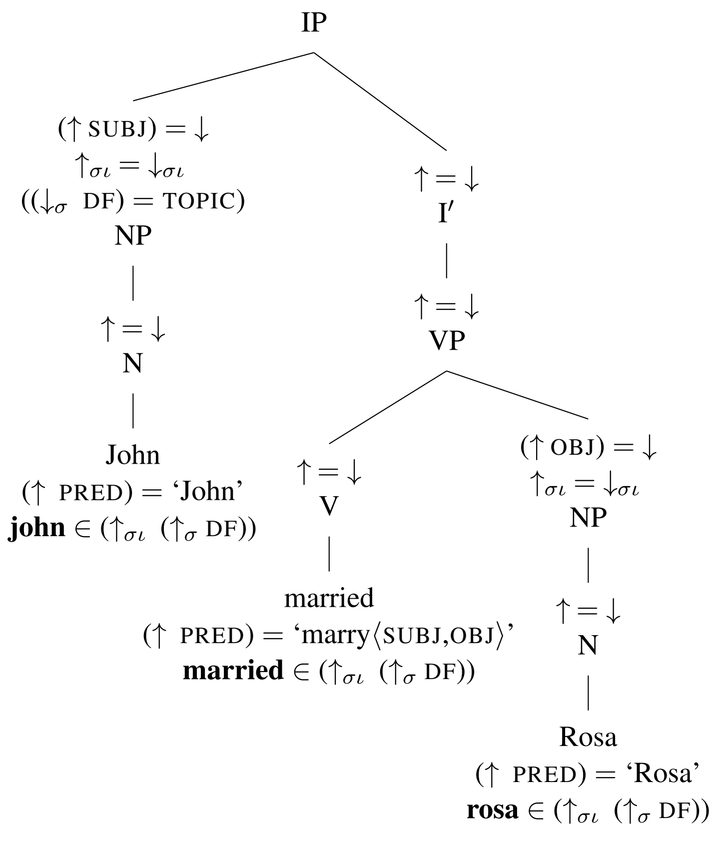
4 Analysis

In this section, we present an LFG analysis of FQs. Since information structure plays a crucial role in constraining the distribution of FQs, we adopt the standard LFG projection architecture, in which different types of linguistic information are encoded in distinct structures. Following Butt & King (1996, 2000) and Choi (1999), we assume that a sentence is partitioned into four discourse functions (DFs) in information structure: TOPIC, FOCUS, BACKGROUND and COMPLETIVE. Further, as formulated in Dalrymple & Nikolaeva (2011), DF is present in s(emantic)-structure and can be specified in various ways, such as by phrase-structure position, prosody or morphological marking. The specification of a value for the semantic structure feature DF determines the membership of the information structure roles. Dalrymple & Nikolaeva (2011) show correspondences between c-structure, f-structure, s-structure and information structure for sentence (26-A), in which *John* is a topic and *married Rosa* is a focus. These correspondences are shown in (27). The Spec,IP position is a subject position, but it is also a default topic position, so the value of DF is optionally determined by the c-structure configuration, i.e. $(\downarrow_\sigma\ \text{DF}) = \text{TOPIC}$. The discourse prominence feature and other linguistic features reinforce this TOPIC assignment. The boldface notation is an abbreviation for meaning constuctors, e.g., **john** is an abbreviation for *john*: $\uparrow_\sigma$. Since **john** $\in (\uparrow_{\sigma\iota}\ (\uparrow_\sigma\ \text{DF}))$ is specified in the lexical entry of *John*, the meaning constructor, **john**, becomes a member of TOPIC in the information structure, namely **john** $\in (\uparrow_{\sigma\iota}\ (\uparrow_\sigma\ \text{TOPIC}))$. Contrary to topic specification, the specification of FOCUS is determined not by c-structure annotation, but by pragmatic context as in (29), which makes the meaning constructors, **married** and **rosa**, become members of FOCUS in the information structure. Thus, the resultant information structure can be represented as in (30).

(26) Q: What did John do?

A: John married Rosa.
TOPIC FOCUS

(27)



(28)
$$m : \begin{bmatrix} \text{PRED} & \text{`marry}\langle\text{SUBJ,OBJ}\rangle\text{'} \\ \text{SUBJ} & s : \begin{bmatrix} \text{PRED} & \text{`John'} \end{bmatrix} \\ \text{OBJ} & o : \begin{bmatrix} \text{PRED} & \text{`Rosa'} \end{bmatrix} \end{bmatrix}$$

(29)
$$s_\sigma : \begin{bmatrix} \text{DF} & \text{TOPIC} \end{bmatrix}$$
$$m_\sigma : \begin{bmatrix} \text{DF} & \text{FOCUS} \end{bmatrix}$$
$$o_\sigma : \begin{bmatrix} \text{DF} & \text{FOCUS} \end{bmatrix}$$

(30)
$$m_{\sigma\iota} : \begin{bmatrix} \text{TOPIC} & \{\ \textbf{john}\ \} \\ \text{FOCUS} & \left\{ \begin{matrix} \textbf{married} \\ \textbf{rosa} \end{matrix} \right\} \end{bmatrix}$$

(cf. Dalrymple & Nikolaeva, 2011, 84–5)

4.1 VP-adjunct quantifiers

As discussed above, an FQ functions as a focus, so the DF specification of FOCUS is given in the c-structure positions associated with FQs. Firstly, as (13) shows, a QP can be adjoined to VP, so we propose a c-structure rule as in (31). Crucially, QP is annotated with $(\downarrow_\sigma\ \text{DF}) = \text{FOCUS}$. Further, the constituent following the QP becomes a background, so the annotation $(\downarrow_\sigma\ \text{DF}) = \text{BACKGROUND}$ is given to the adjoined VP.

(31)
$$\text{VP} \longrightarrow \begin{matrix} \text{QP} \\ \downarrow \in (\uparrow \text{ADJ}) \\ \uparrow_{\sigma\iota} = \downarrow_{\sigma\iota} \\ (\downarrow_\sigma\ \text{DF}) = \text{FOCUS} \end{matrix} \quad \begin{matrix} \text{VP} \\ \uparrow = \downarrow \\ \uparrow_{\sigma\iota} = \downarrow_{\sigma\iota} \\ (\downarrow_\sigma\ \text{DF}) = \text{BACKGROUND} \end{matrix}$$

Consider the Japanese example (1d), repeated here as (32).

(32) kodomo-tati wa *minna* eiga o tanosinda.
children-PL TOPIC all movie ACC enjoy.PAST
'The children all enjoyed the movie.' (Japanese)

Since the quantifier is adjoined to VP and annotated as $\downarrow \in (\uparrow \text{ADJ})$, it is not a modifier of the quantified noun in the f-structure. In terms of semantics, however, an FQ is clearly related to a noun. In glue semantics, a quantifier relates an individual x to two propositions $R(x)$ (restrictive meaning) and $S(x)$ (scope meaning) (Dalrymple et al., 1997; Dalrymple, 2001), so this specification must be encoded in the lexical entry of quantifiers. This can be achieved by using a local name. For instance, we propose the lexical entry for Japanese quantifier *minna* 'all' as in (33), in which the local name %t is introduced. Since the quantifier modifies the topic noun, %t must be equal to the grammatical function corresponding to that noun. For the GF specification of %t, the inside-out path (ADJ $\in\uparrow$) is required to refer to the outer f-structure, and in that f-structure the off-path constraint, $(\rightarrow_\sigma \text{ DF}) = \text{TOPIC}$, picks up the GF that corresponds to an s-structure whose value of DF is TOPIC.

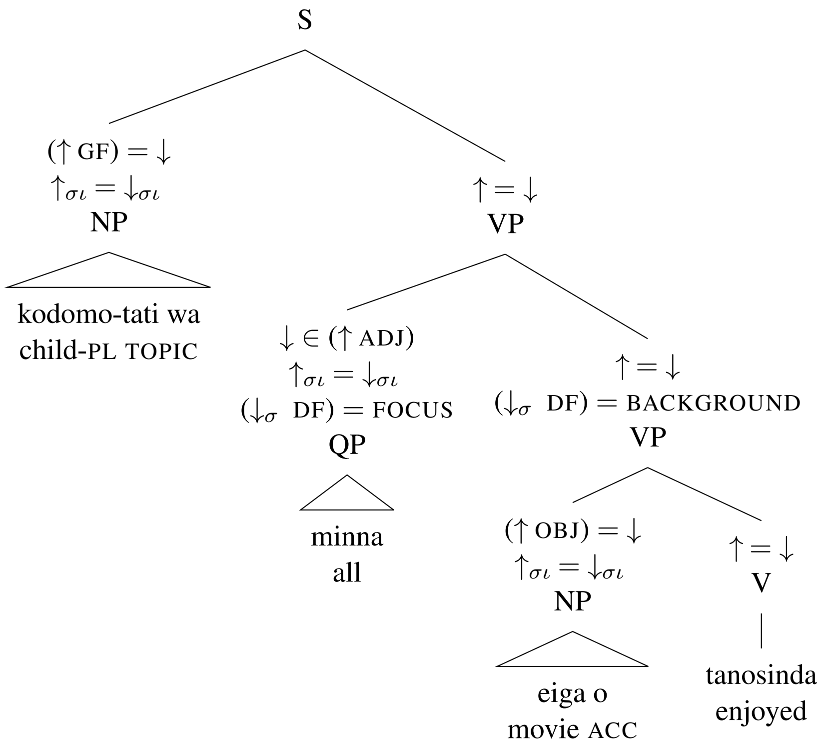
(33) *minna* Q $(\uparrow \text{ PRED}) = \text{'all'}$

$$\lambda R.\lambda S.all(x, R(x), S(x)) :$$
$$[((\%t)_\sigma \text{ VAR}) \multimap ((\%t)_\sigma \text{ RESTR})]$$
$$\multimap [\forall H.[(\%t)_\sigma \multimap H] \multimap H]$$
$$((\text{ADJ} \in \uparrow) \underset{(\rightarrow_\sigma \text{ DF}) = \text{TOPIC}}{\text{GF}}) = \%t$$
$$\mathbf{all} \in (\uparrow_{\sigma\iota}\ (\uparrow_\sigma \text{ DF}))$$

For the sentence (32), the c-structure (34), the f-structure (35), the s-structure (36) and the information structure (37) can be postulated. The subject NP is not encoded purely by c-structure position in Japanese, so the GF value of the sentence-initial NP is unspecified (cf. Matsumoto, 1996). The VP-internal NP is an object.[2] In (34), the topic status of the sentence-initial NP is overtly encoded by the topic marker *wa*, so the DF value of the subject NP becomes TOPIC in the s-structure. As discussed in Dalrymple & Nikolaeva (2011, 78–79), topic NPs are characterised by various semantic properties such as identifiability and activation in the discourse, so those features are encoded in the s-structure as well by such features as STATUS and ACTV. Those semantic feature specifications rule out the occurrence of a noun that lacks topic-worthiness like indefinite nouns. Due to the specifications in the c-structure rule (31), the DF value in the s-structure corresponding to the QP becomes FOCUS. The presence of the QP makes the following elements BACKGROUND, so despite the immediately pre-verbal position, the object NP becomes part of BACKGROUND. As a result, in the information structure in (34), the meaning constructor of the modified NP is a member of TOPIC, that of the FQ is a member of FOCUS, and those of the remaining elements are in BACKGROUND. Hence, the topic–comment information partitioning is correctly encoded in this structure.

[2]According to Sells (1990) an object NP can appear either under VP or under S in Japanese. For expository purpose, we only show structures in which an object NP appears under VP.

(34)



(35)

$$e : \left[\begin{array}{ll} \text{PRED} & \text{‘enjoy}\langle\text{SUBJ,OBJ}\rangle\text{’} \\ \text{SUBJ} & s : \left[\text{PRED} \quad \text{‘child’}\right] \\ \text{OBJ} & o : \left[\text{PRED} \quad \text{‘movie’}\right] \\ \text{ADJ} & \left\{ a : \left[\text{PRED} \quad \text{‘all’}\right] \right\} \end{array}\right]$$

(36)

$$s_\sigma : \left[\begin{array}{l} \text{STATUS} \quad \text{IDENTIFIABLE} \\ \text{ACTV} \quad \text{ACTIVE} \\ \text{VAR} \quad [\,] \\ \text{RESTR} \quad [\,] \\ \text{DF} \quad \text{TOPIC} \end{array}\right]$$

$$o_\sigma : \left[\text{DF} \quad \text{BACKGROUND}\right]$$

$$a_\sigma : \left[\text{DF} \quad \text{FOCUS}\right]$$

$$e_\sigma : \left[\text{DF} \quad \text{BACKGROUND}\right]$$

(37)

$$e_{\sigma\iota} : \begin{bmatrix} \text{TOPIC} & \{\ \textbf{children}\ \} \\ \text{FOCUS} & \{\ \textbf{all}\ \} \\ \text{BACKGROUND} & \left\{ \begin{array}{l} \textbf{enjoyed} \\ \textbf{movie} \end{array} \right\} \end{bmatrix}$$

Based on the s-structure in (36), the meaning of the sentence can be composed as in (38). Thanks to the local name specification in the lexical entry (33), the quantifier correctly consumes the resource of the s-structure corresponding to the topic NP, i.e. s_σ, and the meaning of the entire sentence, i.e. e_σ, is obtained.

(38)

all	$\lambda R.\lambda S.all(x, R(x), S(x))$:	$[(s_\sigma \text{ VAR}) \multimap (s_\sigma \text{ RESTR})]$ $\multimap [\forall H.[s_\sigma \multimap H] \multimap H]$
child	$\lambda x.child(x)$:	$(s_\sigma \text{ VAR}) \multimap (s_\sigma \text{ RESTR})$
enjoy-movie	$\lambda x.enjoy(x, movie)$:	$s_\sigma \multimap e_\sigma$

all, **child**, **enjoy-movie** $\vdash all(x, child(x), enjoyed\text{-}movie(x)) : e_\sigma$

4.2 VP-internal floating quantifiers

An FQ appearing inside of VP as sister to V requires a different analysis. As shown in (16), an FQ can appear under VP alongside object and oblique arguments. Thus, we postulate the c-structure rule that yields English ditransitive constructions as in (39).[3] The QP is an adjunct in the f-structure and the DF value in its s-structure is FOCUS. We assume that the presence of a QP makes the preceding object NP become a topic, so the optional equation for the TOPIC assignment for the DF in the s-structure is specified for the object NP. Similarly, the oblique PP receives the DF of BACKGROUND when the QP precedes it, so the s-structure DF value is optionally specified as BACKGROUND.

(39)

$$\text{VP} \longrightarrow \begin{array}{c} \text{V} \\ \uparrow = \downarrow \end{array} \quad \begin{array}{c} \text{NP} \\ (\uparrow \text{OBJ}) = \downarrow \\ \uparrow_{\sigma\iota} = \downarrow_{\sigma\iota} \\ ((\downarrow_\sigma \text{ DF}) = \text{TOPIC}) \end{array} \quad \begin{array}{c} \text{QP} \\ \downarrow \in (\uparrow \text{ADJ}) \\ \uparrow_{\sigma\iota} = \downarrow_{\sigma\iota} \\ (\downarrow_\sigma \text{ DF}) = \text{FOCUS} \end{array} \quad \begin{array}{c} \text{PP} \\ (\uparrow (\downarrow \text{PCASE})) = \downarrow \\ \uparrow_{\sigma\iota} = \downarrow_{\sigma\iota} \\ ((\downarrow_\sigma \text{ DF}) = \text{BACKGROUND}) \end{array}$$

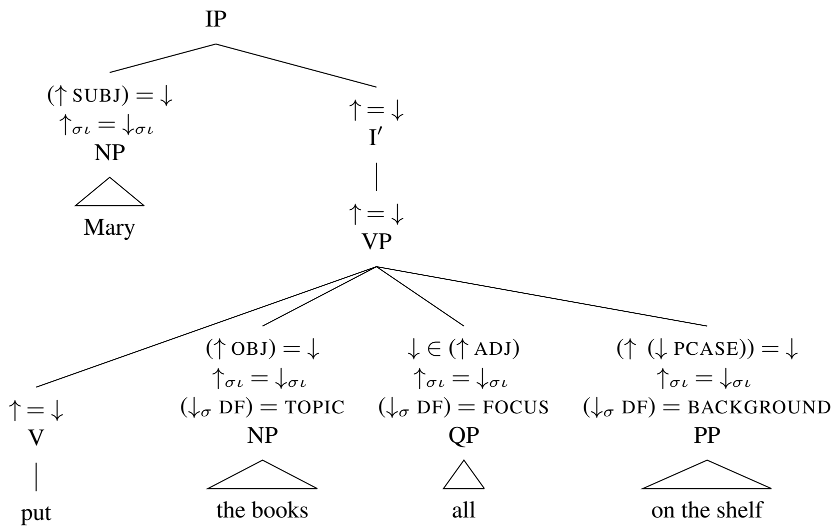
The following structures can be posited for an English ditransitive sentence with an FQ, *Mary put the books all on the shelf*. The quantifier functions as an adjunct of the clause in terms of f-structure, but it establishes the topic–comment structure inside of the VP by having the preceding object NP, *the books*, as TOPIC and the following oblique PP, *on the shelf*, as BACKGROUND.[4] We assume that the elements preceding the object NP are outside of the topic–comment frame, so they

[3]For ease of exposition, we only show the c-structure rule relevant to ditransitive constructions, in which a primary object NP and an oblique PP appear under VP.

[4]Regarding semantic composition of a quantifier and a modified noun, a complication would arise in relation to a determiner. A determiner also relates an individual x to two propositions $R(x)$ and $S(x)$, so the determiner *the* in *the books* and the FQ *all* both relate *books* to the restricted meaning and the scope meaning simultaneously, which causes invalid semantic composition. Since it is beyond the scope of this paper, we leave this issue open.

are given COMPLETIVE roles in the information structure. Thus, the information structure in (43) is constructed.

(40)



(41)

$$p:\begin{bmatrix} \text{PRED} & \text{`put}\langle\text{SUBJ,OBJ,OBL}_{on}\rangle\text{'} \\ \text{SUBJ} & s:\begin{bmatrix}\text{PRED} & \text{`Mary'}\end{bmatrix} \\ \text{OBJ} & o:\begin{bmatrix}\text{PRED} & \text{`book'} \\ \text{DEF} & + \\ \text{NUM} & \text{PL}\end{bmatrix} \\ \text{OBL}_{on} & l:\begin{bmatrix}\text{PRED} & \text{`shelf'} \\ \text{DEF} & + \\ \text{PCASE} & \text{OBL}_{on}\end{bmatrix} \\ \text{ADJ} & \left\{ a:\begin{bmatrix}\text{PRED} & \text{`all'}\end{bmatrix} \right\} \end{bmatrix}$$

(42)

$$s_\sigma:\begin{bmatrix}\text{DF} & \text{COMPLETIVE}\end{bmatrix}$$

$$o_\sigma:\begin{bmatrix}\text{STATUS} & \text{IDENTIFIABLE} \\ \text{ACTV} & \text{ACTIVE} \\ \text{DF} & \text{TOPIC}\end{bmatrix}$$

$$l_\sigma:\begin{bmatrix}\text{DF} & \text{BACKGROUND}\end{bmatrix}$$

$$a_\sigma:\begin{bmatrix}\text{DF} & \text{FOCUS}\end{bmatrix}$$

$$p_\sigma:\begin{bmatrix}\text{DF} & \text{COMPLETIVE}\end{bmatrix}$$

(43)
$$p_{\sigma\iota} : \begin{bmatrix} \text{TOPIC} & \{\ \textbf{the-books}\ \} \\ \text{FOCUS} & \{\ \textbf{all}\ \} \\ \text{BACKGROUND} & \{\ \textbf{on-the-shelf}\ \} \\ \text{COMPLETIVE} & \left\{ \begin{matrix} \textbf{Mary} \\ \textbf{put} \end{matrix} \right\} \end{bmatrix}$$

4.3 Agreement

In our analysis, a quantifier is an adjunct at the clause level in f-structure, so agreement between the quantifier and the modified noun cannot be straightforwardly accounted for. However, agreement beyond NP-internal constituents is not so unusual. For instance, Fitzpatrick (2006) reports that agreement is often found between secondary predicates and nouns as illustrated in (44) and (45).

(44) a. Ella llegó borracha.
she.F.SG arrived drunk-F.SG
'She arrived drunk.'

b. Ellas llegaron borrachas/*os.
they.F.PL arrived drunk-F.PL
'They arrived drunk.' (Spanish; Fitzpatrick 2006, 75)

(45) a. Vadim vernulsja iz bol'nicy zdoroviy.
V.M.SG.NOM returned from hospital healthy.M.SG.NOM
'Vadim returned from the hospital healthy.'

b. Ja zakazala rybu syruju.
I ordered fish.F.SG.ACC raw.F.SG.ACC
'I ordered the fish raw.' (Russian; Fitzpatrick 2006, 76)

In Spanish, the secondary predicates agree with the subject noun in number and gender as in (44). Similarly in Russian, the secondary predicate agrees with the main verb's subject in gender, number and case in (45a), while it agrees with the main verb's object in gender, number and case in (45b).

Note that in our proposal, secondary predicate constructions license quantifier floating—see (10c) and (11b), because the NP predicated by the secondary predicate can be a reference-oriented topic expression. Therefore, the agreement patterns found in (44) and (45) can be regarded as an instance of topic agreement, which is attested in many languages (Polinsky & Comrie, 1999; Bobaljik & Wurmbrand, 2002; Corbett, 2006; Dalrymple & Nikolaeva, 2011). In Ostyak, for example, object agreement is optional when the object can be either a topic or not as shown in (46a). When the topic interpretation is obligatory in the discourse, the verb must agree with the object as in (46b, c).

(46) a. (ma) tam kalaŋ we:l-s-əm / we:l-s-∅-e:m
I this reindeer kill-PAST-1.SG.SUBJ kill-PAST-SG.OBJ-1.SG.SUBJ
'I killed this reindeer.'

b. (What did you do to this reindeer?)
tam kalaŋ we:l-s-e:m / *we:l-s-əm
this reindeer kill-PAST-OBJ/1.SG.SUBJ kill-PAST-1.SG.SUBJ
'I killed this reindeer.'

c. kalaŋ xalśa we:l-s-əlli / *we:l-əs
reindeer where kill-PAST-OBJ/1.SG.SUBJ kill-PAST-1.SG.SUBJ
'Where did he kill the/a reindeer?'

(Ostyak; Dalrymple & Nikolaeva 2011, 142, 146)

If long-distance agreement found between an FQ and a noun is an instance of topic agreement, feature specification of an agreement controller can be simply stated in the lexical entry of a quantifier. For example, the lexical entry for the quantifier *allen* 'all.DAT.PL' used in the German example (1c), repeated here as (47), can be postulated as in (48).

(47) Diesen Studenten habe ich (gestern) *allen* geschmeichelt.
these.DAT.PL students have I (yesterday) all.DAT.PL flattered
'I flattered all of these students yesterday.' (German)

(48) *allen* Q

$$(\uparrow \text{PRED}) = \text{'all'}$$
$$(\%t\ \text{CASE}) = \text{DAT}$$
$$(\%t\ \text{NUM}) = \text{PL}$$
$$\ldots$$
$$((\text{ADJ} \in \uparrow)\ \underset{(\rightarrow_\sigma\ \text{DF}) = \text{TOPIC}}{\text{GF}}\ ) = \%t$$
$$\textbf{all} \in (\uparrow_{\sigma\iota}\ (\uparrow_\sigma\ \text{DF}))$$

The local name %t is already introduced to refer to the GF corresponding to a topic noun, so it can also be used to refer to the agreement controller. The equations, (%t CASE) = DAT and (%t NUM) = PL, co-specify the case and number features of the nouns.

As pointed out in subsection 2.2, long-distance agreement is problematic for the VP-modifier analysis because agreement is in principle determined locally between the agreement controller and the target. Our approach overcomes this problem because the specification of the agreement controller is determined with reference to the information structure role.

5 Conclusion

In the generative literature, quantifier floating has long been regarded as a syntactic phenomenon. A stranding analysis tries to capture the association between the quantifier and the noun by assuming that they form a syntactic constituent at an initial stage of derivation, but the absence of non-stranded constituents and the presence of distinct lexical items for FQs make this approach untenable. A VP-modifier analysis, on the other hand, does not face these problems. However, since the quantifier and a noun do not hold a direct syntactic relation, the analysis requires some additional assumptions to capture the semantic association and agreement between them. The existence of VP-internal FQs are also problematic for this type of approach.

The aftermentioned problems that arise in these different proposal stem from the attempt to capture quantifier floating purely in terms of syntactic structure. Our proposal overcomes these problems by incorporating information structure properties necessary for FQ constructions to arise into the analysis. An FQ requires a reference-oriented topic expression to be present, so that it can function as a focus to form a topic–comment information structure partitioning. The phrase structure distribution of FQs reflects the organisation of the sentence in terms of this information structure pattern. The VP-adjunct configuration encodes the preceding subject NP as a topic, the quantifier as a focus and the remaining VP as a background in an isomorphic way. This analysis also gives a straightforward account for the presence of VP-internal FQs in ditransitive constructions and secondary predicate constructions. They form a topic–comment structure with the object as a topic, the quantifier as a focus and the secondary object/oblique argument or the secondary predicate as background information. Agreement between a quantifier and a noun is also captured in a straightforward manner once it is analysed as an instance of topic agreement.

References

Abeillé, Anne & Danièle Godard. 1998. A lexical approach to quantifier floating in French. In Gert Webelhuth, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and constructional aspects of linguistic explanation*, 81–96. Stanford, CA: CSLI Publications.

Baltin, Mark. 1982. A landing site theory of movement rules. *Linguistic Inquiry* 13. 1–38.

Bobaljik, Jonathan David. 1995. *Morphosyntax: The syntax of verbal inflection*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Bobaljik, Jonathan David. 2003. Floating quantifiers: Handle with care. In Lisa Cheng & Rint Sybesma (eds.), *The second Glot International state-of-the-article book*, 107–148. Berlin: Mouton de Gruyter.

Bobaljik, Jonathan David & Susi Wurmbrand. 2002. Notes on agreement in Itelmen. *Linguistic Discovery* 1(1).

Bošković, Željko. 2004. Be careful where you float your quantifiers. *Natural Language and Linguistic Theory* 22. 681–742.

Butt, Miriam & Tracy Holloway King. 1996. Structural Topic and Focus without movement. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG96 conference*, Stanford, CA: CSLI Publications.

Butt, Miriam & Tracy Holloway King. 2000. Null elements in discourse structure. In K. V. Subbarao (ed.), *Papers from the null seminar*, Delhi: Motilal Banarasidas.

Choi, Hye-Won. 1999. *Optimizing structure in context: Scrambling and information structure*. Stanford, CA: CSLI Publications.

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Dalrymple, Mary. 1999. *Semantics and syntax in lexical functional grammar: The resource logic approach*. Cambridge, MA: MIT Press.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. San Diego: Academic Press.

Dalrymple, Mary, John Lamping, Fernando Pereira & Vijay Saraswat. 1997. Quantification, anaphora, and intensionality. *Journal of Logic, Language, and Information* 6. 219–273. Reprinted in Dalrymple (1999).

Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge: Cambridge University Press.

Dowty, David & Belinda Brodie. 1984. The semantics of "floated" quantifiers in a transformationless grammar. In *Proceedings of the third West Coast Conference on Formal Linguistics*, 75–90. Stanford, CA: Stanford Linguistics Association.

Fitzpatrick, Justin Michael. 2006. *Syntactic and semantic roots of floating quantification*. Cambridge, MA: Masachusetts Institute of Technology dissertation.

Kim, Jong-Bok & Jung-Soo Kim. 2009. English floating quantifier constructions: A non-movement approach. *Language and Information* 13. 57–75.

Kuno, Susumu & Ken-ichi Takami. 2003. Remarks on unaccusativity and unergativity in Japanese and Korean. In William McClure (ed.), *Japanese/Korean linguistics 12*, 280–294. Stanford, CA: CSLI Publications.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Maling, Joan. 1976. Notes on quantifier-postposing. *Linguistic Inquiry* 7. 708–718.

Matsumoto, Yo. 1996. *Complex predicates in Japanese: A syntactic and semantic study of the notion 'word'*. Stanford, CA: Kurosio/CSLI Publications.

Neeleman, Ad & Hans van de Koot. 2008. Dutch scrambling and the nature of discourse templates. *Journal of Comparative Germanic Syntax* 11. 137–189.

Neeleman, Ad & Reiko Vermeulen. 2012. The syntactic expression of information structure. In Ad Neeleman & Reiko Vermeulen (eds.), *The syntax of topic, focus, and contrast: An interface-based approach*, 1–38. Berlin: Mouton de Gruyter.

Polinsky, Maria & Benard Comrie. 1999. Agreement in Tsez. *Folia Linguistica* 33(2). 109–130.

Rochman, Lisa. 2010. Why float? floating quantifiers and focus marking. In Nomi Erteschik-Shir & Lisa Rochman (eds.), *The sound patterns of syntax*, 72–92. Oxford: Oxford University Press.

Sells, Peter. 1990. VP in Japanese: Evidence from *-te* complements. In Hajime Hoji (ed.), *Japanese/Korean linguistics*, 319–333. Stanford, CA: CSLI Publications.

Shlonsky, Ur. 1991. Quantifiers as functional heads: A study of quantifier float in Hebrew. *Lingua* 84. 159–180.

Sportiche, Dominique. 1988. A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry* 19. 425–449.

Quirky subjects in Icelandic, Faroese, and German: A Relational Grammar account

Andreas Pankau
Freie Universität Berlin

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 499–519

Keywords: quirky subjects, Germanic languages, multistratalism, subject tests, Arc Pair Grammar

Pankau, Andreas. 2016. Quirky subjects in Icelandic, Faroese, and German: A Relational Grammar account. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 499–519. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.26.

Abstract

This paper presents a new analysis of quirky subjects according to which quirky subjects bear multiple grammatical relations and hence differ syntactically from regular subjects. This contrasts with the standard analysis of quirky subjects according to which quirky subjects are regular subjects bearing lexical case and therefore differ only morphologically from regular subjects. Based on the behavior of quirky subjects in Faroese and German, I argue that the syntactic account is superior. Faroese shows that the case borne by a quirky subject is not lexical, whereas German shows that quirky subjects are not regular subjects to begin with. The behavior of quirky subjects in Icelandic, on which the standard analysis is based, is argued to be the result of a morphosyntactic peculiarity of Icelandic

1 The Standard Analysis of Quirky Subjects

Quirky subjects is the term used to refer to constructions where a subject bears unexpected ("quirky") case, namely some object case instead of the expected nominative case.[1] Two examples from Icelandic are given in (1).

(1)	a.	Jóni	líkar	þessi	bók.	b.	Þeim	var	hjálpað.
		Jón.DAT	likes	this	book.NOM		they.DAT	was	helped
		'Jón likes this book.'					'They were helped.'		

Example (1a) illustrates a lexically determined quirky subject, example (1b) a quirky subject resulting from passivization. I will refer to the former as *lexical quirkies*, and to the latter as *passive quirkies*. The standard analysis for quirky subjects rests on the idea that they are regular subjects to which non-structural, lexical case is assigned. In (1a) the verb *líka* assigns lexical dative case to its subject *Jón*, in (1b) it is the verb *hjálpa* that assigns lexical dative case to its object. The dative case on the object is preserved after promotion of the object to subject in (1b) because lexically assigned case cannot be overwritten. Calling this analysis the standard analysis is due to its acceptance by virtually all grammar frameworks. Originally developed within LFG on the basis of Icelandic (Zaenen et al. 1985), it is adopted by LFG up to this day (Schätzle et al. 2015; Willgohs & Farrell 2009), as well as by HPSG (Bouma 1992; Sag et al. 1992), GB/Minimalism (Jónsson 1996, 2003; Sigurðsson 1989, 1992; Þráinsson 2007), and Construction Grammar (Barðdal 2006; Barðdal & Eyþórsson 2012).

† Many thanks to the anonymous reviewers of this paper, as well as the referees and the audience of HEADLEX16 (especially Miriam Butt, Joan Maling, and Manfred Sailer) for their comments and suggestions. A warning to the reader. There is only little of HPSG or LFG in this paper. This is due to the fact that the paper was submitted to a pre-conference workshop on the representation of grammatical functions, which was later included into the main session.

[1] For reasons of space, the status of the nominative object will be ignored throughout the paper.

This paper argues that despite this overall agreement across frameworks, the standard analysis is inadequate. It suffers from two defects. The first defect is that it confounds general aspects of the syntax of quirky subjects with language particular properties of Icelandic. I will show that the presence of lexical quirkies neither goes along with the presence of passive quirkies nor with the preservation of quirky case on lexical quirkies. The second defect of the standard analysis is that it treats quirky subjects as subjects only. Yet in some languages quirky subjects show an inconsistent behavior, passing some subjects tests, but not all. The alternative analysis I will argue for is a revised version of the Relational Grammar analysis according to which quirky subjects are underlying subjects but surface objects. This analysis neither entails the existence of passive quirkies nor case preservation nor a consistent behavior of quirky subjects vis-à-vis subject tests. These properties, which are found only in Icelandic, are argued to follow from a language particular property.

The paper is structured as follows. I first review the evidence for the subject status of quirky subjects in section 2. In section 3, I present data from Faroese and German that the standard analysis cannot capture. In section 4, I present the alternative relational analysis and introduce some general technicalities. In section 5, I apply this analysis to Icelandic, Faroese, and German and argue that the differences between the three languages reduce to language particular requirements independent of, but with consequences for, quirky subject constructions. Section 6 concludes this paper.

2 Quirky Subjects and Subjecthood

The reason for analyzing quirky subject constructions as clauses containing a subject instead of impersonal clauses is that quirky subjects pass subject tests. Four well-known subject tests are control, reflexivization, subject-to-subject raising, and subject-to-object raising (ECM). Quirky subjects in Icelandic pass all these tests, cf. (2)-(5).

(2) a. Jón$_i$ vonast til að [PRO$_i$ líka þessi bók].
Jón hopes PREP to like this book
'Jón hopes to like this book.'

b. Ég$_i$ vonast til að [PRO$_i$ vera hjálpað].
I hopes PREP to become helped
'I hope to be helped.'

(3) Henni$_i$ leiðist bókin sín$_i$.
her.DAT bores book REFL
'She finds her book boring.'

(4) Ólafi$_i$ virtist [t$_i$ hafa leiðst].
Ólaf.DAT seemed have bored
'Olaf seemed to be bored.'

(5) a. Hann telur Jóni$_i$ [t$_i$ líka þessi bók].
he believes Jón.DAT like this book
'He believes Jón to like this book.'

b. Hann telur Jóni$_i$ [t$_i$ hafa verið hjálpað].
he believes Jón.DAT have become helped
'He believes Jón to have been helped.'

The raising data in (4) and (5) are of special relevance for the standard analysis. They show that quirky subjects preserve their case under raising, in contrast to nominative case marked subjects, which lose their case under raising, cf. (6).

(6) a. Guðrún saknar Haraldar.
Guðrún.NOM misses Harald
'Guðrún misses Harald.'
b. Ég taldi Guðrúnu$_i$ í barnaskap mínum [t$_i$ sakna Haraldar].
I believed Guðrún.ACC in foolishness my to.miss Harald
'I believed Guðrún in my foolishness to miss Harald.'

The only subject test quirky subjects do not pass is verb agreement. The verb either agrees with the nominative marked nominal or, in case no nominative marked nominal is present, bears default third person marking.

(7) a. Henni höfðu / *hafði ekki líkað hestarnir.
she.DAT had.3.PL had.3.SG not liked horses
'She had not liked the horses.'
b. Mér verður / *verð ekki kalt.
I.DAT will.3.SG will.1.SG not cold
'I will not be cold.'

3 Problems for the Standard Analysis

In this section I present data from quirky subject constructions in Faroese (taken from Barnes 2001: chapter 4) and German which show that their properties cannot be captured by the standard analysis of quirky subjects as regular subjects bearing lexical case.

3.1 Quirky Subjects in Faroese

Similar to Icelandic, Faroese possesses lexical quirkies passing subject tests like reflexivization and control.

(8) a. Mær dámar mjólkina.
I.DAT likes milk
'I like milk.'
b. Kjartani$_i$ dámar væl nýggja bil sin$_i$.
Kjartin.DAT likes well new car REFL
'Kjartin likes his new car.'
c. Hann$_i$ royndi at [PRO$_i$ dáma matin].
he tried to like food
'He tried to like the food.'

Quirky subjects in Faroese differ from those in Icelandic in two ways that are, however, unexpected under the standard analysis. The first difference is that the case of lexical quirkies is not preserved under raising.

(9) a. Jógvani tørvaði ein nýggjan bil.
Jógvan.DAT needed a new car
'Jógvan needed a new car.'
b. Eg helt Jógvan$_i$ [t$_i$ tørva ein nýggjan bil].
I believed Jógvan.ACC need a new car
'I believed Jógvan to need a new car.'

There is a confounding factor that needs to be excluded, namely that example (9b) is based on nominative subjects. This option suggests itself because quirky subjects tend to be replaced by nominative subjects in present day Faroese (Jónsson & Eyþórsson 2005: 227; Þráinsson et al. 2004: §5.4.2.1 & §7.6.2). Example (10) shows this for the verb *dáma* from example (8).

(10) Eg dámi ikki tvøst.
I.NOM like not whale.meat
'I don't like whale meat.'

But this option can be excluded because the verb *tørva* from (9) only allows dative-marked subjects even in modern Faroese (Þráinsson et al. 2004: 255). The second difference is that Faroese does not possess passive quirkies. Instead, lexically case marked objects are promoted to nominative subjects.

(11) √Eg / *mær verði hjálpin.
I.NOM I.DAT become helped
'I am helped.'

There is again a confounding factor because not all dative marked objects can be promoted to nominative marked subjects (Þráinsson et al. 2004: §5.4.4).

(12) a. Teir takkaðu honum.
they thanked he.DAT
'They thanked him.'
b. Honum bleiv takkað.
he.DAT was thanked
c.*Hann bleiv takkaður.
he.NOM was thanked
'He was thanked.'

(13) a. Tey trúðu henni.
they believed she.DAT
'They believed her.'
b. Henni bleiv trúð.
she.DAT was believed
c.*Hon varð trúð.
she.NOM was believed
'She was believed.'

This could be taken as evidence that Faroese has some passive quirkies but what seems more likely is that it only shows that certain datives fail to undergo passivization, a situation well-known from German.

(14) a. Sie danken ihm.
they thanked he.DAT
'They thanked him.'

(15) a. Sie glauben ihr.
they believed she.DAT
'They believed her.'

b. Ihm wird gedankt.
he.DAT was thanked
c.*Er bekam gedankt.
he.NOM became thanked
'He was thanked.'

b. Ihr wurde geglaubt.
she.DAT was believed
c.*Sie bekam geglaubt.
she.NOM became believed
'She was believed.'

The problem then posed by Faroese is that it possesses quirky, non-nominative subjects whose properties differ in an unexpected way from Icelandic: their case can be overwritten. Moreover, Faroese lack passive quirkies and promotes dative marked objects to nominative subjects in passives. Consequently, the dative borne by quirky subjects is not lexical. But if dative case is not lexical, then quirky subjects cannot be defined as subjects bearing lexical case.

3.2 Quirky Subjects in German

German too possesses constructions that look like lexical quirkies.

(16) Mir gefällt der Mann.
I.DAT likes the man
'I like the man.'

The first problem with German is that subject tests give conflicting results for the subject status of the dative nominal in (16). It cannot be controlled nor undergo raising to object (no matter what case it bears after raising) indicating that it is not a subject, cf. (17), but it can control itself an empty subject and bind a reflexive, indicating that it *is* a subject, cf. (18).

(17) a.*Ich$_i$ versuche [PRO$_i$ der Mann zu gefallen].
I try the man to like
'I try to like the man.'
b.*Ich sehe ihm$_i$ / ihn$_i$ [t$_i$ der Mann gefallen].
I see he.DAT he.ACC the man like
'I see that he likes the man.'

(18) a. Jedem$_i$ gefiel das Buch [ohne PRO$_i$ es gelesen zu haben].
everyone.DAT liked the book without it read to have
'Everyone liked the book without having read it.'
b. Ihnen$_i$ gefällt es miteinander$_i$.
they.DAT likes it with.each.other
'They enjoy each other.'

The two sets of tests really test for subjects. The ungrammaticality of the examples in (17) is irreducible to a semantic constraint requiring the controlled subject to be agentive. This becomes apparent when *gefallen* is replaced with the synonymous verb *mögen*, whose subject is equally non-agentive but bears nominative case. Sentences based on *mögen* instead of *gefallen* are fine.

(19) a. Ich$_i$ versuche [PRO$_i$ den Mann zu mögen].
I try the man to like
'I try to like the man.'
b. Ich sehe ihn$_i$ [t$_i$ den Mann mögen].
I see he.ACC the man like
'I see that he likes the man.'

Likewise, the grammaticality of the examples in (18) is not due to their dative marking for datives usually cannot control empty subjects or bind reflexives.

(20) a.*Ich helfe jedem$_i$ [ohne PRO$_i$ danach gefragt zu haben].
I help everyone.DAT without thereafter asked to have
'I help everyone without that he asked for that.'
b.*Ich habe den Ärzten$_i$ einander$_i$ empfohlen.
I have the doctors.DAT each.other.ACC recommended
'I have recommended the doctors to each other.'

The second problem posed by German is that clauses resembling passive quirkies in German (cf. 21) pass no subject test whatsoever (cf. 22).

(21) Jedem wurde geholfen.
everyone.DAT became helped
'Everyone was helped.'

(22) a.*Er$_i$ hofft [PRO$_i$ geholfen zu werden].
he hopes helped to become
'They hope to be helped.'
b.*Ich sehe jedem$_i$ [t$_i$ geholfen werden].
I see everyone.DAT helped become
'I see that everyone is helped.'
c.*Jedem$_i$ wird geholfen [ohne PRO$_i$ es zu wollen].
everyone.DAT becomes helped without it to want
'Everyone is helped without wanting it.'
d.*Den Ärzten$_i$ wurde einander$_i$ empfohlen.
the doctors.DAT became each.other.ACC recommended
'The doctors were recommended to each other.'

The contrast between (22) and (17)-(18) indicates that constructions superficially resembling passive quirkies are impersonal clauses. This is a problem for the standard analysis of quirky subjects because it equates subject with the first argument. Since the examples in (21) and (16) both contain such a first argument bearing lexical case, both should behave identically.[2] This mismatch has not gone unnoticed in the literature and two strategies have been adopted to deal with it. According to one strategy, German lacks quirky subjects (Bayer 2004; Haider 2010; Müller 2008; Sigurðsson 2002). According to the other,

[2] This implicit assumption is found in section 5 of Zaenen et al. (1985), where it is argued on the behavior of passive quirkies only that German lacks quirky subjects altogether.

German does have quirky subject construction identical to Icelandic (Barðdal 2006; Barðdal & Eyþórsson 2003). Neither strategy is satisfactory. The first cannot explain why lexical quirkies pass some subject tests, whereas the second cannot explain why they don't pass all subject tests. Moreover, both strategies fail to account for the absence of passive quirkies in German.

3.3 Interim Conclusion and Outlook

The main result from this section is that a language can have quirky subjects but not behave like Icelandic. Faroese has lexical quirkies whose case is not preserved under raising, and it lacks passive quirkies. German too lacks passive quirkies and its lexical quirkies don't pass all subject tests. Such a state of affairs is a serious problem for the standard analysis because the Icelandic pattern is a consequence of the analysis of quirky subjects. To solve this problem, two routes can be taken. The first route is to take the properties of quirky subject constructions in Icelandic as definitional. So in order for some quirky subject like construction to count as a real quirky subject construction, it has to behave like Icelandic quirky subjects. Otherwise, it is not a quirky subject construction. This line of reasoning is not only implicitly assumed in most of the literature on German, it is also explicitly adopted in Willgohs & Farrell (2009: 640). The other route is to reject the premise that Icelandic is the prime example for quirky subject constructions and entertain the possibility that the properties of the Icelandic quirky subject construction result from the interaction of language particular properties with universal aspects of quirky subjects. It is the second route that I take in this paper.[3]

4 An Arc Pair Grammar Analysis of Quirky Subjects

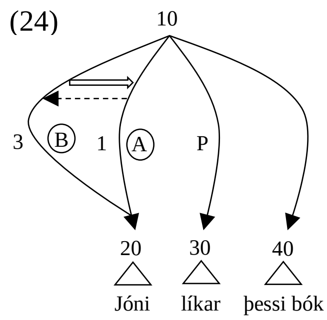
4.1 Quirky Subject Constructions as Inversions

My analysis of quirky subjects is couched within the Arc Pair Grammar framework (Aissen 1987; Johnson & Postal 2013; Pankau 2013), a successor of Relational Grammar (Blake 1990; Perlmutter 1983; Perlmutter & Rosen 1984; Postal & Joseph 1990). The analysis is a modified version of the original Relational Grammar analysis, according to which quirky subjects result from an operation called *inversion*, cf. (23) & (24).[4]

(23) *Inversion Analysis of Quirky Subjects*
Quirky subjects are underlying subjects and surface indirect objects

[3] The idea that Icelandic quirky subjects are not prototypical instances of quirky subjects was already foreseen in Davies (1988), which also suggested that Icelandic subject tests are sensitive to working 1s instead of final 1s (cf. §4.2). Unfortunately, the paper did not present an analysis for Icelandic comprising these insights. The present paper can be seen as taking this second step.
[4] This term is due to Harris (1980), who adopted it from traditional Georgian grammar; cf. Moore & Perlmutter (2000) for an overview of the research on quirky subjects in Relational Grammar.

(24)



Arc Pair Grammar inherits the idea of Relational Grammar that grammatical relations are primitive. The grammatical relation of a constituent is indicated through an *R-sign* attached to the edge the constituent appears at (1=subject, 2=direct object, 3=indirect object, P=predicate). The resulting objects are called *arcs*. In (24), the nominal *Jóni* is a subject qua its R-sign *1* and is said to *head* a *1-arc*, the verb *líkar* is the predicate qua its R-sign *P* and *heads* a *P-arc*; and so forth. Unlabeled edges indicate arcs whose relational status is ignored. Circled letters are of no linguistic relevance but simply names for arcs. The structure in (24) also shows that Arc Pair Grammar features multidominance, called *overlapping*. So arc A and arc B overlap. In (24), this expresses that *Jóni* is both a subject and an indirect object.[5] One innovation of Arc Pair Grammar is the idea that there are two metarelations between arcs, namely *Sponsor* and *Erase*, represented by a dotted arrow and a double arrow, respectively. Sponsor expresses the idea that an arc depends on the existence of another arc. In (24), that A, the 1-arc, sponsors B, the 3-arc, means that the nominal *Jóni* is first a subject and then an indirect object. Erase expresses the idea that the morphological invisibility of an arc is due to the visibility of another arc. In (24), this means that B, the 3-arc, determines case marking and not A, the 1-arc. Not all arcs are sponsored and not all arcs erased. Unsponsored arcs are called *initial* arcs, arcs not sponsoring any further arc are called *final* arcs[6], and unerased arcs are called *output* arcs. The P-arc in (24) is both an initial, a final, and an output arc; the 1-arc is only an initial arc; and the 3-arc is a final and an output arc. The sponsor pair (A, B) is special in that it involves overlapping arcs. The sponsored arc is then called *successor* whereas the sponsoring arc is called *predecessor*. If the two arcs share the same tail node, the sponsored arc is a *L(ocal)-successor* and the sponsoring arc a *L(ocal)-predecessor*. If not, the sponsored arc is a *F(oreign)-successor* and the sponsoring arc a *F(oreign)-predecessor*. In (24), B is an L-successor of A, and A an L-predecessor of B.

My relational analysis of quirky subjects differs from the traditional one in one detail Relational Grammar assumed that this R-sign of the final object are is always *3* so that quirky subjects are always indirect objects and always bear dative case. This constraint is too strict because in Icelandic (25), Faroese (26), and German (27), quirky subjects can also bear other object cases.[7]

[5] Arc Pair Grammar is multistratal but not transformational. Although it assumes that some constituent can bear multiple relations, this is expressed in a *single* object via overlapping arcs.

[6] For reasons that I lack space to elucidate, I deviate here from standard assumptions (Johnson & Postal 1980; Postal 2010) according to which final arcs are also output arcs.

[7] In Faroese, the genitive is extinct in the modern spoken language, and is in decline in German, so that genitive marked quirky subjects are absent from both languages.

(25) a. Bátinn rak á land.
boat.ACC drifted to land
'The boat drifted to shore.'

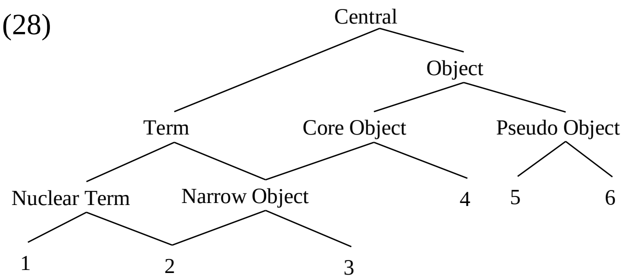
b. Verkjanna gætir ekki.
pains.GEN is.noticeable not
'The pains are not noticeable.'

(26) Meg nøtrar í holdið.
I.ACC shudders in flesh
'I shudder.'

(27) Mich friert/dürstet.
I.ACC freezes/is.thirsty
'I am cold/thirsty.'

The set of surface relations borne by quirky subjects must hence also include object relations other than the indirect object relation. The proposal I make regarding the class of object relations borne by quirky subjects builds on a modified version of Postal's (2010: 72) taxonomy of *primitive* object relations.

(28)



Based on this taxonomy, the following class of *defined* object relations can be given.

(29) Strict Object = Central - Nuclear Term

This set contains all central relations minus the nuclear term relations, that is, the resulting set contains indirect objects (=3), subobjects (=4), semiobjects (=5), and quasiobjects (=6). Semiobjects will be ignored throughout this paper. Quasiobjects correspond to genitive marked objects, subobjects correspond to non-adverbial accusative marked NPs. Due to their accusative case, subobjects are often conflated with direct objects. But they differ from direct objects in that they cannot be passivized[8] nor form middles in English (Postal 2010: 57-60) or German (Pankau 2013: 232).

(30) a.*Milk is not given by snakes. b.*Milk gives frequently.

(31) a.*Milch wird täglich gegeben.
milk becomes daily given
'They give milk daily.'

b.*Milch gibt sich leicht.
milk gives REFL easy.
'It is easy to give milk.'

Moreover, subobjects in German cannot undergo raising in constructions with the raising verbs *sein* or *gehören* implying a necessity (Pankau 2013: 235-6).

[8] The constraint on non-passivizability applies to all subobjects in German, be they single objects or objects in double object constructions (Pankau 2013: 234). In English, they resist passivization only as single objects but not in double object constructions (Postal 2010: chapter 7).

(32) a.*Milch ist zu geben.
milk is to give
'Milk needs to be given.'

b.*Milch gehört gegeben.
milk belongs given
'Milk needs to be given.'

Since quirky subjects bearing accusative case also fail to undergo object raising and raising with *sein* and *gehören*, I analyze them as subobjects as well.

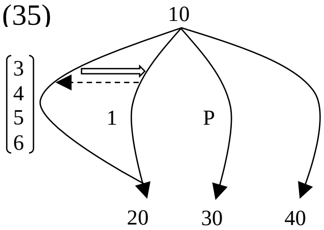
(33) a.*Ich bin leicht zu dürsten.
I am easy to be.thirsty
'It is easy for me to be thirsty.'

b.*Ich bin zu dürsten.
I am to be.thirsty

c.*Ich gehöre gedürstet.
I belong been.thirsty
'I need to be thirsty.'

Accordingly, I suggest the following revised inversion analysis for quirky subjects, cf. (34) & (35).

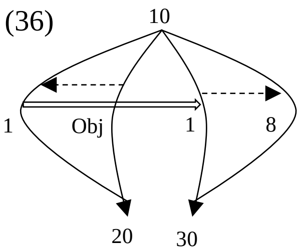
(34) *Revised Inversion Analysis of Quirky Subjects*
Quirky subjects are initial subjects and final strict objects

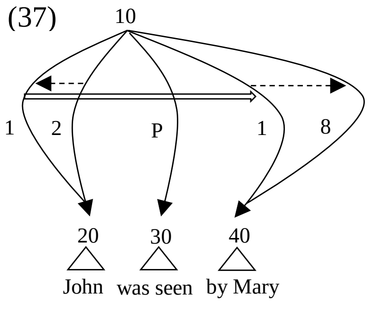
(35)


The crucial difference between my analysis and the standard analysis is that my analysis characterizes quirky subject constructions as constructions involving a change in grammatical relation and not by some exceptional case assignment. This is in sharp contrast to the standard analysis, where only exceptional case assignment is involved.

4.2 Passives, Working Nuclear Terms, Laws and Rules

In order to develop the relational analyses for the three language, some background ideas of Arc Pair Grammar are needed.
The first concerns the structure of passive clauses. I adopt the most recent Arc Pair Grammar treatment of passive clauses (Postal 2010). According to this analysis, passive clauses involve advancement of some object to subject and demotion of the initial subject to a special relation called *chômeur* (=8), as shown in (36). Applying this idea to the English passive clause *John was seen by Mary* results in the structure in (37).

(36)


(37)


The second idea needed is the notion of *working nuclear term*. This notion allows one to refer to nominals that are nuclear term arcs but not necessarily final nuclear terms, cf. (38).

(38) *Working Nuclear Term*
A working nuclear term is any final central arc R-sponsor-linked to a nuclear term arc

The definition mentions the defined relation *R-sponsor-linked* holding between two arcs. If two arcs A and B are sponsor-linked, then either A sponsors B or B sponsors A. The prefix 'R' indicates the ancestral of any relation (Johnson & Postal 1980: 25), turning it into a reflexive and transitive relation. Consequently, if A is R-sponsor-linked to B, then either A is sponsor-linked to B, or A is identical to B, or A is sponsor-linked to C and C is sponsor-linked to B, and so on. Since nuclear term arcs comprise 1-arc and 2-arcs, the notion working nuclear term consequently comprises *working 1s* and *working 2s*. In the relational literature, the necessity of working 1s has long been recognized (Dziwirek 1994; González 1988; Legendre 1994; Perlmutter 1984; Rosen & Wali 1989), whereas working 2s have not figured prominently (Berinstein 1986). In this paper, only working 1s are relevant. Crucially, an instance of a working 1 is the 3-arc in (24): it is a final central arc, namely a 3-arc, and it is R-sponsor-linked to a 1-arc, namely to its 1-arc predecessor. It is also important to note that working 1s also comprise final 1s: a final 1 is a final central arc and it is R-sponsor-linked to a 1, namely to itself (via reflexivity of the ancestral). In contrast to this, the demoted subject in a passive clause (cf. 36 & 37) does *not* head a working 1: the final 8-arc it heads is not a central arc, whereas the 1-arc it heads is a central arc, but not a final one. So although working 1s pick out a broader class of subjects than final 1s, they do *not* pick out *any* 1-arc sponsoring another arc. Grammatical relation changing operations affecting subjects will therefore not necessarily result in working 1s.
The third idea needed is the distinction between *laws* and *rules*. Laws and rules are formally identical – both are stated as implications – but differ in scope: rules are language specific whereas laws apply to all languages. Two important laws are given in (39) & (40).

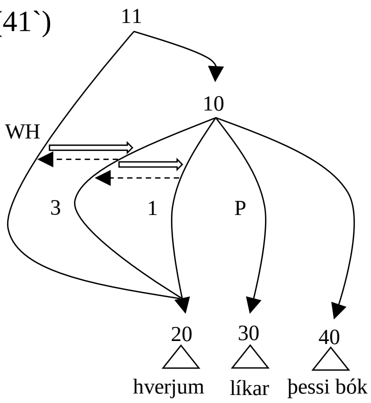
(39) *Unique Eraser Law*
If A is erased by B and by C, then B = C

(40) *Single Mother Law*
No constituent can head more than one unerased arc

The first law says that no arc can have more than one eraser. The second law guarantees that a nominal bearing multiple grammatical relations will surface only with one of these relations. The third law required regulates case marking. In connection to (24) I said that the 3-arc determines case marking because it is unerased. But this is not fully correct. Consider in this respect example (41) whose structure is provided in (41`).

(41) Hverjum líkar þessi bók?
who.DAT likes this book
'Who likes this book?'

(41`)



In (41), the quirky subject is wh-extracted. Extractions in Arc Pair Grammar are analyzed as unique grammatical relations, called *overlay* relations. In (41`), the interrogative pronoun *hverjum* heads an initial 1-arc sponsoring a 3-arc and an overlay arc with the R-sign *WH* erasing the 3-arc. But even though the 3-arc is erased, it still determines dative case on the interrogative pronoun. So case marking references a specific type of unerased arc, which is called *shallow arc*.

(42) *Shallow Arc*
A shallow arc is any argumental arc that is not erased or erased by an overlay successor

Argumental arcs comprise object and oblique arcs, but crucially not overlay arcs. Shallow arcs are then defined as argumental arcs that are either unerased or erased by a successor that is an overlay arc. The 3-arc in (41`) satisfies the definition of shallow arc: although erased it is erased by a successor, it is erased by an overlay arc, namely the WH-arc. With this, the third law can be given.

(43) *Case Marking Law*
Case marking is determined by shallow arcs only

The one rule needed in this paper applies to Icelandic and Faroese, both having a rather fixed word order regulated by surface grammatical relations (Þráinsson 2007; Þráinsson et al. 2004). Since surface grammatical relations correspond to final arcs in Arc Pair Grammar, I suggest the fallowing rule.

(44) *Final Arc Word Order Rule (Icelandic & Faroese)*
Final arcs determine word order

5 Quirky Subjects Reconsidered

My relational analysis for the behavior of quirky subjects in Icelandic, Faroese, and German rests on two ideas. First, only lexical quirkies are typical quirky subjects. Case preservation and the presence of passive quirkies are properties independent of quirky subjects proper and due to language particular rules. Second, subject tests do not reference necessarily final 1s, but also working 1s or output 1s. Languages then differ with respect to which type of subjecthood (working 1, final 1, output 1 etc.) is referenced by which subject test.

5.1 Lexical Quirkies

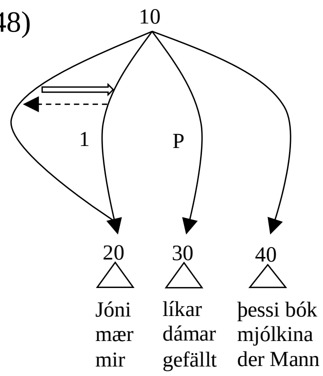
The structure for lexical quirkies in Icelandic (45), Faroese (46), and German (47) is identical in all three languages: they all feature inversion, cf. (48).

(45) Jóni líkar þessi bók.
Jón.DAT likes this book
'Jón likes this book.'

(46) Mær dámar mjólkina.
I.DAT likes milk
'I like milk.'

(47) Mir gefällt der Mann.
I.DAT likes the man
'I like the man.'

(48)



The analysis in (48) is straightforward. Each clause features inversion as presented in section 4.1, that is, an initial subject is demoted to a strict object. Since the structure for lexical quirkies is identical across the three languages, something else must be responsible for the different behavior vis-à-vis subject tests. The analysis I suggest for this is that more subject tests in Icelandic and Faroese reference working 1s than in German.

(49) *Subject Tests in Icelandic & Faroese*
- (i) A reflexive is anteceded by a nominal heading a working 1
- (ii) A controller nominal is a nominal heading a working 1
- (iii) A controlled nominal is a nominal heading a working 1
- (iv) A raising target is a nominal heading a working 1
- (v) Finite verbs agree with a nominal heading an output 1

(50) *Subject Tests in German*
- (i) A reflexive is anteceded by a nominal heading a working 1
- (ii) A controller nominal is a nominal heading a working 1
- (iii) A controlled nominal is a nominal heading a final 1
- (iv) A raising target is a nominal heading a final 1
- (v) Finite verbs agree with a nominal heading an output 1

Recall from section 4.2 that working 1s comprise final 1s and final strict object arcs R-sponsor-linked to a 1-arc. The former option corresponds to regular nominative marked subjects, as they head final 1-arcs. The latter option corresponds to quirky subjects because they head final strict object arcs R-sponsor-linked to a 1-arc. Consequently, the 3-arc in (48) counts as a working 1 as well. Since reflexivization is sensitive to nominals heading a working 1 in all three languages, reflexives can be anteceded by a regular nominative subject but also by a quirky subject. Similarly, a controller nominal is required to head a working 1 in all three languages, so both regular nominative subjects and quirky subjects make licit controller nominals in the three languages. However, whereas a controlled nominal can head a working 1 in Icelandic and Faroese, it has to head a final 1 in German. This captures that only regular nominative subjects can be controlled nominals in German, whereas also quirky subjects can be controlled nominals in Icelandic and Faroese. Similarly for raising: only nominals heading a final 1 are licit raising targets in German, thereby excluding quirky subjects from undergoing raising. Importantly, this account solves the problem why raising of a lexical quirky is impossible in German no matter what case the raised lexical quirky bears after raising (cf. 17b): the lexical quirky is simply not a licit raising target to begin with. In Icelandic and Faroese, on the other hand, nominals heading a working 1 can be raised and hence also quirky subjects can undergo raising. Note that verb agreement is sensitive to nominals heading output 1s in all three languages, capturing that only regular nominative subjects trigger verb agreement in all three languages.

5.2 Case Preservation under Raising

I just said that both Icelandic and Faroese allow for raising of quirky subjects because raising is sensitive to nominals heading working 1s in these two languages. If so, then something else must be responsible for the fact that the case of the quirky subject is preserved under raising in Icelandic (cf. 51) but not Faroese (cf. 52).

(51) Hann telur Jóni$_i$ [t$_i$ líka þessi bók].
he believes Jón.DAT like this book
'He believes Jón to like this book.'

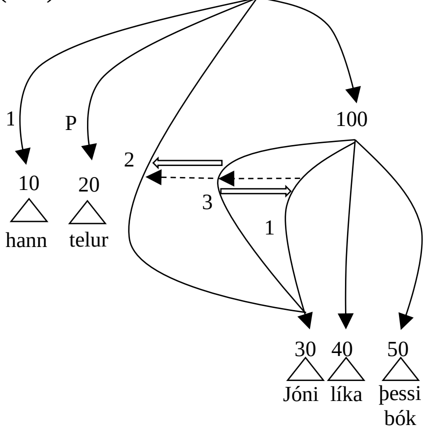
(52) Eg helt Jógvan$_i$ [t$_i$ tørva ein nýggjan bil].
I believed Jógvan.ACC need a new car
'I believed Jógvan to need a new car.'

What I suggest is that Icelandic puts an additional constraint on strict objects.

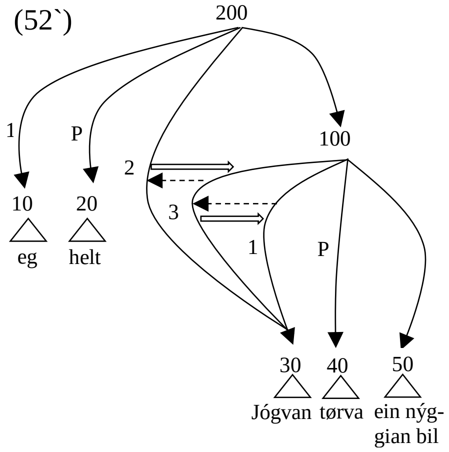
(53) *Icelandic Strict Object Rule*
Strict objects arcs must be shallow arcs

Consider now the respective structures for subject-to-object raising in the two languages.

(51`)



(52`)



Raising in Arc Pair Grammar is analyzed as involving two overlapping central arcs such that the higher arc is an F-successor of the lower arc, that is, sponsored by the lower arc. In the case of subject-to-object raising, the higher arc is a 2-arc and the lower arc is some 1-arc. Both Icelandic and Faroese allow for nominals heading working 1s as raising targets. The quirky subjects in (51`) and (52`) are licit raising targets because they both head a working 1, namely the 3-arc sponsored by the 1-arc. In the general case, successors erase their predecessors. This can be seen in the Faroese raising structure (52`): the 2-arc erases the 3-arc. Nothing additional is required for Faroese. In Icelandic, however, the Icelandic Strict Object Rule prohibits erasure of the 3-arc by the 2-arc. If the 3-arc were erased, then it would cease to be a shallow arc. But the Icelandic Strict Object Rule requires all strict object arcs and hence also 3-arcs to be shallow arcs. Moreover, that one of the two arcs has to be erased follows from the Single Mother Law (cf. 40). Recall from section 4.2 that the Case Marking Law (cf. 43) identifies shallow arcs as being responsible for case marking, whereas the Final Arc Word Order Rule (cf. 44) identifies final arcs as determining word order. In the Faroese structure (52`), the quirky subject heads three arcs, a 1-arc, a 3-arc and 2-arc. The 1-arc is not a final arc (it sponsors the 3-arc) nor is it a shallow arc (it is erased by the 3-arc, a non-overlay successor). Similarly for the 3-arc: it is neither a final arc (it sponsors the 2-arc) nor a shallow arc (it is erased by the 2-arc, a non-overlay successor). Only the 2-arc is both a shallow arc and a final arc: it is unerased and doesn’t sponsor another arc. In Faroese therefore the 2-arc determines word order according to the Final Arc Word Order Rule and accusative case marking on the quirky subject according to the Case Marking Law. In the Icelandic raising structure (51`), the quirky subject also heads three arcs. But contrary to Faroese, two different arcs are identified by the Case Marking Law and the Final Arc Word Order Rule. The 2-arc is still a final arc and determines word order, but it is not a shallow arc because it is erased by the 3-arc. The erasure of the 2-arc by the 3-arc also changes the status of the 3-arc, which is now a shallow arc and hence determines case marking. Regarding the quirky subject, this means that it behaves as a direct object with respect to word order but as an indirect object with respect to case marking. So the mismatch in Icelandic between the morphological marking of the quirky subject and its positioning is not the result of a special type of case that cannot be overwritten, but simply results from the interaction of a language particular rule, the Icelandic Strict Object Rule, and other independent laws and rules, the Case Marking Law and the Final Arc Word Order Rule.

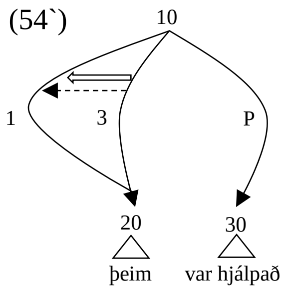
5.3 Passive Quirkies

The Icelandic Strict Object Rule is not only responsible for case preservation under raising in Icelandic, it is also responsible for the peculiarity that Icelandic possesses passive quirkies. Recall from section 4.2 that passives involve advan-cement from some object relation to subject. In case an indirect object[9] in Icelandic undergoes passivization, the Icelandic Strict Object Rule demands that the 3-arc defining the indirect object erase its 1-arc successor. Given the absence of the Icelandic Strict Object Rule in Faroese, the 1-arc erases the 3-arc in Faroese passives. The resulting structure for the passive structures in (54) & (55) for Icelandic and Faroese, respectively, are given in (54`) & (55`).

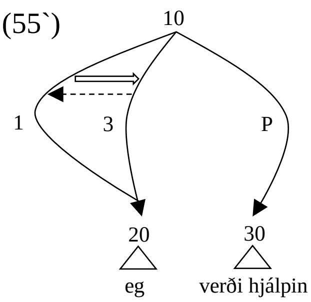
(54) Þeim var hjálpað.
they.DAT was helped
'They were helped.'

(55) Eg verði hjálpin.
I.NOM become helped
'I am helped.'

(54`)



(55`)



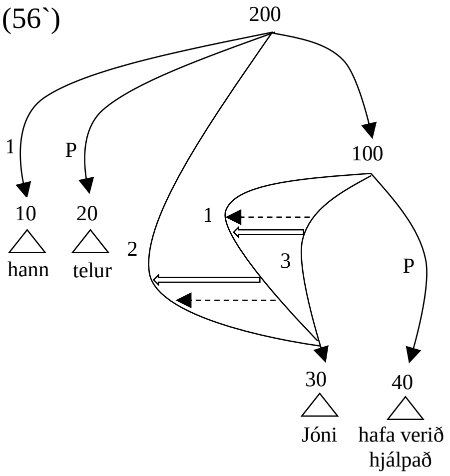
In the Faroese passive (55`), the 1-arc successor erases its 3-arc predecessor, this being the general case for successor-predecessor pairs. In the Icelandic passive (54`), however, this is impossible because the Icelandic Strict Object Rule requires 3-arcs to always be shallow arcs. Hence the 3-arc erases its 1-arc successor. Again, that one of the two has to be erased follows from the Single Mother Law. Note that passive quirkies have a very different structure from lexical quirkies. Passive quirkies are initial strict objects and final subjects,

[9] Passivization of indirect objects is also found outside the Germanic languages, for example in Imbabura Quechua (Postal 1986, Jake 1983) and Ancient Greek (Feldman 1978).

whereas lexical quirkies are initial subjects and final strict objects. Crucially, this difference is irrelevant for most subject tests in Icelandic. Since passive quirkies head a final 1-arc, they also head a working 1 and are predicted to pass all subject tests in Icelandic referencing working 1s. So raising of passive quirkies in Icelandic is possible as they head a final 1 and final 1s are subsumed under working 1s. (56`) illustrates this for a raised passive quirky (cf. 56).

(56) Hann telur Jóni$_i$ [t$_i$ hafa verið hjálpað].
he believes Jón.DAT have become helped
'He believes Jón to have been helped.'

(56`)



The erasure of the 1-arc by the 3-arc follows from the Icelandic Strict Object Rule. The erasure of the 1-arc by the 2-arc follows from the Unique Eraser Law and the Single Mother Law. If the 2-arc erased the lower 1-arc, the 1-arc would have two erasers; and if neither the 2-arc nor the 1-arc were erased, the Single Mother Law would be violated. Similar to raising of a lexical quirky, the passive quirky in (56`) heads three arcs. Of these three, the 3-arc determines case marking, whereas the 2-arc determines word order.

6 Discussion of Alternatives

There appear to exist two alternative ways to handle the data presented in this paper that do not invoke a change of grammatical relations.

The first is to distinguish two types of lexical case, strong and weak lexical case (cf. Holmberg & Platzack 1995: 273). Weak lexical case would differ from strong lexical case in that it can be overwritten and hence be lost under raising and passivization. Icelandic would then possess strong lexical case, whereas Faroese would have weak lexical case. Although this approach successfully captures the differences between Icelandic and Faroese, it fails to handle the German data. On the hand one, German has passives where dative objects are promoted to nominative subjects indicating that lexical case is weak, but German lacks raising of quirky subjects indicating that lexical case is strong. Obviously lexical case cannot be both strong and weak and the same time, as required by this state of affairs. On the other hand, this account offers no solution to the problem why quirky subjects in German pass fewer subject tests than quirky subjects in Icelandic and Faroese.

The other alternative relocates the relation changing operation to the lexicon (Dukes 1999; Manning & Sag 1999) by distinguishing ARG-ST from VALENCE lists. Lexical rules then allow a flexible mapping between these two lists. Alternatively, lexical rules map one lexical entry onto another and preserve the argument structure information of the input lexical entry on the ARG-ST list of the output lexical entry. The distinction between initial and final grammatical relations is then reconstructable through the positions each nominal bears on the respective lists. There are at least three problems with this alternative. On the more conceptual side, preserving information of a related lexical item is hardly compatible with monostratality in the strict sense, namely the idea that the grammatical properties of a sentence and its elements can be described with reference to this sentence alone. For what is preserved on the resulting lexical entry is information about the behavior of a lexical entry used in a *distinct* sentence. The second problem is that grammatical relations are defined via their positions on a list so that subjects of unaccusative and unergative predicates become indistinguishable. Adding an extra device to capture this difference (for example through a D(ESIGNATED) A(RGUMENT) list, cf. Müller 2008) only fixes the defect without actually solving the problem that created it. The third and most serious problem is that this alternative can only reconstruct nominals bearing at most two grammatical relations. However, the Relational Grammar literature documents cases where a nominal bears three grammatical relations. One example for this is described in Jake (1983: 209-217) for quirky subjects in Imbabura Quechua. The interaction between passivization and raising shows that quirky subjects are initial subjects, then demote to indirect object and finally advance to direct object. There is clearly no non-*ad hoc* way of capturing this three-way distinction with two lists only.

7 Conclusion

I have argued in this paper for a syntactic account of quirky subjects according to which they undergo a grammatical relation changing operation from subject to strict object. This account is superior to accounts that locate quirky subjects at the morphology-syntax interface or in the lexicon. Two conclusions can be drawn from this. First, the dichotomy between lexical and structural case is illusory. Nominals bear whatever case they have to bear with respect to one of their grammatical relation. As illustrated by Faroese and German, the question which grammatical operation can affect which nominal is solely determined by the grammatical relations of that nominal and not by its case. Second, the idea that a nominal can bear multiple grammatical relations has also far reaching consequences for subject tests. As argued, subject tests are sensitive to the types of grammatical relations a nominal bears in addition to its bearing a subject relation. Similar to the case of German discussed in this paper, this invites for a reevaluation of claims in the literature as to whether or not some nominal counts as a subject.

References

Aissen, Judith. 1987. *Tzotzil Clause Structure*. D. Reidel Publishing Group.

Barðdal, Jóhanna. 2006. Construction-specific properties of syntactic subjects in Icelandic and German. *Cognitive Linguistics* 17(1), 39-106.

Barðdal, Jóhanna & Þórhallur Eyþórsson. 2003. Icelandic vs. German: Oblique Subjects, Agreement and Expletives. *Proceedings of CSL 39*, 755-773.

Barðdal, Jóhanna & Þórhallur Eyþórsson. 2012. Reconstructing Syntax: Construction Grammar and the Comparative Method. In: Boas, Hans C. & Ivan Sag (eds.), *Sign-Based Construction Grammar*, 257-308. CSLI Publications.

Barnes, Michael. 2001. *Faroese Language Studies*. Novus forlag.

Bayer, Josef. 2004. Non-nominative Subjects in Comparison. In: Bhaskararao, Peri & Karumuri Venkata (eds.), *Non-nominative Subjects, Vol. 1*, 49-76, Benjamins.

Berinstein, Ava. 1986. Indirect object constructions and multiple levels of syntactic representation. *Proceedings of CLS* 22, 36-50.

Blake, Barry J. 1990. *Relational Grammar*. Routledge.

Bouma, Gosse. 1992. A Lexicalist Account of Icelandic Case Marking. *Proceedings of the 14th conference on computational linguistics*, Vol. 1, 94-100.

Davies, William D. 1988. Against functional control. *Lingua* 76, 1-20.

Dukes, Michael. 1999. Kinds of Objecthood in Chamorro Grammar. In: Webelhuth, Gert, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, 15-30, CSLI Publications.

Dziwirek, Katarzyna. 1994. *Polish Subjects.* Garland.

Feldman, Harry. 1978. Passivizing on datives in Greek. *LI* 9(3), 499-502.

González, Nora. 1988. *Object and Raising in Spanish*. Garland.

Haider, Hubert. 2010. *The Syntax of German.* Cambridge University Press.

Harris, Alice. 1981. *Georgian Syntax*. Cambridge University Press.

Holmberg, Anders & Christer Platzack. 1995. *The Role of Inflection in Scandinavian Syntax*. Oxford University Press.

Jake, Janice L. 1983. *Grammatical Relations in Imbabura Quechua.* PhD Thesis, University of Illinois at Urbana-Champaign.

Johnson, David E. & Paul M. Postal. 1980. *Arc-Pair Grammar.* Princeton University Press.

Jónsson, Jóhannes Gísli. 1996. *Clausal Architecture and Case in Icelandic.* PhD Thesis, UMass Amherst.

Jónsson, Jóhannes Gísli. 2003. Not so Quirky: On Subject Case in Icelandic. In: Brandner, Ellen & Heike Zinsmeister (eds.), *New Perspectives in Case Theory*, 129-165, CSLI Publications.

Jónsson, Jóhannes Gísli & Þórhallur Eyþórsson. 2005. Variation in subject case marking in Insular Scandinavian, *Nordic Journal of Linguistics* 28(2), 223-245.

Legendre, Géraldine. 1994. *Topics in French Syntax*. Garland.

Manning, Christopher D. & Ivan A. Sag. 1999. Dissociations between ARG-ST and Grammatical Relations. In: Webelhuth, Gert, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, 63-78, CSLI Publications.

Moore, John & David M. Perlmutter. 2000. What does it take to be dative subject?, *NLLT* 18(2), 373-416.

Müller, Stefan. 2008. *Head-Driven Phrase Structure Grammar*. Stauffenburg.

Pankau, Andreas. 2013. *Replacing Copies. The Syntax of Wh-Copying in German.* LOT Publications.

Perlmutter, David M. (ed.). 1983. *Studies in Relational Grammar 1*. The University of Chicago Press.

Perlmutter, David. M. 1984. Working 1s and Inversion in Italian, Japanese, and Quechua. In: Perlmutter, David M. & Carol Rosen (eds.), *Studies in Relational Grammar 2*, 292-330, The University of Chicago Press.

Perlmutter, David M. & Carol Rosen (eds). 1984. *Studies in Relational Grammar 2,* Chicago: The University of Chicago Press.

Postal, Paul M. 1986. *Studies of Passive Clauses*. SUNY Press.

Postal, Paul M. 2010. *Edge Based Clausal Syntax. A Study of (mostly) English Object Structure.* MIT Press.

Postal, Paul M. & Brian D. Joseph (eds.). 1990. *Studies in Relational Grammar 3*, The University of Chicago Press.

Rosen, Carol & Kashi Wali. 1989. Twin passives, inversion and multi-stratalism in Marathi, *NLLT* 7(1), 1-50.

Sag, Ivan, Lauri Karttunen & Jeffrey Goldberg. 1992. A Lexical Analysis of Icelandic Case. In: Sag, Ivan & Anna Szabolcsi (eds.), *Lexical Matters*, 301-318, CSLI Publications.

Schätzle, Christin, Miriam Butt & Kristina Kotcheva. 2015. The Diachrony of Dative Subjects and the Middle in Icelandic: A Corpus Study. *Proceedings of the LFG15 Conference,* 358-377, CSLI Publications.

Sigurðsson, Halldór. 1989. *Verbal Syntax and Case in Icelandic.* PhD Thesis, University of Lund.

Sigurðsson, Halldór. 1992. The case of quirky subjects. *Working Papers in Scandinavian Syntax* 49(1), 1-26.

Sigurðsson, Halldór. 2002. To Be an Oblique Subject: Russian vs. Icelandic. *NLLT* 20(4), 691-724.

Þráinsson. 2007. *The Syntax of Icelandic.* Cambridge University Press.

Þráinsson, Höskuldur, Hjalmar P. Petersen, Jógvan í Lon Jacobsen & Zakaris Svabo Hansen. 2004. *Faroese: An Overview and Reference Grammar*. Føroya Fróðskaparfelag.

Willgohs, Beatriz & Patrick Farrell. 2009. Case and grammatical functions in Imbabura Quechua: An LFG approach. *Proceedings of the LFG09 Conference*, 624-643, CSLI Publications.

Zaenen, Annie, Joan Maling & Höskuldur Þráinsson. 1985. Case and Grammatical Functions: The Icelandic Passive. *NLLT* 3(4), 441-483.

Integrating a rich external valency dictionary with an implemented XLE/LFG grammar

Agnieszka Patejuk

Institute of Computer Science, Polish Academy of Sciences

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 520–540

Keywords: valence, grammatical functions, grammar engineering, phraseology, coordination

Patejuk, Agnieszka. 2016. Integrating a rich external valency dictionary with an implemented XLE/LFG grammar. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 520–540. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.27.

Abstract

This paper shows how Walenty, a valency dictionary of Polish, was automatically converted in order to be used with an XLE/LFG grammar of Polish, discussing issues such as the grammatical function assignment under unlike category coordination and imposing constraints for lexicalised dependents.

1 Introduction

This paper discusses how Walenty, an innovative valency dictionary of Polish, was used in an LFG grammar of Polish implemented in XLE (Patejuk & Przepiórkowski, 2012b). It begins with introducing the distinctive features of Walenty that make it attractive from the perspective of use in an LFG grammar (§2), then it proceeds to presenting the procedure for interpreting and converting valency specifications from Walenty to the LFG formalism (§3). It describes the procedure of assigning the grammatical function to dependents, taking unlike category coordination into account, and it shows how relevant constraints are formalised using available LFG mechanisms, covering issues such as structural case assignment in Polish, the handling of passive voice and the treatment of optional arguments. Finally, it offers a detailed formalisation of imposing constraints on lexicalised dependents which is capable of accounting for embedding of such lexicalised specifications.

2 Walenty: an innovative valency dictionary of Polish

Walenty (Przepiórkowski et al., 2014b) is currently the largest and most precise valency dictionary of Polish – in October 2016 it contained 85,210 schemata for 16,195 lemmata. Unlike many other dictionaries, it contains not only schemata for verbs, but also for nouns, adjectives and adverbs. For reasons of space, this paper focuses on verbal schemata exclusively: there are over 65,400 schemata for 12,028 lemmata, which gives 5.4 schemata per lemma on average.

Walenty is available on an open source licence (CC BY-SA 4.0) in a variety of formats: plain text, XML and PDF. While XML is used as the input for conversion, schemata are presented below in plain text format in order to save space.

2.1 Grammatical function labels

Walenty explicitly identifies the subject position (`subj`) – understood as the argument that drives verbal agreement, regardless of its category (so it takes into account non-canonical subjects) – and the object (`obj`) – defined as the argument which can become the subject under passive voice, also regardless of its category (and case marking, if the passivisable argument is nominal).

[†]This research is partially supported by the Polish Ministry of Science and Higher Education within the CLARIN ERIC programme 2016–2018 (http://clarin.eu/). The author is grateful to anonymous reviewers for their valuable comments which made it possible to improve this paper.

The verb MANIPULOWAĆ ‘manipulate’ is an example of a predicate taking a nominal object marked for the instrumental case (`inst`), as shown in (2), while (3) demonstrates that this argument can become the subject under passive voice. The schema for MANIPULOWAĆ is given in (1) – it contains a subject marked for the structural case (`str`, see §2.3) and an instrumental object:

(1) `subj{np(str)} + obj{np(inst)}`

(2) Manipulowała mną i swoim późniejszym mężczyzną.
manipulated I.INST and SELF.INST later.INST man.INST
‘She manipulated me and her later man.’ (http://nkjp.pl)

(3) Młodzi ludzie byli manipulowani przez starsze osoby.
young.NOM people.NOM were manipulated.NOM by elder persons
‘Young people were manipulated by elder people.’ (http://nkjp.pl)

Since Walenty does not distinguish grammatical functions other than subject and object as defined above,[1] other positions are not labelled with any grammatical function and it must[2] be assigned during conversion – this is discussed in §3.1.

2.2 Syntactic positions as sets

In Walenty a syntactic schema is a list of positions (separated by +) modelled as sets of categories that can realise a given position. The set contents, enclosed in curly brackets (`{...}`), with particular elements separated by semi-colons (`;`), are specified according to the coordination test of Szupryczyńska 1996: if two or more categories can be coordinated within one position, then it is a multi-element set. For instance, in (4) the subject of BAWIĆ ‘amuse’ consists of a nominative noun phrase (*ta gra*) coordinated with a clause (*że tylu ludzi [...] dało się nabrać*), so the corresponding `subj` position in the schema in (5) is a two-element set containing the `np(str)` and `cp(że)` categories. When such coordination is not possible, a singleton set is used – the `subj` in (7), the schema for AKLIMATYZOWAĆ SIĘ ‘acclimatise’, contains only `np(str)`: a nominal marked for structural case (see §2.3). Note that the second argument of (5), `np(str)`, is a singleton set; since this position is not marked as an `obj`, it is assumed that it cannot passivise.

(4) Trochę bawiła mnie ta gra i że tylu ludzi [...]
a little amused me.ACC this.NOM game.NOM and that so many people
dało się nabrać
let REFL take in
‘This sham and (the fact) that so many people let themselves be cheated amused me a little.’ (http://nkjp.pl)

(5) `subj{np(str);cp(że)} + {np(str)}`

(6) `subj{np(str)} + {np(str)}`
`subj{cp(że)} + {np(str)}`

[1]See Patejuk & Przepiórkowski 2016 for arguments supporting such a solution.
[2]Unless an alternative approach to grammatical functions is adopted: see Patejuk & Przepiórkowski 2016 and Przepiórkowski 2016.

(7) `subj{np(str)} + {xp(locat)}`

Modelling syntactic positions as sets is innovative in that it explicitly accounts for the coordination of unlike categories. Other valency dictionaries would use separate valency schemata for different categories, as in (6), one with a nominal subject (`np(str)`) and another one with a clausal subject (`cp(że)`), which can be interpreted in two ways. If it is an `XOR` (exclusive `OR`) specification of the subject (either a nominal phrase or a clause), the possibility of having a coordinated unlike category subject such as in (4) is ruled out. If is is interpreted as an `OR` specification, it allows for such coordination at the cost of overgenerating (allowing such coordination when it is not possible). Such problems can be avoided by adopting the solution proposed in Walenty, where syntactic positions are modelled as sets which correspond to an `OR` specification, as in (5), which means that the given position can be filled by any single set element (only nominal or only clausal) or by any coordination of these elements, which accounts for unlike category coordination. If the position can be filled in more than one way but the relevant elements cannot be coordinated, the `XOR` specification is obtained by creating separate schemata with singleton sets corresponding to the relevant argument, as in (6).

Note that Walenty uses container categories such as `xp(locat)` in (7) – though it is a singleton set in the schema, it is equivalent to a multi-element set containing all the categories listed in its corresponding list of realisations (see §2.5).

2.3 Arguments marked for structural case

Walenty provides an explicit account of structural case in Polish. Unlike lexical case, which is stable in the sense that it is independent of the syntactic context, structural case is understood here as a case which may take different values depending on the syntactic environment – such arguments have the `str` value of case. The information supplied by the valency dictionary is to be processed by the grammar so as to assign an appropriate case in the given context.

When a subject is marked for structural case, its case marking may be realised in three ways.[3] The first possibility is the nominative case, the most prototypical value – it is appropriate for subjects of finite verb forms which are not non-agreeing numerals. The second value is the accusative case – it is possible when the subject of a finite verb form is a non-agreeing numeral. Finally, the third possible value is the genitive case – this is the case with the subject of gerunds.

When an object (passivisable or not) bears structural case, there are two possible values: accusative or genitive, depending on the availability of sentential negation and part of speech of the head assigning case. Gerunds require the genitive case from their structural objects regardless of negation. With other verbal forms, genitive is required when the verb assigning structural case is in the scope of sentential negation; this phenomenon is known as genitive of negation (GoN). If negation is local, GoN is obligatory, while with non-local negation (present higher in the verb

[3]Note that this does not apply to the subject of adjectival participles (determined by agreement) and infinitives (determined by control).

chain), GoN is optional (Patejuk & Przepiórkowski, 2014b). As a result, accusative is required as structural case when negation is not present at all and it is possible when it is non-local to the predicate assigning structural case.[4]

2.4 Control, raising, predicative elements

Walenty explicitly accounts for raising and control by using `controller` and `controllee` labels to establish relations between respective arguments. In (8), the schema for the verb BAĆ SIĘ 'fear, be afraid', the subject position (`subj`) containing a nominal marked for structural case (`np(str)`) is labelled as the `controller`, while the set containing the infinitival complement (`infp(_)`) is labelled as `controllee` – such notation expresses subject control, whereby the subject of BAĆ SIĘ controls the subject of the infinitive.[5]

(8) `subj,controller{np(str)} + controllee{infp(_)}`

These labels are also used to mark control relations with predicative arguments – the argument that controls the predicative element is marked as `controller`, while the predicative element is marked as `controllee`. Apart from representing what the predicative element refers to, such information may be used for ensuring proper agreement where it is applicable – as in (9), the schema for the verb UCHODZIĆ 'pass (as)', where the predicative adjective inside the prepositional phrase (`prepadjp(za,acc)`) agrees in number and gender with the subject:

(9)
```
subj,controller{np(str)}
+ controllee{prepadjp(za,acc)}
```

2.5 Semantically defined arguments

Walenty introduces a class of `xp` arguments – defined by their semantics rather than category: these include ablative, adlative, locative (see (7)), etc., arguments. For each type of `xp`, there is a defined list of its realisations (see (10),[6] with translations on the right), which results in economic, readable and coherent schemata.

(10)
```
xp(locat)-->                    xp(locat)-->
      advp(locat)                     advp(locat)
      [...]                           [...]
      cp(int[gdzie])                  cp(int[where])
      [...]                           [...]
      prepnp(koło,gen)                prepnp(near,gen)
      prepnp(między,inst)             prepnp(between,inst)
      prepnp(nad,inst)                prepnp(above,inst)
      prepnp(na,loc)                  prepnp(on,loc)
      [...]                           [...]
```

While using a plain `xp` means that all its realisations are possible with a given schema (so the given `xp` corresponds to a set containing all its realisations), sometimes it is the case that a given predicate does not allow all the realisations, though

[4]Some predicates in Polish allow partitive objects – these are covered in Walenty and treated as a variant of structural case, where genitive is additionally allowed in partitive use.

[5]It is assumed that in verb control it is always the subject of the `controlee` that is controlled.

[6]The `int` parameter in `cp(int[gdzie])` stands for interrogative – it is an interrogative clause where the interrogative item is GDZIE 'where'.

the realisations it subcategorises for have the required semantics. If the selected realisations were listed as elements of the set corresponding to the relevant argument position without stating that it is an `xp` phrase with specific semantics, the semantic information would be lost, which would be an unwelcome result (as a realisation of `xp`, `prepnp` is treated as semantic, while outside of `xp` it is non-semantic). On the other hand, classifying such an argument as a "plain" `xp` would allow all its realisations, which is also undesired as it leads to overgeneration. Walenty was designed so as to provide maximal precision of the description of valency requirements, so it offers a subtype mechanism which makes it possible to restrict the range of realisations of a given phrase (here, an `xp` phrase) to those that are applicable in the given schema using a list. For example, the schema for KIEŁKOWAĆ 'sprout' in (11) contains an `xp(adl)` phrase whose realisation list is restricted to two prepositional phrases, headed by SPOD 'from underneath' and Z 'from':

(11)
```
subj{np(str)}
  + {xp(abl[prepnp(spod,gen);prepnp(z,gen)])}
```

Elements of the subtype list, enclosed in square brackets (`[...]`), are separated by semicolons (`;`) since they may be coordinated (as in `xp` without subtypes).

The subtype mechanism is also used with clauses (`cp`): while `cp(int)` is an interrogative phrase with any interrogative element defined on the relevant list, this element may be specified using the subtype list, as in (10) (cf. fn. 6).

2.6 Lexicalised dependents

Last but not least, Walenty is one of the few valency dictionaries that include a rich phraseological component (Przepiórkowski et al., 2014a) – it explicitly specifies lexicalised arguments and constraints imposed on them, with the possibility of embedding such constraints arbitrarily deep, as in (13), the schema for the verb WITAĆ 'welcome' used in (12):

(12) Oni witali ją z (szeroko) *(otwartymi) ramionami.
they.NOM welcomed she.ACC with widely open.INST.PL arm.INST.PL
'They welcomed her with (widely) open arms.' (== very warmly)

(13)
```
subj{np(str)} + obj{np(str)} + {xp(mod);
lex(prepnp(z,inst),pl,XOR('ramię','ręka'),ratr1(
  {lex(adjp(agr),agr,agr,pos,'otwarty',atr1(
    {lex(advp(mod),pos,'szeroko',natr)}))}))}
```

There are three arguments in (13), out of which the first two, subject and object, are not lexicalised. The last one is a two-element set containing `xp(mod)` and a lexicalised (`lex`) prepositional nominal phrase (`prepnp`) with the preposition Z 'with' which requires instrumental case (`inst`) from the nominal which must in turn be specified for plural number (`pl`) and must be a form of either RAMIĘ 'arm' or RĘKA 'hand' (`XOR` specification). This lexicalised nominal must be modified by exactly one dependent (`ratr1`), an embedded `lex` specification follows: an agreeing adjectival phrase (`adjp`) headed by OTWARTY 'open', which may in turn be optionally (`atr1`) modified by a lexicalised adverbial phrase (`advp`) headed by SZEROKO 'widely', which must not be modified (`natr`).

3 Interpreting Walenty and converting it to LFG

Since Walenty uses its own formalism, it is not tied by the constraints of any particular grammar formalism and it can be used with any grammar or grammar engineering platform, provided that the grammar writer is able to interpret the specifications provided in Walenty and convert them to the relevant formalism. This section shows how this can be done for LFG on the basis of selected phenomena.

The conversion is done automatically using a Python script which ensures consistency and coherence. While schemata from Walenty are presented in plain text format, the script relies on the XML format of Walenty.

The general idea of converting a valency schema to LFG constraints is very simple: each argument must be assigned a grammatical function and then appropriate constraints relevant to this argument must be imposed.

3.1 Selecting the grammatical function

Since the grammatical function must be chosen to apply relevant constraints, let us start with the procedure of selecting the grammatical function. As mentioned in §2.1, only two grammatical functions are specified in Walenty: the subject (SUBJ) and the passivisable object (OBJ). The remaining grammatical functions are not specified in Walenty, so they are assigned using the following mapping:

- OBJ$_\theta$: thematic/secondary object – nominal, it does not passivise,
- OBL (oblique): non-semantic prepositional phrase,[7]
- OBL$_\theta$ (thematic oblique): semantic prepositional phrase,
- COMP: closed clausal complement,
- XCOMP: open infinitival complement,
- XCOMP-PRED (predicative complement): open predicative nominal or adjective (possibly embedded in a prepositional phrase).[8]

The specification outlined above works perfectly as long as the relevant argument position contains only one realisation (it is a singleton set in Walenty). If this is not the case and unlike category coordination is possible, as in (14), where a clause is coordinated with a prepositional phrase (see the last position in (15), the schema for the verb PYTAĆ 'ask'), the choice of the grammatical function becomes problematic because different categorial realisations of the relevant argument position seem to correspond to different grammatical functions.

(14) Pytali, [jakie będą pieniądze] oraz [o to, czy zmienią się polskie
asked what will be money and about this PART change REFL Polish
szkoły].
schools
'They asked what money will be there and whether Polish schools will change.' (http://nkjp.pl)

[7] If there is more than one such phrase, a numeric index is appended, yielding OBL2, OBL3, etc.

[8] As an alternative, the closed PREDLINK grammatical function could be used.

(15) `subj{np(str)} + obj{np(gen)}`
`+ {prepnp(o,acc); cp(int); prepncp(o,acc,int)}`

Typically, a coordinated phrase corresponds to one grammatical function in f-structure, so a common grammatical function should be chosen. Since, according to the mapping provided above, a prepositional phrase (`prepnp`, `prepncp`) should be assigned the OBL grammatical function and an interrogative clausal complement (`cp(int)`) – the COMP grammatical function, which of these two grammatical functions should be assigned to their coordination in (14)? An analogous problem has been discussed in LFG literature in the context of OBJ as a candidate grammatical function under coordination: Dalrymple & Lødrup 2000 suggest COMP should be treated as an elsewhere grammatical function, used when only the clausal complement is possible and it cannot be coordinated with a different category. The current conversion of Walenty is inspired by this solution – it uses the ranking of grammatical functions defined in (16) to choose the common grammatical function from the set of candidates: the conversion script assigns each realisation of the relevant argument position (each element of the set) the corresponding ranking and then the highest ranked grammatical function candidate is chosen.

(16)

#	GF
4	OBL-<SEM: ABL, ADL, DUR, INSTR, LOCAT, MOD, PERL, TEMP. . . >
3	OBL
2	OBJ-<CASE: DAT, GEN, INST, STR>
1	COMP, XCOMP

According to the ranking in (16), if an argument position can be realised as a non-semantic prepositional phrase (OBL) or as a clause (COMP), as in (15), it should be assigned the OBL grammatical function. The XCOMP and COMP are the lowest ranked grammatical functions: they are only selected when the clause or infinitive are the only realisations in the set corresponding to the relevant argument position.

3.2 Imposing constraints

Once the grammatical function corresponding to a given syntactic position (the entire set) has been chosen, appropriate constraints are imposed for each realisation of the relevant syntactic position defined in the schema (each element of the set).

The method of formalising constraints corresponding to a given argument position depends on one crucial factor – whether the given position involves unlike category coordination or not. When such coordination is not involved, it is sufficient to use plain constraints such as in (17) and (18). However, when unlike category coordination is allowed, which requires that the argument GF must either have the attribute ATTR1 with V1 as its value, or the attribute ATTR2 whose value is V2, it cannot be formalised using the disjunction of two plain constraints such as in (19), because, instead of yielding the logical `OR` specification, the result will be the undesired `XOR` specification.

(17) $(\uparrow \text{GF ATTR1}) =_c \text{V1}$ (18) $(\uparrow \text{GF ATTR2}) =_c \text{V2}$

(19) $(\uparrow \text{GF ATTR1}) =_c \text{V1} \ \vee \ (\uparrow \text{GF ATTR2}) =_c \text{V2}$

As explained in Przepiórkowski & Patejuk 2012, when a plain disjunctive constraint is used, it is evaluated once (one disjunct is chosen) and applied to all conjuncts, as formalised in (20a). By constrast, the interpretation which is needed to handle unlike category coordination is the one formalised in (20b) – it evaluates the relevant statement for each conjunct separately, so it is possible that different conjuncts satisfy different constraints.

(20) a. $\forall x \in (\uparrow \text{GF}) A(x) \ \vee \ \forall x \in (\uparrow \text{GF}) B(x)$ (actual)

b. $\forall x \in (\uparrow \text{GF}) [A(x) \ \vee \ B(x)]$ (intended)

The solution to this problem described in Przepiórkowski & Patejuk 2012 relies on the use of off-path constraints in order to obtain the effect shown in (20b). In short: constraints to be imposed on a given argument are converted to their off-path equivalent and they are attached to the PRED attribute of the relevant argument – this attribute is distributive by definition, which ensures that the disjunctive off-path constraint will be distributed to each conjunct and evaluated separately.

There is a crucial difference in the formalisation of off-path constraints between LFG theory and XLE implementation: unlike in recent theoretical LFG works (including recent versions of Dalrymple 2001 and Bresnan 2001), off-path constraints are non-constructive in XLE,[9] which means that they can only be constraining or existential, but they cannot be defining – they cannot introduce new attribute-value pairs to the f-structure. As a result, constraints placed on certain attributes must be formalised as constraining equations – rather than defining ones – regardless of whether the constraint is off-path or plain, for the sake of consistency.

3.2.1 Structural case assignment

As mentioned in §2.3, Walenty provides information about the requirement of structural case and the grammatical function of the relevant argument, which is processed by the grammar, taking the syntactic context into account, in order to set the appropriate values of case. As discussed in Patejuk & Przepiórkowski 2014b for verbal heads in Polish, the structural object is marked for accusative case in the absence of negation and genitive case if negation is present – the proposed formalisation (see (21)) uses plain constraints, so it is not compatible with unlike category coordination. A formalisation of structural case assignment that does take this into consideration and uses off-path constraints is provided in Patejuk & Przepiórkowski 2014a (compare (22)):[10]

(21)
$$\begin{array}{rl}\text{STRCASE(GF)} \equiv & [[\neg((\text{XCOMP}^* \uparrow)\ \text{NEG}) \ \wedge \ (\uparrow \text{GF CASE}) =_c \text{ACC}] \ \vee \\ & \quad [((\text{XCOMP}^* \uparrow)\ \text{NEG}) =_c + \ \wedge \\ & \quad [[(\uparrow \text{NEG}) =_c + \ \wedge \ (\uparrow \text{GF CASE}) =_c \text{GEN}] \ \vee \\ & \quad [\neg(\uparrow \text{NEG}) \ \wedge \ (\uparrow \text{GF CASE}) \in_c \{\text{ACC, GEN}\}]]]]\end{array}$$

[9] See the relevant part of the XLE documentation: http://www2.parc.com/isl/groups/nltt/xle/doc/notations.html#N4.1.5b

[10] See the corresponding papers for a detailed discussion of relevant constraints.

(22) $(\uparrow \text{GF} \qquad\qquad \text{PRED} \qquad\qquad)$

$$[[\neg((\text{XCOMP}^* \ \text{GF} \leftarrow) \ \text{NEG}) \wedge (\leftarrow \text{CASE}) =_c \text{ACC}] \vee$$
$$[((\text{XCOMP}^* \ \text{GF} \leftarrow) \ \text{NEG}) =_c + \wedge$$
$$[[((\text{GF} \leftarrow) \ \text{NEG}) =_c + \wedge (\leftarrow \text{CASE}) =_c \text{GEN}] \vee$$
$$[\neg((\text{GF} \leftarrow) \ \text{NEG}) \wedge (\leftarrow \text{CASE}) \in_c \{\text{ACC}, \text{GEN}\}]]]]$$

Such constraints are placed in the lexical entry of the relevant verb – though it is less economic than placing their equivalents in c-structure rules, it allows for an appropriate treatment of implicit arguments,[11] unlike category coordination (whereby only some conjuncts are marked for structural case) and dependent sharing (whereby the shared dependent is assigned structural case by only some of the coordinated verbs). While the constraints are complex, they may be assigned to a template and then short template calls may be used in particular lexical entries, which is considerably more economic.

3.2.2 Complex constraints for clausal phrases

Walenty has two complementiser types (`żeby2` and `gdy`) which are realised as different complementisers (ŻE or ŻEBY for `żeby2`; GDY or GDYBY for `gdy`) depending on the syntactic context – this phenomenon may be thought of as similar to structural case assignment. If this analogy is accepted, the remaining complementiser types (such as `żeby`, `że`, `jeśli` and many more) may be considered lexical since they have the same form (ŻEBY, ŻE and JEŚLI, respectively) regardless of the syntactic environment, which includes factors such as the availability of negation (discussed below for `żeby2`) and mood (`gdy` is sensitive to conditional mood).

The clausal phrase `cp(żeby2)` is different from `cp(żeby)` since the former can be realised in two ways: always as ŻE and as ŻEBY only in scope of sentential negation.[12] Consider the following examples, which illustrate the schema for the verb WYOBRAZIĆ 'imagine' provided in (25):

(23) Ja *(**nie**) mogę sobie wyobrazić, **żeby** ktoś mógł zrobić coś takiego.
I NEG can SELF.DAT imagine that sb could do.INF sth such
'I cannot imagine that somebody could have done something like this.'
(http://nkjp.pl)

(24) Ja (**nie**) mogę sobie wyobrazić, **że** ktoś mógł zrobić coś takiego.
I NEG can SELF.DAT imagine that sb could do.INF sth such
'I can (not) imagine that somebody could have done something like this.'

(25)
```
subj{np(str)} + {np(str);cp(int);cp(żeby2);
ncp(str,int);ncp(str,żeby2)}
+ {lex(np(dat),_,'siebie',natr)}
```

[11] Assigning case to implicit arguments using c-structure rules would require placing such constraints on the verb directly – which has the same effect as placing them in the lexical entry.

[12] ŻEBY may also be used as the realisation of `cp(żeby2)` in generally negative contexts such as in (i), where the verb WYOBRAZIĆ 'imagine' takes `cp(żeby2)` as one of its arguments:

(i) **Z trudem** mogę sobie wyobrazić, **żeby**...
with difficulty can SELF.DAT imagine that
'It is only with difficulty that I can imagine that...'

The plain constraint corresponding to `cp(żeby2)`, one of the realisations of the second argument of (25), is provided in (26) – it states that the complementiser ŻEBY is only possible when negation is present, as in (23), while ŻE is possible at all times (there are no constraints on the value of NEG), as in (24):

(26) $[(\uparrow \text{NEG}) =_c + \;\wedge\; (\uparrow \text{GF COMP-FORM}) =_c \text{ŻEBY}] \;\vee$
$(\uparrow \text{GF COMP-FORM}) =_c \text{ŻE}$

3.2.3 Passive voice

Another issue that is worth discussing is the method of handling passive voice. LFG grammars typically use a lexical rule, but an alternative method is used when converting Walenty – passive versions of schemata are created using the script.

In XLE the passive lexical rule manipulates the assignment of grammatical functions using string substitution: OBJ $\rightarrow$ SUBJ – the active object becomes the passive subject; SUBJ $\rightarrow$ OBL-AG/NULL – the active subject becomes the passive oblique agent or it is dropped. Such a rule is capable of changing control relations: ($\uparrow$ OBJ)=($\uparrow$ XCOMP SUBJ) is the control equation used the active verb TEACH, whereby the object of TEACH is at the same time the subject of the infinitival complement of TEACH, while in the passive it is ($\uparrow$ SUBJ)=($\uparrow$ XCOMP SUBJ) – because the active OBJ becomes the passive SUBJ.

Unfortunately, when applied to constraints which depend on the grammatical function, such a lexical rule has undesired effects. When the active verb takes an object marked for structural case (accusative or genitive, see §3.2.1), the case constraint will be imposed on the passive subject (simplifying, typically nominative), which results in ungrammaticality. It is, however, easy to introduce such changes in the process of conversion: when the passive version of the relevant schema is created, the script first changes the assignment of grammatical functions and then imposes the constraints, which results in changing all the appropriate constraints.

Furthermore, this method makes it possible to introduce more complex additional constraints where it is appropriate: for instance, when the active subject may only be a clause, it could not be the complement of the OBL-AG *by*-phrase because the preposition requires a certain value of case from its complement, which is normally a nominal. In this situation, a correlative pronoun might be added in the passive, resulting in a well-formed *by*-phrase.

3.2.4 Argument reduction

The next issue that must be considered when converting Walenty is the issue of argument reduction: by design, Walenty only provides maximal schemata (listing all possible arguments), but at the same time it assumes that all arguments are optional – in Polish most arguments may be dropped in the sense that they are not expressed. This is illustrated below: the schema for the verb DOWODZIĆ 'command' provided in (27) contains two arguments – a subject and a passivisable object. (28) shows that both arguments can be realised lexically, but they may also be omitted.

(27) `subj{np(str)} + obj{np(inst)}`

(28) (Mój ojciec) dowodził (siłami republikańskimi).
my.NOM father.NOM commanded forces.INST republican.INST
‘My father commanded republican forces.’ (http://nkjp.pl)

When performing the conversion, one must decide how to interpret this phenomenon in an implemented grammar. One way is to assume that the argument is present but it is not realised lexically – in this way the argument is represented syntactically, the relevant grammatical function attribute is present, but its value is ‘PRO’ – see the f-structure in (29), which corresponds to (28) with the object dropped. The alternative approach is to assume that the relevant argument is removed, that it is not present in the f-structure of the verb – this solution involves the creation of reduced frames, which have fewer arguments than the maximal frame, as in (30), where the object is removed from the list in PRED attribute.

(29)
$$\begin{bmatrix} \text{PRED} & \text{‘COMMAND}\langle\boxed{1},\boxed{2}\rangle\text{’} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{‘FATHER’} \\ \text{CASE} & \text{NOM} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 3 \\ \text{ADJ} & \left\{\begin{bmatrix} \text{PRED} & \text{‘MY’} \end{bmatrix}\right\} \end{bmatrix} \\ \text{OBJ} & \boxed{2}\begin{bmatrix} \text{PRED} & \text{‘PRO’} \\ \text{CASE} & \text{INST} \end{bmatrix} \end{bmatrix}$$

(30)
$$\begin{bmatrix} \text{PRED} & \text{‘COMMAND}\langle\boxed{1}\rangle\text{’} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{‘FATHER’} \\ \text{CASE} & \text{NOM} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 3 \\ \text{ADJ} & \left\{\begin{bmatrix} \text{PRED} & \text{‘MY’} \end{bmatrix}\right\} \end{bmatrix} \end{bmatrix}$$

The proposed method of interpreting Walenty uses a hybrid solution – it divides arguments into two classes: obligatory (must be present in syntactic representation) and optional (can be removed from syntactic representation).

First, if the absence of an argument changes the meaning of the predicate – as in the case of lexicalised arguments and the SIĘ marker, which can be reflexive, reciprocal or inherent (in the last case it carries no semantic information, but it is required syntactically as in BAĆ SIĘ in (8), which means ‘to fear’, not ‘to fear oneself’) – then the argument is assumed to be obligatory and it must be lexical (overtly expressed).

The second diagnostic is whether there is syntactic evidence that the relevant argument is syntactically active even though it has no surface realisation. There is evidence which supports implicit subjects and implicit controllers. As shown below using subscript indices, in Polish it is the subject which binds[13] the SIEBIE anaphor (see (31)) and controls participles (see (32)). If the subject were removed from the schema, sentences with no lexical subject could not be parsed (because the subject position would have no value, resulting in incompleteness) and would be expected to be ungrammatical, counter to fact:

[13] With the exception of reciprocal predicates – in (i) *sobie* is bound by the object, *sąsiadów*:

(i) Przedstawił sobie$_s$ (nawzajem) sąsiadów$_s$.
introduced SELF reciprocally neighbours
‘He introduced the neighbours to one another.’

(31) (Antek$_a$) opowiedział Erykowi$_e$ o sobie$_{a/*e}$.
Antek.NOM told Eryk.DAT about SELF
'(Antek) told Eryk about himself.'

(32) Wychodząc$_{a/*e}$, (Antek$_a$) pocieszał Eryka$_e$.
leaving Antek.NOM comforted Eryk.ACC
'Leaving, (Antek) was comforting Eryk.'

The second group of arguments which may be implicit are controllers of infinitives and predicative complements – the reason for having implicit arguments is the same as for controlling participles: the subject of the controlled element is structure-shared with the controller, so the controller must be present in the f-structure. In this case, however, the controller may be different than the subject, see the examples below:

(33) Dowódca kazał (nam wszystkim) uciekać.
commander.NOM ordered us.DAT all.DAT escape.INF
'The commanding officer ordered us all to run away'. (http://nkjp.pl)

(34) Antek zawsze uczyni (Eryka) szczęśliwym.
Antek always make Eryk happy
'Antek will always make (Eryk) happy.'

According to the schema in (35), the controller of the infinitival complement of the verb KAZAĆ 'order' in (33) is the dative nominal. By contrast, the schema in (36) specifies the passivisable object marked for structural case as the controller of the predicative complement of the verb UCZYNIĆ 'make' in (34).

(35) `subj{np(str)} + controller{np(dat)}`
`+ controllee{cp(żeby); infp(_)}`

(36) `subj{np(str)} + obj,controller{np(str)}`
`+ controllee{adjp(inst)}`

In (33)–(34) controllers may have no lexical realisation, they are nevertheless required by syntax (controlled phrases must have controllers), so they are represented in the f-structure representation as implicit arguments ('PRO' is the value of their PRED attribute) – the f-structure in (37) corresponds to (33), while (38)[14] provides a representation of (34) without the lexical object:

(37)
$$\begin{bmatrix} \text{PRED} & \text{'ORDER}\langle\boxed{1},\boxed{2},\boxed{3}\rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{'COMMANDER'} \\ \text{CASE} & \text{NOM} \end{bmatrix} \\ \text{OBJ}_\theta & \boxed{2}\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{CASE} & \text{DAT} \end{bmatrix} \\ \text{XCOMP} & \boxed{3}\begin{bmatrix} \text{PRED} & \text{'RUN\_AWAY}\langle\boxed{2}\rangle\text{'} \\ \text{SUBJ} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

(38)
$$\begin{bmatrix} \text{PRED} & \text{'MAKE}\langle\boxed{1},\boxed{3}\rangle\boxed{2}\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{'ANTEK'} \\ \text{CASE} & \text{NOM} \end{bmatrix} \\ \text{OBJ} & \boxed{2}\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{CASE} & \text{ACC} \end{bmatrix} \\ \text{XC-PRED} & \boxed{3}\begin{bmatrix} \text{PRED} & \text{'HAPPY}\langle\boxed{2}\rangle\text{'} \\ \text{SUBJ} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

Such implicit arguments are introduced optionally (in brackets), so as not to block lexical realisations of the relevant argument – see (39):

(39) $((\uparrow \text{GF PRED}) = \text{'PRO'} \wedge (\uparrow \text{GF CASE}) = \text{CASE\_VALUE})$

[14] For typesetting reasons, XC-PRED is used in (38) instead of XCOMP-PRED.

(40) $((\uparrow \text{GF PRED}) = \text{'PRO'} \wedge [(\uparrow \text{GF CASE}) = \text{ACC} \vee (\uparrow \text{GF CASE}) = \text{GEN}])$

The first conjunct in (39) introduces an implicit argument (PRO) as the value of GF, while the second one assigns case to this argument – in accordance with respective constraints from Walenty. When the implicit argument is marked for structural case, the constraint in (40) is used instead.[15] There is no need to introduce the implicit subject – this is done by the grammar rules (at the level of c-structure).

When none of the criteria presented above is satisfied, the relevant argument is assumed to be optional and it may be reduced – this is done by removing it from the PRED attribute and removing the respective constraints that apply to it. Removing arguments in such a way requires care because controllers must not be removed unless the corresponding controllee is removed. However, once the controllee is removed, the `controller` label is removed from the controller and then it can also be reduced (unless is it a subject – as explained above, it is assumed that subjects do not undergo reduction).

An alternative approach to argument reduction would be to introduce implicit PRO arguments for all arguments, but this would result in implicit clauses and prepositional phrases, which would introduce a lot of additional ambiguity – many predicates take both and a parse would be created for each such argument. Besides, there seems to be no syntactic evidence for introducing such implicit arguments.

3.2.5 Lexicalised dependents: formalisation of modification patterns[16]

Only one aspect of the formalisation of lexicalised dependents is discussed in this section, namely the constraints corresponding to the modification pattern defined in Walenty – these include:

- `natr`: no further modification,
- `atr[(...)]`: modification allowed (optional), it may be iterated,
- `atr1[(...)]`: at most one modifier allowed,
- `ratr[(...)]`: modification required (obligatory), it may be iterated,
- `ratr1[(...)]`: exactly one modifier required.

Apart from `natr` which precludes any modification,[17] the modification pattern symbol can be followed by a list of dependents (its optionality is expressed using square brackets: `[...]`) whose elements are separated by `+`, as in "top level" Walenty schemata. They may be non-lexicalised or lexicalised – in the latter case an embedded `lex` specification is used, it can be arbitrarily deep.

The **`natr`** modification pattern, forbidding any dependents, is converted as the negative constraint in (41) (plain) or in (42) (off-path),[18] where `PATH` is the f-

[15] The values of CASE introduced by (40) are constrained by equations discussed in §3.2.1.

[16] Since §3.2.5 and §3.2.6 are implementational, they use XLE notation. See http://www2.parc.com/isl/groups/nltt/xle/doc/notations.html#N0A for the complete notation mapping used by XLE.

[17] Though the word "modification" is used, the "modifier" is understood as any dependent: it may either be an argument or an adjunct – this is not specified in Walenty as it is assumed that this is restricted by the lexical entry of the lexicalised dependent.

[18] Off-path counterparts of subsequent plain statements may be provided without comments.

structure path leading to the lexicalised dependent, while `GFALL` is defined as in (43)[19] – as the disjunction of all grammatical functions used in the grammar.[20]

(41) `~(PATH GFALL)`

(42) `(PATH PRED: ~(<- GFALL))`

(43) `GFALL = {SUBJ|OBJ|OBL(-?*)|(X)COMP|(X)ADJUNCT|POSS}`

The **`ratr`** specification requires a dependent which may be constrained (an embedded argument list is provided then) or not. In the latter case it is assumed that it may be any dependent allowed by the particular head – the constraint in (44) uses the `GFDEP` variable, which is resolved to the disjunction of grammatical functions allowed by the given head.

(44) `(PATH GFDEP)` (45) `(PATH PRED: (<- GFDEP))`

On the other hand, when the dependent is given explicitly (as an element of the embedded list), the schematic constraint in (46) is used, where `GFDEP` is the grammatical function of the particular dependent, chosen according to its specification, while `ATTR` stands for the relevant attribute and `v` for its required value.

(46) `(PATH GFDEP ATTR)=c v`

(47) `(PATH PRED: (<- GFDEP ATTR)=c v)`

When there is more than one element on the list of possible modifiers, the following constraints are used: the first line in (48) is the disjunctive constraint where particular disjuncts contain existential equations requiring the grammatical functions which correspond to particular dependents on the list inside `ratr`. Its purpose is to satisfy this modification requirement by ensuring that at least one of the listed required dependents is present. The following lines contain disjunctive constraints for each of the dependents on the list (`GFDEP1` for the first one, etc.) which ensure that either the dependent corresponding to the given grammatical function is present and it satisfies the relevant requirements (the positive first disjunct – it corresponds to (46)) or that it is not present (the negative second disjunct).

(48)
```
{(PATH GFDEP1) | (PATH GFDEP2) | ...}
{(PATH GFDEP1 ATTR)=c v | ~(PATH GFDEP1)}
{(PATH GFDEP2 ATTR)=c v | ~(PATH GFDEP2)}
...
```

(49)
```
(PATH PRED:
       {(<- GFDEP1) | (<- GFDEP2) | ...}
       {(<- GFDEP1 ATTR)=c v | ~(<- GFDEP1)}
       {(<- GFDEP2 ATTR)=c v | ~(<- GFDEP2)}
       ...)
```

Finally, it is necessary to block all dependents other than those specified in `ratr` – in (50) the `GFDEP` variable corresponds to a disjunction of all grammatical functions allowed in `ratr`, while `GFALL` corresponds to all grammatical functions possible with the given head.

(50) `~(PATH GFALL - GFDEP)`

[19] The actual version accepted by XLE does not use regular expressions.

[20] The expansion of `GFALL` could be narrowed down to grammatical functions possible with the given head, reducing the number of disjuncts.

(51) `(PATH PRED: ~(<- GFALL - GFDEP))`

The **`ratr1`** specification is a modified version of `ratr` – the difference is that the former limits the number of required dependents to exactly one. As with `ratr`, the `ratr1` can be constrained (using an embedded argument list) or not.

When `ratr1` is not constrained using a list, the constraints in (52) and (53) are used – note that these are a conjunction of `ratr` constraints: the first conjunct is the positive constraint shown in (44) or (45), while the second conjunct is the negative constraint in (50) or (51).

(52)
```
(PATH GFDEP)
~(PATH GFALL - GFDEP)
```

(53)
```
(PATH PRED:
      (<- GFDEP)
      ~(<- GFALL - GFDEP))
```

If the given head can take more than one dependent, more complex constraints must be used: (54) and (55) are disjunctive constraints, where each disjunct corresponds to one category allowed by the given head – the first disjunct requires the `GFDEP1` grammatical function and blocks all grammatical functions other than `GFDEP1`, the second disjunct is analogous.

(54)
```
{
(PATH GFDEP1)
~(PATH GFALL - GFDEP1)
|
(PATH GFDEP2)
~(PATH GFALL - GFDEP2)
|
...
}
```

(55)
```
(PATH PRED:
      {
      (<- GFDEP1)
      ~(<- GFALL - GFDEP1)
      |
      (<- GFDEP2)
      ~(<- GFALL - GFDEP2)
      |
      ...
      })
```

However, when the dependent list is given in `ratr1`, the constraints on particular dependents are imposed as described for `ratr` – when there is only one element on the list, the constraints in (46) or (47) are used for imposing positive requirements for the given phrase and (50) or (51) are used for blocking all other dependents. The following constraints result:

(56)
```
(PATH GFDEP ATTR)=c v
~(PATH GFALL - GFDEP)
```

(57)
```
(PATH PRED:
      (<- GFDEP ATTR)=c v
      ~(<- GFALL - GFDEP))
```

When the list of dependents contains more than one element, the following constraints are used, where, as in (54) and (55), each disjunct corresponds to one element on the list of dependents – the difference is that each disjunct constrains the relevant dependent appropriately:

(58)
```
{
(PATH GFDEP1 ATTR)=c v
~(PATH GFALL - GFDEP1)
|
(PATH GFDEP2 ATTR)=c v
~(PATH GFALL - GFDEP2)
|
...
}
```

(59)
```
(PATH PRED:
      {
      (<- GFDEP1 ATTR)=c v
      ~(<- GFALL - GFDEP1)
      |
      (<- GFDEP2 ATTR)=c v
      ~(<- GFALL - GFDEP2)
      |
      ...
      })
```

Since the **`atr`** specification expresses the optionality of a certain requirement, it either requires a certain dependent using the appropriate `ratr` specification or it blocks any dependents using the `natr` specification. For this reason, instead of rewriting all the constraints, only general schemata are provided here: (60) is appropriate for plain `atr` constraints, while (61) is its off-path counterpart – `<ratr_constraint>` is the placeholder for the relevant `ratr` constraint (`atr`, like `ratr`, may be followed by an embedded list specifying dependents or unspecified), which may be disjunctive or not, and `<natr_constraint>` is the placeholder for the `natr` constraint.

(60)
```
{
<ratr_constraint>
|
<natr_constraint>
}
```

(61)
```
(PATH PRED:
       {
       <ratr_constraint>
       |
       <natr_constraint>
       })
```

To give an example, (62) is the plain version of constraints for `atr` with a specified list of dependents, where the list contains more than one dependent – the last disjunct is the negative constraint corresponding to the `natr` specification in (41), while the remaining disjuncts are taken from the corresponding `ratr` specification in (48). (63) is the off-path counterpart of (62) – its last disjunct corresponds to (42), while the remaining ones correspond to the off-path version of `ratr` specification given in (49).

(62)
```
{
{(PATH GFDEP1) | (PATH GFDEP2) | ...}
{(PATH GFDEP1 ATTR)=c v | ~(PATH GFDEP1)}
{(PATH GFDEP2 ATTR)=c v | ~(PATH GFDEP2)}
...
|
~(PATH GFALL)
}
```

(63)
```
(PATH PRED:
      {
      {(<- GFDEP1) | (<- GFDEP2) | ...}
      {(<- GFDEP1 ATTR)=c v | ~(<- GFDEP1)}
      {(<- GFDEP2 ATTR)=c v | ~(<- GFDEP2)}
      ...
      |
      ~(<- GFALL)
      )
```

Finally, the treatment of **`atr1`** is fully analogous to `atr` discussed above – since `atr1` expresses that a certain optional dependent can occur only once, it is formalised as a disjunction of the appropriate `ratr1` constraint and the `natr` constraint, as presented in the general schemata provided below:

(64)
```
{
<ratr1_constraint>
|
<natr_constraint>
}
```

(65)
```
(PATH PRED:
       {
       <ratr1_constraint>
       |
       <natr_constraint>
       })
```

3.2.6 Lexicalised dependents: an example

Let us now consider an example on the basis of the schema for WITAĆ 'welcome' in (13) discussed in §2.6, repeated as (66) for the sake of convenience:

(66)
```
subj{np(str)} + obj{np(str)} + {xp(mod);
lex(prepnp(z,inst),pl,XOR('ramię','ręka'),ratr1(
  {lex(adjp(agr),agr,agr,pos,'otwarty',atr1(
    {lex(advp(mod),pos,'szeroko',natr)}))}))}
```

It consists of 3 positions, of which the last one contains a non-lexicalised `xp(mod)` phrase which can be coordinated with a lexicalised (`lex`) `prepnp` phrase – according to the ranking in (16), the entire position is assigned the OBL-MOD grammatical function (see §3.1). Since the position involves coordination (the set contains two phrases, one of which, `xp(mod)`, is additionally a container category, see §2.5), off-path constraints must be used.

Note that constraints for lexicalised (`lex`) phrases consist of two parts: the non-lexicalised constraints appropriate for the given base category (the first parameter of `lex`) and lexicalised constraints – these two constraint types are marked using comments below (enclosed in quotes: `"..."`).

Let us discuss the constraints for the last position of (66) in detail, stepwise: first, only fragments of relevant constraints are presented (all partial constraints use the same off-path anchor: the `PRED` attribute of the `OBL-MOD` grammatical function) and placeholders such as <constraints_for_...> are used for the rest of the relevant constraint (discussed later as the next fragment). Finally, the entire constraint, consisting of fragments discussed earlier, is presented.

For the prepositional nominal phrase `prepnp(z,inst)`, base category constraints include the preposition form (assigned in `PRED` since it is a semantic[21] preposition) and case required from the nominal element (since the preposition is semantic, the nominal is analysed as its `OBJ`). Furthermore, the specification of lexicalised constraints such as number, lemma and modification pattern of the prepositional nominal phrase (`prepnp`) applies to its nominal component – it must be plural, its lemma can either be `ramię` or `ręka` (it is constrained using a two element list with the `XOR` operator)[22] and it requires exactly one modifier (`ratr1`) which is specified further as another lexicalised phrase, an adjectival phrase: `adjp`.

(67)
```
(^ OBL-MOD PRED:
    {
    <constraints_for_xp(mod)>
    |
    "base category constraints: prepnp(z,inst)"
    (-> FN)=c z (<- OBJ CASE)=c inst
    "lexicalised constraints"
    (<- OBJ NUM)=c pl
    (<- OBJ PRED FN)=c ramię
       <constraints_for_ratr1>
    })
```

[21] It is semantic because the grammatical function assigned to the entire position is `OBL-MOD`.

[22] The constraint in (67) handles `ramię` as the lemma, an analogous constraint is used for `ręka`.

The adjp dependent of the nominal (np – this is because constraints apply to the nominal inside the prepositional phrase) is assigned[23] the ADJUNCT grammatical function (GFDEP is resolved to ADJUNCT in ratr1 specification). The only base category constraint for adjp restricts the value of its _CAT attribute to one of the following categories: adj is an adjective, ppas and pact are adjectival participles, passive and active. When it comes to lexicalised constraints, the values of case, number and gender are specified as agreeing (agr), so no constraints are introduced – such agreement is handled by the general grammar rules. The adjective is specified for positive degree (pos), so a DEGREE constraint is used. The lemma of adjp must be otwarty – a simple PRED specification is used here. Finally, the modification pattern of adjp is specified as atr1 – it may optionally take exactly one dependent, which is another lexicalised phrase, an advp(misc).

```
(68) (^ OBL-MOD PRED:
       "base category constraints: adjp(agr)"
       (<- OBJ ADJUNCT CHECK _CAT)$c {adj ppas pact}
       "lexicalised constraints"
       (<- OBJ ADJUNCT DEGREE)=c positive
       (<- OBJ ADJUNCT PRED FN)=c otwarty
       ~(<- OBJ GFALL - ADJUNCT)
          <constraints_for_atr1>)
```

The constraints for the advp(misc) in (69) include the base category constraint restricting its _CAT to adv, followed by lexicalised constraints on degree (pos) and lemma (szeroko) and natr as its modification pattern (see (70)):

```
(69) (^ OBL-MOD PRED:
       {
       "base category constraints: advp(misc)"
       (<- OBJ ADJUNCT ADJUNCT CHECK _CAT)=c adv
       "lexicalised constraints"
       (<- OBJ ADJUNCT ADJUNCT DEGREE)=c positive
       (<- OBJ ADJUNCT ADJUNCT PRED FN)=c szeroko
       ~(<- OBJ ADJUNCT GFALL - ADJUNCT)
          <constraints_for_natr>)
       |
       ~(<- OBJ ADJUNCT GFALL)
       })
(70) (^ OBL-MOD PRED: ~(<- OBJ ADJUNCT ADJUNCT GFALL))
```

The entire (except for the placeholder for xp(mod) constraints, used for reasons of space) constraint for the last argument of (13) is given in (71):[24]

```
(71) (^ OBL-MOD PRED:
       {
       <constraints_for_xp(mod)>
       |
       "base category constraints: prepnp(z,inst)"
       (-> FN)=c z (<- OBJ CASE)=c inst
       "lexicalised constraints"
       (<- OBJ NUM)=c pl
```

[23]For reasons of space, the entire mapping for dependents of non-verbal predicates cannot be presented here. As is standard in LFG and ParGram, the adjp dependent of np is an ADJUNCT, the advp dependent of adjp is also an ADJUNCT.

[24]All instances of GFALL variable used in (71) could replaced with unique, indexed variables such as GFALL_ADJP, GFALL_ADVP so that they have an expansion which is appropriate for a given head–dependent pair of categories, as explained in fn. 20.

```
(<- OBJ PRED FN)=c ramię
   "<constraints_for_ratr1>"
   "base category constraints: adjp(agr)"
   (<- OBJ ADJUNCT CHECK _CAT)$c {adj ppas pact}
   "lexicalised constraints"
   (<- OBJ ADJUNCT DEGREE)=c positive
   (<- OBJ ADJUNCT PRED FN)=c otwarty
   ~(<- OBJ GFALL - ADJUNCT)
      "<constraints_for_atr1>"
      {
      "base category constraints: advp(misc)"
      (<- OBJ ADJUNCT ADJUNCT CHECK _CAT)=c adv
      "lexicalised constraints"
      (<- OBJ ADJUNCT ADJUNCT DEGREE)=c positive
      (<- OBJ ADJUNCT ADJUNCT PRED FN)=c szeroko
      ~(<- OBJ ADJUNCT GFALL - ADJUNCT)
         "<constraints_for_natr>)"
         ~(<- OBJ ADJUNCT ADJUNCT GFALL)
      |
      ~(<- OBJ ADJUNCT GFALL)
      }
})
```

4 Conclusion

This paper presented how valency information from Walenty, currently the largest and most precise valency dictionary of Polish, can be used in an LFG grammar of Polish, presenting selected issues in more detail, together with a full formalisation.

The quality of the performed conversion of Walenty is evaluated and improved by building an LFG structure bank of Polish: human annotators manually disambiguate structures produced by the grammar which uses the lexicon with converted valency information from Walenty – see Patejuk & Przepiórkowski 2014c.

It is perhaps worth noting that the work presented here from the implementational side also supported theoretical work on the definition of grammatical functions in LFG (Patejuk & Przepiórkowski, 2016, 2014a) and formal issues such as imposing constraints in LFG under unlike category coordination (Przepiórkowski & Patejuk, 2012; Patejuk & Przepiórkowski, 2012a).

References

Arnold, Doug, Miriam Butt, Berthold Crysmann & Tracy Holloway King (eds.). 2016. *The proceedings of the HeadLex16 conference*. Stanford, CA: CSLI Publications.

Bresnan, Joan. 2001. *Lexical-functional syntax* Blackwell Textbooks in Linguistics. Blackwell.

Butt, Miriam & Tracy Holloway King (eds.). 2014. *The proceedings of the LFG'14 conference*. Stanford, CA: CSLI Publications.

Dalrymple, Mary. 2001. *Lexical-Functional Grammar*. Academic Press.

Dalrymple, Mary & Helge Lødrup. 2000. The grammatical functions of complement clauses. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'00 conference*, University of California, Berkeley: CSLI Publications.

Patejuk, Agnieszka & Adam Przepiórkowski. 2012a. A comprehensive analysis of constituent coordination for grammar engineering. In *Proceedings of the 24rd international conference on computational linguistics (COLING 2012)*, Mumbai, India.

Patejuk, Agnieszka & Adam Przepiórkowski. 2012b. Towards an LFG parser for Polish: An exercise in parasitic grammar development. In *Proceedings of the eighth international Conference on Language Resources and Evaluation, LREC 2012*, 3849–3852. Istanbul, Turkey: ELRA.

Patejuk, Agnieszka & Adam Przepiórkowski. 2014a. Control into selected conjuncts. In Butt & King (2014) 448–460.

Patejuk, Agnieszka & Adam Przepiórkowski. 2014b. Structural case assignment to objects in Polish. In Butt & King (2014) 429–447.

Patejuk, Agnieszka & Adam Przepiórkowski. 2014c. Synergistic development of grammatical resources: a valence dictionary, an LFG grammar, and an LFG structure bank for Polish. In *Proceedings of the thirteenth workshop on Treebanks and Linguistic Theories (TLT13)*, Tübingen, Germany.

Patejuk, Agnieszka & Adam Przepiórkowski. 2016. Reducing grammatical functions in LFG. In Arnold et al. (2016).

Przepiórkowski, Adam. 2016. How *not* to distinguish arguments from adjuncts in LFG. In Arnold et al. (2016).

Przepiórkowski, Adam, Elżbieta Hajnicz, Agnieszka Patejuk & Marcin Woliński. 2014a. Extended phraseological information in a valence dictionary for NLP applications. In *Proceedings of the workshop on lexical and grammatical resources for language processing (LG-LP 2014)*, 83–91. Dublin, Ireland: Association for Computational Linguistics and Dublin City University.

Przepiórkowski, Adam, Elżbieta Hajnicz, Agnieszka Patejuk, Marcin Woliński, Filip Skwarski & Marek Świdziński. 2014b. Walenty: Towards a comprehensive valence dictionary of Polish. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the ninth international Conference on Language Resources and Evaluation, LREC 2014*, 2785–2792. Reykjavík, Iceland: ELRA.

Przepiórkowski, Adam & Agnieszka Patejuk. 2012. On case assignment and the coordination of unlikes: The limits of distributive features. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'12 conference*, 479–489. Stanford, CA: CSLI Publications.

Szupryczyńska, Maria. 1996. Problem pozycji składniowej. In Krystyna Kallas (ed.), *Polonistyka toruńska uniwersytetowi w 50. rocznicę utworzenia UMK. Językoznawstwo*, 135–144. Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika.

Reducing grammatical functions in LFG

Agnieszka Patejuk

Institute of Computer Science, Polish Academy of Sciences

Adam Przepiórkowski

Institute of Computer Science, Polish Academy of Sciences and
Institute of Philosophy, University of Warsaw

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 541–559

Keywords: grammatical functions, obliqueness, functional hierarchy, coordination

Patejuk, Agnieszka & Adam Przepiórkowski. 2016. Reducing grammatical functions in LFG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 541–559. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.28.

Abstract

The aim of this paper is to reexamine the rich repertoire of grammatical functions assumed in LFG and provide novel arguments for the claim, voiced earlier for example in Alsina et al. 2005, that most of them are redundant. We also demonstrate that a textbook LFG test for the sameness of grammatical functions of different predicates fails on closer scrutiny. Constructively, we propose a more constrained approach to grammatical functions, which has the advantage of formalising the grammatical function hierarchy, assumed in LFG analyses of diverse phenomena but apparently not previously formalised.

1 Introduction

While LFG emphasises that grammatical functions (GFs) are first-class linguistic entities, not defined via tree-configurational or any other primitives, there is surprisingly little agreement on the definition of particular grammatical functions. The only function investigated in some depth is SUBJ, with a proposal of Falk 2006 to decompose it into two separate (but co-extensive in many of the familiar languages) functions: the most prominent argument of a verb and the argument that is accessible cross-clausally.[1] In practice, most subjects are easy to identify as those arguments which agree with the verb, although in many languages this test is limited to nominative arguments, and in some languages it is complicated by the existence of object agreement.

However, as the discussion in Dalrymple 2001, pp. 19–24, makes clear, there is no single cross-linguistically valid definition of object in LFG, not even one relating to passivisation; rather, as put in Dalrymple & Nikolaeva 2011, p. 24, "[d]iagnostics targeting nonsubject grammatical functions, specifically objects, also [i.e., as in the case of subjects] vary from language to language". Even less agreement is to be expected on GFs other than subject and (direct) object. However, if definitions of GFs are language-dependent, and there are no universal properties of, say, objects, it makes limited sense to assume a "universally available inventory of grammatical functions" (Dalrymple, 2001, p. 9); rather, LFG assumes an inventory of *names* of GFs, which have somewhat different meanings in the case of different languages.

The aim of this paper is to discuss further problems with the LFG approach to grammatical functions. In particular, we show that the way they are understood in actual LFG analyses is largely redundant (Section 2) and we substantiate proposals to reduce this redundancy (Section 3). We also show that a test aimed at identifying

[†]This research is partially supported by the Polish Ministry of Science and Higher Education within the CLARIN ERIC programme 2016–2018 (http://clarin.eu/). The authors are grateful to Mary Dalrymple and Alex Alsina for their comments on an early version of this paper, as well as to both anonymous reviewers for detailed remarks, which led to many improvements in form and content.

[1]See Sag 2007 and references therein for related work within HPSG.

the same grammatical functions of different predicates, based on dependent sharing, does not stand up to scrutiny (Section 4). On the basis of these considerations, we propose to minimise the role of (names of) GFs in LFG (Section 5).

2 GFs are redundant

The following (names of) grammatical functions are commonly assumed in LFG (Dalrymple 2001, p. 9, Bresnan et al. 2015, pp. 97–100): SUBJect, OBJect, OBJ_θ, COMP, XCOMP, OBLique_θ, ADJunct and XADJunct. In fact, the names with the θ subscript do not refer to specific grammatical functions such as SUBJ or OBJ, but they "represent families of relations indexed by semantic roles, with the θ subscript representing the semantic role associated with the argument" (Dalrymple, 2001, p. 9). In this paper, we concentrate on the governable grammatical functions SUBJ, OBJ, OBJ_θ, OBL_θ, COMP and XCOMP, i.e., grammatical functions of arguments of predicates (as opposed to the modifier functions ADJ and XADJ).

Subjects and (direct) objects are usually defined in a way independent of their morphosyntax. For example, while prototypical subjects in Indo-European languages are nominative NPs, not all such nominative NP dependents are subjects, and common tests such as ability to be controlled and being the sole binder of anaphors may identify as subjects NPs bearing cases other than nominative (as in the well-known case of quirky subjects in Icelandic). Moreover, coordination may provide evidence for non-NP subjects (see Section 3 below).

Similarly, if passivisation is used as the main test for objecthood, objects defined this way are not simply co-extensive with, say, accusative NPs: in many languages not all accusative dependents of active forms become subjects in the passive, and in various languages some of the arguments bearing other cases may passivise (see Section 4 for an example from Polish). Evidence from passivisation, psych-verbs, the contrast between unaccusative and unergative predicates, etc., also makes it clear that subjects and objects cannot be defined in terms of thematic roles they bear.[2] Hence, subject and object(s) may indeed be defined in a way that makes these grammatical functions primitive.

However, the same cannot be said about other argument GFs, which, in the usual LFG practice, are often conglomerates of independent syntactic (categorial) and semantic (thematic) properties. In the case of English, once we exclude subjects and (direct) objects, nominal arguments are often assumed to deterministically map into OBJ_θ, prepositional arguments – into OBL_θ, finite clauses (CPs) – into COMPs, and infinitival clauses – into XCOMPs:

(1)

XP:	NP	PP	CP	InfP
GF:	OBJ_θ	OBL_θ	COMP	XCOMP

The claim that XCOMP is often assumed to correspond directly to InfP may seem controversial since other categories – in particular, predicative NPs, APs and

[2]See, e.g., Dowty 1989, 1991 and references therein.

PPs – may in theory also map to this open complement function. However, this theoretical possibility is rarely taken advantage of in practice; two other analyses of such predicative complements are discussed in Dalrymple et al. 2004, including one involving a closed grammatical function, PREDLINK. Moreover, in implemented LFG/XLE grammars, a distinct grammatical function, XCOMP-PRED, is often used for open predicative complements. Hence, the correlation between XCOMP and InfP is rather strong in the actual LFG practice.

A mapping similar to (1) is also often assumed for languages other than English,[3] although in the case of languages with nominal morphology richer than in English, values of grammatical cases may also play a role, as in an analysis of Russian, where OBL_{GOAL} arguments may be bare (adpositionless) nominals marked for the dative case (King, 1995, p. 180). A clear illustration of this kind of morphosyntactic redundancy may be found in Nordlinger 1998, an LFG analysis of Australian languages based on the idea of constructive case, where grammatical functions are explicitly defined on the basis of morphological cases; since, as discussed in Nordlinger 1998, pp. 69–84, case features are required in such languages independently of grammatical functions, the question arises whether in such languages different GF features are really required independently of morphological case.

Grammatical functions, as understood in LFG, are redundant not only with respect to morphosyntax. While morphosyntax often determines the choice between OBJ_θ and OBL_θ, the particular value of $_\theta$, say, BENEFICIARY or INSTRUMENT), is redundant with respect to another level of representation, namely, s-structure, which is currently assumed to contain semantic attributes such as BENEFICIARY, PATH or INSTRUMENT (Asudeh & Giorgolo, 2012; Asudeh et al., 2013, 2014).

Let us also note in passing that, in some LFG analyses, the indices in OBL_θ do not always refer to thematic roles, but may also refer to specific (non-semantic) prepositions heading the PPs. This practice not only introduces further redundancy (as information about the form of these prepositions is already present both at c-structure and elsewhere at f-structure), but also clashes with the view that LFG provides a "universally available inventory of grammatical functions". For example, OBL_{OF} (for an argument of the adjective AWARE, the noun RELATIVE, and – more generally – nominal gerunds; Dalrymple 2001, pp. 82, 249, Bresnan et al. 2015, p. 316), with the English preposition OF as the index, cannot be assumed to be a part of the universal linguistic endowment.

3 Reducing the redundancy

It is clear that not all finite clauses bear the COMP grammatical function. One of the arguments for treating at least some CPs as subjects or objects concerns the

[3] For example: "Since the recipient/goal of the Urdu ditransitive verb *de* 'give' is marked with dative case, and never with a postposition, I assume that it is not an oblique, but an indirect object (OBJ_θ)" (Butt, 1995, pp. 163–164).

possibility to coordinate them with uncontroversial SUBJs and OBJs (Sag et al., 1985, p. 165):[4]

(2) a. The implications frightened many observers.
    b. That Himmler appointed Heydrich and the implications thereof frightened many observers.
    c. That Himmler appointed Heydrich frightened many observers.

(3) a. Pat remembered the appointment.
    b. Pat remembered the appointment and that it was important to be on time.
    c. Pat remembered that it was important to be on time.

Thus, given that the phrase *the implications* is an uncontroversial subject in (2a) and that such an NP may be coordinated in this position with a CP, as shown in (2b), the finite clause should also be assumed to be the subject in (2c). Analogously, given that the phrase *the appointment* is an uncontroversial object in (3a) and that it may be coordinated in this position with a CP, as shown in (3b), the finite clause should also be assumed to be the object in (3c). Since CPs may be subjects and objects, should COMP be retained as a separate GF at all?

Dalrymple & Lødrup (2000) show that two different kinds of (non-subject) clausal complements exist in languages such as English, German and Swedish, and propose retaining COMP as the grammatical function of those CP arguments which are not OBJs (or SUBJs). However, Alsina et al. (2005) convincingly argue that the grammatical differences between different CP arguments may be accounted for without recourse to COMP; instead, English CPs treated in Dalrymple & Lødrup 2000 as COMPs should be analysed as OBLiques, which is sufficient to distinguish them from SUBJects or OBJects.

In passing, Alsina et al. (2005, p. 41) also postulate that "XCOMP should probably go the same way as COMP", but provide no arguments for this position (apart from mentioning that "XCOMP may be considered a special case of COMP"). While this move would be more far-reaching than getting rid of COMP, as it would obliterate the distinction between closed and open GFs, we believe it is sanctioned by the same kind of evidence that led to the acceptance of CP subjects and objects, namely, evidence from the coordination of unlikes.

Consider the following attested examples involving the control verb TEACH:[5]

(4) I taught him manners and to respect his elders.[6]

(5) . . . they taught me patience and to not take everything for granted.[7]

(6) Cooking has taught me patience, perseverance and to be creative.[8]

(7) It was my mother who taught me right from wrong, and to be careful who I

[4] Sag et al. (1985, pp. 164–165) mention that not all speakers accept (2c) and (3c).

[5] Another English example of this kind is given in Patejuk & Przepiórkowski 2014a, p. 456.

[6] https://bellamiataurus.com/tag/strengthineverknewihad/

[7] http://blog.girlscouts.org/2016/06/the-golden-girls-of-troop-520.html

[8] http://www.thekitchn.com/what-cooking-taught-me-about-being-happy-204508

surrounded myself with.[9]

(8) You taught me about disappointment and to recognize when something wasn't right…[10]

In all these examples, a closed constituent (e.g., *manners* in (4)) is coordinated with an open constituent with a controlled subject (e.g., *to respect his elders* in the same example).[11] Should this coordinated argument be assigned a closed grammatical function (probably an OBJ$_\theta$, with some appropriate index), or the open grammatical function XCOMP?

Similar examples may be found for other uncontroversial control verbs, e.g. WANT:

(9) The majority want peace and to live a comfortable life…[12]

(10) I just want friends and to be happy.[13]

(11) Adult learners want respect and to be seen as capable learners.[14]

(12) Really I just want a mask and to wear this to an elegant ball.[15]

(13) The survey suggests that unlike Boomers who want their objectives and to be left alone to execute, Gen Y wants an almost constant stream of feedback.[16]

Obviously, such constructions are not limited to English; for example, Patejuk & Przepiórkowski 2014a discuss – and provide an analysis for – similar examples in Polish, including the following (originally from Kallas 1993, p. 92):

(14) Nie chciał pić ani kanapki. (Polish)
NEG wanted drink.INF nor sandwich.GEN
'He didn't want to drink nor (did he want) a sandwich.'

While all the above examples involve coordination of a broadly nominal element (a PP in the case of (8)) and an apparent XCOMP, in this order, (15) below il-

[9]http://www.inc.com/joe-desena/6-lessons-my-mother-taught-me-about-business.html

[10]https://whisperedthingsiwillscream.wordpress.com/2016/03/22/you-taught-me-more-than-happily-ever-after-could-have/

[11]The external anonymous reviewer suggests that such examples could perhaps be analysed as cases of non-constituent coordination, i.e., as cases of sentential coordination. For example, (4) "would then get an f-structure generally shaped as the f-structure for *I taught him manners and I taught him to respect his elders* (but with appropriate reentrancies)" (citing the review). However, this alternative analysis seems to suffer from the kind of data originally discussed in Partee 1970 and more recently in Kubota & Levine 2015 (see also references therein), involving the distribution of quantification over coordination. For example, in case of *Two different people taught him manners and to respect his elders*, the f-structure representation analogous to *Two different people taught him manners and two different people taught him to respect his elders* would probably require a much more complicated syntax–semantics mapping in order to get the intended reading where one person taught him manners and another one taught him to respect his elders.

[12]http://www.newyorker.com/magazine/2014/12/01/quiet-german

[13]http://www.healthguidance.org/entry/15944/1/Can-Maladaptive-Daydreaming-Be-Treated.html

[14]https://ala.asn.au/adult-learning/the-principles-of-adult-learning/

[15]https://pl.pinterest.com/pin/127226758198429442/

[16]http://www.forbes.com/sites/tykiisel/2012/05/16/gimme-gimme-gimme-millennials-in-the-workplace/

lustrates coordination of an apparent XCOMP and an apparent COMP; and (16)–(17) are similar examples from Polish, involving the verb CHCIEĆ 'want', which in Polish may combine with various categories, including CPs:

(15) I hope to be successful and that you all will always be with us.[17]

(16) Publiczność chce skakać i żeby było głośniej. (Polish)
audience.NOM wants jump.INF and that is louder
'The audience wants to jump and that it be louder.'[18]

(17) Musimy to zmienić, jeśli chcemy być konkurencyjni na tamtejszych
must this change if want be competitive on those
rynkach i aby rósł nasz eksport. (Polish)
markets and that grow our export
'We must change this if we to want be competitive in those markets and that our export grows.'[19]

Examples such as (4)–(17) undermine the distinction between closed and open grammatical functions.[20] In their analysis of cases like (14), Patejuk & Przepiórkowski (2014a) treat the coordinated argument as OBJ, explicitly allowing control into OBJ, if this syntactic position is occupied by an open constituent. Similarly, Arka & Simpson (1998) convincingly argue for the possibility of control into SUBJ in Balinese. Hence, there is ample justification for Alsina et al.'s (2005) postulate to remove XCOMP from LFG's repertoire of grammatical functions.[21]

Getting rid of COMP and XCOMP would also be beneficial for the LFG linking theory, i.e., Lexical Mapping Theory (Bresnan & Kanerva, 1989), as traditionally LMT has nothing to say about these grammatical functions. Furthermore, attempts to include COMP and XCOMP in the purview of LMT have either assumed that these functions are actually OBLs (Zaenen & Engdahl, 1994, p. 198), or mapped arguments to (X)COMP on the basis of both thematic and categorial information (Butt, 1995, pp. 168–169), rather than on the basis of the ±r and ±o annotations, as in the standard LMT.

[17] http://www.visedal.org/sonia.html

[18] http://poznan.wyborcza.pl/poznan/1,36037,19099237,kosmiczna-odyseja-czyli-wodecki-plus-mitch-mitch-pszczolki.html

[19] National Corpus of Polish (http://nkjp.pl/; Przepiórkowski et al. 2011, 2012)

[20] Note that the problem would persist even if these apparent XCOMP constituents were analysed as COMPs with obligatory anaphoric control – a mechanism would still be needed to ensure such obligatory control into just one of the conjuncts.

[21] Once the XCOMP grammatical function is removed under the proposed analysis, the question arises (also in comments from the internal anonymous reviewer) about the treatment of predicative items, sometimes analysed via XCOMP. As already mentioned in Section 2 above, an alternative LFG analysis is also available, involving the closed grammatical function PREDLINK. The closed analysis of predicative complements has the advantage of accounting for cases where the predicative item has a subject of its own (as in the case of gerunds or clauses), while appropriate control can be ensured using, e.g., a dedicated CONTROLLER attribute inside the predicative item, as proposed in Patejuk & Przepiórkowski 2014a. This PREDLINK analysis carries over to the current account, with the only difference that such predicative complements do not have a dedicated grammatical function but are treated as obliques (rather than objects, as predicative complements do not passivise).

In summary, given that:

- COMP and XCOMP are superfluous,
- the θ indices in OBJ_θ and OBL_θ are redundant,
- SUBJect and OBJect(s) are perhaps the only truly primitive grammatical functions,

a three-way distinction presents itself between subjects, objects and other dependents. This is essentially the system proposed (but not amply justified) in Alsina 1996, where the "other dependents", i.e. obliques, also include adjuncts (in line with the proposal of Przepiórkowski 2016). Section 5 suggests ways of formalising this idea that eschews certain technical problems with the formalisation of Alsina 1996.

Let us end the current section by emphasising that the division of dependents into the three classes should be understood as fully independent of their categorial status. In particular, it cannot be maintained, even in the case of English, that NPs are only subjects and objects, and PPs are only obliques. In the case of languages with sufficiently rich case systems, evidence that obliques may also be realised as NPs is provided by coordination of NPs and PPs, as in the following Polish examples:

(18) Owinął dziecko w koc i ręcznikiem. (Polish)
wrapped baby in blanket.ACC and towel.INST
'He wrapped the baby in a blanket and with a towel.'[22]

(19) Gola dedykuję dla rodziców i sympatii Iwonie.
goal.ACC dedicate.1.SG to parents.GEN and girlfriend.DAT Iwona.DAT
(Polish)
'I dedicate this goal to my parents and my girlfriend Iwona.'[23]

However, NPs are not limited to subjects and objects even in languages like English and may even play the role of typical adjuncts – i.e. obliques, given the approach of Alsina 1996 – as in the following examples from Larson 1985, p. 595, with oblique NPs emphasised:

(20) I saw John *that day*.

(21) I saw John *someplace you'd never guess*.

(22) John was headed *that way*.

(23) Max pronounced my name *every way imaginable*.

Conversely, it is also easy to find PP subjects and objects. Multiple examples in English and Polish are provided in Jaworska 1986a,b, including examples of (raised) subjects in (24) and examples of objects in (25), which become subjects in the passive voice, cf. (26) (Jaworska, 1986b, pp. 355–356):

(24) a. *Between six and seven* seems to suit her fine.
b. *Across the road* appeared to be swarming with bees.

[22] Kosek 1999, p. 43
[23] National Corpus of Polish

(25) a. The campaigners planned *until Christmas* in detail.
b. The new tenants are reclaiming *behind the garage*.

(26) a. *Until Christmas* was planned in detail.
b. *Behind the garage* is being reclaimed by the new tenants.

We conclude that only two grammatical functions need to be distinguished: SUBJect and OBJect(s). (All other dependents, including adjuncts, may be called OBLiques, as in Alsina 1996.) In addition, even in the case of English, no assumptions should be made about the morphosyntactic makeup of grammatical functions.

4 GFs and dependent sharing

The conclusion of the previous section seems to be directly contradicted by the contrast in (27)–(28), originally from Barbara Partee's dissertation (Hall, 1965, p. 66); in LFG, this contrast is claimed to show that dependents shared between two coordinated verbs "must bear the same grammatical function in both conjuncts" (Dalrymple, 2001, p. 366):

(27) John washes and polishes his car in the garage.

(28) *John washes and keeps his car in the garage.

While in (27) the locative phrase is an adjunct to both WASH and POLISH, in (28) it is an adjunct to WASH, but an (oblique) argument to KEEP; hence the ungrammaticality, on the assumption – rejected in Alsina 1996, in the current paper and in Przepiórkowski 2016 – that argument obliques and adjuncts are different grammatical functions.

However, closer inspection shows that this apparent test for the sameness of grammatical functions of different predicates regularly contradicts dominant LFG analyses. For example, a locative phrase is syntactically required in the case of verbs such as RESIDE (McConnell-Ginet, 1982, p. 166), so it must be treated as its argument, if one adopts the prevailing view that required dependents are arguments. On the other hand, in the case of DIE, such a locative phrase is a prototypical optional adjunct. Hence, the following attested sentences should be ungrammatical, and for the same reason as (28):

(29) If a person resided and died in a foreign country and had assets in US, can the estate be probated in US?[24]

(30) Prime Minister Sir Winston Churchill resided and died in Number 28 on the street called Hyde Park Gate...[25]

(31) We assessed data on Medical Examiner-certified suicide victims aged 65 years or older from 2001 through 2004 who had resided and died in New York City...[26]

[24] http://www.avvo.com/legal-answers/if-a-person-resided-and-died-in-a-foreign-country--206311.html

[25] http://www.apeksdevelopments.co.uk/famous-hyde-park-residents-throughout-history/

[26] http://europepmc.org/abstract/MED/19210947

Another problematic case is illustrated with the following examples:

(32) I will devour this cake.

(33) I will give Mary this cake.

(34) I will either devour or give Mary the carrot cake my mother baked yesterday.

In (32), *this cake* is the passivisable OBJ, while in (33) it is an OBJ_θ, as the OBJ position is taken by the passivisable *Mary* (Dalrymple, 2001, p. 22). However, these two supposedly different grammatical functions may be shared, as (34) illustrates.

The problem also occurs in languages other than English. For example, Patejuk 2015, p. 51, discusses the following examples from Polish:

(35) Marek manipuluje i wysługuje się Marysią. (Polish)
Marek.NOM manipulates and uses REFL Marysia.INST
'Marek manipulates and uses Marysia.'

(36) Marysia lubi ale też boi się Marka. (Polish)
Marysia.NOM likes but also fears REFL Marek.ACC/GEN
'Marysia likes but at the same time is afraid of Marek.'

The natural definition of object in Polish is as the passivisable argument; if so, in both examples the non-subject argument (*Marysią* in (35) and *Marka* in (36)) bears the OBJ function only in relation to one of the conjoined verbs (to *manipuluje* 'manipulates' and to *lubi* 'likes', respectively). This again violates the claim that shared dependents must bear the same grammatical function in relation to conjoined verbs. One way to attempt to defend this claim would be to revert to the more traditional understanding of the direct object, as the argument in the accusative case. If so, neither of the verbs in (35) takes an OBJ (the shared argument is in the instrumental). However, in (36), one verb, *lubi* 'likes', takes such an accusative object and the other verb, *boi się* 'fears', takes a genitive argument; so the shared argument *Marka* still simultaneously fills two different grammatical function slots.[27] As there is no other reasonable way of defining OBJ in Polish, we must conclude that either it makes no sense (or at least there is no need) to posit OBJ in Polish, or the coordination test based on the contrast from Hall 1965 does not work.

In fact, the latter seems to be the case. Without attempting to provide an exhaustive analysis, let us note that in all the grammatical examples where a dependent bearing different grammatical roles is shared, it has the same (or sufficiently similar) semantic role in relation to the conjoined verbs. In particular, in the *resided and died* examples, the locative phrase, while obligatory in the case of RESIDE and optional in the case of DIE, has the semantic role of event location, the same as the locative *in the garage* in Partee's grammatical (27). On the other hand, while the phrase *in the garage* also expresses location in the case of (28), it arguably bears two rather different semantic roles with respect to WASH and KEEP, namely, event

[27] See Dalrymple & Kaplan 2000 and, especially, Dalrymple et al. 2009 on how *Marka* may be analysed as accusative and genitive at the same time.

location in the case of the former, but participant location in the case of the latter.[28]

Let us finally note that the fact that two predicates may assign different grammatical functions to their shared dependent is not a technical problem for LFG; as verified in the XLE implementation of Polish (Patejuk & Przepiórkowski, 2012, 2014b; Patejuk, 2016), all that is required is the assignment of grammatical functions in c-structure rules via functional uncertainty, as in (37), rather than via separate equations, as in (38):

(37) $(\uparrow \{\textsc{gf}1|\textsc{gf}2\}){=}\downarrow$

(38) $(\uparrow \textsc{gf}1){=}\downarrow \vee (\uparrow \textsc{gf}2){=}\downarrow$

So the (only) conclusion of this section is that shared dependents do not provide a test for the sameness of grammatical functions, contra common LFG assumptions (expressed, e.g., in Dalrymple 2001, p. 366, and in Peterson 2004).

5 Minimising the role of GFs in LFG

Alsina 1996, ch. 2, proposes to represent all dependents of a predicate via just three (types of) attributes: SUBJ, OBJ and OBL. In fact, these attributes are understood there as shorthands for, respectively, the following feature bundles: $\begin{bmatrix}\text{subj} & + \\ \text{obl} & -\end{bmatrix}$, $\begin{bmatrix}\text{subj} & - \\ \text{obl} & -\end{bmatrix}$, and $\begin{bmatrix}\text{subj} & - \\ \text{obl} & +\end{bmatrix}$. It is not clear to us how to extend the formal apparatus of LFG so that not only atomic symbols, but also such feature bundles may act as attributes, so we continue using the atomic values SUBJ, etc., here. In typical f-structures with a propositional content there must be exactly one subject in languages such as English and Catalan (as decreed by the Subject Condition, Alsina 1996, p. 20), but there may be multiple objects and obliques. This creates the obvious formal problem of possible multiple occurrences of the same OBJ or OBL attribute. Alsina (1996, pp. 47–48) solves this problem by indexing such attributes with the identifiers of f-structures which are the values of these attributes. Again, this mechanism does not seem to be a generally assumed part of the LFG apparatus. Below we will provide a formalisation which only assumes the standard LFG machinery.

An obvious solution is to make OBJ and OBL set-valued, on par with ADJ in the usual LFG analyses. As far as we can see, various constraints and analyses of Alsina 1996 may be easily reformulated to accommodate this solution. However, we would like to propose a more radical solution, more scrupulously justified in Przepiórkowski 2016, which also deals with the long-standing problem of the lack of formalisation of the syntactic hierarchy of grammatical functions, assumed to play a role in standard LFG analyses of control (Bresnan 1982, p. 294, Dalrymple 2001, p. 345), binding (Bresnan et al. 2015, chs. 9–10, and references therein)

[28]See, e.g., Koenig et al. 2003 on this distinction, as well as Maienborn & Schäfer 2011 and references to Claudia Maienborn's work therein on the more general distinction between event-external and event-internal modification.

and unbounded dependencies (Dalrymple 2001, p. 412 and references therein). According to the functional hierarchy of LFG, and similar hierarchies assumed by other grammatical theories (cf. the accessibility hierarchy of Keenan & Comrie 1977, the relational hierarchy of Perlmutter 1983, the obliqueness hierarchy of Pollard & Sag 1987, 1994, etc.), the subject syntactically outranks the direct object, which outranks the indirect object, which in turn outranks any oblique dependent. This hierarchy is also assumed in Alsina 1996, p. 253, even though, as in LFG at large, it is not formally represented anywhere in the analysis.

The basic idea of the solution is to represent all dependents of a predicate within a single ordered DEPS list, in a way reminiscent of the use of this attribute in some HPSG analyses (Bouma et al., 2001; Przepiórkowski, 1999). Just as in HPSG, the order of the elements reflects the grammatical function hierarchy. To the extent that some core grammatical functions, i.e., functions which take part in processes such as verbal agreement and passivisation, need to be distinguished and cannot be predicted from the position in the argument list, they could be singled out as values of separate attributes (apart from being present on the argument list),[29] in a language-dependent fashion.

For example, in the case of Polish, it makes sense to distinguish two grammatical functions: SUBJect and OBJect. Polish subjects are typically nominative and, when they are nominative, they agree with the verb, but not all nominative dependents are subjects, and not all subjects are nominative: as is well known, they may in particular be finite and infinitival clauses (Świdziński, 1992, 1993), prepositional phrases (Jaworska, 1986a,b), or accusative numeral phrases (Franks, 1995; Przepiórkowski, 1999). Similarly, given that passivisable arguments cannot be recognised morphosyntactically in Polish (not all accusative arguments passivise; and some genitive, instrumental and maybe even dative arguments do; Zabrocki 1981, pp. 124–125), objects should also be distinguished, when present. Thus, f-structures for the sentences (39) (featuring an instrumental passivisable object) and (40) (featuring an instrumental dependent which does not passivise), which involve the two verbs discussed in the context of (35) above (MANIPULOWAĆ 'manipulate' and WYSŁUGIWAĆ SIĘ 'use (somebody)'), would be as in (41) and (42), respectively.

(39) Marek manipuluje Marysią. (Polish)
Marek.NOM manipulates Marysia.INST
'Marek manipulates Marysia.'

(40) Marek wysługuje się Marysią. (Polish)
Marek.NOM uses REFL Marysia.INST
'Marek uses Marysia.'

[29]This would again follow the HPSG practice; e.g., Heinz & Matiasek 1994 single out the deep subject as the value of the DA (designated argument) attribute, Sag 2007 proposes to encode the argument visible outside of the maximal projection (i.e., roughly, Falk's 2006 PIVOT) as XARG, etc.

(41)
$$\begin{bmatrix} \textsc{pred} & \text{‘}\textsc{manipulate}\text{’} \\ \textsc{tense} & \textsc{pres} \\ \textsc{subj} & \boxed{1} \\ \textsc{obj} & \boxed{2} \\ \textsc{deps} & \left\langle \boxed{1}\begin{bmatrix} \textsc{pred} & \text{‘Marek’} \\ \textsc{case} & \textsc{nom} \end{bmatrix}, \boxed{2}\begin{bmatrix} \textsc{pred} & \text{‘Marysia’} \\ \textsc{case} & \textsc{inst} \end{bmatrix} \right\rangle \end{bmatrix}$$

(42)
$$\begin{bmatrix} \textsc{pred} & \text{‘}\textsc{use}\text{’} \\ \textsc{tense} & \textsc{pres} \\ \textsc{subj} & \boxed{1} \\ \textsc{deps} & \left\langle \boxed{1}\begin{bmatrix} \textsc{pred} & \text{‘Marek’} \\ \textsc{case} & \textsc{nom} \end{bmatrix}, \begin{bmatrix} \textsc{pred} & \text{‘Marysia’} \\ \textsc{case} & \textsc{inst} \end{bmatrix} \right\rangle \end{bmatrix}$$

Moreover, unlike in English, where each verb has a syntactic subject, some Polish verbs arguably do not have any subjects, not even expletive or PRO subjects. One such a verb is MDLIĆ ‘nauseate’, in its use illustrated in (44), to be contrasted with (43), which does involve an agreeing subject:[30]

(43) Zapach kwiatów mdlił mnie. (Polish)
smell.NOM.M.SG flowers.GEN nauseated.M.SG me.ACC
‘The smell of the flowers made me nauseous.’

(44) Mdliło mnie od zapachu kwiatów. (Polish)
nauseated.N.SG me.ACC from smell.GEN.M.SG flowers.GEN
‘I felt nauseous from the smell of the flowers.’

Hence, in the case of the two Polish examples above, the following f-structures result:[31,32]

(45)
$$\begin{bmatrix} \textsc{pred} & \text{‘}\textsc{nauseate}\text{’} \\ \textsc{tense} & \textsc{past} \\ \textsc{subj} & \boxed{1} \\ \textsc{deps} & \left\langle \boxed{1}\begin{bmatrix} \textsc{pred} & \text{‘}\textsc{smell}\text{’} \\ \textsc{deps} & \left\langle \begin{bmatrix} \textsc{pred} & \text{‘}\textsc{flower}\text{’} \\ \textsc{case} & \textsc{gen} \end{bmatrix} \right\rangle \\ \textsc{case} & \textsc{nom} \end{bmatrix}, \begin{bmatrix} \textsc{pred} & \text{‘I’} \\ \textsc{case} & \textsc{acc} \end{bmatrix} \right\rangle \end{bmatrix}$$

[30] The arguments in Babby 2009, ch. 1, for the lack of any grammatical subject of the Russian cognate of this verb carry over to Polish. Other examples of genuinely subjectless verbs and verbal constructions in Polish may be found in Kibort 2006.

[31] We follow here the observation that Glue Semantics makes PRED – and also the principles of Completeness and Coherence – largely redundant (Dalrymple et al. 1993, pp. 13–14; Kuhn 2001, § 1.3.3). In particular, we adopt the practice of Asudeh & Giorgolo 2012 and later work of retaining PRED, albeit with values representing the bare predicate, without its arguments.

[32] The attribute PFORM in (46) is commonly used in implemented LFG/XLE grammars to indicate the form of a non-semantic (‘case-marking’) preposition.

(46)
$$\begin{bmatrix} \text{PRED} & \text{`NAUSEATE'} \\ \text{TENSE} & \text{PAST} \\ \text{DEPS} & \left\langle \begin{bmatrix}\text{PRED} & \text{`I'}\\ \text{CASE} & \text{ACC}\end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{`SMELL'} \\ \text{DEPS} & \left\langle \begin{bmatrix}\text{PRED} & \text{`FLOWER'}\\ \text{CASE} & \text{GEN}\end{bmatrix}\right\rangle \\ \text{PFORM} & \text{OD} \\ \text{CASE} & \text{GEN}\end{bmatrix}\right\rangle \end{bmatrix}$$

In the case of Romance, since subjects are readily identifiable as the first elements of DEPS in f-structures expressing propositional content, only a set-valued attribute OBJect is needed to carry over the analyses of Alsina 1996.[33] Further, since the value of OBJ will identify any objects in DEPS, all other DEPS elements, following the subject and the object(s), if any, must be obliques.

Let us illustrate the analysis with the following two Catalan examples, from Alsina 1996 (with the original glosses left intact):

(47) El mestre fa llegir un poema al nen. (Catalan)
the teacher makes read a poem to-the boy
'The teacher is making the boy read a poem.'[34]

(48) Cauen rocs de la muntanya. (Catalan)
fall.3.PL stones from the mountain
'Stones fall from the mountain.'[35]

Example (47) involves a complex predicate, *fa llegir* 'makes read', with both verbs contributing to the grammatical functions of the clause: the causer, *el mestre* 'the teacher', is the subject, the agent of reading, *al nen* 'the boy', affected by causation, is realised as a dative (hence, indirect) object, and the patient of reading, *un poema* 'the poem', is realised as a non-dative (hence, direct) object (Alsina, 1996, p. 191):[36]

[33]Obviously, it is possible to have a separate SUBJ attribute, also in the case of Romance, whose value would always be structure-shared with the first element of DEPS. This would perhaps be redundant in the case of Catalan, but it would better reflect the idea that all languages have subjects (to the extent that this generalisation is true; see Falk 2006 and references therein) and it could also be beneficial from the point of view of parallel grammar development. Also, Alsina's (1996) supposedly universal distinction between direct and oblique dependents could simply be represented as that between the values of SUBJ and OBJ on the one hand, and all other DEPS elements on the other.

[34]Alsina 1996, p. 190, ex. (6b)

[35]Alsina 1996, p. 130, ex. (19)

[36]In the following f-structures we ignore argument structures, which Alsina (1996) encodes within the values of PRED.

(49)
$$\begin{bmatrix} \textsc{pred} & \text{`CAUSE READ'} \\ \textsc{tense} & \textsc{pres} \\ \textsc{obj} & \left\{\boxed{1}, \boxed{2}\right\} \\ \textsc{deps} & \left\langle \begin{bmatrix}\textsc{pred} & \text{`TEACHER'}\\ \textsc{dat} & -\end{bmatrix}, \boxed{1}\begin{bmatrix}\textsc{pred} & \text{`POEM'}\\ \textsc{dat} & -\end{bmatrix}, \boxed{2}\begin{bmatrix}\textsc{pred} & \text{`BOY'}\\ \textsc{dat} & +\end{bmatrix}\right\rangle \end{bmatrix}$$

If the sentence also contained obliques, they would follow the two objects on DEPS, and they would not have to be explicitly mentioned outside of this list. This is illustrated by the following f-structure for (48), which involves the oblique phrase *de la muntanya* 'from the mountain':

(50)
$$\begin{bmatrix} \textsc{pred} & \text{`FALL'} \\ \textsc{tense} & \textsc{pres} \\ \textsc{obj} & \left\{\boxed{1}\right\} \\ \textsc{deps} & \left\langle \boxed{1}\begin{bmatrix}\textsc{pred} & \text{`STONE'}\\ \textsc{dat} & -\end{bmatrix}, \begin{bmatrix}\textsc{pred} & \text{`FROM'}\\ \textsc{deps} & \left\langle\begin{bmatrix}\textsc{pred} & \text{`MOUNTAIN'}\end{bmatrix}\right\rangle\end{bmatrix}\right\rangle \end{bmatrix}$$

An interesting feature of this f-structure is that it economically reflects the analysis of Alsina 1996, p. 132, in which *rocs* 'stones', the first argument of the unaccusative verb *cauen* 'fall', simultaneously fills two grammatical functions: subject and object. This is represented in (50): the first element of DEPS, as always in the case of propositional f-structures in Catalan, is the subject, but it is also present in the value of OBJ, so it is at the same time an object.

6 Conclusion

It is surprising how ill-defined, redundant and inconsistent the notion of grammatical functions – claimed to be fundamental in LFG – is on closer inspection. In this paper we returned to the basic LFG assumptions and re-examined the need for a repertoire of grammatical functions as first-class theoretical citizens. We reappraised and further substantiated the approach of Alsina 1996, where the only grammatical functions assumed are SUBJect and OBJect (and OBLique, for anything else), but we suggested a different implementation of this general idea, further formalised and illustrated in Przepiórkowski 2016, one that substantially extends an HPSG approach.

References

Alsina, Alex. 1996. *The role of argument structure in grammar* (CSLI Lecture Notes 62). Stanford, CA: CSLI Publications.

Alsina, Alex, Tara Mohanan & KP Mohanan. 2005. How to get rid of the COMP.

In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'05 conference*, University of Bergen, Norway: CSLI Publications.

Arka, I Wayan & Jane Simpson. 1998. Control and complex arguments in Balinese. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'98 conference*, University of Queensland, Brisbane: CSLI Publications.

Arnold, Doug, Miriam Butt, Berthold Crysmann & Tracy Holloway King (eds.). 2016. *The proceedings of the HeadLex16 conference*. Stanford, CA: CSLI Publications.

Asudeh, Ash, Mary Dalrymple & Ida Toivonen. 2013. Constructions with Lexical Integrity. *Journal of Language Modelling* 1(1). 1–54.

Asudeh, Ash, Gianluca Giorgolo & Ida Toivonen. 2014. Meaning and valency. In Butt & King (2014) 68–88.

Asudeh, Ash & Gianluca Giorgolo. 2012. Flexible composition for optional and derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'12 conference*, 64–84. Stanford, CA: CSLI Publications.

Babby, Leonard H. 2009. *The syntax of argument structure*. Cambridge: Cambridge University Press.

Bouma, Gosse, Robert Malouf & Ivan A. Sag. 2001. Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory* 19(1). 1–65.

Bresnan, Joan. 1982. Control and complementation. In Joan Bresnan (ed.), *The mental representation of grammatical relations* MIT Press Series on Cognitive Theory and Mental Representation, 283–390. Cambridge, MA: The MIT Press.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2015. *Lexical-functional syntax* Blackwell Textbooks in Linguistics. Wiley-Blackwell 2nd edn.

Bresnan, Joan & Jonni M. Kanerva. 1989. Locative inversion in Chicheŵa: A case study of factorization in grammar. *Linguistic Inquiry* 20(1). 1–50.

Butt, Miriam. 1995. *The structure of complex predicates in Urdu*. Stanford, CA: CSLI Publications.

Butt, Miriam & Tracy Holloway King (eds.). 2014. *The proceedings of the LFG'14 conference*. Stanford, CA: CSLI Publications.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. San Diego, CA: Academic Press.

Dalrymple, Mary, Angie Hinrichs, John Lamping & Vijay Saraswat. 1993. The resource logic of complex predicate interpretation. In *Proceedings of ROCLING 1993*, 3–21.

Dalrymple, Mary, Helge Dyvik & Tracy Holloway King. 2004. Copular complements: Closed or open? In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'04 conference*, 188–198. Stanford, CA: CSLI Publications.

Dalrymple, Mary, Tracy Holloway King & Louisa Sadler. 2009. Indeterminacy by underspecification. *Journal of Linguistics* 45. 31–68.

Dalrymple, Mary & Ronald M. Kaplan. 2000. Feature indeterminacy and feature resolution. *Language* 76(4). 759–798.

Dalrymple, Mary & Helge Lødrup. 2000. The grammatical functions of complement clauses. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'00 conference*, University of California, Berkeley: CSLI Publications.

Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge University Press.

Dowty, David. 1989. On the semantic content of the notion of 'thematic role'. In Gennero Chierchia, Barbara H. Partee & Raymond Turner (eds.), *Properties, types and meaning: II*, 69–129. Dordrecht: Kluwer.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Falk, Yehuda N. 2006. *Subjects and Universal Grammar*. Cambridge University Press.

Franks, Steven. 1995. *Parameters of Slavic morphosyntax*. New York: Oxford University Press.

Hall, Barbara Corey. 1965. *Subject and object in modern English*: Massachusetts Institute of Technology Ph.D. Thesis.

Heinz, Wolfgang & Johannes Matiasek. 1994. Argument structure and case assignment in German. In John Nerbonne, Klaus Netter & Carl Pollard (eds.), *German in Head-driven Phrase Structure Grammar* (CSLI Lecture Notes 46), 199–236. Stanford, CA: CSLI Publications.

Jaworska, Ewa. 1986a. *Aspects of the syntax of prepositions and Prepositional Phrases in English and Polish*: University of Oxford Ph.D. Thesis.

Jaworska, Ewa. 1986b. Prepositional phrases as subjects and objects. *Journal of Linguistics* 22. 355–374.

Kallas, Krystyna. 1993. *Składnia współczesnych polskich konstrukcji współrzędnych*. Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika.

Keenan, Edward L. & Bernard Comrie. 1977. Noun Phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8. 63–99.

Kibort, Anna. 2006. On three different types of subjectlessness and how to model them in LFG. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'06 conference*, 289–309. Universität Konstanz, Germany: CSLI Publications.

King, Tracy Holloway. 1995. *Configuring topic and focus in Russian*. Stanford, CA: CSLI Publications.

Koenig, Jean-Pierre, Gail Mauner & Breton Bienvenue. 2003. Arguments for adjuncts. *Cognition* 89. 67–103.

Kosek, Iwona. 1999. *Przyczasownikowe frazy przyimkowo-nominalne w zdaniach współczesnego języka polskiego*. Olsztyn: Wydawnictwo Uniwersytetu

Warmińsko-Mazurskiego.

Kubota, Yusuke & Robert Levine. 2015. Against ellipsis: Arguments for the direct licensing of 'noncanonical' coordinations. *Linguistics and Philosophy* 38. 521–576.

Kuhn, Jonas. 2001. Resource sensitivity in the syntax-semantics interface: Evidence from the German Split NP construction. In Detmar Meurers & Tibor Kiss (eds.), *Constraint-based approaches to Germanic syntax*, 177–215. Stanford, CA: CSLI Publications.

Larson, Richard. 1985. Bare NP adverbs. *Linguistic Inquiry* 16. 595–621.

Maienborn, Claudia & Martin Schäfer. 2011. Adverbs and adverbials. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 2, 1390–1420. Berlin: De Gruyter Mouton.

McConnell-Ginet, Sally. 1982. Adverbs and logical form: A linguistically realistic theory. *Language* 58(1). 144–184.

Nordlinger, Rachel. 1998. *Constructive case: Evidence from Australian languages*. Stanford, CA: CSLI Publications.

Partee, Barbara H. 1970. Negation, conjunction, and quantifiers: Syntax vs. semantics. *Foundations of Language* 6. 153–165.

Patejuk, Agnieszka. 2015. *Unlike coordination in Polish: an LFG account*. Cracow: Institute of Polish Language, Polish Academy of Sciences Ph.D. dissertation.

Patejuk, Agnieszka. 2016. Integrating a rich external valence dictionary with an implemented XLE/LFG grammar. In Arnold et al. (2016).

Patejuk, Agnieszka & Adam Przepiórkowski. 2012. Towards an LFG parser for Polish: An exercise in parasitic grammar development. In *Proceedings of the eighth international Conference on Language Resources and Evaluation, LREC 2012*, 3849–3852. Istanbul, Turkey: ELRA.

Patejuk, Agnieszka & Adam Przepiórkowski. 2014a. Control into selected conjuncts. In Butt & King (2014) 448–460.

Patejuk, Agnieszka & Adam Przepiórkowski. 2014b. Synergistic development of grammatical resources: A valence dictionary, an LFG grammar, and an LFG structure bank for Polish. In Verena Henrich, Erhard Hinrichs, Daniël de Kok, Petya Osenova & Adam Przepiórkowski (eds.), *Proceedings of the thirteenth international workshop on treebanks and linguistic theories (TLT 13)*, 113–126. Tübingen, Germany: Department of Linguistics (SfS), University of Tübingen.

Perlmutter, David (ed.). 1983. *Studies in relational grammar*, vol. 1. University of Chicago Press.

Peterson, Peter G. 2004. Coordination: Consequences of a lexical-functional account. *Natural Language and Linguistic Theory* 22. 643–679.

Pollard, Carl & Ivan A. Sag. 1987. *Information-based syntax and semantics, vol-*

ume 1: Fundamentals (CSLI Lecture Notes 13). Stanford, CA: CSLI Publications.

Pollard, Carl & Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago, IL: Chicago University Press / CSLI Publications.

Przepiórkowski, Adam. 1999. *Case assignment and the complement-adjunct dichotomy: A non-configurational constraint-based approach*. Tübingen: Universität Tübingen Ph.D. dissertation. http://nlp.ipipan.waw.pl/~adamp/Dissertation/.

Przepiórkowski, Adam. 2016. How *not* to distinguish arguments from adjuncts in LFG. In Arnold et al. (2016).

Przepiórkowski, Adam, Mirosław Bańko, Rafał L. Górski, Barbara Lewandowska-Tomaszczyk, Marek Łaziński & Piotr Pęzik. 2011. National Corpus of Polish. In Zygmunt Vetulani (ed.), *Proceedings of the 5th language & technology conference: Human language technologies as a challenge for computer science and linguistics*, 259–263. Poznań, Poland.

Przepiórkowski, Adam, Mirosław Bańko, Rafał L. Górski & Barbara Lewandowska-Tomaszczyk (eds.). 2012. *Narodowy korpus języka polskiego*. Warsaw: Wydawnictwo Naukowe PWN.

Sag, Ivan A. 2007. Remarks on locality. In Stefan Müller (ed.), *Proceedings of the HPSG 2007 conference*, 394–414. Stanford, CA: CSLI Publications.

Sag, Ivan A., Gerald Gazdar, Thomas Wasow & Steven Weisler. 1985. Coordination and how to distinguish categories. *Natural Language and Linguistic Theory* 3. 117–171.

Świdziński, Marek. 1992. Realizacje zdaniowe podmiotu-mianownika, czyli o strukturalnych ograniczeniach selekcyjnych. In Andrzej Markowski (ed.), *Opisać słowa*, 188–201. Dom Wydawniczy Elipsa.

Świdziński, Marek. 1993. Dalsze kłopoty z bezokolicznikiem. In Jadwiga Sambor, Jadwiga Linde-Usiekniewicz & Romuald Huszcza (eds.), *Językoznawstwo synchroniczne i diachroniczne*, 303–314. Warsaw: Wydawnictwa Uniwersytetu Warszawskiego.

Zabrocki, Tadeusz. 1981. *Lexical rules of semantic interpretation: Control and NP movement in English and Polish* (Filologia Angielska 14). Poznań: Uniwersytet im. Adama Mickiewicza.

Zaenen, Annie & Elisabet Engdahl. 1994. Descriptive and theoretical syntax in the lexicon. In Beryl T. S. Atkins & Antonio Zampolli (eds.), *Computational approaches to the lexicon*, 181–212. Oxford: Oxford University Press.

How *not* to distinguish arguments from adjuncts in LFG

Adam Przepiórkowski
Institute of Computer Science, Polish Academy of Sciences and
Institute of Philosophy, University of Warsaw

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 560–580

Keywords: argument–adjunct distinction, dependents, argument structure, adjuncts, modifiers

Przepiórkowski, Adam. 2016. How *not* to distinguish arguments from adjuncts in LFG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 560–580. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.29.

Abstract

The paper briefly reexamines arguments for the argument–adjunct dichotomy, commonly assumed in contemporary linguistics, showing that they do not stand up to scrutiny. It demonstrates that – perhaps surprisingly – LFG currently only assumes this dichotomy in its f-structure feature geometry, and does not rely on it in any crucial way. Building on this observation, the paper presents a way of getting rid of this dichotomy altogether.

1 Introduction

A great number of tests for the argument–adjunct distinction have been proposed over the last almost 60 years (since Tesnière 1959), many quickly discarded. As noted by Tutunjian & Boland 2008, p. 633, "[t]he sheer number of these tests underlines the fact that no single test is entirely satisfactory". While the vast majority of linguists share the sentiment that "[t]he distinction between arguments and adjuncts is crucial in linguistics" (Needham & Toivonen, 2011, p. 402), some have long noted that it is difficult to make it operational, e.g.: "The problem of how to differentiate between complements and adjuncts has not yet been solved satisfactorily" (Vater, 1978, p. 21) or "No single criterion for this distinction has been found yet and it is rather doubtful that it can be found in the future" (Sawicki, 1988, p. 17).

The most common escape strategy, exemplified also by recent LFG work (Needham & Toivonen, 2011; Asudeh & Giorgolo, 2012; Toivonen, 2013; Asudeh et al., 2014), is to make this a three- or more-way distinction, with a separate class (or classes) for difficult or borderline cases. An extreme exemplar of this strategy is Somers 1984, which splits dependents into six classes: *integral complements*, *obligatory complements*, *optional complements*, *middles*, *adjuncts* and *extraperipherals*. This strategy brings us a little closer to the position defended in this paper, i.e., that dependents form a continuum which may be divided in various ways and according to various criteria, but at the prohibitive cost of replacing one vague boundary with two or more even vaguer boundaries.

Before I conclude that – after over half a century of looking for convincing and stable tests for the argument–adjunct dichotomy – the burden of proof is on the proponents of this dichotomy, I examine in Section 2 a few popular tests which are relatively language-independent, theory-independent and stable over time. In Section 3 I show that, perhaps surprisingly, LFG does not really rely on this distinction in any crucial way, but rather assumes it in the f-structure feature geometry. The paper concludes by considering three ways of getting rid of this distinction at f-structure and, hence, in LFG in general.

[†]Many thanks to Agnieszka Patejuk for her comments on various versions of the material presented here, as well as to Mary Dalrymple and two anonymous reviewers of HeadLex16 proceedings for their comments on the previous version of this paper. This research is partially supported by the Polish Ministry of Science and Higher Education within the CLARIN ERIC programme 2016–2018 (http://clarin.eu/).

2 Main tests for the argument–adjunct distinction

The common intuition is that the meaning of arguments is more central to the meaning of the predicate; unfortunately, this intuition has never (to the best of my knowledge) been translated into an operational procedure of splitting dependents into arguments and adjuncts. In fact, this intuition already suggests that the notion of argument is gradable ("more central") rather than categorial. Nevertheless, some more operational tests have been proposed, and this section examines some of the most popular such tests.

2.1 Obligatoriness

If there is one more or less operational test that almost all linguists agree about, it is that obligatory dependents are arguments. However, this is not really a binary classifier; it does not say anything about optional dependents, and in some languages almost all dependents are to some extent optional. Even in English, direct objects – i.e., prototypical arguments – may be syntactically omitted, as in the case of the italicised verbs in the following attested examples:

(1) I lost 20 lbs and nobody has *noticed*. Feeling down about it.[1]

(2) He will tell you everything when he has *finished*.[2]

(3) Make his favorite meal or dessert (if he has already *eaten*) and surprise him when he comes home![3]

And even the most prototypical verbs usually assumed to obligatorily combine with a dependent, such as DEVOUR, are happy without it, under the right circumstances:

(4) He doesn't eat, he *devours*.[4]

The fact that syntactic obligatoriness may be understood in a number of ways and may indeed be a graded notion has been recognised – on the basis of a different kind of evidence than that cited above – within valency theory (Herbst & Roe, 1996). Hence, the notion of *obligatoriness* must be made much more precise than is common in discussions of the argument–adjunct dichotomy, if it is to be operational even in this limited unidirectional (syntactic obligatoriness implies argumenthood, not the other way round) way.

Another problem with the application of this test is the existence of so-called *obligatory adjuncts*, as in the following example from Grimshaw & Vikner 1993, p. 143, which is supposed to be ungrammatical unless at least one of the phrases in the brackets (most of which are uncontroversial adjuncts) appears in the sentence:

(5) *This house was built.

[1] https://www.reddit.com/r/loseit/comments/3ntqsi/i_lost_20_lbs_and_nobody_has_noticed_feeling_down/

[2] https://www.englishforums.com/English/WhenHeHasFinished/bwhml/post.htm

[3] http://love.allwomenstalk.com/sure-ways-to-make-him-happy

[4] http://kitfrazier.com/wordpress/yes-hes-a-mean-cat-but-hes-mine-if-youve-got-him-please-send-him-home-atticusphonehome/

(6) This house was built {yesterday / in ten days / in a bad part of town / only with great difficulty / by a French architect}.

Goldberg & Ackerman (2001), together with Jung (1997) and Szymańska & Śpiewak (2000), convincingly reject the event-semantic analysis of such cases proposed by Grimshaw & Vikner (1993), and offer a pragmatic analysis in terms of Grice's maxim of quantity, arguing that – given a proper context – such adjuncts are not really obligatory. Nevertheless, the existence of this phenomenon supports the observation that obligatoriness is a subtle and possibly graded notion and that the perceived mandatory presence of a dependent may in fact result from a variety of factors.

One way of dealing with such problems with the notion of syntactic obligatoriness is to concentrate on the *semantic obligatoriness* instead, as determined by the dialogue test (Panevová, 1974, pp. 17–19), or its "monologue" version (Fillmore, 1986, p. 96):

(7) He's already *noticed* (#but I have no idea *what* he's noticed).

(8) He's already *finished* (#but I have no idea *what* he's finished).

These examples show that the missing direct objects of the forms of NOTICE and FINISH are semantically obligatory in the sense that they may be syntactically omitted only if they are contextually provided. As semantically obligatory, they are arguments, according to Panevová 1974, 1975. However, it is clear that this test alone cannot determine the argument vs. adjunct status of a dependent, as it would classify the direct object of EAT as an adjunct:

(9) He's already *eaten* (but I have no idea *what* he's eaten).

So, again, this is at best a unidirectional criterion: semantic obligatoriness implies argumenthood, not the other way round. Also, as discussed at length in Przepiórkowski 2016, the applicability of this test is not always straightforward and its results are open to interpretation. Hence, I maintain the conclusion that the notion of obligatoriness – whether understood syntactically, or semantically – has never been operationalised to the extent that would make it usable as a test for argumenthood.

2.2 Iterability

A test assumed in theories as different as LFG and HPSG on one hand and Functional Generative Description (FGD; Sgall et al. 1986)[5] on the other is the iterability test: "[A]djuncts may be iterated freely without any effect on syntactic well-formedness" (Williams, 2015, p. 69). The much cited example showing iterability of adjuncts is (10) from Bresnan 1982c, p. 164, contrasted with (11) from Bresnan 1982c, p. 165, which is supposed to show that instruments are arguments ("[Inst]" added in (11) for the sake of parallelism with (10)):

[5] A critique of tests assumed in FGD, including iterability and specificity, may be found in Przepiórkowski 2016, from which this and especially the ensuing subsection draw heavily.

(10) Fred *deftly* [Manner] handed a toy to the baby *by reaching behind his back* [Manner] *over lunch* [Temp] *at noon* [Temp] *in a restaurant* [Loc] *last Sunday* [Temp] *in Back Bay* [Loc] *without interrupting the discussion* [Manner].

(11) *John escaped from prison *with dynamite* [Inst] *with a machine gun* [Inst].

However, this contrast is ill-conceived, as all [Temp] phrases in (10) are different references to *the same* time of the event, all [Loc] phrases – to *the same* location of the event, and all [Manner] phrases arguably describe aspects of a *single* manner; on the other hand, the two [Inst] phrases in (11) cannot describe the same instrument – dynamite and a machine gun are necessarily *different* entities. Examples such as (10) should rather be compared to the following two examples from Zaenen & Crouch 2009, p. 646, which illustrate the (perhaps more limited) possibility to iterate arguments:

(12) I count on you, on your kindness.

(13) He lives in France, in a small village.

Goldberg (2002, pp. 334–335, 341) argues that also instrumental phrases may be iterated as long as they "concentrically" refer to the same entity, and supports this claim with the following – perhaps more controversial – examples:

(14) With a slingshot he broke the window with a rock.

(15) The robot opened the door with a key with its robotic arm.

In all the cases where two *different* entities are involved they should be expressed via coordination:

(16) Fred will perform [today and tomorrow] / *[today tomorrow].

(17) John escaped from prison [with dynamite and with a machine gun] / *[with dynamite with a machine gun].

(18) I count [on you and on his kindness] / *[on you, on his kindness].

(19) The robot opened the door with [an axe and a crowbar] / *[with an axe, with a crowbar].

It is also easy to construct examples of other iterated arguments, for example, an iterated subject, as in the following Polish example, where the three nominative NPs are understood as referring to the same person:

(20) Ważny urzędnik wczoraj przyszedł, dyrektor
important.NOM official.NOM yesterday came director.NOM
departamentu, bardzo wysoko postawiona osoba...
department.GEN very highly placed person
'An important official came yesterday: the director of a/the department, a very high-ranking person.'

It could be argued that (12)–(13), and maybe also (20), should be analysed as some special construction, maybe a type of apposition. Perhaps so. However, whatever the analysis of such examples of iterated arguments, the burden is on the shoulders of the proponents of the dichotomy to show that this analysis does not carry over to examples of iterated adjuncts, i.e., that iterability does distinguish arguments from adjuncts. Since I am not aware of such an argument, I conclude that iterability, as currently understood, fails to distinguish arguments from adjuncts.

2.3 Specificity

Another common test reflects the intuition that promiscuous types of phrases, happy to combine with all or a great number of different predicates, tend to be adjuncts, while arguments are restricted to smaller classes of predicates.

In Functional Generative Description, specificity is understood strongly: all types of adjuncts are assumed to be able to occur with all verbs (Panevová, 1974, p. 11). Taken literally, the test clearly gives undesirable results, as very few adjunct types may really depend on every verb. For example, McConnell-Ginet 1982, p. 166, notes that WEIGH fails to combine with many typical adverbials:

(21) *Annie weighs 120 pounds {heavily / beautifully / quickly / elegantly}.

(22) *Annie weighs 120 pounds {for her mother / with a fork / in an hour / toward Detroit}.

Even such prototypical types of adjuncts as temporal or locative are subject to exceptions. As shown in Koenig et al. 2003, p. 80, where an experiment consisting in the manual examination of 3909 English verbs is reported, 0.2% (i.e. 8) of them do not combine with temporal dependents and 1.8% (i.e. as many as 70) do not combine with event locations. Such ratios are bound to be much higher in the case of most other dependent types claimed to be adjuncts, e.g., manner or instrument phrases.

It is also clear that the results of this test depend on the granularity of types of dependents. For example, simplifying a little, Koenig et al. (2003) treat as arguments those dependents which may occur with up to 30% of all verbs, and as adjuncts – those which may occur with at least 90% of all verbs. It seems then that agents should count as typical adjuncts. Koenig et al. (2003) avoid this conclusion by splitting this dependent type into more fine-grained semantic roles, as proposed in Dowty 1989, 1991, and showing that each of them occurs with less than 30% of the examined verbs. However, Przepiórkowski 2016 shows that the same reasoning could be applied to Polish durative phrases, i.e. prototypical adjuncts, with the result of classifying them as arguments.

The problem that many intended adjuncts do not really combine with all verbs is duly noted in FGD, but it is played down: "it appears as a rule that such a combination is not grammatically excluded but is unusual due to cognitive or ontological reasons" (Panevová, 1974, fn. 6). Unfortunately, this view makes the test largely unusable in practice, as there is no operational procedure of distinguishing "grammatical unacceptability" from "cognitive or ontological unacceptability". Moreover, it is not clear that such a distinction is justified at all; as shown in Levin 1993, grammatical behaviour of verbs (their diathesis patterns) strongly correlates with their meaning (which may be hard to distinguish from "cognitive or ontological" aspects).

In summary, very few classes of dependent types, if indeed any, "can depend on every verb", and attempts to distinguish reasons for not satisfying this criterion have never, to the best of my knowledge, been translated into an operational test, so the specificity criterion simply does not do the job it was supposed to do.

2.4 Verbal pro-forms

A once popular test concerns the behaviour of *do so* and similar *verbal pro-forms* (Lakoff & Ross, 1976), which apparently may be substituted for a repeated VP (i.e., a verbal projection which contains all complements – that is, non-subject arguments – and perhaps some adjuncts). Multiple arguments against this syntactic status of *do so* – and against using it as a test of argumenthood – may be found in Przepiórkowski 1999a, ch. 7, and in Culicover & Jackendoff 2005, ch. 8, but since the *do so* vampire is occasionally resurrected – recently in a handbook article (Ackema, 2015) – I present the supposed test and arguments against it here.

Consider the following examples (Ackema, 2015, p. 260).

(23) John ate a banana yesterday, and Geraldine did so today.

(24) John ate a banana yesterday, and Geraldine did so, too.

(25) *John ate a banana, and Geraldine did so an apple.

The first example is grammatical, with *did so* referring to *ate a banana*, i.e., the verb and its sole complement. The second example is ambiguous: here *did so* may be understood as being substituted either for *ate a banana*, or for a larger constituent, containing also the adjunct *yesterday*. By contrast, the final example is ungrammatical supposedly because an attempt is made to substitute *did so* for a constituent which is too small, i.e., for *ate* without its complement *a banana*.

Examples showing that *do so* does not have to refer to an existing syntactic constituent are easy to find. Active–passive mismatches like the following are already noted in Bouton 1969, and many more, both from the linguistic literature and from corpora, are cited in Przepiórkowski 1999a:

(26) Because the issue had been discussed so thoroughly in our committee that afternoon, we were asked not to waste time doing so again that night. (*do so* = discuss the issue)

An antecedent of *do so* may also be nominal, as in the following corpus example from Meijs 1984:

(27) Its cord was useless in effect, so I'd no trouble in its removal; on doing so I was dumbfounded by its unexpected contents. (*doing so* = removing it)

In fact, even the weakest requirement of syntactic parallelism seems to be missing, as the antecedent of *do so* may be constructed "on the fly" from different pieces of syntactic structure (Przepiórkowski, 1999a, pp. 303–304):

(28) ... featuring people (like Woody Allen himself) who can't sing and can't dance, but do so anyway. (*do so* = sing and dance)

(29) Fortunately, the first person to die in 1990 and the first couple to file for divorce in 1990 were allowed to do so anonymously. (*do so* = die and file for divorce, respectively)

Currently, a more common analysis is that such verbal pro-forms should not be understood as involved in some syntactic reconstruction process that requires parallelism, but rather as lexical items that have no arguments (apart from the subject) but may combine with the usual adjuncts (Culicover & Jackendoff 2005, ch. 8, Williams 2015, pp. 71–72); the pro-form refers to a VP antecedent, with each possible adjunct acting as "an orphan expression that represents what is *not* the same in the interpretation of the fragment [here: *do so*] and the antecedent" (Culicover & Jackendoff, 2005, p. 289). This last statement is supported by the kind of data apparently first discussed in Miller 1990, 1992; see the perhaps somewhat artificial (30)–(31) from Miller 1992, pp. 96–97, perhaps more natural (32)–(33) from Culicover & Jackendoff 2005, p. 285, and the attested[6] (34):

(30) John kicked Mary and Peter did so to Ann.

(31) John spoke to Mary and Peter did so with Ann.

(32) Robin broke the window (with a hammer) and Mary did the same to the vase.

(33) John turned the hot dog down flat, but he wouldn't have done so with filet mignon.

(34) That's all I've been trying to say: think about what you say next time. Opinion or not, you could hurt someone. Be thankful that you did so to someone who can keep their head and not lash out like you seem to normally do.

In all these examples, the pro-form – *do the same* in (32) and *do so* in the other four sentences – occurs with an apparent prepositional adjunct which, however, corresponds to a prototypical argument in the antecedent (passivisable direct object, in the case of (30) and (32)–(34)).

Given such examples, verbal pro-forms cannot be straightforwardly used to distinguish arguments from adjuncts within dependents of other verbs; at best, one may assume, together with Culicover & Jackendoff 2005 and Williams 2015, that dependents co-occurring with *do so* and similar pro-forms are adjuncts of *do so*, but even this claim is controversial, given that the *with*-dependents in (31) and (33) should probably be classified as themes, and *to*-dependents in the other three examples above – as patients (Culicover & Jackendoff, 2005, p. 285).

3 Argument–adjunct distinction in LFG

Given that the purported argument–adjunct distinction (AAD) is so difficult to pin down, the possibility must be entertained that there is no single fundamental partition of possible dependents of a predicate into two (or three) classes. Would that be a problem for LFG?

3.1 AAD at grammatical levels

Perhaps surprisingly, the only grammatical level where the argument–adjunct distinction surfaces in contemporary LFG is f-structure, and there only as a distinction between the attributes representing (closed and open) adjuncts (ADJ and XADJ) on the one hand and the attributes representing governable functions (SUBJ, OBJ, etc.), on the other.

[6] http://mewkwota.deviantart.com/art/Everyone-Stop-Looking-at-Me-413002782

Traditionally, AAD also has a reflex in semantic forms, i.e., values of PRED: only arguments, not adjuncts, are mentioned there, and the principles of Completeness and Coherence (Dalrymple, 2001, pp. 35–39) make sure that only and all arguments listed in such semantic forms are represented as the values of SUBJ, OBJ, etc. However, as noted already in Dalrymple et al. 1993, pp. 13–14, and Kuhn 2001, § 1.3.3, Glue Semantics (Dalrymple, 1999) makes PRED – and also the principles of Completeness and Coherence – largely superfluous. As a result, in some recent work, PRED values do not mention arguments at all; for example, the lexical entry for *ate* in Asudeh & Giorgolo 2012, p. 73, contains the equation $(\uparrow \text{ PRED}) = \text{`EAT'}$ rather than $(\uparrow \text{ PRED}) = \text{`EAT}\langle\text{SUBJ}, \text{OBJ}\rangle\text{'}$.

Unlike transformational grammar, LFG has never assumed that AAD must be represented in syntactic trees. For example, Kaplan & Bresnan 1982, p. 217, propose the following syntactic rule for an English VP (abbreviated here), according to which prepositional phrases occupy the same c-structure positions, whether they are arguments ($(\uparrow (\downarrow \text{ PCASE})) = \downarrow$) or adjuncts ($\downarrow \in (\uparrow \text{ ADJUNCTS})$):

(35)
$$\begin{array}{llcccc} \text{VP} \longrightarrow & \text{V} & (\quad \text{NP} \quad) & (\quad \text{NP} \quad) & \text{PP}^* & \ldots \\ & & (\uparrow \text{ OBJ}) = \downarrow & (\uparrow \text{ OBJ}_\theta) = \downarrow & \{(\uparrow (\downarrow \text{ PCASE})) = \downarrow \mid & \\ & & & & \downarrow \in (\uparrow \text{ ADJUNCTS})\} & \end{array}$$

There is currently no standard LFG approach to semantic structure, and sometimes its very existence is denied (Andrews, 2010), but none of various approaches to s-structure assumes AAD. This is least obvious in the case of recent approaches, e.g., Asudeh et al. 2014, which – following Findlay 2014 – make the semantic structure a locus of the Lexical Mapping Theory and assume s-level attributes ARG_1, …, ARG_4. However, only a proper subset of arguments fall under the purview of LMT, and remaining arguments correspond to s-structure attributes other than ARG_n, just as in the case of adjuncts. For example, Asudeh et al. 2014, p. 81, propose the following s-structure (and mapping from f-structure) for *Kim drew Godzilla for Sandy*, in which the OBL argument *for Sandy* corresponds to the value of the s-structure attribute BENEFICIARY rather than ARG_n (their Figure 5):

(36)
$$\left[\begin{array}{ll} \text{PRED} & \text{`draw'} \\ \text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`Kim'}\end{array}\right] \\ \text{OBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`Godzilla'}\end{array}\right] \\ \text{OBL} & \left[\begin{array}{ll}\text{PRED} & \text{`for'} \\ \text{OBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`Sandy'}\end{array}\right]\end{array}\right] \\ \text{TENSE} & \text{PAST} \end{array}\right] \xrightarrow{\sigma} t\left[\begin{array}{ll} \text{REL} & \text{draw} \\ \text{EVENT} & ev[\;] \\ \text{ARG}_1 & k[\;] \\ \text{ARG}_2 & g[\;] \\ \text{BENEFICIARY} & s[\;] \end{array}\right]$$

Finally, the resulting logical forms also do not exhibit AAD, as LFG analyses commonly assume the neo-Davidsonian approach to logical forms (Parsons, 1990). For example, the sentences (37a) and (38a) may receive the respective logical forms in (37b) and (38b) (simplified here), which differ only in the name of the main predicate (*sleep* vs. *reside*), even though the locative phrase is a prototypical adjunct in (37) and a clear (obligatory) argument in (38).

(37) a. Peter sleeps in the garage.

b. $\exists e.[sleep(e) \wedge agent(e, peter) \wedge location(e, the\ garage)]$

(38) a. Peter resides in the garage.

b. $\exists e.[reside(e) \wedge agent(e, peter) \wedge location(e, the\ garage)]$

I conclude that the only level of grammatical representation that assumes the AAD is f-structure, namely, the attributes (X)ADJ vs. SUBJ, OBJ, etc.

3.2 AAD in the grammar

Even if grammatical representations do not exhibit AAD, it is possible that processes leading to their construction are sensitive to this distinction. For example, even if (37a) and (38a) have the same (up to the name of the main predicate) c-structures, s-structures and logical forms, perhaps radically different grammatical mechanisms have to be invoked to construct these analogous representations? It turns out that this is not so, especially given recent LFG developments.

Traditionally, arguments of predicates are only specified in lexical entries of these predicates and adjuncts are only added via general syntactic rules. However, some recent analyses (Asudeh et al., 2008, 2013, 2014; Asudeh & Giorgolo, 2012) blur this distinction. According to such analyses, arguments – also their semantic contributions – are adduced via calls to templates such as @AGENT and @PATIENT for the usual (deep) subjects and objects, @BENEFACTIVE for derived benefactive arguments (Asudeh & Giorgolo, 2012; Asudeh et al., 2014), or @TRANSITIVE-OBLIQUE in the analysis of Swedish Direct Motion Construction (DMC; Asudeh et al. 2008, 2013). What is important is that such calls are made not only within lexical entries, but also within grammatical rules – this is exactly the analysis of the Swedish DMC, which is signalled by a special c-structural configuration (Asudeh et al., 2013, §§ 2.2 and 4.1). Similarly, in the case of the analogous English *way*-constructions, as in *Sarah elbowed her way through the crowd*, the argument headed by *way* is added to the f-structure of the head (*elbowed* in the above example) only by virtue of a relevant template call in the lexical entry of *way*; such an argument is never mentioned in the lexical entry of the head verb (Asudeh et al., 2013, § 4.2). As this analysis is analogous (in relevant aspects) to the standard treatment of adjuncts, I conclude that the same grammatical mechanisms are involved in the introduction of arguments and adjuncts, and that the only place where AAD surfaces in contemporary LFG is f-structure, with its distinction between adjunct attributes (X)ADJ and governable functions SUBJ, OBJ, etc.

4 Argument–adjunct *non*-distinction in LFG

I propose three ways of getting rid of the last vestiges of AAD in LFG. The first is very conservative and consists in replacing (X)ADJ with specific "grammatical functions". The second follows (and exceeds) the approach known from HPSG and consists in replacing all specific attributes for arguments and adjuncts with a single

DEPS list. The third combines the former two and has the additional advantage of encoding the functional hierarchy. I only sketch the main ideas of the first two proposals, but I provide more details and a worked example (39) in the case of the third proposal.

(39) John resided in France for two years, in a village called Les Vans.

In the process, I ignore the internal structure of the nominal phrases in this sentence and their quantificational impact – I make the simplifying assumption that all NPs in (39) semantically contribute constants: j in the case of *John*, f in the case of *France*, ty in the case of *two years* and av in the case of *a village called Les Vans*.

4.1 Conservative proposal

The most conservative way to get rid of AAD altogether is to replace the attributes ADJ and XADJ, which are currently sets of adjuncts of various types, with more specific attributes such as LOC(ation), TEMP(oral), DUR(ative), XPART(icipial) (for open participial adjuncts), etc., as illustrated in (40) below.[7]

(40)
$$\left[\begin{array}{ll} \text{PRED} & \text{`RESIDE'} \\ \text{TENSE} & \text{PAST} \\ \text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`JOHN'}\end{array}\right] \\ \text{LOC} & \left\{ \left[\begin{array}{ll}\text{PRED} & \text{`IN'} \\ \text{OBJ} & \text{``FRANCE''}\end{array}\right], \left[\begin{array}{ll}\text{PRED} & \text{`IN'}\\ \text{OBJ} & \text{``A VILLAGE}\ldots\text{''}\end{array}\right] \right\} \\ \text{DUR} & \left[\begin{array}{ll}\text{PRED} & \text{`FOR'}\\ \text{OBJ} & \text{``TWO YEARS''}\end{array}\right] \end{array}\right]$$

This proposal, and the combined analysis of Section 4.3 below, does not necessarily contradict the proposal of Patejuk & Przepiórkowski 2016 that the repertoire of grammatical functions assumed in LFG be strictly limited, perhaps only to SUBJ and OBJ. Rather, I view the set of "extended grammatical functions" SUBJ, OBJ, LOC, DUR, etc., as analogous to "functors" assumed in the FGD approach to valency (Panevová, 1974, 1975): almost all of some 35 FGD functors (Žabokrtský, 2005, pp. 117–118) are defined purely semantically (e.g., LOC(ative), CUAS(e), various temporal functors, etc.), but a couple simply mark grammatical functions. In particular, the perhaps misnamed functor ACT(or) refers to the subject regardless of its semantic relation to the verb, i.e., also in case of non-agentive subjects. Similarly, in the current proposal, SUBJ and OBJ may be regarded as true grammatical functions, and the other "extended grammatical functions" such as LOC and DUR – as indicating syntactic realisations of appropriate semantic roles.

As argued in Section 2.2, there is no clear difference in terms of iterability between such new semantically defined "grammatical functions" and the standard governable functions – it seems that each may be realised as a set of phrases (see Zaenen & Crouch 2009 on OBLs) – but in order to alleviate parenthesis clutter, only

[7] For a related idea, see Nordlinger 1998, pp. 71–72, fn. 26.

those values of extended grammatical functions are represented as sets which have more than one realisation in the sentence. In the running example (39), this only concerns the locative phrases.

Note that this solution does not preserve the AAD. Assuming this dichotomy, one of the locative phrases would have to be treated as an argument (it is syntactically obligatory), and the other – as a typical adjunct. By contrast, they are both members of the LOC value in (40), without any indication of which one is an argument, and which one is an adjunct.

Similarly, once the implicit AAD vanishes, there is no need to assume that all OBLs are arguments, so the agentive *by*-phrase in passive constructions may be represented as $\text{OBL}_{\text{AGENT}}$, without any commitment to its argument/adjunct status. This evades the problem that led Grimshaw 1990 to the postulation of the interim class of “argument adjuncts”, specifically for such *by*-phrases (and possessives, in the nominal domain), and liberates LFG researchers from having to make an arbitrary decision on the status of such agentive PPs.

4.2 HPSG-like proposal

The second possibility consists in replacing all such functional attributes with a single ordered DEP(endent)S list:

(41)
$$\begin{bmatrix} \textsc{pred} & \text{‘}\textsc{reside}\text{’} \\ \textsc{tense} & \textsc{past} \\ \textsc{deps} & \left\langle \begin{bmatrix} \textsc{pred} & \text{‘}\textsc{John}\text{’} \end{bmatrix}, \left\{ \begin{bmatrix} \textsc{pred} & \text{‘}\textsc{in}\text{’} \\ \textsc{deps} & \langle \text{“}\textsc{France}\text{”} \rangle \end{bmatrix}, \begin{bmatrix} \textsc{pred} & \text{‘}\textsc{in}\text{’} \\ \textsc{deps} & \langle \text{“}\textsc{a village\ldots}\text{”} \rangle \end{bmatrix} \right\}, \begin{bmatrix} \textsc{pred} & \text{‘}\textsc{for}\text{’} \\ \textsc{deps} & \langle \text{“}\textsc{two years}\text{”} \rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

This idea seems to mirror the HPSG analysis of Przepiórkowski 1999a, ch. 9, and Bouma et al. 2001, but it goes further. In HPSG, while the final values of DEPS do not distinguish between arguments and (some) adjuncts, the grammar still retains this distinction: arguments appear on DEPS by virtue of the lexical entries of heads, while adjuncts are added to this list via a separate mechanism. Moreover, according to Bouma et al. 2001, only some (post-verbal) adjuncts end up on DEPS. By contrast, I assume that all (at least all event-related, as opposed to speaker-oriented, etc.) prototypical adjuncts appear on DEPS and that they are introduced by the same mechanisms as prototypical arguments (cf. Section 3.2).

4.3 Combined proposal

The above two proposals reflect the fundamental difference between LFG and HPSG: the former implements what Pollard & Sag 1987, p. 118, call “a ‘keyword’ theory of grammatical relations”, where each grammatical function receives a name, and the latter is based on the obliqueness hierarchy (essentially, the accessibility hierarchy of Keenan & Comrie 1977, also reflected by the relational hierarchy of Relational Grammar), with particular grammatical functions defined as particular positions in this hierarchy (the highest element is the subject, etc.). This ordered list of grammatical functions is the locus of the HPSG binding theory (Pollard & Sag, 1994, ch. 6), and plays a role in its control theory (ch. 7).

LFG also assumes such an obliqueness hierarchy, here called *functional hierarchy* (to be distinguished from the *thematic hierarchy*), and refers to it in analyses of control (Bresnan 1982a, p. 294, Dalrymple 2001, p. 345), binding (Bresnan et al. 2015, chs. 9–10, and references therein) and *wh*-movement (Dalrymple 2001, p. 412 and references therein). However, unlike in HPSG, this notion has apparently never been formalised in LFG.

I propose to combine the two approaches in a way that encodes both: particular grammatical functions and the functional hierarchy. The gist of the idea is to represent grammatical functions as a *named list*, where each element of the ordered list is annotated with the appropriate extended grammatical function, and the whole list reflects functional hierarchy:

(42)
$$
\left[\begin{array}{ll}
\textsc{pred} & \text{`}\textsc{reside}\text{'}\\
\textsc{tense} & \textsc{past}\\
\textsc{deps} & \left\langle \begin{array}{l} \textsc{subj}: \left[\textsc{pred}\ \ \text{`}\textsc{John}\text{'}\right], \\ \qquad \textsc{loc}: \left\{ \left[\begin{array}{ll}\textsc{pred} & \text{`}\textsc{in}\text{'}\\ \textsc{deps} & \left\langle \textsc{obj}: \text{``}\textsc{France}\text{''}\right\rangle\end{array}\right], \left[\begin{array}{ll}\textsc{pred} & \text{`}\textsc{in}\text{'}\\ \textsc{deps} & \left\langle \textsc{obj}: \text{``}\textsc{a village}\ldots\text{''}\right\rangle\end{array}\right]\right\},\\ \qquad \textsc{dur}: \left[\begin{array}{ll}\textsc{pred} & \text{`}\textsc{for}\text{'}\\ \textsc{deps} & \left\langle\textsc{obj}: \text{``}\textsc{two years}\text{''}\right\rangle\end{array}\right]\end{array}\right\rangle
\end{array}\right]
$$

According to the f-structure (42), there are four dependents corresponding to three extended grammatical functions: SUBJ, LOC and DUR. The subject outranks all other dependents, and both locative dependents outrank the durative dependent (with the order between these two locatives undefined).

Technically, lists have a standard (Shieber, 1986, p. 29) encoding in feature structures via attributes such as FIRST and REST (or HEAD and TAIL). In the case of named lists, the FIRST (or HEAD) attribute is replaced with the specific name. So (42) above is a shorthand for the complete f-structure (43) on the next page.

The advantages of such a representation are multiple. First of all, it makes it possible to formalise compactly those modules of LFG which assume a functional hierarchy: they only need to make a reference to the order of elements on the DEPS list. Second, it extends the HPSG approach, in which some grammatical functions are already singled out (see Pollard & Sag 1994, ch. 9, and references therein, as well as the use of XARG in Sag 2007, etc.). Such an extension is needed for example to explicitly mark the passivisable object. Note that it is not sufficient to say that a verb has a passive form and assume that the second DEPS element is the passivisable object, as some such objects are optional and, as a result, the second position of a passivisable verb may be occupied by a dependent which is not passivisable.[8] Third, various grammatical phenomena which do not necessarily distinguish adjuncts from arguments, such as case assignment (Przepiórkowski, 1999a,b) and extraction (Bouma et al., 2001), now receive a uniform locus of analysis – this is in fact the main motivation for the introduction of DEPS in HPSG.

(43)
$$\left[\begin{array}{ll}
\text{PRED} & \text{`RESIDE'} \\
\text{TENSE} & \text{PAST} \\
\text{DEPS} & \left[\begin{array}{ll}
\text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`JOHN'}\end{array}\right] \\
\text{REST} & \left[\begin{array}{ll}
\text{LOC} & \left\{ \left[\begin{array}{ll}\text{PRED} & \text{`IN'}\\ \text{DEPS OBJ} & \text{``FRANCE''}\end{array}\right], \left[\begin{array}{ll}\text{PRED} & \text{`IN'}\\ \text{DEPS OBJ} & \text{``A VILLAGE...''}\end{array}\right]\right\} \\
\text{REST} & \left[\begin{array}{ll}\text{DUR} & \left[\begin{array}{ll}\text{PRED} & \text{`FOR'}\\ \text{DEPS OBJ} & \text{``TWO YEARS''}\end{array}\right]\end{array}\right]
\end{array}\right]
\end{array}\right]
\end{array}\right]$$

Standard LFG analyses carry over to this new feature architecture, with two modifications. First, whenever an analysis refers to ADJ, it should now refer to an extended grammatical function representing a specific type of adjunct (e.g., LOC or DUR), or perhaps a disjunction of such grammatical functions. The second modification that is needed is more technical: functional equations accessing grammatical functions must be modified, as these functions are now embedded within DEPS. For example, an equation like: $\downarrow = (\uparrow \text{ OBJ})$ may now – at least in the case of English, with its obligatory subjects – be replaced with: $\downarrow = (\uparrow \text{ DEPS REST OBJ})$. In the case of extended grammatical functions other than subject or object, such equations are more complex, as the number of elements preceding them on the DEPS list is not constant. For example, the specification of a position in a syntactic rule which is occupied by locative or durative phrases may be adorned with an equation such as:

(44) $\downarrow \in (\uparrow \text{ DEPS REST}^* \ \{\text{LOC}|\text{DUR}\})$

Instead of specifying such equations directly, I propose a template (see the Appendix), let us call it GF, that should be used whenever an assignment to an extended grammatical function is to be made. With such a template in hand, the above equation will be shortened to @GF({LOC|DUR}), and typical syntactic rules will look as follows:[9]

[8]The external reviewer suggests defining objects as the second (after subjects) least oblique structurally cased NPs. However, as argued at length in Przepiórkowski 1999a, passivisation in some languages, including Polish, is dissociated from the structural vs. lexical case assignment dichotomy, so I will not follow this suggestion here.

[9]I adopt the usual abbreviatory conventions concerning the omission of head equations $\uparrow=\downarrow$ (Dalrymple 2001, p. 119, Bresnan et al. 2015, p. 106). Obviously, (46) is much simplified, as also other types of constituents may be sisters to I and as PPs may also bear grammatical functions other than LOC or DUR.

(45) IP $\longrightarrow$ NP I$'$
@GF(SUBJ)

(46) I$'$ $\longrightarrow$ I (NP) (PP)*
@GF(OBJ) @GF({LOC|DUR})

Additionally, I make use of abbreviations such as:

(47) $^{\mathrm{D}}$SUBJ $\equiv$ DEPS SUBJ

(48) $^{\mathrm{D}}$OBJ $\equiv$ DEPS REST* OBJ

(49) $^{\mathrm{D}}$LOC $\equiv$ DEPS REST* LOC, etc.

Hence, @GF({LOC|DUR}) is equivalent to: $\downarrow \in$ ($\uparrow$ {$^{\mathrm{D}}$LOC|$^{\mathrm{D}}$DUR}). However, the definition of GF in the Appendix also makes sure that each extended grammatical function appears on DEPS at most once and that they appear in the order reflecting the functional hierarchy.

Let us illustrate this proposal with the running example (39). The main verb receives the following lexical entry, where only the penultimate line – requiring the presence of LOC within DEPS – is non-standard; an entry for an intransitive verb such as *slept* would be analogous, minus this penultimate line:

(50) *resided* I ($\uparrow$ PRED) = 'RESIDE'
@AGENT @PAST
($\uparrow$ $^{\mathrm{D}}$LOC)
$\lambda e.reside(e) : (\uparrow_\sigma \text{ EVENT}) \multimap \uparrow_\sigma$

I assume that the templates for AGENT and PAST are largely analogous to those proposed in Asudeh et al. 2014 (cf. their (48) and (54)):

(51) AGENT := @ARG1
$\lambda P \lambda x \lambda e.\ P(e) \wedge agent(e,x)$:
$[(\uparrow_\sigma \text{ EVENT}) \multimap \uparrow_\sigma] \multimap (\uparrow_\sigma \text{ ARG1}) \multimap (\uparrow_\sigma \text{ EVENT}) \multimap \uparrow_\sigma$

(52) PAST := ($\uparrow$ TENSE) = PAST
$\lambda P \exists e.\ P(e) \wedge past(e)\ :\ [(\uparrow_\sigma \text{ EVENT}) \multimap \uparrow_\sigma] \multimap \uparrow_\sigma$

@ARG1 in (51) invokes the part of LMT responsible for mapping agents to appropriate grammatical functions; as used in the lexical entry (50) for the past form *resided*, it has the same effect as: $(\uparrow {}^{\mathrm{D}}\text{SUBJ})_\sigma = (\uparrow_\sigma \text{ ARG1})$; in the analysis of the running example, this glue resource is contributed by *John*. Hence, the combination of meaning constructors in (50)–(51) would yield:

(53) $\lambda x \lambda e.\ reside(e) \wedge agent(e,x)\ :\ (\uparrow_\sigma \text{ ARG1}) \multimap (\uparrow_\sigma \text{ EVENT}) \multimap \uparrow_\sigma$,

then, after combining with the meaning constructor provided by *John*:

(54) $\lambda e.\ reside(e) \wedge agent(e,j)\ :\ (\uparrow_\sigma \text{ EVENT}) \multimap \uparrow_\sigma$,

and finally, after combining with the meaning constructor in (52):

(55) $\exists e.\ reside(e) \wedge agent(e,j) \wedge past(e)\ :\ \uparrow_\sigma$.

Obviously, this derivation for the ungrammatical sentence *John resided* is blocked by the unsatisfied constraint in the penultimate line of (50), but an analogous derivation would work for *John slept*.

For semantic prepositions, I assume lexical entries such as (56)–(57), analogous to the lexical entry for the benefactive *for* in Asudeh et al. 2014 (their (64)):

(56) *in* P
$(\uparrow \text{PRED}) = \text{`IN'}$
$\%\text{HD} = ({}^{\text{D}}\text{LOC} \uparrow)$
$\lambda x \lambda P \lambda e.\ P(e) \wedge \mathit{location}(e, x) :$
$(\uparrow {}^{\text{D}}\text{OBJ})_\sigma \multimap$
$[(\%\text{HD}_\sigma\ \text{EVENT}) \multimap \%\text{HD}_\sigma] \multimap$
$(\%\text{HD}_\sigma\ \text{EVENT}) \multimap \%\text{HD}_\sigma$

(57) *for* P
$(\uparrow \text{PRED}) = \text{`FOR'}$
$\%\text{HD} = ({}^{\text{D}}\text{DUR} \uparrow)$
$\lambda x \lambda P \lambda e.\ P(e) \wedge \mathit{duration}(e, x) :$
$(\uparrow {}^{\text{D}}\text{OBJ})_\sigma \multimap$
$[(\%\text{HD}_\sigma\ \text{EVENT}) \multimap \%\text{HD}_\sigma] \multimap$
$(\%\text{HD}_\sigma\ \text{EVENT}) \multimap \%\text{HD}_\sigma$

Note the local name %HD, which – in the case of the running example – points to the matrix f-structure shown in (43), but only if the PP headed by the preposition is the value of an appropriate extended grammatical function (LOC in (56) and DUR in (57)). So, while the rule (46) is indeterminate about the grammatical function assigned to the PP, the right function must be assigned for %HD to have a value and for the meaning constructor to be defined.

Once *in* combines with *France*, the following meaning constructor will result (with g referring to the matrix f-structure and f – to *France*):

(58) $\lambda P \lambda e.\ P(e) \wedge \mathit{location}(e, f) : [(g_\sigma\ \text{EVENT}) \multimap g_\sigma] \multimap (g_\sigma\ \text{EVENT}) \multimap g_\sigma.$

It will further combine with the meaning constructor introduced by *reside*, giving:

(59) $\lambda e.\ \mathit{reside}(e) \wedge \mathit{location}(e, f) : (g_\sigma\ \text{EVENT}) \multimap g_\sigma.$

Analogously, taking into consideration semantic contributions of *in a village...* and *for two years* (still ignoring the quantificational impact of the NPs), we get:

(60) $\lambda e.\ \mathit{reside}(e) \wedge \mathit{location}(e, f) \wedge \mathit{location}(e, av) \wedge \mathit{duration}(e, ty) :$
$(g_\sigma\ \text{EVENT}) \multimap g_\sigma.$

Combining this result with semantic contributions of the AGENT template, *John* and the PAST template, we end up with the expected:

(61) $\exists e.\ \mathit{reside}(e) \wedge \mathit{agent}(e, j) \wedge \mathit{location}(e, f) \wedge \mathit{location}(e, av) \wedge$
$\mathit{duration}(e, ty) \wedge \mathit{past}(e) : g_\sigma.$

Note that there is no fundamental difference between the representation of the durative *for two years*, a prototypical adjunct in (39), and the two locative phrases, at least one of which is obligatory and, hence, an argument. In fact, while other approaches would treat one locative phrase as an argument, and the other as an adjunct, here both locative phrases contribute to the value of the same LOC feature in (43), and have fully parallel semantic representations in (61).

5 Conclusion

Given that – after well over half a century of attempts to operationalise the purported argument–adjunct distinction – we do not seem any closer to a coherent and precise characterisation of this dubious dichotomy, it is high time to ask how detrimental it would be for LFG if it were generally conceded one day that AAD is just another linguistic hoax. The surprising answer is: formally, almost not at all. I argued that contemporary LFG makes this distinction only at the level of f-structures, and there only by insisting on the presence of the separate attributes (X)ADJ. I proposed three ways of getting rid of this dichotomy altogether, which do not seem to compromise previous LFG analyses. In particular, the final proposal, which combines the main insights of LFG and HPSG, has the additional advantage of formally encoding the functional hierarchy, which plays a role in LFG analyses of binding, control and *wh*-extraction.

References

Ackema, Peter. 2015. Arguments and adjuncts. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – theory and analysis. International handbook. Volume 1*, 246–274. Berlin: De Gruyter Mouton.

Andrews, Avery D. 2010. Propositional glue and the correspondence architecture of LFG. *Linguistics and Philosophy* 33(3). 141–170.

Asudeh, Ash, Mary Dalrymple & Ida Toivonen. 2008. Constructions with lexical integrity: Templates as the lexicon–syntax interface. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'08 conference*, 68–88. University of Sydney, Australia: CSLI Publications.

Asudeh, Ash, Mary Dalrymple & Ida Toivonen. 2013. Constructions with Lexical Integrity. *Journal of Language Modelling* 1(1). 1–54.

Asudeh, Ash, Gianluca Giorgolo & Ida Toivonen. 2014. Meaning and valency. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'14 conference*, 68–88. Stanford, CA: CSLI Publications.

Asudeh, Ash & Gianluca Giorgolo. 2012. Flexible composition for optional and derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'12 conference*, 64–84. Stanford, CA: CSLI Publications.

Bouma, Gosse, Robert Malouf & Ivan A. Sag. 2001. Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory* 19(1). 1–65.

Bouton, Lawrence F. 1969. Identity constraints on the do-so rule. *Papers in Linguistics* 1(2). 231–247.

Bresnan, Joan. 1982a. Control and complementation. In Bresnan (1982b) 283–390.

Bresnan, Joan (ed.). 1982b. *The mental representation of grammatical relations* MIT Press Series on Cognitive Theory and Mental Representation. Cambridge, MA: The MIT Press.

Bresnan, Joan. 1982c. Polyadicity. In Bresnan (1982b) 149–172.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2015. *Lexical-functional syntax* Blackwell Textbooks in Linguistics. Wiley-Blackwell 2nd edn.

Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler Syntax*. Oxford University Press.

Dalrymple, Mary (ed.). 1999. *Semantics and syntax in Lexical Functional Grammar: The resource logic approach.* Cambridge, MA: The MIT Press.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. San Diego, CA: Academic Press.

Dalrymple, Mary, Angie Hinrichs, John Lamping & Vijay Saraswat. 1993. The resource logic of complex predicate interpretation. In *Proceedings of ROCLING 1993*, 3–21.

Dowty, David. 1989. On the semantic content of the notion of 'thematic role'. In Gennero Chierchia, Barbara H. Partee & Raymond Turner (eds.), *Properties, types and meaning: II*, 69–129. Dordrecht: Kluwer.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Fillmore, Charles J. 1986. Pragmatically controlled zero anaphora. In *Proceedings of the twelfth annual meeting of the Berkeley Linguistics Society*, 95–107. Berkeley: Berkeley Linguistics Society.

Findlay, Jamie. 2014. Mapping theory without argument structure. Unpublished manuscript, University of Oxford.

Goldberg, Adele E. 2002. Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13(4). 327–356.

Goldberg, Adele E. & Farrell Ackerman. 2001. The pragmatics of obligatory adjuncts. *Language* 77. 798–814.

Grimshaw, Jane. 1990. *Argument structure* Linguistic Inquiry Monographs. Cambridge, MA: The MIT Press.

Grimshaw, Jane & Sten Vikner. 1993. Obligatory adjuncts and the structure of events. In Eric Reuland & Werner Abraham (eds.), *Knowledge and language*, vol. II, 143–155. Dordrecht: Kluwer.

Herbst, Thomas & Ian Roe. 1996. How obligatory are obligatory complements? An alternative approach to the categorization of subjects and other complements in valency grammar. *English Studies* 2. 179–199.

Jung, Yunsun. 1997. Obligatory adjuncts. In Ralph C. Blight & Michelle J. Moosally (eds.), *Proceedings of the 1997 Texas Linguistics Society Conference on the syntax and semantics of predication*, vol. 38 Texas Linguistic Forum, 161–171. Austin, TX.

Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Bresnan (1982b) 173–281.

Keenan, Edward L. & Bernard Comrie. 1977. Noun Phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8. 63–99.

Koenig, Jean-Pierre, Gail Mauner & Breton Bienvenue. 2003. Arguments for adjuncts. *Cognition* 89. 67–103.

Kuhn, Jonas. 2001. Resource sensitivity in the syntax-semantics interface: Evidence from the German Split NP construction. In Detmar Meurers & Tibor Kiss (eds.), *Constraint-based approaches to Germanic syntax*, 177–215. Stanford, CA: CSLI Publications.

Lakoff, George & John Robert Ross. 1966. Criterion for Verb Phrase constituency. Mathematical Linguistics and Automatic Translation Report NSF-17 Computation Laboratory, Harvard University Cambridge, MA.

Lakoff, George & John Robert Ross. 1976. Why you can't *do so* into the sink. In James D. McCawley (ed.), *Notes from the linguistic underground*, vol. 7 Syntax and Semantics, 101–111. New York: Academic Press. Originally appeared as

Lakoff & Ross (1966).
Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
McConnell-Ginet, Sally. 1982. Adverbs and logical form: A linguistically realistic theory. *Language* 58(1). 144–184.
Meijs, Willem. 1984. 'You can do so if you want to': Some elliptic structures in Brown and LOB and their syntactic description. In Jan Aarts & Willem Meijs (eds.), *Corpus linguistics: Recent developments in the use of computer corpora in english language research*, 141–162. Amsterdam: Rodopi.
Miller, Philip H. 1990. Pseudogapping and *Do So* substitution. In Michael Ziolkowski, Manuela Noske & Karen Deafon (eds.), *Papers from the 26th regional meeting of the chicago linguistic society*, 293–303. Chicago, IL: Chicago Linguistic Society.
Miller, Philip H. 1992. *Clitics and constituents in phrase structure grammar*. New York: Garland.
Needham, Stephanie & Ida Toivonen. 2011. Derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'11 conference*, 401–421. Stanford, CA: CSLI Publications.
Nordlinger, Rachel. 1998. *Constructive case: Evidence from Australian languages*. Stanford, CA: CSLI Publications.
Panevová, Jarmila. 1974. On verbal frames in Functional Generative Description. Part 1. *The Prague Bulletin of Mathematical Linguistics* 22. 3–40.
Panevová, Jarmila. 1975. On verbal frames in Functional Generative Description. Part 2. *The Prague Bulletin of Mathematical Linguistics* 23. 17–52.
Parsons, Terence. 1990. *Events in the semantics of English: A study in subatomic semantics*. Cambridge, MA: The MIT Press.
Patejuk, Agnieszka & Adam Przepiórkowski. 2016. Reducing the 'F' in 'LFG': Arguments for fewer grammatical functions in Lexical Functional Grammar. In Doug Arnold, Miriam Butt, Berthold Crysmann & Tracy Holloway King (eds.), *The proceedings of the HeadLex16 conference*, Stanford, CA: CSLI Publications.
Pollard, Carl & Ivan A. Sag. 1987. *Information-based syntax and semantics, volume 1: Fundamentals* (CSLI Lecture Notes 13). Stanford, CA: CSLI Publications.
Pollard, Carl & Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago, IL: Chicago University Press / CSLI Publications.
Przepiórkowski, Adam. 1999a. *Case assignment and the complement-adjunct dichotomy: A non-configurational constraint-based approach*. Tübingen: Universität Tübingen Ph.D. dissertation.
Przepiórkowski, Adam. 1999b. On case assignment and 'adjuncts as complements'. In Gert Webelhuth, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and constructional aspects of linguistic explanation*, 231–245. Stanford, CA: CSLI Publications.
Przepiórkowski, Adam. 2016. Against the Argument–Adjunct Distinction in Func-

tional Generative Description. *The Prague Bulletin of Mathematical Linguistics* 106. 5–20.
Sag, Ivan A. 2007. Remarks on locality. In Stefan Müller (ed.), *Proceedings of the HPSG 2007 conference*, 394–414. Stanford, CA: CSLI Publications.
Sawicki, Lea. 1988. *Verb-valency in contemporary Polish* (Tübinger Beiträge zur Linguistik 319). Tübingen: Gunter Narr Verlag.
Sgall, Petr, Eva Hajičová & Jarmila Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. Dordrecht: Reidel.
Shieber, Stuart M. 1986. *An introduction to unification-based approaches to grammar* (CSLI Lecture Notes 4). Stanford, CA: CSLI Publications.
Somers, Harold L. 1984. On the validity of the complement-adjunct distinction in valency grammar. *Linguistics* 22. 507–530.
Szymańska, Izabela & Grzegorz Śpiewak. 2000. Deconstructing the so-called obligatory adverbial modification – an excercise in data interpretation. In Bożena Rozwadowska (ed.), *PASE papers in language studies: Proceedings of the 8th annual conference of the Polish Association for the Study of English*, 321–337. Wrocław: Aksel s.c.
Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Paris: Klincksieck.
Toivonen, Ida. 2013. English benefactive NPs. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'13 conference*, 503–523. Stanford, CA: CSLI Publications.
Tutunjian, Damon & Julie E. Boland. 2008. Do we need a distinction between arguments and adjuncts? Evidence from psycholinguistic studies of comprehension. *Language and Linguistics Compass* 2(4). 631–646.
Vater, Heinz. 1978. On the possibility of distinguishing between complements and adjuncts. In Werner Abraham (ed.), *Valence, semantic case and grammatical relations*, vol. 1 Studies in Language Companion Series (SLCS), 21–45. Amsterdam: John Benjamins.
Williams, Alexander. 2015. *Arguments in syntax and semantics*. Cambridge: Cambridge University Press.
Zaenen, Annie & Dick Crouch. 2009. OBLs hobble computations. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of the LFG'09 conference*, 644–654. Trinity College, Cambridge, UK: CSLI Publications.
Žabokrtský, Zdeněk. 2005. *Valency lexicon of Czech verbs*: Charles University, Prague Ph.D. Thesis.

Appendix: Encoding the functional hierarchy

This appendix contains a definition of a template, GF, used to assign extended grammatical functions. Its sole argument passes the name of the function (SUBJ, LOC, etc.). This grammatical function is assigned to ↓, but it would be easy to define a more general two-argument template which also takes a path to the f-structure to be assigned the grammatical function.

The template is a disjunction of statements like the following, with the effect that ↓ is assigned the grammatical function OBJ if OBJ is indeed the function F passed via the template call (and similarly for other functions):

(62) $\mathrm{F} =_c \textsc{obj} \ \wedge \ \downarrow \in (\uparrow \textsc{deps} \quad \textsc{rest}^* \quad \mathrm{F})$

As just described, an equivalent template could be defined within one line. Splitting the definition into separate statements for each grammatical function is necessary to ensure that each grammatical function appears only once within DEPS, that the list contains no gaps (i.e., no REST without a grammatical function attribute at the same level of f-structure) and in order to encode the functional hierarchy, apparently not formally encoded in LFG so far. To this end, off-path constraints are attached to REST attributes in the path, to the effect that all of them must have accompanying attributes corresponding to less oblique grammatical functions. For example, in the case of OBJ, the full path must be either DEPS OBJ (e.g., in the case of prepositions, but – arguably – also in the case of some verbs in Russian or Polish) or DEPS REST OBJ, but then REST must have a sister attribute for a grammatical function less oblique than OBJ, i.e., for SUBJ.

This functional hierarchy is defined via the following abbreviations, each specifying the set of grammatical functions less oblique than the given one; for example, HGFS-OBLSO specifies the set of functions less oblique than $\textsc{obl}_{\textsc{source}}$, i.e.: SUBJ, OBJ and $\textsc{obj}_{\textsc{theme}}$:[10]

(63) $\textsc{hgfs-obj} \equiv \textsc{subj}$

(64) $\textsc{hgfs-objth} \equiv \{\textsc{hgfs-obj} | \textsc{obj}\}$

(65) $\textsc{hgfs-oblso} \equiv \{\textsc{hgfs-objth} | \textsc{obj}_{\textsc{theme}}\}$

(66) $\textsc{hgfs-oblgo} \equiv \{\textsc{hgfs-oblso} | \textsc{obl}_{\textsc{source}}\}$, etc.

Given such abbreviations, the GF template may be defined as follows:[11]

(67)
$$
\begin{array}{llllll}
\textsc{gf}(\mathrm{F}) := & \{\ \mathrm{F} =_c \textsc{subj} & \wedge \downarrow \in (\uparrow \textsc{deps} & & \mathrm{F}) & | \\
& \ \ \mathrm{F} =_c \textsc{obj} & \wedge \downarrow \in (\uparrow \textsc{deps} & \underset{(\leftarrow \{\textsc{hgfs-obj}\})}{\textsc{rest}^*} & \mathrm{F}) & | \\
& \ \ \mathrm{F} =_c \textsc{obj}_{\textsc{theme}} & \wedge \downarrow \in (\uparrow \textsc{deps} & \underset{(\leftarrow \{\textsc{hgfs-objth}\})}{\textsc{rest}^*} & \mathrm{F}) & | \\
& \ \ \mathrm{F} =_c \textsc{obl}_{\textsc{source}} & \wedge \downarrow \in (\uparrow \textsc{deps} & \underset{(\leftarrow \{\textsc{hgfs-oblso}\})}{\textsc{rest}^*} & \mathrm{F}) & | \\
& \ \ \mathrm{F} =_c \textsc{obl}_{\textsc{goal}} & \wedge \downarrow \in (\uparrow \textsc{deps} & \underset{(\leftarrow \{\textsc{hgfs-oblgo}\})}{\textsc{rest}^*} & \mathrm{F}) & \ldots\}
\end{array}
$$

Note that such off-path constraints ensure that the DEPS list contains no gaps (whenever a function is assigned to some element of the list, all previous elements must also have associated grammatical functions) and that each grammatical function may occur at most once (it is assigned only if all grammatical functions assigned to previous DEPS elements are irreflexively less oblique).

[10] I do not make any substantive linguistic claims here about the relative order of various extended grammatical functions in the obliqueness hierarchy. Also, this formalisation assumes mostly traditional LFG grammatical functions, including $\textsc{obj}_{\textsc{theme}}$, $\textsc{obl}_{\textsc{source}}$, etc., rather than more semantically defined "functions", as suggested in the main text.

[11] I assume here that all extended grammatical functions are in principle iterable (cf. Section 2.2), so all functional attributes are set-valued.

On the argument structure of raising-to-subject with passive predicates in Swedish

Rickard Ramhöj
University of Gothenburg

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 581–598

Keywords: raising-to-subject, raising-to-object, passive predicates, Swedish, evidentiality

Ramhöj, Rickard. 2016. On the argument structure of raising-to-subject with passive predicates in Swedish. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 581–598. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.30.

Abstract

This paper concerns the argument structure analysis of raising-to-subject with passive predicates in Swedish and other Germanic languages. Support is given for the analysis in which the raising-to-subject construction constitutes a regular passive, the passive counterpart of active raising-to-object. The fact that there does not seem to be an active counterpart for certain predicates, such as the predicate *say*, as well as the fact that raising-to-subject does not seem to be possible with the periphrastic passive in Swedish is attributed to certain semantic restrictions on the raising-to-object construction and the periphrastic passive construction, respectively.

1 Introduction

This paper gives an LFG-analysis of the argument structure of raising-to-subject with passive predicates in Swedish and other Germanic languages, giving support to the traditional standpoint in which passive raising-to-subject constitutes the passive counterpart to active raising-to-object. Furthermore, the paper discusses the reasons for the differences between the Germanic languages in which passive construction can be combined with raising-to-subject, concluding that what passive construction is used in a language corresponds to the general restrictions on the available passive constructions in that language.

The outline of the paper is as follows. Section 2 gives a background to the problem under discussion. Section 3 concerns related studies on raising-to-subject with passive predicates in the Germanic languages, focusing on an HPSG analysis of the construction in Danish. Section 4 discusses why the construction is not possible with the periphrastic passive in Danish and Swedish, and Section 5 why certain verbs only occur in the raising-to-object construction when the object is a reflexive pronoun. The conclusions drawn from sections 4 and 5 are then formalised within the Lexical Mapping Theory of LFG in Section 6. Section 7 summarizes the main conclusions of the study.

2 Background

In Swedish, as well as in several of the other Germanic languages, there is a construction in which the subject of a passive predicate has a thematic role only in relation to an embedded infinitival predicate. In (1-a), an example is given with the passive predicate *sägs* 'be said'. Example (1-b) illustrates that raising-to-subject is typically only available with the morphological passive in Swedish, formed with a suffix *-s*, and not with the periphrastic passive, formed with the auxiliary *bliva* 'become' in combination with a past participle.

[†]I thank the reviewers for their constructive feedback on an earlier version of the paper. I also thank the participants of the HeadLex16 conference in Warsaw for valuable comments and suggestions.

(1) a. Hon sägs vara en utpräglad målskytt.
She says.PASS be a specialized goal-scorer
'She is said to be a specialized goal scorer'.
(http://www.vf.se/node/315601)

b. *Hon blir sagd vara en utpräglad målskytt.
She becomes said be a specialized goal-scorer
'She is said to be a specialized goal scorer.'
[constructed]

The subject *hon* in (1-a) does not have a thematic role in relation to the passive predicate *sägs* 'be said'. Instead it has a thematic role in relation to the embedded predicate *vara en utpräglad målskytt* 'be a specialised goal scorer'.

The fact that there is no thematic role associated with the subject is reinforced by the possibility of a non-referential subject.

(2) Det sägs att hon är en utpräglad målskytt.
EXPL say.PASS that she is a specialised goal-scorer
'It is said that she is a specialised goal scorer.'
[constructed]

In (2), the non-referential *det* 'it' constitutes the subject of the passive predicate *sägs* 'is said'.

If we assume that the subject of the passive predicate corresponds to the object of an active predicate, which is the traditional view on passives (e.g. Teleman et al., 1999, 360), the active correspondent to the passive sentence in (1-a) would be the subject-to-object raising construction in (3).

(3) *Folk säger henne vara en utpräglad målskytt.
People say her be a specialized goal-scorer
'People say that she is a specialized goal scorer.'
[constructed]

The object *henne* 'her' in (3) would correspond to the subject *hon* 'she' in (1-a). However, there is a problem about this correspondence. The sentence in (3) is unacceptable, which has led to claims that raising-to-subject with passive predicates should be analysed as a non-canonical passive, where the subject of the passive does not correspond to the object of any perceivable active construction (Ørsnes, 2011; Ørsnes & Müller, 2013).

Even though the sentence in (3) is unacceptable, the predicate *säga* 'say' is not always unacceptable in the raising-to-object construction. When the object is a reflexive pronoun, raising-to-object seems possible, as exemplified in (4). Note that it is also possible to insert the emphatic *själv* 'self', showing that this is no intrinsic reflexive.

(4) Hon säger sig/sig själv vara en utpräglad målskytt.
She says PRO.REFL be a specialized goal-scorer
'She considers herself to be a specialized goal scorer.'
[constructed]

In (4), the object is the reflexive pronoun *sig* or *sig själv*, which is coreferential with the subject referent. This referent is also associated with the subject of the embedded predicate *vara en utpräglad målskytt* 'be a specialised goal scorer'.

The verb *säga* 'say' thus only occurs in raising-to-object when the object is a reflexive pronoun. Other verbs, such as *anse* 'consider', occur in both raising-to-subject and raising-to-object, without the object being restricted to reflexive pronouns. This is shown in (5).

(5) a. Folk anser henne vara en utpräglad målskytt.
People consider her be a specialized goal-scorer
'People consider her to be a specialized goal scorer.'
[constructed]

b. Hon anses vara en utpräglad målskytt.
She consider.PASS be a specialized goal-scorer
'She is considered to be a specialized goal scorer.'
[constructed]

In both (5-a) and (5-b), the pronoun *henne/hon* 'her/she' does not have a thematic role in relation to the main clause predicate *anse/anses* 'consider/be considered', but instead to the embedded predicate *vara en utpräglad målskytt* 'be a specialised goal scorer'.

As pointed out in the introduction, the paper concerns two questions relating to the data shown above. The first question concerns the relationship between passive raising-to-subject and active raising-to-object. Do these form an active-passive alternation for predicates such as *säga* 'say', or is passive raising-to-subject derived from another type of sentence? The second question concerns the reasons why it is only the morphological passive that occurs in raising-to-subject in Swedish (and possibly also Danish). As will be seen, in other languages, such as English, raising-to-subject is also possible with a periphrastic passive.

3 Related studies

Raising-to-subject with passive predicates is found in several, but not all, Germanic languages. Examples are given in (6) for Danish, Dutch, English, Norwegian and German, respectively. All but German exhibit the construction.

(6) a. Han påstås at være bortrejst
he claim.PASS to be away
'He is claimed to be away.'
(Ørsnes, 2011: 24)

b. De boten worden geacht over een dag of vier Kaap Hoorn te
the boats became considered over one day or four Cape Horn to
ronden.
round
'The boats are predicted to round Cape Horn in about four days.'
(Noël & Colleman, 2010: 158)

c. Melvin was believed to be an addict (by everyone).
(Postal, 1974: 56)

d. pasienten (. . .) kan tenkes å være gravid.
patient.DEF can think.pass to be pregnant
'The patient might be assumed to be pregnant.'
(Lødrup, 2008: 175)

e. *Obama wird behauptet die Wahlen zu gewinnen
Obama is claimed the elections to win
'Obama is claimed to win the elections.'
(Ørsnes, 2011: 23)

The construction seems to be most productive in English and Swedish (Postal, 1974; Lyngfelt, 2010), but it is also found in Norwegian (Lødrup, 2008) and Danish (Ørsnes, 2011). In Dutch, raising-to-subject with passive predicates is restricted to a considerably more limited number of verbs than the other languages (Noël & Colleman, 2010). Noël & Colleman (2010, 161-162) list the predicates *geacht worden* ('be considered/supposed to'), *verondersteld worden* ('be supposed to') and *verwacht worden* ('be expected to'). In Standard German, the construction is said not to be found at all (Reis, 1973; Ørsnes, 2011).

There are basically two ways of analysing raising-to-subject with passive predicates, concerning the relationship between passive raising-to-subject and active raising-to-object. One way is to analyse raising-to-subject with passive predicates as a regular passive in which the subject of the passive predicate corresponds to the object of an active sentence. In that case, the subject of the raising-to-subject sentence corresponds to the object of the raising-to-object sentence. This is unproblematic for predicates such as Swedish *anse*, as exemplified in (4). However, it is then necessary to explain why there does not seem to be an active correspondent in connection to certain other predicates, such as Swedish and English *säga*/*say*. Such an explanation will be given in this paper, based on the semantic restrictions on the raising-to-object construction.

If raising-to-subject with passive predicates is not seen as a regular passive, the alternative[1] is to see it as an irregular passive, deriving not from active raising-to-object, but from another type of sentence containing a finite clausal complement. For those who take this stance (e.g. Ørsnes, 2013), it is necessary to explain why it seems that only the languages that allow raising-to-object allow raising-to-subject

[1]There is also the analysis, which occurs within constructionist approaches, in which the relevant sentence is not derived from an active sentence, but constitutes a construction in its own right (Noël & Colleman, 2010; Lyngfelt, 2010). This approach will not be further discussed here.

with passive predicates. The question is also what accounts for the impossibility of raising-to-subject with passive predicates in languages such as German if the construction is simply derived from sentences with verbs taking finite complements, which are possible in German. One study that seeks to answer this question is Ørsnes (2011, 2013). In the rest of this section, Ørsnes' analysis will be presented, and it will be shown that, although descriptively accurate, Ørsnes' approach lacks in explanatory force.

Ørsnes (2013) gives an HPSG analysis of raising-to-subject with passive predicates for Danish. Ørsnes has a lexical rule in which the input is verbs taking a finite clausal complement, where the subject is a so-called Designated Argument (i.e. restricted to agents and experiencers). The output of the rule is raising predicates, where the verb has the *s*-form and where the complement is a full infinitive. There is thus no connection made between passive raising-to-subject and active raising-to-object. Ørsnes' rule is governed by four constraints: (i) Passivization, (ii) the Subject Condition, (iii) the Raising Principle, and (iv) the Participle Principle. Passivization suppresses the most prominent argument, letting a less prominent argument (NB not restricted to the object) be linked to the subject function. The Subject Condition requires all verbal predicators to have a subject. The Raising Principle says that, if a referential subject is not assigned a thematic role by a verb, it must be structure-shared with the unexpressed subject of an embedded predicate. Finally, the Participle Principle is a language-specific constraint, which says that 'a past participle can only be formed from verb with a subject which is not raised' Ørsnes (2013, 331).

The Participle Principle is particularly relevant for the present study as it is supposed to account for the differences we see in the realization of raising-to-subject with passive predicates in different languages, in Ørsnes' case Danish, English and German. The idea is that the subject of the past participle cannot occur with the past participle alone. Ørsnes describes the subject of the past participle as being 'blocked' (Ørsnes, 2013, 333). Instead, there has to be an auxiliary, where, in the case of active sentences, the subject of the past participle occurs as the subject of the auxiliary. For Danish and German, Ørsnes holds that it is only in the case of non-raised subjects that the subject can be 'blocked'. Raising-to-subject is thus incompatible with the formation of past participles in Danish and German. For English, any kind of subject can be blocked in the formation of past participles, which means that raising-to-subject is possible with periphrastic passives (Ørsnes 2013: 333-334).

In section 3.1, I will show that there is a different way from Ørsnes' to account for the unacceptability of the raising-to-subject construction with periphrastic passives in Danish, and its acceptability in English. It will be seen that the semantic restrictions associated with the various passive constructions in the SVO languages (English and the Scandinavian languages) govern the possibility of raising-to-subject with periphrastic passives.

Furthermore, in opposition to Ørsnes, I will claim that raising-to-subject and raising-to-object do form an active-passive alternation. In section 3.2, it will be shown that there is a semantic restriction on raising-to-object tied to evidentiality, which makes certain verbs incompatible with the construction unless the object is a reflexive pronoun. There will thus be no need for a specific non-canonical passive formation rule yielding raising-to-subject with passive predicates, as there is in Ørsnes (2013).

4 Raising-to-subject with the periphrastic passive

As was mentioned in the background section, raising-to-subject is only available for the morphological passive and not for the periphrastic passive in Swedish, which is also claimed to be the case in Danish (Ørsnes, 2013). In the case of Danish, Ørsnes proposes that this is due to the so called Participle Principle, which precludes raising-to-subject in conjunction with past participles in Danish and German.

The Participle Principle is slightly problematic as a syntactic principle. Assuming that the empirical support is correct, i.e. that there is no raising-to-subject with periphrastic passives in Danish,[2] it is nonetheless strange that this constraint would apply to Danish and German, presumably Swedish as well, but not to the closely related language Norwegian. In Norwegian, raising-to-subject is undoubtely possible with the periphrastic passive,[3] and seems to be the only choice in the past tense. In the first part of this section, data from Norwegian on the possibility of raising-to-subject with periphrastic passives are presented. Then, in the second part of the section, my account of the relationship between raising-to-subject and the periphrastic passive is given.

4.1 Raising-to-subject with passive predicates in Norwegian

Norwegian makes use of a periphrastic passive and a morphological passive, just like Danish and Swedish. However, the distribution and use of the passives differ between the three languages. In Norwegian newspaper text, Laanemets (2012, 92) finds about a 50-50 distribution between the periphrastic and the morphological passive. In Danish, the distribution is said to be 60 % use of the morphological passive and 40 % use of the periphrastic passive. In Swedish, she finds a distribution of 97 % use of the morphological *s*-passive and 3 % use of the periphrastic *bliva*-passive. It can thus be seen that, in terms of frequency, the morphological

[2] One reviewer claims that it is easy to find examples of raising-to-subject with periphrastic passives in Danish, and gives examples such as the following:

(i) Sidstnævnte rige blev sagt at være ”ledet eller bistå af” frimureri.
latter realm was said to be led or assisted by freemasonry
‘The latter realm was said to be “led or assisted by” freemasonry.’

[3] The fact that raising-to-subject is possible with periphrastic passives in Norwegian was pointed out to me by Helge Lødrup at the HEADLEX16 conference.

passive is the default, unmarked alternative in Swedish, but not in Danish or Norwegian. Furthermore, in Norwegian, the morphological passive is only used in the present tense, while for the past tense, the periphrastic passive is the only option (Laanemets, 2012, 97).

Interestingly, raising-to-subject with the periphrastic passive is found in both the present and past tense in Norwegian. In the present tense, where both passive constructions are possible, we find both passive constructions represented in raising-to-subject. Google searches for the strings *sies å være* and *blir sagt å være* both yield numerous credible sentences, two of which are given in (7).

(7) a. Jeg blir sagt å være relativt lettlest
I become said to be relatively easy-to-read
'I am said to be relatively easy to read.'
(http://vgd.no/)

b. Jeg sies å være selvopptatt ...
I say.PASS to be selfcentered
'I am said to be self centered.'
(http://www.klassekampen.no/)

The sentence in (7-a) contains the periphrastic passive *blir sagt* ('is said'), and the sentence in (7-b) contains the morphological passive *sies* ('is said').[4]

In the past tense, we only find the periphrastic passive represented in raising-to-subject in Norwegian. This corresponds to the fact that the morphological passive is generally not used in the past tense in Norwegian. A Google search for the string *ble sagt å være* yields numerous sentences. Two examples are given in (8).

(8) a. Kvinnens påkledning ble sagt å være i strid med god moral og sekularisme
the-woman's dress became said to be in battle with good morals and secularism
'The way the woman dressed was said to stand in opposition to good morals and secularism.'
(https://www.minervanett.no/)

b. Hun ble sagt å være beskytteren ...
She became said to be the-guardian
'She was said to be the guardian ....'
(https://no.wikipedia.org/)

The examples in (8) show that raising-to-object is possible with the periphrastic *bliva*-passive in Norwegian.

[4]Even though both passive constructions can be found in conjunction with raising-to-subject in the present tense in Norwegian, it should be mentioned that the morphological passive is the dominant construction here, just as it is in Swedish and Danish. A reviewer points out that, in the Norwegian NoWaC corpus, there are 5780 hits for the string *sies å være* 'is said to be', but only 21 hits for the periphrastic *bli/blir sagt å være* 'become/becomes said to be'.

From the above data, it can be seen that the possibility of raising-to-subject for the morphological passive and the periphrastic passive follows the general restrictions on the passive construction in terms of tense in Norwegian. Ørsnes would have to assume that any kind of subject can be blocked in the participle formation of Norwegian, just as they can be in English. However, no explanation is then given for what other properties of English and Norwegian would make this the case. In the next section, a different account is given for the possibility or impossibility of raising-to-subject for the periphrastic passive.

4.2 Raising-to-subject and the semantic restrictions on the two passives

As mentioned previously, it seems as if raising-to-subject is only possible with the morphological passive in Danish and Swedish, while it is possible with the periphrastic passive in English and Norwegian. The claim made in the present section is that the possibility of the periphrastic passive in these languages is a result of the general semantic restrictions on the periphrastic passive in the respective language.

If we start by considering English, there is only one passive construction available, which is the periphrastic passive using the auxuliary *be* (disregarding other auxiliaries) in conjunction with a past participle. As a result of the fact that there is only one passive construction, there are not the same semantic restrictions on that passive construction as we will see for for instance Swedish where there is a choice between the periphrastic and the morphological passive.

Similar to the situation for English, in the past tense in Norwegian, only the periphrastic passive is possible. The fact that only one passive construction is available means that there are no semantic restrictions on the construction. It thus follows that raising-to-subject with periphrastic passives is possible in the past tense in Norwegian.

For the present tense in Norwegian, the past and present tense in Danish and all tenses in Swedish, both the morphological and the periphrastic passives are available options (Laanemets, 2012, 97). The reason why the morphological passive is preferred in all three languages when both passive constructions are available seems to be a result of the general semantic restrictions on the periphrastic passive in these languages, as well as what passive construction can be considered the default passive in the respective language. One semantic restriction, which seems to hold relatively well for all three languages, is the tendency for the periphrastic passive not to be used for generic statements (Laanemets, 2012, 111). Engdahl (2000) uses the following pair of sentences to exemplify the difference between the two passives.

(9) a. Dörren öppnas utåt
door.DEF open.PASS outward
'The door opens outward.'

b. *Dörren blir öppnad utåt
door.DEF becomes opened outward
'The door opens outward.'

In the case where we have a sign on a door saying that the door opens outward, only the morphological passive is possible in Swedish. The periphrastic passive here seems incompatible with a generic statement of this kind. Furthermore, as discussed in Engdahl (1999, 2000, 2006), the subject referent of the periphrastic passive in Swedish tends to be in control of the event in some way and tends to be animate. In (9), we have a generic statement with an inanimate subject, which is in no way in control of the event.

In the case of raising-to-subject with passive predicates, we have a similar situation as in the case of the sentence in (9). Consider our sentence from the introduction again, here repeated as (10).

(10) Hon sägs vara en utpräglad målskytt.
She says.PASS be a specialized goal-scorer
'She is said to be a specialized goal scorer.'
(http://www.vf.se/node/315601)

The sentence in (10) expresses a generic statement and not a specific event. The subject, although it is animate, is not in control of the event. It thus seems as if the periphrastic *bliva*-passive is incompatible with the raising-to-subject construction in Swedish.

5 The relationship between raising-to-subject and raising-to-object

As pointed out in section 3, one of the issues for the approach in which raising-to-subject with passive predicates is a regular passive construction concerns the faulty correspondence between predicates occurring in passive raising-to-subject and active raising-to-object, respectively. In this section, it will be argued that there is a reason for this seemingly faulty correspondence in the form of one restriction concerning evidentiality on raising-to-object.

5.1 The restrictions on raising-to-object

As mentioned above, it seems as if not all predicates that occur in passive raising-to-subject have an active raising-to-object correspondent. This is particularly the case for Danish, where the raising-to-object construction is said to be marginal at best (Ørsnes, 2013). For Swedish, Lyngfelt (2010) found 47 passive verbs participating in raising-to-subject in the PAROLE[5] corpus of written Swedish. Out of these 47 verbs, Lyngfelt claims that only 24 verbs participate in the raising-to-object construction. If there is an active-passive alternation between raising-to-object and raising-to-subject in the case of all 47 verbs, we need an explanation why only 24 are found in the raising-to-object construction. Such an explanation seems to be possible to provide in terms of a particular restriction concerning evidentiality on raising-to-object. Lyngfelt claims that there are certain situations in which raising-to-object structures are facilitated, in particular cases where the object is a reflexive pronoun. As will be seen below, in a corpus investigation of the relationship between raising-to-subject and raising-to-object, which also includes additional material not contained in the PAROLE-corpus, most verbs that Lyngfelt only found in raising-to-subject structures can also be found in raising-to-object structures when the object constitutes a reflexive pronoun. One group of verbs seems to be more or less completely restricted to reflexive pronoun objects when they occur in the raising-to-object construction.

5.2 Raising-to-object and evidentiality

As mentioned, the reason for the above-mentioned restriction to reflexive pronouns for many predicates in the raising-to-object structure will here be provided in terms of evidentiality. Linguistic evidentiality concerns 'the explicit encoding of a source of information or knowledge (i.e. evidence) which the speaker claims to have made use of for producing the primary proposition' (Diewald & Smirnova, 2010, 1). Evidentiality can be encoded either lexically or grammatically. An example of the lexical encoding of evidentiality in Swedish is for instance the verb *lär*, which is claimed to express the fact that the truth of the proposition is based on indirect evidence (de Haan, 2007, 143). An example of the grammatical/syntactic encoding of evidentiality is described in Asudeh & Toivonen (2012), where copy raising in English and Swedish is discussed. Asudeh & Toivonen (2012) claim that, in a copy raising sentence such as the one in (11), the subject constitutes the perceptual source for the proposition expressed.

(11) Han verkar som om han är lugnare nu.
he seems as if he is calmer now
'He seems like he is calmer now.'
(Teleman et al., 1999, 56)

In 11, the sentence thus expresses the evidential fact that the proposition that the referent of *he* is calmer now is based on the speaker perceiving the referent of *he*.

In the case of the raising-to-object construction, the construction likewise seems to express evidentiality. Interestingly, in comparison to the copy raising construction mentioned above, where the subject expresses the perceptual source, in

[5]PAROLE is one of the corpora within the corpus collection Korp (Borin et al., 2012). It contains approximately 19 million words of primarily newspaper texts (but also novels, magazines and web material), and is annotated with morphosyntactic information.

the case of the raising-to-object construction, the subject also seems to express the source for the primary proposition of the utterance. However, rather than being a perceptual source, it is here a purely evidential source. There seems to be a requirement for the raising-to-object construction that the subject referent constitutes the evidential source for the truth of the proposition expressed in the complement subclause. We can call this restriction the Evidential Source Requirement:

(12) The Evidential Source Requirement:
The subject referent of the raising-to-object construction is required to express the evidential source for the truth of the proposition expressed in the complement subclause.

In the current section, support is given for the Evidential Source Requirement in terms of corpus data from written Swedish. First, the corpus data is presented, and then it is shown how this data provides support for the Evidential Source Requirement.

In order to see what verbs are represented for raising-to-subject and raising-to-object, a search for raising-to-subject with passive predicates was performed in the same corpus used by Lyngfelt, PAROLE. A search query was constructed, yielding all sentences containing a verb ending an an -s, immediately followed by a verb ending an -a.[6]

The search query yielded 4,674 hits. Out of these hits, 56 verbs were found occurring in the subject with infinitive construction;[7] 16 of these were object control verbs and 40 raising-to-object verbs. For some verbs that were not found in the raising-to-object construction in PAROLE, supplementary Google searches were made, yielding credible examples for all but two verbs.[8] Arguably, the reason that Lyngfelt found only 24 out of 47 verbs participating in raising-to-object is the size of the PAROLE corpus.

Out of the 40 raising-to-object verbs, 15 occur more or less exclusively with reflexive pronoun objects. Many of the verbs that Lyngfelt found only occurring in the raising-to-subject construction also occur in the raising-to-object construction when the object is a reflexive pronoun.

In (13) and (14), the raising-to-object verbs are listed, both those restricted to reflexive objects and those that are not.

[6]The search string used is [word = ".s" & pos = "VB"] [word = ".a" & pos = "VB"]. As can be seen, it does not cover cases where there is one or more words intervening between the passive verb and the infinitive verb. Furthermore, it does not cover infinitives that do not end in an -a, such as Swedish *få* 'get'.

[7]Verbs such as *hoppas* 'hope' and *ryktas* 'be rumoured', which always occur in the *s*-form, were excluded.

[8]The two verbs that were not found in the object control or raising-to-object are *karaktärisera* 'characterise', *upplysa* 'inform'. Possibly, these would be found as well given a large enough material.

(13) Verbs not restricted to reflexive pronouns in the raising-to-object construction:
ange 'state', anse 'consider', anta 'assume', beräkna 'calculate', förklara 'declare', förmoda 'assume', förutsätta 'presume', konstatera 'point out', misstänka 'suspect', påstå 'suggest', tippa 'bet', tro 'believe', uppfatta 'perceive', döma 'judge', erkänna 'admit', räkna 'count', spå 'foretell', uppleva 'experience', uppskatta 'estimate'.

(14) Verbs restricted to reflexive pronoun objects in the raising-to-object construction:
rapportera 'report', säga 'say', tänka 'imagine', uppge 'state', vänta 'await', antyda 'suggest', avse 'intend', besluta 'decide', befara 'fear', betrakta 'regard', förutspå 'foretell', förutse 'anticipate', förutsätta 'presume', utlova 'promise', förvänta 'expect'.

The fact that the verbs which under other circumstances do not participate in the raising-to-object construction do so when the object is a reflexive pronoun provides support for the Evidential Source Requirement. For verbs such as *anse* 'consider' or *uppleva* 'experience', it is part of the lexical semantics of these predicates that the subject referent constitutes the evidential source for the proposition expressed in the complement subclause. For *anse* 'consider', the evidential source is tied to the opinion of the subject referent. For *uppleva* 'experience', the evidential source is tied to the experience of the subject referent. However, for a predicate such as *säga* 'say', which is one of the predicates restricted to a reflexive pronoun object when occurring in raising-to-object, there does not seem to be any specification for the evidential source for the truth of the proposition expressed in the complement subclause. If somebody says something, we do not know the basis for the truth or falsity of what is said. The same holds for predicates such as *rapportera* 'report' or *uppge* 'state'. Consider then what happens when the object in the raising-to-object construction for a predicate such as *säga* 'say' is a reflexive pronoun. It seems that when you say (or state, or report) something about yourself, then you also express that you are the evidential source for whatever is said (or stated, or reported). The presence of an evidential source in the lexical semantics of the relevant predicates thus seems to govern whether they are restricted to a reflexive pronoun object or not.[9]

In the next section, we will see how the conclusions drawn from sections 4 and 5 can be formalised within the Lexical Mapping Theory of LFG.

[9]Some of the more infrequent predicates, such as *förutspå* 'foretell', could be analysed as being lexically specified for an evidential source, but is nevertheless not found without a reflexive pronoun object. The meaning of *förutspå* 'foretell' is also approximately the same as for the predicate *spå* 'foretell', which is found with other types of objects. It is possible that the predicate *förutspå* 'foretell' is not restricted to reflexive pronoun objects, but that the infrequency of the predicate clouds the facts.

6 Argument structure analysis

For the argument structure analysis presented here, the revised Lexical Mapping Theory (LMT) of Kibort (2007, 2014) and Kibort & Maling (2015) is made use of. Similar to the mapping theory of Bresnan & Kanerva (1989) (one recent realisation in Bresnan et al. (2016)), it is based on the use of the features [±r] and [±o]. The argument function SUBJ is [-r, -o], OBJ is [-r, +o], OBL_θ is [+r, -o], and OBJ_θ is [+r, +o]. Kibort's approach differs from previous versions of the LMT in two primary assumptions. Firstly, Kibort assumes a universally available syntactic subcategorisation frame with fixed argument positions:

(15) Subcategorisation frame:

⟨ arg1,	arg2,	arg3 . . . ,	arg4 . . . ⟩
[-o]/[-r]	[-r]	[+o]	[-o]

Every predicate subcategorises based on the subcategorisation frame above. The features associated with each argument position governs what argument functions the predicate can take. The arg1[-o] slot, for instance, can map to either SUBJ or OBL_θ. What makes the arg1[-o] slot typically map to SUBJ is the so called Subject Default:

(16) Subject Default:
The first argument compatible with the SUBJ function is mapped to SUBJ.

Apart from the Subject Default, there is only one mapping principle, namely that '[t]he ordered arguments are mapped to the available functions compatible with their intrinsic marking' (Kibort, 2014).

As pointed out above, I assume that raising-to-subject with passive predicates, including predicates such as *säga* 'say', constitutes the passive counterpart of active raising-to-object. Given this assumption, the argument structures for raising-to-object *säga* and raising-to-subject *sägs*, respectively, can be represented as follows:

(17) Argument structure for raising-to-object *säga* 'say':

	1 (agent)	∅	**4** (proposition)
	|	|	|
säga 'say'	arg1	arg2	arg4
	[–o]	[–r]	[–o]
	|	|	|
	SUBJ	OBJ	XCOMP

(18) Argument structure for passive raising-to-subject *sägs* 'say':

	1 (agent)	$\emptyset$	**4** (proposition)
	|	|	|
sägs 'is said'	arg1	arg2	arg4
	[–o]	[–r]	[–o]
	[+r]		
	|	|	|
	(OBL_{agent})	SUBJ	XCOMP

The active raising-to-object predicate *säga* 'say' takes three arguments, arg1[-o], arg2[-r] and arg4[-o].[10] These three arguments are realised in a sentence such as the one in (4): *hon säger sig vara en utpräglad målskytt* 'she says REFL be a specialised goal scorer'. Following the Subject Default, the first argument is mapped to SUBJ. The two other arguments are mapped to the only compatible functions OBJ and XCOMP.[11] For the passivised *sägs* 'be said', arg1[-o] is assigned a [+r]-feature, meaning that it can only be mapped to OBL_{agent}. Following the Subject Default, arg2[-r] is mapped to SUBJ. The only compatible function for arg4[-o] is then OBJ. The two argument structures in (16) and (17) thus represent a standard active-passive alternation.

As can be seen, the OBJ of the active predicate, which is associated with the semantically empty arg2[-r], corresponds to the SUBJ of the passive predicate, which is also associated with the semantically empty arg2[-r]. As arg2[-r] is not associated with any semantic marker, the grammatical function associated with this argument position is either structure shared with an argument embedded within the XCOMP or realized as an expletive (in Swedish, *det*). An example of the construction with an expletive was given in (2). In this sense, raising-to-subject *sägs* 'is said' is like any other raising predicate, alternating with a construction with a subject *it* and a finite clausal complement.

Given the argument structure above, the lexical entries for active *säga* 'say' and passive *sägs* 'be said' can be represented as follows.

(19) Lexical entry for the active raising-to-object predicate *säger* 'says':

säger 'says' V ($\uparrow$ PRED) = '*say* $\langle$SUBJ, XCOMP$\rangle$, OBJ'
($\uparrow$ TENSE) = PRESENT
($\uparrow$ OBJ) = ($\uparrow$ XCOMP SUBJ)
($\uparrow$ OBJ PRON-TYPE) $=_c$ REFLEXIVE

[10] There is a separate argument structure for the non-raising predicate *säga* 'say', which only takes two arguments, associated with the roles agent and proposition, respectively.

[11] Following Zaenen & Engdahl (1994), the function XCOMP is assumed to be equivalent to OBL_{prop}.

(20) Lexical entry for the passive raising-to-subject predicate *sägs* 'is said':

sägs 'is said' V $\{(\uparrow \text{PRED}) = \text{`}\textit{say}\ \langle \text{OBL}_{agent}, \text{XCOMP}\rangle, \text{SUBJ'} \mid$
$(\uparrow \text{PRED}) = \text{`}\textit{say}\ \langle \text{XCOMP}\rangle, \text{SUBJ'}\}$
$(\uparrow \text{TENSE}) = \text{PRESENT}$
$(\uparrow \text{SUBJ}) = (\uparrow \text{XCOMP SUBJ})$
$(\uparrow \text{XCOMP VFORM}) =_c \text{BARE-INF}$

Note the f-description saying that the object of active *säger* 'says' is restricted to reflexive pronouns: $(\uparrow \text{OBJ PRON-TYPE}) =_c \text{REFLEXIVE}$. As argued above, the fact that the object of this particular predicate is restricted to reflexive objects is not an arbitrary syntactic fact. It is instead derived from the Evidential Source Requirement. There should thus be a semantic constraint formulated in the lexical entry saying that the referent of the subject constitutes the evidential source for the truth of the proposition expressed in the XCOMP. How such a constraint should be formulated in detail will have to be left for future research, however. It is unclear what the status of the role evidential source is in the current theory. As the subject is already associated with the agent role, associating the subject with another role would possibly violate the principle of coherence.

How the restriction of the raising-to-subject construction to the periphrastic passive should be formalised will also not be specified in detail here.

7 Conclusion

This paper has provided an analysis of the argument structure of raising-to-subject with passive predicates in Swedish and other Germanic languages. It has been shown how the traditional analysis in which raising-to-subject with passive predicates constitutes the passive counterpart of active raising-to-object can be given support, despite the fact that certain predicates, such as *säga*, only seem to participate in the passive alternative and not the active. It is claimed that there is an Evidential Source Requirement for the raising-to-object construction, where the subject referent is required to express the evidential source for the truth of the proposition expressed in the complement subclause. Furthermore, it is claimed that there are semantic restrictions precluding the periphrastic passive from participating in the raising-to-subject construction in Danish and Swedish. In Swedish, the raising-to-subject construction is incompatible with the restrictions on the periphrastic passive, which is typically used for specific rather than generic events, where the subject referent can be characterised as being in control. In Norwegian, it can be shown that both passive constructions participate in raising-to-subject. Work remains in specifying how the semantic restrictions on raising-to-object and the periphrastic passive could be given a more detailed formulation within LFG.

References

Asudeh, Ash & Ida Toivonen. 2012. Copy raising and perception. *Natural Language and Linguistic Theory* 30(2). 321–380.

Borin, Lars, Markus Forsberg & Johan Roxendal. 2012. Korp – the corpus infrastructure of Språkbanken. In *Proceedings of LREC 2012*, Istanbul: ELRA.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-Functional Syntax*. Oxford/Malden, MA: Blackwell.

Bresnan, Joan & Jonni Kanerva. 1989. Locative inversion in Chicheŵa: a case study of factorization in grammar. *Linguistic Inquiry* 20(1). 1–50. Reprinted in Tim Stowell and Eric Wehrli (eds.), Syntax and Semantics No. 26: Syntax and the Lexicon, pp. 53–101. New York: Academic Press.

Diewald, Gabriele & Elena Smirnova. 2010. Evidentiality in European languages: the lexical-grammatical distinction. In Gabriele Diewald & Elena Smirnova (eds.), *Linguistic Realization of Evidentiality in European Languages*, 1–15. Berlin/New York: Walter de Gruyter.

Engdahl, Elisabet. 1999. *The choice between bli-passive and s-passive in Danish, Norwegian and Swedish* Manuscript. Gothenburg: University of Gothenburg.

Engdahl, Elisabet. 2000. Valet av passivform i modern svenska. In Linda Jönsson, Viveca Adelswärd, Ann Cederberg, Per A. Pettersson & Caroline Kelly (eds.), *Svenskans beskrivning 24*, Linköping University Electronic Press.

Engdahl, Elisabet. 2006. Semantic and syntactic patterns in swedish passives. In Benjamin Lyngfelt & Torgrim Solstad (eds.), *Demoting the agent*, 21–45. Amsterdam: John Benjamins.

de Haan, Ferdinand. 2007. Raising as grammaticalization: the case of Germanic *seem*-verbs. *Rivista di Linguistica* 19(1). 129–150.

Kibort, Anna. 2007. Extending the Applicability of Lexical Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG07 Conference*, Stanford, CA: CSLI Publications.

Kibort, Anna. 2014. Mapping out a Construction Inventory with (Lexical) Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG14 Conference*, Stanford, CA: CSLI Publications.

Kibort, Anna & Joan Maling. 2015. Modelling the Syntactic Ambiguity of the Active vs. Passive Impersonal in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG15 Conference*, Stanford, CA: CSLI Publications.

Laanemets, Anu. 2012. *Passiv i moderne dansk, norsk og svensk: et korpusbaseret studie af tale- og skriftsprog*. Tartu: University of Tartu Press.

Lødrup, Helge. 2008. Raising to object in Norwegian and the derived object constraint. *Studia Linguistica* 62(2). 155–181.

Lyngfelt, Benjamin. 2010. Om subjekt med infinitiv, särskilt i passiv. In Ann-Catrine Edlund & Ingmarie Mellenius (eds.), *Svenskans beskrivning 31*, Umeå: Umeå universitet, Institutionen för språkstudier.

Noël, Dirk & Timothy Colleman. 2010. Believe-type raising-to-object and raising-to-subject verbs in English and Dutch. *International Journal of Corpus Linguistics* 15(2). 157–182.

Ørsnes, Bjarne. 2011. Passives and evidentiality: Danish reportive passives and their equivalents in German. *Acta Linguistica Hafniensia: International Journal of Linguistics* 43(1). 21–59.

Ørsnes, Bjarne. 2013. The Danish Reportive Passive as a Non-Canonical Passive. In Artemis Alexiadou & Florian Schäfer (eds.), *Non-Canonical Passives*, 315–336. Amsterdam: John Benjamins.

Ørsnes, Bjarne & Stefan Müller. 2013. Passive in Danish, English and German. In Stefan Müller (ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar*, 140–160. Stanford, CA: CSLI Publications.

Postal, Paul M. 1974. *On Raising. One Rule of English Grammar and Its Theoretical Implications*. Cambridge, MA: The MIT Press.

Reis, Marga. 1973. Is there a rule of subject-to-object raising in German? *Chicago Linguistic Society: Papers from the 9th regional meeting* .

Teleman, Ulf, Staffan Hellberg & Erik Andersson. 1999. *Svenska Akademiens grammatik 4. Satser och meningar*. Svenska Akademien.

Zaenen, Annie & Elisabet Engdahl. 1994. Descriptive and Theoretical Syntax in the Lexicon. In B. T. S. Atkins & Antonio Zampolli (eds.), *Computational Approaches to the Lexicon*, 181–212. Oxford: Oxford University Press.

Categorematic unreducible polyadic quantifiers in Lexical Resource Semantics

Frank Richter

Goethe Universität Frankfurt a.M.

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 599–619

Keywords: constraint-based semantics, polyadic quantifiers, unreducible quantifiers, compositionality

Richter, Frank. 2016. Categorematic unreducible polyadic quantifiers in Lexical Resource Semantics. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 599–619. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.31.

Abstract

Early work on quantification in natural languages showed that sentences like *Every ape picked different berries*, on the reading that the sets of berries picked by any two apes are not the same, can be logically represented with a single polyadic quantifier for the two nominal phrases. However, since that quantifier cannot be decomposed into two quantifiers for the two nominal phrases, a compositional semantic analysis of this reading is not possible under standard assumptions about syntax and semantics. This paper shows how a constraint-based semantics with Lexical Resource Semantics can define a systematic syntax-semantics interface which captures the reading in question with a polyadic quantifier.

1 Introduction

One of an impressive series of fundamental contributions by Richard Montague to semantic theory was the consistent semantic treatment of nominal phrases as quantifiers, sets of sets of objects.[1] Although this proposal has been challenged and might not be the most adequate solution even for some of the cases it was originally designed for, an initial analysis of nominal phrases as (generalized) quantifiers (Barwise & Cooper, 1981) or a close variant thereof is still a fruitful and sound methodological strategy. It is this perception that forms the nucleus of the following proposal of treating a well-known but particularly challenging reading of nominal phrases with the adjective *different* as in *different berries* in terms of polyadic generalized quantifiers. *Polyadic* quantification is the necessary slight deviation from 'ordinary' monadic quantification in the present proposal that allows me to stay very close to the original spirit. The deviation is necessary because the relevant reading has been shown not to be amenable to the classic Montagovian treatment. Polyadic quantification also seems like the most conservative conceivable semantic modification available to provide an analysis of the data.

In fact, the proposed interpretation of *sentence internal* readings with *different* is not new at all. With certain modifications, benefiting from insights of subsequent literature on *different*, it is lifted straight from (Keenan, 1992). What is entirely new and the main topic of this paper is its full integration in an explicit syntax-semantics interface of a phrase structure grammar without an extra layer of elaborate LF syntax. Doing this has been impossible before in other frameworks of semantic composition for reasons to be explicated below.

I will be concerned with one particular reading of (1), namely (2-a).

[†]I thank three plus two anonymous reviewers and David Lahm for insightful comments, and audiences at Heinrich-Heine-Universität Düsseldorf, Goethe Universität Frankfurt a.M. and HeadLex16 in Warsaw for discussions of ideas presented in this paper. Janina Radó helped with proofreading.

[1]Intensionality will be ignored. Two-sorted Type Theory (Ty2, Gallin (1975)) with a type for worlds will be assumed for compatibility with other work in Lexical Resource Semantics, and to make clear that including possible worlds is desirable.

(1) [$_{\text{S}}$ [$_{\text{NP}}$Every ape] [$_{\text{VP}}$picked [$_{\text{NP}}$different berries]]]

(2) a. The berries that any one of the apes picked were different from the berries each other ape picked.
 b. Every ape picked berries that were different from the ones mentioned before.
 c. Every ape picked various/many berries.

Following the broad characterization by Brasoveanu (2011) (and ignoring further possible distinctions), I call (2-a) the *sentence internal reading* and (2-b) an *external reading*; (2-c) is a third reading in which the *different* phrase shows no apparent contextual dependency similar to the other two readings.

In light of results proving that reading (2-a) cannot be obtained from two independent nominal phrases as in the syntactic analysis in (1), I develop a constraint-based syntax-semantics interface in Lexical Resource Semantics (LRS) that produces the reading in question. Instead of viewing the two nominal phrases as two monadic generalized quantifiers, they are semantically treated as a single categorematic polyadic quantifier which is unreducible to individual monadic quantifiers. At the same time the HPSG syntax with two unrelated nominal phrases is left intact, and lexical underspecification makes lexical redundancy unnecessary.

The crucial insight is that a constraint-based semantics can give a systematic and purely semantic account of the internal reading, unlike other sufficiently precise theories of a syntax-semantics interface. For reasons of space, the present discussion will mostly be confined to an explication of the semantic constraint system and the semantic lexical specification of the adjective *different* and interacting determiners that constitute the syntax-semantics interface. Broader empirical considerations and an extension of the proposal beyond a small set of core data have to await a later occasion.

2 *Different* with Polyadic Quantification

Figure 1 summarizes essential terminology and notation for the following exposition.

The nominal phrase *two apes* is seen as a monadic quantifier, Lindström type $\langle 1\rangle$, that takes a set, or unary relation (of type $\langle et\rangle$), as an argument and returns a truth value, t, depending on the state of the world. The binary quantifier *two apes, all berries* (Lindström type $\langle 2\rangle$) takes a binary relation as argument and returns a truth value – true if *two apes, all berries* holds of the relation, and false otherwise. In principle, this schema continues for polyadic quantifiers from any number of nominal phrases, with the Lindström type reflecting the number of NPs and the corresponding functional type.

Instead of combining NPs into larger units, we can also combine determiners to obtain new quantifiers. While *two* is a monadic generalized quantifier, taking two sets and returning a truth value, *two, all* forms a polyadic generalized quantifier

Figure 1: Some quantifiers with their Lindström types and functional types

Quantifier	Lindström type	Functional type
(two apes)	$\langle 1\rangle$	$\langle\langle et\rangle\, t\rangle$
(two apes, all berries)	$\langle 2\rangle$	$\langle\langle e\,\langle et\rangle\rangle\, t\rangle$
(two apes, every girl, many berries)	$\langle 3\rangle$	$\langle\langle e\,\langle e\,\langle et\rangle\rangle\rangle\, t\rangle$
($\text{NP}_1, \dots, \text{NP}_n$)	$\langle n\rangle$	$\dots$
(two)	$\langle 1, 1\rangle$	$\langle\langle et\rangle\,\langle\langle et\rangle\, t\rangle\rangle$
(two, all)	$\langle 1^2, 2\rangle$ $= \langle 1, 1, 2\rangle$	$\langle\langle et\rangle\,\langle\langle et\rangle\,\langle\langle e\,\langle et\rangle\rangle\, t\rangle\rangle\rangle$
(two, every, many)	$\langle 1^3, 3\rangle$ $= \langle 1, 1, 1, 3\rangle$	$\langle\langle et\rangle\,\langle\langle et\rangle\,\langle\langle et\rangle\,\langle\langle e\,\langle e\,\langle et\rangle\rangle\rangle\, t\rangle\rangle\rangle\rangle$

that takes as arguments two sets (of type $\langle et\rangle$) and a binary relation, type $\langle e\,\langle et\rangle\rangle$, and returns a truth value. The entire construct is of Lindström type $= \langle 1, 1, 2\rangle$, or short $\langle 1^2, 2\rangle$, and the complete functional type can be read off the table. Again this construction can be continued, this time for any number of determiners.[2]

The semantic object language of the HPSG grammar covering (1) will be a language of Ty2 with categorematic polyadic and polymorphic quantifiers with the functional types shown in Figure 1. Determiners like *two* or *every* will receive a semantic specification that permits their realization in the semantic representation of a proposition either as a monadic quantifier as shown for *two* in Figure 1 (this is the 'normal' case), or as a component of a polyadic quantifier as shown for *two, all* or *two, every many* – this will be a special case that presupposes the existence of a triggering element in the syntactic neighborhood. In our discussion, *different* will be the only relevant triggering element, but there can of course be others, such as negative quantifiers in negative concord constructions. Semantically, *different* will be translated by a constant much like the determiners *every* or *few*, but it will come with special contextual restrictions.

2.1 Intended Semantics

(Keenan, 1992, p. 202–203) observed that sentences of the structural type (3), for our purposes essentially a paraphrase of (1) when only considering reading (2-a), can be given a plausible semantics that cannot be reduced to some combination of two Fregean (or type $\langle 1\rangle$) quantifiers as independent translations of the two NPs.

(3) Different apes picked different berries.

This means that Keenan's (1992) quantifiers *different apes, different berries* as well

[2]For an historical account of the relevant concepts, see Westerståhl (1989). Lindström's paper is Lindström (1966). Keenan & Westerståhl (1997) has more on polyadic quantification in natural languages.

as *every ape, different berries* put such a specific, fine-grained condition on the binary picking relation that no combination of monadic quantifiers can express it. In linguistic terms, the condition cannot be obtained by having one of the two quantifiers take scope over the other, which would be expressed technically by combining the two putative component quantifiers by the operation of *iteration*.[3]

Before I can define the semantics of quantifiers containing *different*, I need a notational convention for certain sets of elements in a relation. E is the set with the elements in the discourse, E^2 ($= E \times E$) is the binary Cartesian product relation over E.

Notational Convention: Given a set E and a binary relation R, $R \subseteq E^2$, for each $x \in E$, I write Rx for the set of objects x bears R to: $Rx = \{y|(x,y) \in R\}$.

For example assume a binary relation pick that signifies who picks what, and assume further that pick contains pairs of apes and berries. The set of berries ape a picks can now simply be notated as $\mathsf{pick}\, a = \{b|(a,b) \in \mathsf{pick}\}$.

Keenan & Westerståhl (1997) develop a semantics in terms of polyadic quantification for many non-monadic examples in (Keenan, 1992), including quantifiers with *different*. I adapt their formulation[4] to the present discussion, and symbolize *different* with Δ.

Definition 1: Semantics of a quantifier containing Δ
For $\mathcal{Q}$ a polyadic quantifier of type $\langle 1^2, 2\rangle$ containing Δ, $A, B \subseteq E$, $R \subseteq E^2$, and H a quantifier of type $\langle 1, 1\rangle$, the interpretation of $\mathcal{Q}$ is as follows:
$\mathcal{Q}(A, B, R) = 1$ iff there is an A', $A' \subseteq A$ such that

> $H(A, A') = 1$, and
> for all $x, y \in A'$: $\quad (x \neq y) \Rightarrow (B \cap Rx \neq B \cap Ry)$.

Applied to (1), $\mathcal{Q}$ symbolizes *every, different*, with A the set of apes, B the set of berries, R the picking relation, and H the universal quantifier (*every*). The sentence is true iff there is a subset A' of the set of apes such that A is also a subset of A' ($H(A, A')$), i.e. A' actually equals the set of apes A; and for any pair x, y of distinct apes in A', the set of berries that one of them, say x, picks does not equal the set of berries that the other one, y, picks.

Various elaborations of this semantics are conceivable, such as demanding that A contain more than one element, or that the intersection of Rx and Ry be empty (no overlap in the set of berries picked by distinct apes). I will not pursue the topic of alternative formulations but want to stress in closing that H is meant to also comprise quantifiers other than *every*, including numerals such as *two*, or the

[3]Two tests for checking type $\langle 2\rangle$ quantifiers for reducibility are presented in (Keenan, 1992) with a generalization to quantifiers of type $\langle n\rangle$ in (Dekker, 2003). (Iordăchioaia, 2010, p. 37–38) uses one of Keenan's (1992) tests to show very carefully why *Two boys in my class date different girls* is unreducible. This proof can easily be adapted to our semantics of my sentence (1).

[4]Based on Iordăchioaia's (2010, p. 27) variant of (Keenan & Westerståhl, 1997, p. 877).

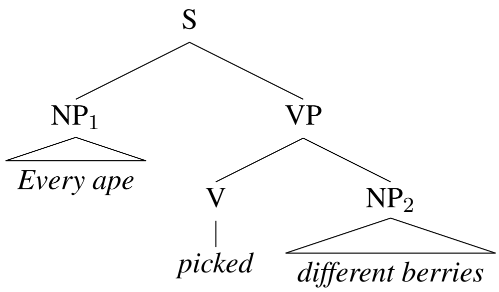
higher-order quantifiers *many* and *few*.[5]

Keenan & Westerståhl (1997) speculate that the polyadic quantifiers of natural languages are not arbitrary, they are built in certain regular ways from monadic quantifiers, and their semantics for quantifiers containing *different* is one of those regular patterns. If this is correct, a systematic syntax-semantics interface should be able to account for them, even if their makeup does not follow the typical pattern of iterated quantifiers that are usually in focus.

2.2 A Challenge for a Syntax-Semantics Interface

The non-reducibility results on *quantifiers containing* Δ mean that we are far from being done with a linguistic analysis of (1) if we adopt the semantic analysis above. In particular, it is impossible to interpret the two NPs *every ape* and *different berries* independently as generalized quantifiers and obtain that semantics. Even worse for syntactic theories like LFG and HPSG, a standard compositional semantic analysis using that semantics is impossible with the syntactic structure provided by the usual structural analysis of the sentence in these two frameworks, shown in the following tree:

(4)



In order to construct the intended polyadic quantifier in syntax, at the very least some additional syntactic movement would be necessary so that components of NP_1 and NP_2 form suitable syntactic units in the resulting representation. Neither HPSG nor LFG envisage this type of LF representation, as there is no genuinely syntactic evidence for its existence. Early versions of HPSG employed semantic representations expressed in feature logic that included a Cooper storage mechanism for storing and retrieving quantifiers in syntactic trees. Due to their limitations, quantifier storage mechanisms were superseded by theories that constrain the composition of logical contributions of syntactic constituents with *dominance constraints* or *subterm constraints* on semantic representations. I will show in the next section that one of these semantic composition theories, LRS, provides the necessary tools to keep the syntactic analysis above exactly as it is and still derive the semantic analysis of Keenan & Westerståhl (1997), and with it an analysis

[5](Keenan & Westerståhl, 1997, p. 877, fn. 18) contains a noteworthy remark that the monotonicity properties in the second argument of the quantifier H necessitates a refinement of the semantics they provide: as it stands, it does not yield the desired result when replacing *every* by *no*. They suggest an improved semantics which I avoid here for ease of exposition.

that is only a minimal variation of the idea that nominal phrases are interpreted as quantifiers.

3 An LRS Analysis

Three steps will be taken to prepare the syntax-semantics interface of LRS for an analysis of (1): First, the quantifiers in the logical object language receive a categorematic representation with sets (and relations) as arguments. Instead of representing *Every woman walks* as $\forall x(woman'_{\langle et\rangle}(x) \rightarrow walk'_{\langle et\rangle}(x))$, I use $\forall_{\langle\langle et\rangle\langle\langle et\rangle t\rangle\rangle}(\lambda x.woman'_{\langle et\rangle}(x), \lambda x.walk'_{\langle et\rangle}(x))$. $\forall_{\langle\langle et\rangle\langle\langle et\rangle t\rangle\rangle}$ is still a monadic quantifier, which leads to the second step: Polymorphic polyadic quantifiers in Ty2 are a generalization of these monadic quantifiers and will be presented in lexical entries that also illustrate how type polymorphism can be captured as an effect of lexical underspecification. The lexical specification of *different* is a special case that still fits the general pattern. In the third and final step, the LRS principles that govern the space of admissible semantic compositions must be generalized from monadic quantification to polyadic quantification. The essential architecture of LRS will not be reviewed here for reasons of space. For a compact and yet comprehensive introduction to all aspects of LRS relevant in the discussion below, the reader may want to consult (Iordăchioaia & Richter, 2015, pp. 626–632).

3.1 Revising the Representation of Quantifiers

Quantificational determiners in LRS introduce an appropriate constant for the corresponding logical quantifier ($\forall$ for *every*, most' for *most*), a variable that is bound in two lambda abstracts and used as a hook to nouns and verbs at the syntax-semantics interface, and a few restrictions on how these expressions of logical syntax enter into a larger quantificational expression, typically into the semantic representation of a sentence.

At the syntax-semantics interface, determiners indirectly identify syntactic valencies of a noun and a verb whose logical representations are functors of the variable the determiner introduces, and the determiners ensure (again indirectly) that the correct argument of the nominal and verbal predicates contain that variable.[6] Mostly as a matter of expository convenience when discussing essentially first-order examples, the formulae in the restrictor and scope of quantifiers have usually been represented as expressions of type t (Richter & Kallmeyer, 2009; Iordăchioaia & Richter, 2015). For better notational compatibility with generalized quantifier theory, it is preferable to switch the arguments to type $\langle et\rangle$, obtaining the usual monadic quantifiers of type $\langle\langle et\rangle\,\langle\langle et\rangle\,t\rangle\rangle$ (Lindström type $\langle 1, 1\rangle$). For the semantic aspects of the lexical specification of a quantifier such as two', we obtain (5):

[6]For present purposes, I am only considering verb-argument structures.

(5)
$$\begin{bmatrix} \text{PHON} & \langle \textit{two} \rangle \\ \text{SS LOC CONT} & \begin{bmatrix} \text{INDEX DR} & \boxed{4a}\ x \\ \text{MAIN} & \boxed{4b}\ \mathsf{two'} \end{bmatrix} \\ \text{LRS} & \begin{bmatrix} \text{EXC} & \textit{me} \\ \text{INC} & \boxed{4}\ \mathsf{two'}(\lambda x.\alpha, \lambda x.\beta) \\ \text{PARTS} & \langle \boxed{4}, \boxed{4a}\ x, \boxed{4b}\ \mathsf{two'}, \boxed{4c}(\lambda x.\alpha), \boxed{4d}\ (\lambda x.\beta), \boxed{4e}\ \mathsf{two'}(\lambda x.\alpha) \rangle \end{bmatrix} \end{bmatrix}$$
$$\&\ x \triangleleft \alpha\ \&\ x \triangleleft \beta$$

The determiner *two* introduces a fresh variable (x) as INDEX DR value. The MAIN value, i.e. the central lexical semantic contribution of the word, is the quantifier's constant, in this case $\mathsf{two'}$. The IN(TERNAL-)C(CONTENT) contains the core expression that everything else in the syntactic projection of the determiner would outscope, and as such it is the most intuitive semantic representation of the word: $\mathsf{two'}$ applied to its two arguments, a lambda abstract $\lambda x.\alpha$, with as yet unknown α, and a second lambda abstract $\lambda x.\beta$, with as yet unknown β. The two subterm restrictions in the last line of (5) demand that x occur in α and in β. In simple constructions, α will ultimately contain the logical representation of the nominal projection that the determiner combines with, and β will contain the semantics of the verbal projection that the completed NP combines with (modulo operators with wider scope than the NP, which will not be subterms of β). The PARTS list enumerates all individual contributions to logical syntax by the determiner, including various function applications, e.g. $\mathsf{two'}(\lambda x.\alpha)$ ($\mathsf{two'}$ applied to the lambda abstract that is the quantifier's restrictor).

Nothing has to be changed in the usual syntactic analysis and LRS system of a grammar to derive nominal phrases such as *two apes*:

(6)
$$\begin{bmatrix} \text{PHON} & \langle \textit{two, apes} \rangle \\ \text{SS LOC CONT} & \begin{bmatrix} \text{INDEX DR} & \boxed{4a}\ x \\ \text{MAIN} & \boxed{5a}\ \mathsf{ape'} \end{bmatrix} \\ \text{LRS} & \begin{bmatrix} \text{EXC} & \boxed{4}\ \mathsf{two'}(\lambda x.\alpha, \lambda x.\beta) \\ \text{INC} & \boxed{5}\ \mathsf{ape'}(x) \\ \text{PARTS} & \left\langle \begin{array}{l} \boxed{4}, \boxed{4a}\ x, \boxed{4b}\ \mathsf{two'}, \boxed{4c}\ (\lambda x.\alpha), \boxed{4d}\ (\lambda x.\beta), \\ \boxed{4e}\ \mathsf{two'}(\lambda x.\alpha), \boxed{5}, \boxed{5a}\mathsf{ape'} \end{array} \right\rangle \end{bmatrix} \end{bmatrix}$$
$$\&\ \boxed{5} \triangleleft \alpha\ \&\ x \triangleleft \alpha\ \&\ x \triangleleft \beta$$

The noun *apes* is the syntactic head of the noun phrase in (6) and selects the SYNSEM value of its determiner (5) by an appropriate valence attribute. The phrase *two apes* shares its INDEX DR and MAIN values with its head daughter, the noun *apes* (not depicted here individually). Similarly, the EX(TERNAL-)C(ONTENT) and the INC values are identified along the syntactic head path, while the PARTS list of the phrase contains all and only the elements of the PARTS lists of its two daughters.

The clause of the SEMANTICS PRINCIPLE that restricts the combination of determiners with nominal projections plays a crucial role.[7] It dictates that the INC

[7] A variant of that clause for polyadic quantification is spelled out below in (16).

value of the non-head daughter *two* ($\boxed{4}$) and the head daughter's EXC value be identical, hence the EXC of the phrase is also $\boxed{4}$, because EXC values are identical along syntactic head paths. Finally the INC of the head daughter, $\boxed{5}$, is a (possibly improper) subterm of the restrictor of the quantifier, expressed in (6) by $\boxed{5} \triangleleft \alpha$. Note that this makes sure that x occurs in α, as is lexically required by *two* in (5).

3.2 Lexical Specification of Polyadic Quantifiers

Keenan & Westerståhl's (1997) semantics for a 'polyadic quantifier containing *different*' can be reformulated straightforwardly for languages of Ty2 extended by a syntactic construct for this class of complex quantifiers. The new quantifiers are notated as $\mathcal{Q}$ below:

Definition 2: For H a quantifier of type $\langle 1,1\rangle$ and $\mathcal{Q} = (H, \Delta)$ a polyadic quantifier of type $\langle 1^2, 2\rangle$ containing Δ, x, y variables of type e, α, β expressions of type $\langle et\rangle$, and ρ an expression of type $\langle e\,\langle et\rangle\rangle$, the interpretation of $\mathcal{Q}$ is as follows:
$[\![(H,\Delta)(\alpha,\beta,\rho)]\!]^{M,g} = 1$ iff there is an A', $A' \subseteq [\![\alpha]\!]^{M,g}$, such that

$$[\![H(\alpha)]\!]^{M,g}(A') = 1, \text{ and}$$
$$\forall \mathsf{e}_1, \mathsf{e}_2 \in A' \colon \mathsf{e}_1 \neq \mathsf{e}_2 \Rightarrow [\![\beta]\!]^{M,g} \cap [\![\rho]\!]^{M,g}(\mathsf{e}_1) \neq [\![\beta]\!]^{M,g} \cap [\![\rho]\!]^{M,g}(\mathsf{e}_2).$$

This definition presupposes a syntax where Δ is an appropriate logical constant for *different*, and for any type $\langle 1,1\rangle$ quantifier H, (H, Δ) is a well-formed syntactic expression of the functional type $\langle\langle et\rangle\,\langle\langle et\rangle\,\langle\langle e\,\langle et\rangle\rangle\,t\rangle\rangle\rangle$ (see Figure 1).

Given such an extended syntax and semantics of Ty2, the necessary logical lexical specifications for the adjective *different* are simple:

(7)
$$\begin{bmatrix} \text{PHON} & \langle \textit{different} \rangle \\ \text{SS LOC} & \begin{bmatrix} \text{CAT HEAD MOD LOC CONT} & \begin{bmatrix} \text{INDEX DR} & \boxed{1a}\, y \\ \text{MAIN} & \zeta \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{INDEX DR} & \boxed{1a}\, y \\ \text{MAIN} & \boxed{1b}\, \Delta \end{bmatrix} \end{bmatrix} \\ \text{LRS} & \begin{bmatrix} \text{EXC} & \textit{me} \\ \text{INC} & \boxed{1}\, (\gamma,\Delta)(\sigma_1, \lambda y.\beta, \ldots \lambda y.\rho) \\ \text{PARTS} & \left\langle \begin{array}{l} \boxed{1}, \boxed{1a}\, y, \boxed{1b}\, \Delta, \boxed{1c}\, (\gamma,\Delta), \boxed{1d} (\lambda y.\beta), \boxed{1e}\, (\lambda y.\rho), \\ \boxed{1f}\, (\gamma,\Delta)(\sigma_1), \boxed{1g}\, (\gamma,\Delta)(\sigma_1, \lambda y.\beta) \end{array} \right\rangle \end{bmatrix} \end{bmatrix}$$
$$\& \; y \triangleleft \beta \;\&\; y \triangleleft \rho \;\&\; \zeta \triangleleft \beta$$

The semantic specification is particularly perspicuous in the INC: *different* contributes a quantifier with Δ whose other component, γ, is lexically undetermined. It applies to two restrictor arguments, the first one lexically unknown (σ_1), whereas the second is a lambda abstract whose variable y is the DR value of *different* and shared with the DR value of the element it modifies (2nd line in (7)). The MAIN value, ζ, of a nominal projection that *different* modifies will end up in β of this restrictor, as required by the subterm statement $\zeta \triangleleft \beta$ in the last line of (7). The last argument of (γ, Δ), a binary relation, is largely undetermined except for the lambda abstraction over y, which is the second lambda abstraction in the expression, as indicated by the leading dots.

3.3 Principles of Semantic Composition

The final steps in capturing reading (2-a) of sentence (1) based on the syntactic analysis (4) concern the principles of semantic composition. In LRS all principles of the semantic combinatorics that are dependent on specific syntactic constellations are expressed as clauses of the SEMANTICS PRINCIPLE. We will have to reconsider adjective-noun combinations (for combining *different* with the nominal projection it modifies), and we have to be careful about determiner-noun combinations and NP-VP combinations, because these have to be generalized from monadic quantifiers, for which the relevant clauses of the SEMANTICS PRINCIPLE were originally defined, to the polyadic case. All of these will in fact be minor adaptations of existing clauses.

The first phrase to be examined is the noun phrase *different berries*, in which the adjective *different* forms a head-adjunct phrase with the count noun *berries*, which is the syntactic head of the phrase.

(8)
$$\begin{bmatrix} \textsc{phon} & \left\langle \textit{different, berries} \right\rangle \\ \textsc{ss loc cont} & \begin{bmatrix} \textsc{index dr} & \boxed{1a}\ y \\ \textsc{main} & \boxed{2a}\ \mathsf{berry}' \end{bmatrix} \\ \textsc{lrs} & \begin{bmatrix} \textsc{exc} & \boxed{1}\ (\gamma,\Delta)(\sigma_1, \lambda y.\beta, \ldots \lambda y.\rho) \\ \textsc{inc} & \boxed{2}\ \mathsf{berry}'(y) \\ \textsc{parts} & \left\langle \begin{array}{l} \boxed{1}, \boxed{1a}\ y, \boxed{1b}\ \Delta, \boxed{1c}\ (\gamma,\Delta), \\ \boxed{1d}\ (\lambda y.\beta), \boxed{1e}\ (\lambda y.\rho), \boxed{1f}\ (\gamma,\Delta)(\sigma_1), \\ \boxed{1g}\ (\gamma,\Delta)(\sigma_1, \lambda y.\beta), \boxed{2}, \boxed{2a}\ \mathsf{berry}' \end{array} \right\rangle \end{bmatrix} \end{bmatrix} \ \&\ \boxed{2a} \triangleleft \beta \ \&\ y \triangleleft \rho$$

According to the SEMANTICS PRINCIPLE, in head-adjunct phrases the EXC of the non-head (*different*) is a subterm of the EXC of the head (*berries*). Now observe that the NP *different berries* as it occurs in sentence (1) is the maximal projection of the noun *berries* (NP_2 in (4)), and NP_2 is a non-head daughter of a verbal projection (VP). Moreover, *different* is the maximal projection of the adjectival non-head *different* in the NP *different berries*. By the INCONT PRINCIPLE and clause (a) of the EXCONT PRINCIPLE[8] it follows that the INC value of *different* equals its own EXC value, i.e. both the INC and the EXC in (7) are $\boxed{1}$ in our sentence. But with $\boxed{1}$ being the largest logical expression on the PARTS list of *different berries* in the sense that all other elements on the PARTS list are subterms of $\boxed{1}$, $\boxed{1}$ must actually equal the EXC value of the non-head *different berries*, as depicted in (8); it cannot be a proper subterm of that EXC. The INDEX DR value and the MAIN value of the

[8](Richter & Kallmeyer, 2009, p. 47) present the two principles as follows:
INCONT PRINCIPLE:
In each *lrs*, the INCONT value is an element of the PARTS list and a component of the EXCONT value.
The EXCONT PRINCIPLE, Clause (a):
In every phrase, the EXCONT value of the non-head daughter is an element of the non-head daughter's PARTS list.

phrase are inherited from its head daughter, *berries*, by the CONTENT PRINCIPLE, as is the INC, $\boxed{2}$ (due to the LRS PROJECTION PRINCIPLE). Also note that the subterm restriction $\zeta \triangleleft \beta$ from (7) becomes more specific in NP_2, because we now know that the relevant MAIN value of the modified head is berry', which is why it is now required that $\boxed{2a}$, which is berry', be a component of the relevant second restrictor of the complex quantifier (γ, Δ).

The noun phrase *different berries* (8) forms a head-complement phrase with the verb *picked* shown in (9). Semantically, this head-complement phrase is an instance of combining a generalized quantifier as interpretation of the NP (recognizable through the form of its EXC) with the semantics of a verbal projection, represented here by the INC of (9).

(9)
$$\begin{bmatrix} \text{PHON} & \langle \textit{picked} \rangle \\ \text{SS LOC} & \begin{bmatrix} \text{CAT VAL} & \begin{bmatrix} \text{SUBJ} & \langle \text{NP}_{\boxed{4a}} \rangle \\ \text{COMPS} & \langle \text{NP}_{\boxed{1a}} \rangle \end{bmatrix} \\ \text{CONT MAIN} & \boxed{3b}\ \mathsf{pick}' \end{bmatrix} \\ \text{LRS} & \begin{bmatrix} \text{EXC} & \boxed{0} \\ \text{INC} & \boxed{3}\ \mathsf{pick}'(\boxed{4a}, \boxed{1a}) \\ \text{PARTS} & \langle \boxed{3}, \boxed{3a}\ \mathsf{pick}'(\boxed{1a}\, y), \boxed{3b}\ \mathsf{pick}' \rangle \end{bmatrix} \end{bmatrix}$$

(10) is the clause of the SEMANTICS PRINCIPLE that imposes combinatory restrictions on the VP *picked different berries* shown in (11).

(10) SEMANTICS PRINCIPLE, Clause 2,
combinations of quantified NPs with a verbal projection
(adapted from (Richter & Kallmeyer, 2009, p. 65))
In each *headed-phrase*, for some $n \geq 2$, if the non-head is a quantified NP with an EXC value of the form *generalized-quantifier*$(\sigma_1, \ldots, \sigma_n, \rho)$, the INC value of the head is a subterm of ρ.[9]

(11)
$$\begin{bmatrix} \text{PHON} & \langle \textit{picked, different, berries} \rangle \\ \text{SS LOC CONT MAIN} & \boxed{3b}\ \mathsf{pick}' \\ \text{LRS} & \begin{bmatrix} \text{EXC} & \boxed{0} \\ \text{INC} & \boxed{3}\ \mathsf{pick}'(\boxed{4a}, \boxed{1a}\ y) \\ \text{PARTS} & \left\langle \begin{array}{l} \boxed{3}, \boxed{3a}\ \mathsf{pick}'(\boxed{1a}\, y), \boxed{3b}\ \mathsf{pick}', \\ \boxed{1}\ (\gamma, \Delta)(\sigma_1, \lambda y.\beta, \ldots \lambda y.\rho), \boxed{1a}\ y, \boxed{1b}\ \Delta, \\ \boxed{1c}\ (\gamma, \Delta), \boxed{1d}\ (\lambda y.\beta), \boxed{1e}\ (\lambda y.\rho), \boxed{1f}\ (\gamma, \Delta)(\sigma_1), \\ \boxed{1g}\ (\gamma, \Delta)(\sigma_1, \lambda y.\beta), \boxed{2}\ \mathsf{berry}'(y), \boxed{2a}\ \mathsf{berry}' \end{array} \right\rangle \end{bmatrix} \end{bmatrix}$$

$\& \ \boxed{2a} \triangleleft \beta \ \& \ \boxed{3} \triangleleft \rho$

The VP in (11) inherits its MAIN value and its EXC and INC values from its syntactic verbal head, (9), as dictated by the CONTENT PRINCIPLE and the LRS PROJECTION PRINCIPLE. As always, the PARTS list contains all elements of the PARTS

[9] In the feature-logical encoding in a sort hierarchy, *generalized-quantifier* is a supersort of all quantifier symbols.

lists of the two syntactic daughters. Moreover, complying with (10), the INC of the verb *picked*, $\boxed{3}$, must be in the scope ρ of the quantifier in the EXC of *different berries*, $\boxed{1}$, in the phrase *picked different berries*. In (11) this is shown as the last conjunct in the conjunction of subterm conditions in the last line ($\boxed{3} \triangleleft \rho$).

The difference between (10) and the corresponding standard clause of the SEMANTICS PRINCIPLE in LRS is in the flexible number of restrictors σ_n of the generalized quantifier, and in the different functional types of the restrictors and the scope, which is not visible above because the types are not mentioned in the descriptions. A similar variant of Clause 2 of the SEMANTICS PRINCIPLE is already foreshadowed in the version of this clause in (Iordăchioaia & Richter, 2015, p. 631) in an analysis of Romanian negative concord constructions. The representations of the Romanian counterparts of n-words like *nobody* and *nothing* also have undetermined numbers of restrictors.

The last missing step in the analysis of sentence (1), adding the subject NP *every ape* to the verb phrase *picked different berries* shown in (11), involves all central techniques for the integration of complex unreducible polyadic quantifiers in the specification of the syntax-semantics interface. At its core the successful treatment of unreducible quantifiers hinges on one particular property of lexical semantic resources in LRS, namely on the fact that distinct lexical elements in the same utterance may contribute the same piece of semantic representation structure. This feature of LRS has been used extensively in accounts of negative concord in languages such as French, Polish, and Romanian.[10] To illustrate the point, consider (12) from a variety of American English with negative concord:

(12) This ain't no half-assed sub shop.

LRS captures the observation that (12) clearly has only a single negation reading by assuming that while *ain't* and *no* both lexically contribute a negation operator to the semantics of the sentence, the semantic constraint system of this variety of English forces those two negation operators to be the same negation operator. In essence, the single negation operator in the semantic representation of the sentence is connected to two distinct lexical units.

The parallelism of this type of treatment of negative concord to our construction with *different* NPs becomes more apparent in examples from Romanian negative concord such as (13-a) (Iordăchioaia & Richter, 2015, p. 610 (3c), pp. 638–639), because just like in our sentence with two NPs, *every ape* and *different berries*, the Romanian negative concord construction also involves two nominal phrases, *no student* and *no book* which, *prima facie*, seem to be independent negative quantifiers.[11]

[10] For Polish, see Richter & Sailer (2004); for a typological discussion of French, Polish and German Richter & Sailer (2006); and for Romanian Iordăchioaia & Richter (2015).

[11] In the current discussion I ignore the behavior and influence of the negative verbal prefix *nu*, because this leads to orthogonal morpho-syntactic and semantic considerations.

(13) a. Niciun student nu a citit nicio carte.
no student not has read no book
No student read any book.

b. *no((*x*,* y*), student*$'(x)$*, book*$'(y)$*, read*$'(x, y)$*)*

c. niciun student: *no((*x*,* ν_2*), student*$'(x)$*,* β*,* ρ*)*

d. nicio carte: *no((*ν_1*,* y*),* α*, book*$'(y)$*,* ρ'*)*

e. nu a citit: *no((*ν_1'*,* ν_2'*),* α'*,* β'*, read*$'(\nu_1', \nu_2')$*)*

Despite the apparent independence of the two NPs, Iordăchioaia & Richter (2015) analyzes the negative concord reading of the sentence with a single polyadic negative quantifier as shown in (13-b).[12] (13-c)-(13-d) sketch in a highly informal notation what the semantic contributions in LRS of the two NP quantifiers and the negated verb are to the final semantic representation (13-b): *niciun student* contributes a variable x and a restrictor *student*$'(x)$ whose linear position in the expression corresponds to the position of x in the initial sequence of variables: x is the first of two variables, and *student*$'(x)$ is the first of two restrictor arguments. This correspondence can again be observed in (13-d), *nicio carte*, which contributes variable y (in the second position of the sequence of variables) and the linearly second restrictor, *book*$'(y)$. The verb form in (13-e) contributes the material in the scope of the negative quantifier. In order to obtain the representation (13-b), the three sets of semantic contributions have to be matched up, simultaneously obeying all other restrictions on possible representations of the sentence from all LRS composition principles. The only way to do this is to conclude that, despite first appearances, there is in fact only one polyadic negative quantifier in this sentence, and this single polyadic negative quantifier is contributed by all three syntactic units, with the various syntactic slots of the polyadic quantifier filled in by each of the three expressions in turn. Variable x from (13-c) fills the unspecified slots ν_1 and ν_1', y fills the slots ν_2 and ν_2', *student*$'(x)$ equals α and α', and so on. The semantic constraint system of the syntax-semantics interface is specified in such a way that (13-b) is the only solution for the negative concord reading, modulo variable names.[13]

The present syntax-semantics interface for polyadic quantifiers with Δ essentially proceeds along the same lines. There are two differences: (1) the syntactic representation format of polyadic quantifiers is changed, and (2) the NPs that contribute to the same quantificational expression do not contribute the same quantifier constant, instead they contribute different pieces to the same complex quantifier. Again simplifying an LRS constraint set by projecting it onto pseudo-logical expressions in a somewhat *ad hoc* notation that is meant to highlight the underlying ideas, (14) gives an informal sketch of the semantic contributions of the two NPs and the verb to the semantic representation (14-b) of sentence (14-a). In contrast

[12] As already mentioned earlier, Iordăchioaia & Richter (2015) assumes a different syntactic form for polyadic quantifiers than the present study.

[13] In some contexts, the sentence also has a double negation reading with two negative quantifiers. This is also captured by the grammar, but not shown here.

to the Romanian negative concord construction, the verb is not quantificational.

(14) a. Every ape picked different berries.
b. $(\forall, \Delta)(\lambda x.\mathsf{ape}'(x), \lambda y.\mathsf{berry}'(y), \lambda x \lambda y.\mathsf{pick}'(x, y))$
c. every ape: $(\forall, \gamma)(\lambda x.\mathsf{ape}'(x), \sigma_2, \lambda x \lambda \nu_2.\rho)$
d. different berries: $(\gamma', \Delta)(\sigma_1, \lambda y.\mathsf{berry}'(y), \lambda \nu_1 \lambda y.\rho')$
e. picked: $\mathsf{pick}'(\nu_1', \nu_2')$

As we saw already in the lexical specification of *different* and its combination with the noun *berries*, *different* produces an 'incomplete' complex quantifier whose first component must be contributed by some other quantifier. In (14-d) this open slot is marked γ'. To obtain a well-formed semantic expression, the open slot must be filled by some appropriate other lexical resource in the utterance. The principles of the semantic combinatorics of course determine which syntactic units are eligible for providing semantic resources for that slot. In sentence (14-a) there is only one candidate, the determiner *every*, which is contained in (14-c) as the specifier of the NP *every ape*. Lexically *every* must be appropriately underspecified in such a way that one of its realizations contains the universal quantifier, $\forall$, in the first slot of a complex quantifier whose second slot is open (γ in (14-c)). Since $\forall$ is the first component of the complex quantifier, its restrictor must also be in the first restrictor slot, while the second restrictor slot, σ_2, must ultimately be contributed by whatever fills the second syntactic quantifier slot. Note that there is also a lambda abstraction in corresponding linear order: Since $\forall$ is the first component quantifier and its restrictor fills the first slot, the relevant lambda abstraction λx also comes first in (14-c) (and second in (14-d), where it is λy). The other lambda abstraction slot does not contain a concrete variable but an open slot for a variable, ν_2 in (14-c) and ν_1 in (14-d). These two variable slots are filled by the variables from the other contributor to the quantifier, and ν_2 ultimately becomes y whereas ν_1 becomes x in the representation for the sentence, (14-b). Finally, the role of *picked*, (14-e), is to contribute the binary relation in the scope of the lambda abstractions, marked ρ and ρ', respectively.

It should already be clear from the Romanian negative concord example (13) how the semantic constraint set sketched in (14) can be resolved: If we assume that *every ape* and *different berries* contribute the same complex quantifier, (14-c) and (14-d) fill each other's open slots, and (14-e) turns the scope of the resulting completed complex quantifier into a well-formed expression, as shown in (14-b).

In the remainder of this section, I introduce the lexical specifications and discuss the clauses of the SEMANTICS PRINCIPLE that achieve what (14) illustrates. First of all, a lexical entry for *every* is needed whose semantic resources are specified in such a way that the NP quantifier (14-c) emerges as one possible result of combining *every* with *ape*. Doing so with complete formal rigor requires a rather technical relational specification in order to ensure the linear correspondence of the contributed component to the complex quantifier, the restrictor's position, and the order of the lambda abstraction in the scope (first quantifier component, first restrictor, first lambda abstraction). Instead of spelling out a relational constraint over tuples of term structures, I employ a more informal notation that indicates with sufficient clarity what a fully explicit specification would require.

(15)
$$\begin{bmatrix} \text{PHON } \langle \textit{every} \rangle \\ \text{SS LOC CONT} \begin{bmatrix} \text{INDEX DR } \boxed{4a}\, x \\ \text{MAIN } \boxed{4b}\, \forall \end{bmatrix} \\ \text{LRS} \begin{bmatrix} \text{INC} & \boxed{4}\, (\ldots, \forall_n, \ldots)(\ldots, (\lambda x.\alpha)_n, \ldots, \ldots (\lambda x)_n \ldots .\beta) \\ \text{PARTS} & \left\langle \begin{matrix} \boxed{4a}\, x, \boxed{4b}\, \forall, \boxed{4c}\, (\ldots, \forall, \ldots), \boxed{4d} (\lambda x.\alpha), \boxed{4e}\, (\lambda x \ldots .\beta), \ldots, \\ \boxed{4f}\, (\ldots, \forall, \ldots)(\ldots, \lambda x.\alpha), \ldots, \boxed{4} \end{matrix} \right\rangle \end{bmatrix} \end{bmatrix}$$
$$\& \; x \triangleleft \alpha \; \& \; x \triangleleft \beta$$

The notation with subscripts n in the INC is to be read as designating the position of the respective expressions: A quantifier in position n of the complex quantifier goes with a restrictor in position n of a sequence of restrictors of equal length, and with the nth lambda abstraction in a sequence of lambda abstractions. In addition, the lambda abstraction in the quantifier's scope binds the same variable as the lambda abstraction in the designated restrictor.

When *every* becomes the specifier in a nominal phrase with the nominal head *ape*, there is an underspecified complex quantifier for any finite number of restrictors, with the monadic constant $\forall$ contributed by *every* in one of the available positions of the complex quantifier, and the relation ape' of *ape* in the corresponding restrictor slot. But when *every ape* occurs in sentence (1), it is clear that only one choice of representation has a chance to be resolved to a well-formed logical expression: a complex quantifier with $\forall$ in first position, and ape' a subterm of the first restrictor, because this form will match the open slots in the quantificational expression of *different berries*. For any other choice, the two quantifiers from the two NPs cannot be identical, because they would either be inconsistent or there would be syntactic slots in the resulting complex quantifier that cannot be filled by syntactic contributions of lexical elements in the sentence. If the two contributed quantifiers are not identical, one of them must outscope the other. But then they would again have slots that do not contain expressions that are contributed by other lexical elements in the sentence. Such representations are ruled out by one of the standard LRS principles of semantic composition (EXCONT PRINCIPLE, Clause (b): All components of the logical representation of an utterance are contributed by some lexical element in the utterance).

The combination of quantificational determiners with nominal projections is semantically restricted by Clause 1 of the SEMANTICS PRINCIPLE. Here is a modified version for polyadic quantifiers:

(16) SEMANTICS PRINCIPLE, Clause 1,
combinations of quantificational determiners with a nominal projection
(adapted from (Richter & Kallmeyer, 2009, p. 65))
In each *headed-phrase*, for some $n \geq 1$,
if the non-head is a quantificational determiner with INC value of the form

$(\ldots, H_n, \ldots)(\ldots, (\lambda x.\alpha)_n, \ldots, \ldots \lambda x_n \ldots . \rho)$ in which H is a quantifier constant on the PARTS list of the non-head,
then the INC of the head is a subterm of α, and the INC value of the non-head daughter is identical with the EXC value of the head daughter.

Given Clause 1 of the SEMANTICS PRINCIPLE and focusing on the only version of *every ape* that can be consistently combined with *different berries*, we obtain (17) for the semantics of NP_1:

(17)
$$\begin{bmatrix} \text{PHON} & \langle \textit{every, ape} \rangle \\ \text{SS LOC CONT} & \begin{bmatrix} \text{INDEX DR} & \boxed{4a}\, x \\ \text{MAIN} & \boxed{5a}\, \mathsf{ape}' \end{bmatrix} \\ \text{LRS} & \begin{bmatrix} \text{EXC} & \boxed{4}\, (\forall, \psi)(\lambda x.\alpha, \sigma_2, \lambda x.\kappa) \\ \text{INC} & \boxed{5}\, \mathsf{ape}'(x) \\ \text{PARTS} & \left\langle \begin{matrix} \boxed{4}, \boxed{4a}\, x, \boxed{4b}\, \forall, \boxed{4c}\, (\forall, \psi), \boxed{4d}\, (\lambda x.\alpha), \boxed{4e}\, (\lambda x.\kappa), \\ \boxed{4f}\, (\forall, \psi)(\lambda x.\alpha), \boxed{4g}\, (\forall, \psi)(\lambda x.\alpha, \sigma_2), \boxed{5}, \boxed{5a}\, \mathsf{ape}' \end{matrix} \right\rangle \end{bmatrix} \end{bmatrix}$$
$$\& \boxed{5} \triangleleft \alpha \;\&\; x \triangleleft \kappa$$

The INDEX DR and MAIN values are inherited from the syntactic head of the NP (CONTENT PRINCIPLE), and the INC and EXC values are identical to the respective values of the syntactic head, *ape*. But the EXC of *ape* is also identical to the INC of *every* (according to Clause 1 of the SEMANTICS PRINCIPLE), which guarantees it is the quantifier, $\boxed{4}$. Finally, the INC $\boxed{5}$ of *ape* is a subterm of the lambda abstract in the first restrictor of the quantifier (again due to Clause 1 of the SEMANTICS PRINCIPLE).

For the last step, we have to consider Clause 2 of the SEMANTICS PRINCIPLE, (10), the NP *every ape*, (17), and the VP *picked different berries*, (11). According to Clause 2, the INC of *picked* (identical to all INC values along its syntactic head projection), must be a subterm of the scope of *every ape*, just as it must be in the scope of the quantifier of *different berries*, as we saw earlier. Initially this restriction leaves three alternatives, of which we can quickly rule out the first two: If *every ape* outscopes *different berries* or *vice versa*, there are slots in the two quantifiers (ψ in *every ape* and γ in *different berries*) that are not filled with syntactic contributions of other words, excluding these two possibilities as semantic representations of the utterance. This leads to assuming identity of the two quantificational expressions ($\boxed{1} = \boxed{4}$) from NP_1 and NP_2 as the only consistent solution of the semantic constraint set. This means that $\psi = \Delta$, $\gamma = \forall$, $\sigma_1 = \lambda x.\mathsf{ape}'(x)$, $\sigma_2 = \lambda y.\mathsf{berry}'(y)$, and that the scope argument of the polyadic quantifier takes the form $\lambda x \lambda y.\mathsf{pick}'(x, y)$. The result is shown in (18), omitting the long PARTS list which enumerates all and only the subterms of $\boxed{1}$.

(18)
$$\begin{bmatrix} \text{PHON} & \langle \textit{every, ape, picked, different, berries} \rangle \\ \text{SS LOC CONT MAIN} & \boxed{3b}\, \mathsf{pick}' \\ \text{LRS} & \begin{bmatrix} \text{EXC} & \boxed{1}\, (\forall, \Delta)(\lambda x.\mathsf{ape}'(x), \lambda y.\mathsf{berry}'(y), \lambda x \lambda y.\mathsf{pick}'(x, y)) \\ \text{INC} & \boxed{3}\, \mathsf{pick}'(\boxed{4a} x, \boxed{1a} y) \end{bmatrix} \end{bmatrix}$$

The EXC of an utterance contains its logical representation. According to **Definition 2** and the EXC value of (18), *Every ape picked different berries* is true iff any two apes picked unequal sets of berries.

4 Extensions

The main point of this paper is to show that a constraint-based semantics is capable of providing a systematic syntax-semantics interface for the semantics going back to (Keenan, 1992) of sentences with *different* such as (1) and (3). While Iordăchioaia & Richter (2015) show in their analysis of Romanian negative concord constructions that polyadic quantification can be used to give a parsimonious semantic account of difficult data that previously required additional syntactic assumptions about covert LF movement, the present account of *different* goes one step further. The negative polyadic quantifiers in the analysis of Romanian negative concord are reducible polyadic quantifiers in the sense that they can be decomposed into iterations of monadic quantifiers. Thus it is possible to have an analysis of that class of constructions in terms of traditional compositional semantics that assumes the same readings as the account in terms of polyadic quantifiers, even if only at the price of complicating syntax significantly. By contrast, the polyadic quantifiers in the present account of internal readings with *different* are (in-)famous for being unreducible to iterations of monadic quantifiers. It is for that very reason that previous accounts of 'compositional' semantics could not employ the elegant and direct semantics of **Definition 1**.

Seeing the feasibility of a syntax-semantics interface with unreducible polyadic quantification naturally leads to the question how the present approach fits into the general landscape of constructions with *different* that have been discussed in the literature, especially in (Beck, 2000; Brasoveanu, 2011; Bumford & Barker, 2013), and how its coverage can be extended. In this section, I offer a few initial thoughts on aspects of constructions with *different* in light of the current approach.

The main example (1) of the preceding discussion is similar to *Q-bound different* in Beck (2000), except that, similar to Keenan & Westerståhl's (1997) examples, I do not follow Beck's (2000) distinction between singular and plural *different*. One of the reasons is that it seems attractive to subsume under the general characterization of 'quantifiers containing Δ' other quantifiers not mentioned in (Beck, 2000), or (Brasoveanu, 2011):

(19) a. Two apes picked different berries.
b. Five apes picked different berries.
c. Few apes picked different berries.
d. Many apes picked different berries.
e. Most apes picked different berries.
f. No apes picked different berries.[14]

[14]See footnote 5 above for necessary refinements of the present semantics.

This extension is certainly in the spirit of the original idea of complex quantifiers, but it also requires a more careful future study of properties of monadic quantifiers that can enter into this construction.

For any theory of sentence internal *different*, it is crucial to determine in which syntactic and semantic environment *different* must find its licensing second quantifier component. The lexical specification in (7) determines two aspects of the distribution of *different*: It needs to find a second quantifier of Lindström type $\langle 1, 1 \rangle$ for the empty first slot of the complex quantifier that it introduces. Moreover, this other quantifier must be structured in such a way that it matches the structure of *different*'s complex quantifier. With the latter property, it also fills the open first restrictor slot. This means that being able to contribute an underspecified complex quantifier structure of this form is a necessary property of determiners (or nominal lexical elements such as *everyone*) to which *different* can attach, and the ability to form such a quantifier determines what can be a felicitous licenser of sentence internal *different*.

In addition to form, scope also plays a role as a licensing condition. Without any further refinements of the theory, candidate licensing quantifiers for *different* must be in *different*'s scope domain. Since the possibility of being in the same scope domain of course depends to a great extent on the scope properties of the NP with *different* itself, this is not a very strong restriction. For example, it is possible that scope islands are weak for *different* NPs, in which case they can find their licenser outside of sentences that are scope islands to other quantifiers. For a first impression on the issues at stake, consider the following examples from (Bumford & Barker, 2013, p. 360):

(20) a. Every boy gave every girl a different poem.
b. Every boy gave every girl he liked a different poem.
c. Every boy said [every girl read a different poem].

Assuming that *a different poem* is analyzed similarly in all relevant aspects to our earlier plural NPs with *different*, there are two quantifiers, *every boy* and *every girl*, in (20-a) that can license the internal reading, because both are in the same scope domain with *a different poem*. Of course, the *different* NP requires via the mechanisms discussed above that one of the two universal quantifiers comes with a semantic representation that can be identified with *different*'s representation.

(20-b) is an example that Bumford & Barker (2013) consider problematic for Brasoveanu (2011), since in the latter approach only the most local distributivity operator (associated with the licensing universal quantifier) can function as the licenser of the *different* NP.[15] The pronoun in the relative clause attached to *every girl* in (20-b) requires that its binding operator outscope *every girl*. As a consequence, *every boy* cannot take immediate scope over *a different poem*, and, if only

[15] See Lahm (2016) for further discussion of this issue, and for problems with the alternative solution proposed by Bumford & Barker (2013). Correctly identifying the licensing NPs and their domain is an open problem.

the closer of the two quantifiers can function as a licenser then only the intervening *every girl* is able to do so. However, both internal readings are available for (20-b): The sentence may mean that every boy chose a different poem for any girl he likes, i.e. no two girls he likes receive the same poem from him. But a girl who is liked by two boys might receive the same poem from the two. Alternatively every boy may give the same poem to every girl he likes, but no two boys give the same poem to the girls they like.

This constellation of readings is compatible with the present approach since the complex quantifier comprising the universal quantifier and *different* arising from *every boy* and *a different poem* can outscope *every girl he liked*, and it is also possible that a monadic quantifier *every boy* outscopes a complex polyadic quantifier arising from *different poem* and *every girl he liked*.

Yet another aspect of licensing domains is illustrated with (20-c), where one of the potential licensers of *a different poem* is in a matrix clause. Without specific additional restrictions, the combinatorics of LRS allows quantifiers in embedded clauses to take scope outside of their clause. It follows that *every boy* and *a different poem* can be in the same scope domain (the matrix clause), and that they can form a complex quantifier in the matrix clause. On the other hand, *every girl* and *a different poem* may form a complex quantifier in the embedded clause. The present theory is so far silent about possible *de re*/*de dicto* readings of *a different poem* that come from its structural position in the embedded clause and the possibility of interpreting it in the matrix clause. Lahm (2016) argues that there is a *de re*/*de dicto* ambiguity of *different* phrases with licensers in this constellation and shows that his theory of the behavior of sentence internal *different* in terms of a restriction on Skolem functions can account for it.[16]

As a final set of examples, consider the following pattern:

(21) a. Every ape picked different berries.
b. Every ape picked two different berries.
c. Every ape picked a different berry.

(21-a)–(21-c) all have sentence internal readings. The last two examples show that at least some specifiers can be added to *different* phrases while retaining the sentence internal reading, and that this phenomenon goes beyond the indefinite singular construction often discussed in the literature. A first conceivable representation of (21-b) that extends the polyadic theory of *different* is shown in (22):

(22) $(\forall, (\mathsf{two}', \Delta))(\lambda x.\mathsf{ape}'(x), \lambda y.berry'(y), \lambda x \lambda y.\mathsf{pick}'(x, y))$

The idea here is to follow the lead of Keenan & Westerståhl (1997) and assume

[16]Another example by Lahm (p.c.) highlights the related issue of intervening modalities between the *different* phrase and its putative licenser as a challenge to the polyadic quantifier analysis:

(i) John and Mary want to live in different cities.

that polyadic quantifiers in natural languages are constructed following certain systematic but rather limited patterns. (22) tentatively extends the polyadic quantifier for *different* by adding the quantifier constant of the specifier to Δ. **Definition 2** can then be modified in a relatively straightforward way through imposing an additional restriction on the restrictor set $[\![\beta]\!]$ that is due to the extra quantifier, parallel to the condition H imposes on $[\![\alpha]\!]$.

The considerations in this section suggest that the polyadic perspective on *different* emphasizes the importance of taking a closer look at its distribution. Recent methods of corpus research could offer interesting new insights that would help to see the advantages and disadvantages of current theories.

5 Conclusion

I demonstrated that the constraint-based semantics of Lexical Resource Semantics can give an explicit semantic account of sentence internal readings with *different* with categorematic unreducible polyadic quantifiers. The present paper mainly focused on the lexical semantic specifications and on the combinatoric principles expressed in the clauses of the SEMANTICS PRINCIPLE that lead to this result. The result is theoretically significant because such a semantics cannot be obtained with the kind of minimalistic syntactic structure assumed in so-called surface-oriented frameworks like LFG and HPSG in combination with the flavor of compositional semantics that is widely adopted in linguistics.

In Section 4 I began to explore new perspectives on the data which are opened up by the polyadic analysis of sentence internal *different*, indicating that the new analysis exhibits a few promising properties. At this early stage there are of course also many open questions: Plural semantics was ignored as well as consequences of intensional contexts, there was no satisfying treatment of singular indefinites, and the readings (2-b) and (2-c) of the ambiguous main example (1) were not addressed. They should soon come into view when turning to the question of assigning an appropriate semantics to the mysterious constant Δ of Lindström type $\langle 1, 1\rangle$, the core of my logical representation of *different*.

References

Barwise, Jon & Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophie* 4. 159–219.

Beck, Sigrid. 2000. The semantics of *different*: Comparison operator and relational adjective. *Linguistics and Philosophy* 23(2). 101–139.

Brasoveanu, Adrian. 2011. Sentence-internal *different* as quantifier-internal anaphora. *Linguistics and Philosophy* 34(2). 93–168.

Bumford, Dylan & Chris Barker. 2013. Association with distributivity and the problem of multiple antecedents for singular *different*. *Linguistics and Philosophy* 36(5). 355–369.

Dekker, Paul. 2003. ‘Meanwhile, within the Frege boundary’. *Linguistics and Philosophy* 26. 547–556.

Gallin, Daniel. 1975. *Intensional and higher-order modal logic*. Amsterdam: North-Holland.

Iordăchioaia, Gianina. 2010. *Negative concord with negative quantifiers: a polyadic quantifier approach to Romanian negative concord*: University of Tübingen dissertation. http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-48224.

Iordăchioaia, Gianina & Frank Richter. 2015. Negative concord with polyadic quantifiers. The case of Romanian. *Natural Language and Linguistic Theory* 33(2). 607–658.

Keenan, Edward L. 1992. Beyond the Frege boundary. *Linguistics and Philosophy* 15. 199–221.

Keenan, Edward L. & Dag Westerståhl. 1997. Generalized quantifiers in linguistics and logic. In Alice ter Meulen & Johan van Benthem (eds.), *Handbook of logic and language*, 837–893. Elsevier Science B.V.

Lahm, David. 2016. “Different” as a restriction on Skolem functions. In *Proceedings of SALT 26*, To appear.

Lindström, Per. 1966. First order predicate logic with generalized quantifiers. *Theoria* 32. 186–195.

Richter, Frank & Laura Kallmeyer. 2009. Feature logic-based semantic composition: A comparison between LRS and LTAG. In Anders Søgaard & Petter Haugereid (eds.), *Typed feature structure grammars*, 32–92. Frankfurt a.M.: Peter Lang.

Richter, Frank & Manfred Sailer. 2004. Basic concepts of Lexical Resource Semantics. In *ESSLLI 2003 – Course Material I*, vol. 5 Collegium Logicum, Kurt Gödel Society Wien.

Richter, Frank & Manfred Sailer. 2006. Modeling typological markedness in semantics: The case of negative concord. In Stefan Müller (ed.), *Proceedings of the 13th international conference on Head-driven Phrase Structure Grammar*, 305–325. CSLI Publications.

Westerståhl, Dag. 1989. Quantifiers in formal and natural languages. In Dov Gabbay & Franz Guenthner (eds.), *Handbook of philosophical logic*, vol. IV, chap. IV.1, 1–131. D. Reidel Publishing Company.

Two cases of prominent internal possessor constructions

Sandy Ritchie
SOAS

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 620–640

Keywords: possession, prominence, internal possessors, Chimane, Tsimane', Maithili

Ritchie, Sandy. 2016. Two cases of prominent internal possessor constructions. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 620–640. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.32.

Abstract

This paper outlines a new analysis of the syntactic structure and discourse function of a ‘prominent internal possessor construction’ (PIPC) in Chimane (unclassified, Bolivia) and compares it with an existing analysis of a different kind of PIPC found in Maithili (Indo-Aryan, India/Nepal). PIPCs in Chimane and Maithili involve an apparently non-local agreement relation between verbs and possessors which are internal to possessive NPs. In Chimane, it is argued that internal possessors are able to control object agreement via a clause-level ‘proxy’ of the internal possessor – see also Ritchie (under review). The paper goes on to compare this construction with PIPCs in Maithili, and shows that speakers use PIPCs in discourse to indicate the information structure role of the internal possessor. In the case of Chimane, it seems that internal possessors which bear the secondary topic role are more likely to control object agreement, while in Maithili, other semantic and information structural features of internal possessors are at play. The contributions of the various levels of sentence structure are modelled using the LFG architecture developed in Dalrymple & Nikolaeva (2005; 2011).

1 Introduction

Many languages have a means of syntactically promoting possessors.[1] This typically involves ‘raising’ of the possessor to an argument function, with concomitant demotion of the phrase headed by the possessed nominal. The motivation for possessor promotion is usually to indicate the prominent semantic and information structure (IS) role of the possessor.

This paper considers two cases of ‘prominent internal possessor constructions’ or PIPCs in two genetically unrelated languages: Chimane (or Tsimane’, unclassified, Bolivia) and Maithili (Indo-Aryan, India/Nepal). PIPCs are functionally similar to EPCs in that they are typically employed to signal the se-

[1] I would like to thank the UK Arts and Humanties Research Council for grants for PhD research (grant no. AH/J500410/1) and the Prominent Possessors project (grant no. AH/M010708/1, Principal Investigator Irina Nikolaeva). I would also like to thank my Chimane consultants in Bolivia: Benjamin Caity, Cupertino Caity, Santa Caity, Berthi Cayuba, Leonilda Plata, Dino Nate and Manuel Roca, our Maithili consultants in London: Dilip Mahaseth and Pushkar Patel, and especially Yogendra Yadava for his expert advice and guidance on the Maithili data. Thanks also to Irina Nikolaeva, Oliver Bond, Greville Corbett, Lutz Marten, Teresa Poeta, Charlotte Hemmings and the reviewers of the HeadLex16 proceedings for their comments, help and suggestions. Any remaining errors are my own. Abbreviations used: 1 = first person, 2 = second person, 3 = third person, ACC = accusative, BEN = benefactive, APPL = applicative, CAUS = causative, CLF = verbal classifier, F = feminine, FOC = focus, GEN = genitive, H = honorific, HRSY = hearsay evidential, INTR = intransitive, M = masculine, MH = midhonorific, NH = nonhonorific, NOM = nominative, NONNOM = non-nominative, O/OBJ = object, PASS = passive, POSS = possessor, PST = past, PTCP = participle, REFL = reflexive, S/SUBJ = subject, SG = singular.

mantic/IS prominence of possessors. However, as their name suggests, PIPCs differ from EPCs in one fundamental regard: the prominent possessor in a PIPC remains internal to the phrase headed by the possessed nominal. Despite this, possessors in PIPCs can participate in phrase-external syntax, for example by controlling agreement on the verb.

Prominent internal possessors (PIPs) in both Chimane and Maithili can control verbal agreement. However, there are several differences between the two languages which are revealing with respect to their underlying structure. In Maithili, possessors internal to possessive phrases functioning as both subjects and objects can control verbal agreement, as illustrated in (1):

(1) a. *dekha-l-thun*
saw.-PST-3H.2MH
'He (honorific) saw you (mid-honorific).'

b. *tohər ba:bu Mohan-ke dekha-l-thun*
you.NH.GEN father Mohan-ACC see-PST-3H.2NH
'Your (non-honorific) father saw Mohan.'

c. *o tora: ba:p-ke dekha-l-thun*
he.H you.NH.ACC father-ACC see-PST-3H.2NH
'He saw your (non-honorific) father.'

(Stump & Yadav 1988: 306, 309, 317)

In (1a), the subject controls the 'primary' agreement (on the left in the gloss) and the object controls the 'secondary' agreement.[2] Primary agreement in Maithili is always controlled by the subject, while secondary agreement can be controlled by a number of non-subject elements including internal possessors. Control of secondary agreement by a possessor internal to a subject can be seen in (1b), while in (1c), the possessor internal to the object NP controls the secondary agreement. Stump & Yadav provide evidence to show that these agreement controlling possessors are internal to the possessive phrase.

In a similar way, agreement is also possible between the verb and internal possessors in Chimane. In this case, however, the agreement pattern is more restricted; it can only occur between possessors internal to object NPs, and must be accompanied by an additional applicative-like verbal suffix.[3]

(2) a. *Juan täj-je-' un mu' Sergio-s.*
Juan(M) touch-CLF-3SG.F.O hand(F) the.M Sergio(M)-F
'Juan touched Sergio's hand.'

[2]There are no overt arguments in (1a) but it is assumed here that the morphology on the verb functions as agreement morphology and not as incorporated pronouns.

[3]The verbs in the examples in (2) also feature verbal classifiers labelled as CLF. These are suffixes which obligatorily occur on most verbal roots to create inflectable stems. They have various meanings related to subject control and transitivity – see Sakel (2004; 2007).

b. *Juan täj-je-bi-te un mu' Sergio-s.*
Juan(M) touch-CLF-APPL-3SG.M.O hand(F) the.M Sergio(M)-F
'Juan touched Sergio's hand.'

c. **Juan täj-je-te un mu' Sergio-s.*
Juan(M) touch-CLF-3SG.M.O hand(F) the.M Sergio(M)-F
('Juan touched Sergio's hand.')

d. **Mu' vojity=yụ nạij-tye-ye / nạij-bi-ye mi.*
the.M brother(M)=my see-CLF-1SG.2SG see.CLF-APPL-1SG.2SG you
('My brother saw you.')

(Ritchie 2015)

(2a) shows object agreement with the feminine head of the patient-like possessive NP. In (2b), the verb appears to exhibit object agreement with the internal possessor, and it also exhibits the suffix *-bi*. (2c) shows that this agreement pattern is not possible if the *-bi* suffix is not present, and (2d) shows that it is not possible for possessors internal to subject NPs to control agreement on the verb, whether or not the *-bi* suffix is present. Just as in Maithili, syntactic tests can be used to show that the possessor in (2b) is internal.

This kind of configuration presents a challenge for linguistic theories, as it has hitherto been assumed that verbal agreement can only be controlled by the head of a noun phrase, and not by non-head subconstituents (cf. e.g. the Control Agreement Principle in Gadzar & Pullum 1982; Gadzar et al. 1985 and similar constraints in LFG). A further complication is the observed variance in languages which exhibit PIPCs. Maithili allows agreement with possessors internal to both subject and object (as well as other) arguments, while in Chimane the pattern is restricted to only occurring with possessors internal to objects. What these differences suggest is that PIPCs are not a homogeneous phenomenon and require different types of analysis for different languages.

A detailed analysis of the syntax of the Chimane PIPC can be found in Ritchie (under review). In Section 2, I will briefly summarise the analysis developed there and go on to show how information structure can be integrated with this analysis using the multi-level LFG architecture developed in Dalrymple & Nikolaeva (2011). In Section 3, I show how a different approach is required for the syntax and discourse function of the Maithili PIPC, principally following the analysis developed in Dalrymple & Nikolaeva (2005). A summary and areas for further research are identified in Section 4.

2 Prominent internal possessors in Chimane

This section provides a summary of the analysis of Chimane PIPCs set out in Ritchie (2015) and Ritchie (under review).

2.1 Possessive noun phrase

In possessive noun phrases, specifiers and modifiers, including possessors, agree with the gender of the head noun:

(3) a. *mọ' Juan-si' där-si' ococo*
the.F Juan(M)-F big-F frog(F)
'Juan's big frog'

b. *mu' Juan-tyi' där-tyi' ịtsiquij*
the.M Juan(M)-M big-M jaguar(M)
'Juan's big jaguar'

Possessors exhibit the same nominal agreement suffixes as adjectives and can co-occur with determiners in the phrase, suggesting that they function as modifiers (Lyons 1986). Heads and modifiers in the NP cannot precede determiners:

(4) a. **ococo mọ' mu'-si'*
frog(F) the.F his-F
('his frog')

b. **mu'-si' mọ' ococo*
his-F the.F frog(F)
('his frog')

c. **där-si' mọ' ococo*
big-F the.F frog(F)
('the big frog')

Apart from this restriction, the other constituents can occur in any order. This suggests that (i) the determiner occupies a higher structural position in the NP, and (ii) the rest of the NP has a flat structure.

There is also a type of bound possessor expression. Pronominals which do not exhibit agreement with the head noun must attach to the right of some NP constituent; there is a preference for them to attach to the right-most element of the NP, though they can also attach to other elements within the NP. Compare the lack of marking and positional restriction on the bound possessor in (5a) with the 'free' one in (5b):

(5) a. *ococo=mu'* / **mu' ococo*
frog(F)=his / his frog(F)
'his frog'

b. *ococo mu'-si'* / *mu'-si' ococo*
frog(F) his-F / his-F frog(F)
'his frog'

Bound possessors can also co-occur with free possessors:

(6) a. *Juan-si' ococo=mu'*
Juan(M)-F frog(F)=his
'Juan's frog'

b. *mu'-si' ococo=mu'*
his-F frog(F)=his
'his frog'

Bound possessors which co-occur with free possessors will be termed 'doubling possessors' because they are anaphorically controlled by and therefore double the features of the free possessor.

2.2 Clausal syntax

There is no case marking of core arguments in Chimane. Subject and objects can be identified by a number of other properties, most prominently the fact that they can control subject and object agreement on the verb. Depending on the combination of subject and object in transitive clauses, one or two suffixes indicate person, number, gender and clusivity features of the two arguments.

In double object constructions, the non-patient-like argument controls object agreement on the verb. In (7a), the monotransitive verb *tu-* 'bring' exhibits object agreement with the feminine patient-like argument. When a primary object argument expressing a beneficiary is added to the argument structure of this verb by the benefactive applicative *-ye*, as in (7b), the verb exhibits agreement with this argument:

(7) a. *Judyeya' mọ' qui jejmitidye' tu-i-'=in.*
and the.F so cooked.food(F) bring-CLF-3SG.F.O=they
'And they brought hot food.'

b. *Judyeya' qui ca jejmitidye' tu-ye-te=in.*
and so HRSY cooked.food(F) bring.CLF-BEN.APPL-3SG.M.O=they
'And they brought him hot food.'

These examples show that Chimane exhibits secundative alignment with respect to agreement between verbs and patient- and non-patient-like arguments.

The object which controls agreement on the verb in a double object construction will be termed the primary object, while the other object is the secondary object. The primary object is the direct object of a monotransitive verb or indirect object of a ditransitive verb, while the secondary object is the direct object of a ditransitive verb (e.g. Bresnan 1982; Dryer 1986).

2.3 Evidence that PIPs are internal in Chimane

An obligatory property of PIPs is that they cannot control agreement on the verb if they do not exhibit nominal agreement with the possessed noun.[4]

(8) *Yụ nạij-bi-te mọ' ococo Juan(*-si')*
I see.CLF-APPL-3SG.M.O the.F frog(F) Juan(M)-F
'I saw Juan's frog.'

This is a strong indication that possessors in PIPCs are internal to the possessive phrase, as only internal modifiers exhibit nominal agreement with the head of the phrase.

Constituency tests involving insertion of a clause-level adverb like 'yesterday' between the possessor and possessed noun do not provide clear evidence that the possessor is internal or external to the possessive NP, as Chimane exhibits free word order and discontinuous NPs are a possibility, as shown in (9):

[4] The verb in (8) does not feature an overt verbal classifier due to a process of morphophonological deletion. The phonological form of the stem *-tye* is similar to the agreement suffix *-te* and is therefore deleted due to a morphophonological rule, thus **nạijtyebite* is realised as *nạijbite*. See also Sakel (2007).

(9) *Yụ-ty na are'-yi pa'tyi'.*
my-M FOC hurt-CLF.M.S fingernail(M)
'My fingernails hurt.'

Another test is passivization. Chimane PIPs can function as the subject of the passive verb, as shown in (10a), where the passive verb exhibits subject agreement with the feminine possessor. However, just as in examples where the possessive phrase functions as the object of the verb, like (8), here again the possessor must exhibit nominal concord with the head noun in this construction, as shown by the ungrammaticality of (10b):

(10) a. *Maria-ty vojity=mọ' ja'-ĉat-bu-ti-' (Juan)*
Maria(F)-M brother(M)=she PASS-hit-APPL-PASS-F.S Juan(M)
Maria's brother was hit (by Juan).'

b. **Maria vojity ja'-ĉat-bu-ti-' (Juan)*
Maria(F) brother(M) PASS-hit-APPL-PASS-F.S Juan(M)
Maria's brother was hit by Juan.'

This indicates that possessors cannot function as independent arguments in syntactic processes such as passivization, which constitutes further evidence that possessors in PIPCs are internal to the possessive phrase.

2.4 Mediated locality

One potential analysis of Chimane PIPCs is that the PIP has a representation or 'proxy' in the clause which stands in for it and functions as the object, and this is what enables the possessor to control object agreement. This idea has not been developed for PIPCs but appears in some analyses of long-distance agreement (LDA) constructions (e.g. Polinsky 2003).

The PIPC in Chimane exhibits one particular feature which may support this type of analysis. It is a common (though optional) feature of PIPCs that the PIP is doubled by a bound possessor:

(11) *Mi nạij-bi-te ococo Juan-si' (=mu').*
you see.CLF-APPL-3SG.M.O frog(F) Juan(M)-F =him
'You saw Juan's frog.'

If the doubling possessor in (11) is an overt expression of a proxy of the internal possessor which functions as the object of the verb, then it is possible to predict that this element can only occur in PIPCs and not in the corresponding internal possessor construction (IPC) in which the possessed noun controls object agreement. This prediction is borne out; the bound pronominal cannot easily occur in the default IPC. Its insertion in the IPC equivalent of (11) is considered strange or ungrammatical:

(12) *Mi nạij-tye-' ococo Juan-si' (?*=mu').*
you see-CLF-3SG.F.O frog(F) Juan(M)-F =him
'You saw Juan's frog.'

This seems to indicate that the doubling possessor might be an overt expression of an external clause-level proxy of the internal possessor in the PIPC.

Another argument in favour of a mediated locality-type analysis of the Chimane PIPC is the fact that agreement between the possessor and the verb only occurs with objects. As shown in (2d), repeated in (13), possessors internal to subjects cannot control agreement on the verb:

(13) **Mu' vojity=yụ nạij-tye-ye / nạij-bi-ye mi.*
the.M brother(M)=my see-CLF-1SG.2SG see.CLF-APPL-1SG.2SG you
('My brother saw you.')

This is also the case for all other argument and non-argument functions apart from objects. This shows that Chimane exhibits a restricted paradigm of agreement between verbs and internal possessors, and these restrictions seem to be syntactic in nature. Chimane PIPCs are akin to applicative constructions in that a non-argument in the default counterpart of the construction functions as the object in the applicative construction. If the PIPC is akin to applicative constructions in these respects, then it also seems plausible to assume that the doubling possessor represents a clause-level proxy of the internal possessor which functions similarly to an applied object in an applicative construction.

Evidence that the PIPC may be similar to applicative double object constructions comes from its use with ditransitive verbs. Recall from Section 2.2 that the non-patient-like argument controls object agreement in double object constructions. In cases of ditransitive verbs featuring the *-bi* suffix, this argument appears to correspond to the possessor. The following example comes from a description of a picture of a girl giving a monkey its baby back after taking it away:

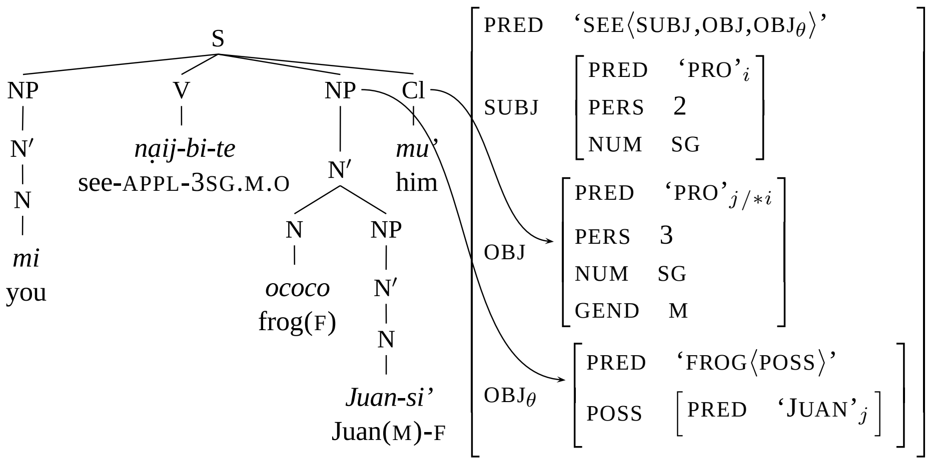
(14) *Ji'-cañ-e-bi-baj-te qui ạva'.*
CAUS-return-CLF-APPL-again-3SG.M.O so baby(F)
'So she [the girl] gives it [the monkey] back its baby.'

In all the examples of PIPCs discussed so far, PIPs control object agreement despite being internal to the single patient-like argument of a semantically monotransitive verb. In (14), the possessor appears to correspond to the recipient-like argument of the semantically ditransitive verb. This suggests that the possessor is an object of this verb as well as functioning as the possessor of the possessive patient-like argument, and that the possessor and possessive phrase may be associated with different object functions. This configuration with ditransitives may indicate that in fact all verbs in PIPCs subcategorize for both a primary and secondary object function, with the possessor bearing the primary object function and the possessed noun the secondary object function.

The first point to note about this analysis is that the external representation of the PIP is not necessarily overtly expressed. Many of the examples of PIPCs presented do not feature the doubling possessor. Therefore, I will follow Bresnan & Mchombo (1987) in assuming that the agreement morphology on the verb functions as an incorporated pronoun and bears the grammatical function instead of the doubling possessor in such cases. A second point is the correspondence between PIPs and their external representations. When overt, the doubling possessor must match the features of the PIP one-to-one. Therefore, this element will be analysed as being anaphorically controlled by the PIP.

Taking all these features of the analysis of the PIPC together, it is possible to provide a formal representation of the construction using the LFG architecture. The type of PIPC which is represented in (15) is that featuring the doubling possessor. At c-structure, this element is analysed following Dalrymple (2001) as a non-projecting Cl(itic) function. As argued above, the doubling possessor is an overt realization of a clause-level proxy of the internal possessor which functions as the object of the verb. Therefore, it occurs directly under the clausal head at c-structure, and maps to the OBJ function at f-structure. The possessive phrase bears the OBJ_θ function, but otherwise has the same internal structure as other possessive phrases. The anaphoric control of the proxy object by the internal possessor is shown by the indices in the f-structures of these two elements. The necessarily disjoint relationship between the subject and object is also indicated by indices. Example (11) is repeated here as (15):

(15) *Mi naịj-bi-te ococo Juan-si'=mu'.*
you see.CLF-APPL-3SG.M.O frog(F) Juan(M)-F=him
'You saw Juan's frog.'



The analysis in (15) explicitly shows the syntactically ditransitive nature of the Chimane PIPC. However, it is still unclear from this analysis what the conditions are under which the PIPC occurs in discourse.

2.5 Discourse function of PIPCs in Chimane

With the assumptions made in Section 2.4, we can now consider the contribution of semantics and information structure in determining when PIPCs are used in discourse. In order to uncover the discourse motivation for using PIPCs, I designed some picture description tasks.

In the animacy and alienability task, participants were shown a series of pictures which depicted a person acting on another person, animal or inanimate object, and a third person or animal looking angry about the situation. It was explained to the participants that the person, animal or object who was being acted on was the possession (i.e. the kin or alienable possession) of the person or animal who looked angry. The participant was then asked a question about the picture which either topicalized the possession or the possessor.

In the sibling story task, the participants were shown a series of pictures which developed the story of a brother and sister and their interactions with their parents and possessions and with animals and their possessions. The task was designed to elicit many examples of possessive constructions.

There are two results in the animacy and alienability task which may indicate that topicality of the possessor is an influencing factor in the decision to use the PIPC. In their descriptions of situations in which people act on other people's kin, the participants used the PIPC more often when the possessor was topicalized in the question. The results are shown in Table 1 (the topic in the question is highlighted in each case):

Situation	Question	PIPC	%
man grabs man's sister	Why is **the man** angry?	9/16	56
	What's happening to **the man's sister**?	0/16	0
woman hits woman's son	Why is **the woman** angry?	12/16	75
	What's happening to **the woman's son**?	4/16	25

Table 1: Animacy and alienability task – topicality of the possessor

The results in Table 1 show that the participants preferred to use PIPCs to describe a person acting on another person's kin when the second person (i.e. the possessor) is topicalized in the question. More data is needed to show if this is a significant tendency, but these results do seem to provide an initial indication that topicality may be an influencing factor on the speaker's choice between the PIPC and the IPC.

Another type of evidence which may support the argument that PIPCs are used to express topicality comes from examples of the use of IPCs in the sibling story task. In the story, after the brother and sister interact with their mother and father, they leave for the forest. After this episode, they find a canoe. It was explained to the participants that this canoe belonged to the children's father. Some of the participants used transitive constructions with possessive objects to describe this situation. In such cases, they always used IPCs rather than PIPCs:

(16) *Aty jọba-'=in nạij-te covamba jen'-tyi'=in.*
now leave-F.S=they see.CLF-3SG.M.O canoe(M) father(M)-M=their
'Now they're leaving and see their father's canoe.'

The use of IPCs in this discourse context is revealing because of the topicality of the possessor. The possessor referent (the children's father) is not topical at this point in the discourse; several events separate this mention of him from the last mention, and this may be why the participants selected the IPC rather than PIPC to describe the situation.

These results provide some initial indication that PIPCs are preferred when the possessor is topical. This proposal is similar to the observations made for PIPCs in other languages, but no formal analysis of this type of construction involving information structure has been proposed. However, differential object marking (DOM), which is a related phenomenon as it also involves variability in morphosyntactic marking of arguments, has been formally analysed by Dalrymple & Nikolaeva (2011). They develop a formal architecture involving interacting levels of sentence grammar. It is the interactions which constrain DOM. Specifically, topicality constrains differential marking of objects. Topical objects are marked, while non-topical objects are not. If the observations given about the use of PIPCs in Chimane discourse are correct, then this constraint-based approach can also be applied to these constructions.

Unlike DOM constructions, in which the object either bears the topic role or not, in PIPCs, it is also necessary to consider the role of the possessed noun. In situations in which the possessed noun is marked, as in IPCs, it will be assumed that both the possessed noun and the possessor share a single information structure role which applies to the entire possessive phrase. In situations where the PIP bears the topic role, it is assumed here that the possessed noun bears a completive information structure role, as it is discourse-new but not in focus (Dalrymple & Nikolaeva 2011).

Before moving to the analysis of some specific examples of PIPCs, a formalization of the proposed general constraint is required. In their analysis of topical non-subject agreement in Tabassaran, Dalrymple and Nikolaeva offer the following formalization of the general constraint that any nonsubject element which bears a topic role will control agreement on the verb (assuming that agreement is an explicit signal of topicality):

(17) $((\uparrow [\text{GF}-\text{SUBJ}])_\sigma\ \text{DF}) = \text{TOPIC}$ (Dalrymple & Nikolaeva 2011: 122)

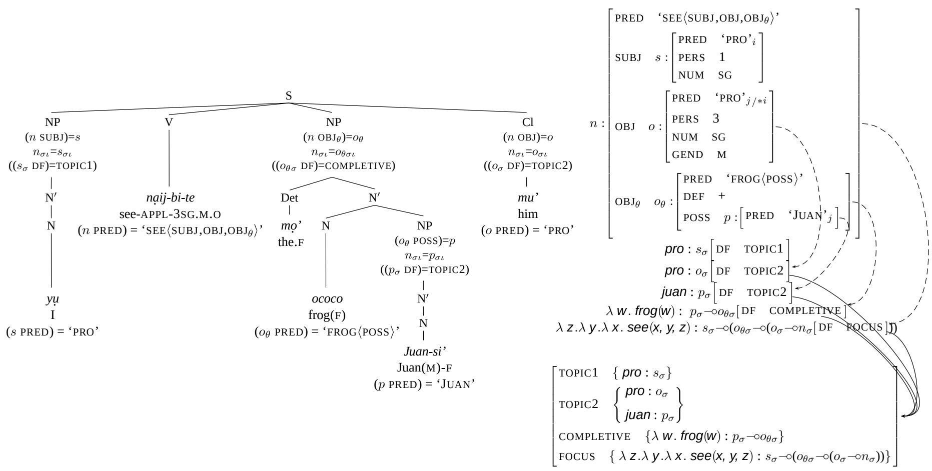
Adapting this kind of general formalization, the proposed analysis of the Chimane PIPC can be restated in the following terms: whichever element bears the secondary topic role at information structure will bear the object function at functional structure. This constraint can be represented as follows:

(18) $((\uparrow \text{OBJ})_\sigma\ \text{DF}) = \text{TOPIC2}$

This constraint entails that in contexts in which the internal possessor bears the secondary topic role, it (or rather its proxy) must bear the object function. This also implies that the possessive phrase must bear a different grammatical function. Using this general constraint, it is now possible to provide a formal analysis of the syntax and information structure of PIPCs.

In (19), the possessor bears the secondary topic role, but since the possessed noun is not topical in this construction, it instead bears a completive information structure role. It is possible that it could bear another role, but the important point is that in this construction, the IS roles of the possessor and possessed noun are different. It is this difference in IS prominence which triggers the valency change and enables the external proxy of the possessor to bear the object function. Thus the conditions under which the verb agrees with the possessor in the PIPC, which was left unresolved in Section 2.4, are now clarified.

(19) *Yụ nạij-bi-te mọ' ococo Juan-si'=mu'*
I see.CLF-APPL-3SG.M.O the.F frog(F) Juan(M)-F=him
'I saw Juan's frog.'



The representation in (19) demonstrates the integration of the c-structure and f-structure of the PIPC, in which the external proxy of the internal possessor bears the object function, with the s-structure and i-structure of the construction, at which it is specified that the possessor and its external proxy bear the secondary topic role while the possessive phrase bears a completive role (see Ritchie (2015) for more details on this analysis).

3 Prominent internal possesssors in Maithili

Maithili exhibits SOV word order and both case marking on nominals and agreement on verbs. Nominals and pronominals are distinguished for nominative, accusative/dative and genitive case. It also has a system of honorificity involving four levels in the second person and two in the third person. The levels are High-Honorific (HH), Honorific (H), Mid-Honorific (MH), and Non-Honorific (NH) (Yadav 1996).

3.1 Possessive noun phrase

Possessors in possessive NPs generally stand in the genitive, as in (20a), but when the possessive phrase bears certain grammatical functions, for example the object function, possessors can also stand in the dative/accusative case, as in (20b):

(20) a. *tohər bap əe-l-thun*
you.MH.GEN father come-PST.3H.2MH
'Your (MH) father (H) came.'

b. *həm tora beta-ke dekha-l-iau*
I you.NH.ACC son-ACC see-PST-1.2NH
'I saw your (NH) son.' (Stump & Yadav 1988: 309)

The order of elements in the possessive NP is fixed as possessor-possessed noun. The opposite order is ungrammatical (all of the following examples come from a recent paper by Yadava et al. 2016):

(21) **həm nokər-ke tohər pita-l-iau*
I servant-ACC you.NH.GEN hit-PST-1.2NH
('I hit your (NH) servant (NH).')

Possessors can also co-occur with determiners. In such cases, the possessor can either precede or follow the determiner:

(22) a. *i tohər nokər əe-l-əi*
this you.NH.GEN servant come-PST-3NH
'*This* servant (NH) of yours (NH) came.'

b. *tohər i nokər əe-l-au*
you.NH.GEN this servant come-PST-2NH.NONNOM
'This servant (NH) of *yours* (NH) came.'

Note the difference in emphasis indicated in the translations. When the determiner precedes the possessor, the focus is on the determiner. When the possessor precedes the determiner, the focus is on the possessor.

3.2 Clausal syntax

There are three verbal agreement paradigms: the nominative intransitive paradigm, the non-nominative intransitive paradigm and the cross-reference paradigm. Major features of nominals referenced by the paradigm are person and honorific grade.

Nominative intransitive agreement is controlled by the sole argument of an intransitive verb, while non-nominative intransitive agreement is controlled by non-nominative subjects on intransitive verbs, for example dative subjects. (23) shows examples of nominative and dative subjects of intransitive verbs:

(23) a. *tu ae-l-æ*
you.NH come-PST-2NH
'You (NH) came.'

b. *tora bukh lagh-l-au*
you.NH.ACC hungry feel-PST-2NH.NONNOM
'You (NH) were hungry.'

The cross-reference paradigm consists of verbal agreement suffixes which cross-reference two referents in the clause: the 'primary' and 'secondary' referents. The primary referent is nearly always (but does not necessarily have to be) the subject. The secondary referent is the second most prominent referent in the clause, which can be the object but also obliques and, crucially, possessors internal to a number of terms and non-terms.

Secondary agreement is possible with single objects of monotransitive verbs, patient-like and non-patient-like objects of ditransitive verbs, and possessors internal to all of these:

(24) a. *həm tora pita-l-iau*
I you.NH.ACC hit-PST-1.2NH
'I hit you.' (single object of monotransitive verb)

b. *həm tohər nokər-ke pita-l-iau*
I you.NH.GEN servant-ACC hit-PST-1.2NH
'I hit your servant.' (possessor internal to object)

c. *həm tora bəcha de-l-ie / de-l-iau*
I you.NH.ACC baby give-PST-1.3NH give-PST-1.2NH
'I gave you the baby.' (direct or indirect obj. of ditransitive verb)

d. *həm tohər guruji-ke bəcha de-l-iau*
I you.NH.GEN teacher-ACC baby give-PST-1.2NH
'I gave the baby to your teacher.' (poss. internal to indirect object)

e. *həm tohər bəcha guruji-ke de-l-iau*
I you.NH.GEN baby teacher-ACC give-PST-1.2NH
'I gave your baby to the teacher.' (poss. internal to direct object)

Secondary agreement is also possible with oblique arguments, and with possessors internal to obliques:

(25) a. *həm tohər sange khana pakau-l-iau*
I you.NH.GEN with food cook-PST-1.2NH
'I cooked with you.' (oblique)

b. *həm tohər guruji-ke sange khana pakau-l-iau*
I you.NH.GEN teacher-ACC with food cook-PST-1.2NH
'I cooked with your teacher.' (possessor internal to oblique)

Agreement is also possible with possessors internal to subjects of intransitives. In such cases, the possessor triggers non-nominative agreement on the verb:

(26) *tohər nokər əel-əi / əel-au*
you.NH.GEN servant come-PST-3NH come-PST-2NH.NONNOM
'Your (NH) servant came.'

Possessors internal to subjects of transitive verbs can also control secondary agreement. In such cases, primary agreement is controlled by the possessed noun. This means both elements of the possessive subject are referenced:

(27) *tohər bhai həmra pita-l-kho*
you.NH.GEN brother me hit-PST-3NH.2NH
'Your brother hit me.'

These examples show that in Maithili, agreement is possible with possessors internal to (i) subjects of intransitive verbs, (ii) subjects of transitive verbs, (iii) direct and indirect objects of mono- and ditransitive verbs, and (iv) obliques.

3.3 Evidence that PIPs are internal in Maithili

One piece of evidence which suggests that possessors which control secondary agreement on the verb are internal to the phrase headed by the possessed noun is the fact that they cannot be separated from the possessed noun by a clause-level element. For example, it is not possible for a clause-level adverb to occur between the possessor and possessed noun.

(28) **həm tohər khail nokər-ke pita-l-iau*
I you.NH.GEN yesterday servant-ACC hit-PST-1.2NH
'I hit your servant yesterday.'

Another test which may indicate that possessors in PIPCs are internal to the possessive phrase is the fact that they are not accessible to passivization. The only possible passive for (1c), which is repeated here in (29a), is (29b), where the subject corresponds to the entire possessive phrase which bears the object function in (29a). Example (29a) cannot have a passive variant such as (29c) whose subject is the former possessor:

(29) a. *o tora: ba:p-ke dekha-l-thun*
he.H your.NH.ACC father-ACC see-PST-3H.2NH
'He saw your (NH) father (H).'

b. *tohar ba:p dekhal ge-l*
your father see.PST.PTCP go-PST.3NH
'Your (NH) father (H) was seen.'

c. **tõ ba:p(-ke) dekhal ge-le*
you.NOM father-ACC see.PST.PTCP go-PST.2NH
'Your (NH) father (H) was seen.'

(Stump & Yadav 1988: 317)

An example like (29c), where the possessor stands in the nominative and the auxiliary exhibits agreement with the possessor, is ungrammatical. Stump & Yadav argue that this shows that the possessor which controls secondary agreement in (29a) does not bear an argument function in the clause, but is internal to the possessive phrase headed by the possessed noun.

3.4 Trigger-happy agreement

The examples in Section 3.3 show that it is not only internal possessors which can control secondary agreement. Direct objects of ditransitive verbs and obliques can also control this agreement. This suggests that the controller of secondary agreement does not necessarily correspond to an unrestricted argument. Instead, secondary agreement can be controlled by whichever potential controller is most semantically or information structurally prominent in a given discourse context. The specific semantic and/or IS factors which determine the agreement controller are not immediately apparent, and different studies have argued for different factors (see Section 3.5). However, in terms of the syntax of the construction in which the possessor controls agreement, all that needs to be said is that there is no difference in the structure of the PIPC versus the default IPC. In both cases, the possessor is internal to the possessive NP. In the case that the possessor controls secondary agreement, one or a combination of prominent semantic and/or IS features of the possessor means that it 'wins out' in the competition for control of secondary agreement over other potential controllers. This type of pragmatically determined agreement system has been termed 'trigger-happy' by Comrie (2003), as the agreement target (in this case the verb) can have more than one potential trigger or controller.

The f-structure of this kind of configuration has been analysed by Dalrymple & Nikolaeva (2005) as in (30):

(30) *o* *tora:* *ba:p-ke* *dekha-l-thun*
he.H you.NH.ACC father-ACC see-PST-3H.2NH
'He (H) saw your (NH) father.'

$$\left[\begin{array}{ll} \text{PRED} & \text{`SEE}\langle\text{SUBJ,OBJ}\rangle\text{'} \\ \text{SUBJ} & \left[\begin{array}{ll} \text{PRED} & \text{`HE'} \\ \text{PERS} & 3 \end{array}\right] \\ \text{OBJ} & \left[\begin{array}{ll} \text{PRED} & \text{`FATHER'} \\ \text{PERS} & 3 \\ \text{POSS} & \left[\begin{array}{ll} \text{PRED} & \text{`YOU'} \\ \text{PERS} & 2 \end{array}\right] \end{array}\right] \end{array}\right]$$

Based on Dalrymple & Nikolaeva (2005: 87)

However, just as with Chimane, it is not possible to fully explain why PIPs are able to control agreement on the verb in Maithili without also considering the semantic and/or information structure features of the possessor.

3.5 Discourse function of PIPCs in Maithili

Based on the observations about secondary agreement in Maithili presented here, it must be stated that in Maithili, agreement controllers do not correspond one-to-one with grammatical functions, as they do in many other languages. Instead, predicate-'argument' agreement is conditioned by something else.

So far, analyses of Maithili have claimed that it is the functional prominence of internal possessors which enables them to control secondary agreement. Stump & Yadav (1988) claim that topicality is the main motivating factor for using a PIPC in discourse. Bickel et al. (1999) cite the interaction between the pragmatic concepts of 'face' and 'empathy' as the primary motivation for speakers' choice between potential controllers of secondary agreement. Comrie (2003) and Dalrymple & Nikolaeva (2005) argue that the possessors will control secondary agreement when they bear the secondary topic role at information structure. Finally, Dalrymple & Nikolaeva (2011) argue that some notion of contrast may also be involved in conditioning the choice. In a more recent proposal, Yadava et al. (2016) argue that the motivation for the alternation is to index a combination of semantic and information structural features of possessors.

The semantic feature referenced by secondary agreement is the honorific grade of the possessor, or more specifically 'face versus empathy' (Bickel et al. 1999). Potential controllers which are higher in honorific grade will control secondary agreement, even if they are more 'lowly' in their syntactic status. For example, in a social context in which you are referring to an honoured person's non-honorific possessions, and that person is present in the situation, it is infelicitous for the verb to show agreement with their non-honorific possession over them. This is despite the fact that the honorific referent is an internal possessor and the non-honorific possessed noun is the head of the object NP:

(31) *tu hunak nokər-ke pit-l-ahunh* / **pit-l-ahi*
you he.H.GEN servant-ACC hit-PST-2NH.3H hit-PST-2NH.3NH
'You (NH) hit his (H) servant.' (Honorific possessor is present)

It is only felicitous to use the the variant in which the possessed noun controls secondary agreement if the honorific possessor is absent from the situation:

(32) *tu hunak nokər-ke *pit-l-ahunh* / *pit-l-ahi*
you he.H.GEN servant-ACC hit-PST-2NH.3H hit-PST-2NH.3NH
'You (NH) hit his (H) servant.' (Honorific possessor is absent)

This kind of judgement indicates that the need to respect honoured people is one of the factors motivating the choice between potential agreement controllers.

Agreement with honorific referents can also be 'overridden' if another potential controller is focussed. If a non-honorific possessor is focussed, it is possible for it to control secondary agreement:

(33) a. *tu kəkər sikshak-ke pit-l-ahunh*
you who.GEN teacher hit-PST-2NH.3H
'Whose teacher (H) did you hit?'

b. *həm tohər sikshak-ke pit-l-iau*
I you.NH.GEN teacher hit-PST-1.2NH
'I hit your (NH) teacher (H).' (teacher absent, possessor in focus)

This indicates that it is possible for focus to override honorificity, enabling non-honorific possessors to control secondary agreement over honorific possessed nouns.

Adapting the analysis of Maithili developed in Dalrymple & Nikolaeva (2005), it is possible to integrate a level of information structure with the f-structure given in Section 3.4.

(34) *o* *tora:* *ba:p-ke* *dekha-l-thun*
he.H you.NH.ACC father-ACC see-PST-3H.2NH
'He saw your (NH) father.'

Functional structure:

$$\begin{bmatrix} \text{PRED} & \text{`SEE}\langle\text{SUBJ,OBJ}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{`HE'}\\ \text{PERS} & 3\end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`FATHER'} \\ \text{PERS} & 3 \\ \text{POSS} & \begin{bmatrix}\text{PRED} & \text{`YOU'}\\ \text{PERS} & 2\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

Information structure:

$$\begin{bmatrix} \text{TOPIC} & \begin{bmatrix}\text{PRED} & \text{`HE'}\\ \text{STATUS} & \text{H}\end{bmatrix} \\ \text{FOCUS} & \begin{bmatrix}\text{PRED} & \text{`YOU'}\\ \text{STATUS} & \text{NH}\end{bmatrix}\end{bmatrix}$$

Based on Dalrymple & Nikolaeva (2005: 87)

In a similar way to Chimane, the integration of a level of information structure into the analysis motivates the choice of the internal possessor as the controller of secondary agreement in Maithili.

4 Summary and further questions

In this paper, two different types of prominent internal possessor constructions have been presented. Certain features of the constructions indicate that despite their functional similarity, they have very different underlying structures. The fact that possessor agreement is restricted to only occurring with objects in Chimane, and only in the presence of applicative-like verbal morphology, leads us to conclude that a valency-increasing process is the best analysis for the construction. In this respect, the Chimane PIPC is close to EPCs proper, albeit with the added complication that it is not the possessor itself which is 'raised' to argument status but a clause-level representation of the possessor which is inserted into the argument structure of the verb by the applicative.

Maithili presents a very different syntactic profile. In this case, possessors internal to a number of different terms and non-terms can control agreement, and they share this property with other non-terms including obliques. This suggests that instead of some valency-changing process akin to that in Chimane, the best explanation for the PIPC in Maithili is that verbal agreement does not reference grammatical functions, but rather semantically or information structurally prominent referents.

Despite these very different syntactic profiles, there is a sense in which

PIPCs in Chimane and Maithili constitute the same type of construction. That sense is functional: in both cases, internal possessors take on the syntactically prominent property of controlling agreement on the verb when they are also semantically or information structurally prominent. Some initial evidence has been presented here to show that in Chimane, topical possessors are more likely to control verbal agreement. In Maithili, possessors which are higher in honorific grade and which bear the focus role are more likely to control secondary agreement, and these two features can interact, with focussed non-honorific possessors able to override honorific non-focussed possessors.

Further questions

On the analysis of Chimane: it is not clear what is the anaphoric binding domain of the negative constraint which specifies that the agreeing possessor is disjoint in reference from the subject. Further data is required to test this. It is also not clear how to capture this constraint in the formal representation.

More generally on the study of PIPCs crosslinguistically: two types have been presented in this paper, but there may also be other types of which have not yet been identified. For example, in his analysis of Jarawara (Arawan), Dixon (2000) claims that possessors which control verbal agreement may take on the function of the head of the possessive NP. Another type of explanation may be that PIPs occur in a more peripheral position within the NP than their non-PIP counterparts, and it is this more peripheral position which makes them 'visible' to the phrase-external syntax.

References

Bickel, Balthasar, Walter Bisang & Yogendra P. Yadava. 1999. Face vs. Empathy: The Social Foundation of Maithili Verb Agreement. *Linguistics* 37(3). 481–518.

Bresnan, Joan. 1982. *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.

Bresnan, Joan & Sam A. Mchombo. 1987. Topic, Pronoun, and Agreement in Chichewa. *Language* 63(4). 741–782.

Comrie, Bernard. 2003. When Agreement Gets Trigger Happy. *Transactions of the Philological Society* 101(2). 313–337.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. New York: Academic Press.

Dalrymple, Mary & Irina Nikolaeva. 2005. Nonsubject Agreement and Discourse Roles. In *Oxford Working Papers in Linguistics, Philology, and Phonetics*, 73–94. University of Oxford: Centre for Linguistics and Philology.

Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and Information Structure*. Cambridge: Cambridge University Press.

Dixon, Robert M. W. 2000. Categories of the Noun Phrase in Jarawara. *Journal of Linguistics* 36(3). 487–510.

Dryer, Matthew S. 1986. Primary Objects, Secondary Objects, and Antidative. *Language* 62(4). 808–845.

Gadzar, Gerald, Ewan Klein, Geoffrey K. Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Oxford: Blackwell.

Gadzar, Gerald & Geoffrey K. Pullum. 1982. Generalized Phrase Structure Grammar: A Theoretical Synopsis. In *Proceedings of the Indiana University Linguistics Club*, Bloomington, IN: Indiana University.

Lyons, Christopher. 1986. The Syntax of English Genitive Constructions. *Journal of Linguistics* 22(1). 123–143.

Polinsky, Maria. 2003. Non-Canonical Agreement Is Canonical. *Transactions of the Philological Society* 101(2). 279–312.

Ritchie, Sandy F. 2015. *Internal Possessor Prominence in Chimane*. London: SOAS, University of London Doctoral thesis.

Ritchie, Sandy F. under review. Agreement with the Internal Possessor in Chimane: A Mediated Locality Approach .

Sakel, Jeanette. 2004. *A Grammar of Mosetén*. Berlin: Mouton de Gruyter.

Sakel, Jeanette. 2007. The Verbness Markers of Mosetén Complex Predicates. In Bernhard Wälchli & Matti Miestamo (eds.), *New Trends in Typology: Broadening the Horizons and Redefining the Foundations*, 315–338. Berlin: Mouton de Gruyter.

Stump, Gregory T. & Ramawatar Yadav. 1988. Maithili Verb Agreement and the Control Agreement Principle. In Diane Brentari, Gary Larson & Lynn MacLeod (eds.), *Papers from the 24th Annual Regional Meeting of the Chicago Linguistic Society, Part Two: Parasession on Agreement in Grammatical Theory*, 304–321. Chicago: Chicago Linguistic Society.

Yadav, Ramawatar. 1996. *A Reference Grammar of Maithili* (Trends in Linguistic Documentation 11). Berlin: Mouton de Gruyter.

Yadava, Yogendra P., Sandy Ritchie, Irina Nikolaeva & Oliver Bond. 2016. The Syntax of Internal Possessors in Maithili. Paper Presented at the Workshop on Prominent Internal Possessors. SOAS, London, 22-23 September 2016.

Definite meaning and definite marking

Manfred Sailer
Goethe University Frankfurt a. M.

Assif Am-David
Goethe University Frankfurt a. M.

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 641–661

Keywords: definite article, Lexical Resource Semantics, Yiddish, Papiamentu, Esperanto

Sailer, Manfred & Assif Am-David. 2016. Definite meaning and definite marking. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 641–661. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.33.

Abstract

We look at definite marking in Esperanto, Papiamentu, and Yiddish considering three semantically definite contexts: the referential use of proper names and unique nouns, as well as anaphoric definites. Based on the typological study of languages with multiple definite articles in Am-David (2016), we argue for a three-dimensional decomposition of the ι-operator: an individual denotation, an existence presupposition, and a uniqueness conventional implicature. We present an HPSG encoding of this system and model the central aspects of the definite marking systems of our three object languages.

1 Introduction

In this paper, we will look at the way in which three languages express definiteness in three selected contexts. We will not be able to do justice to the rich literature on definiteness in linguistics and beyond, but rather concentrate on the discussion of the data and the development of a particular encoding. We will argue that definite meaning should be expressed as a combination of an asserted content, a presupposition, and a conventional implicature. The object languages of this paper will be Esperanto, Papiamentu, and Yiddish. These languages have been chosen as they represent different types of systems of definite marking, they are very well documented, and there is easily accessible data (corpora and internet).

We will provide general information on the definiteness systems of our three object languages in the introduction. In Section 2, we will show how they mark definiteness in three semantically definite environments: the referential use of proper names and unique nouns as well as anaphoric definites. In Section 3, we will look at one of the ways in which definite meaning is analyzed in the literature and identify a basic problem of the standard use of the ι (*iota*) operator for this purpose. We will present our own, three-dimensional, semantics of definiteness in Section 4. We will then move to the HPSG part of the paper. We will sketch the relevant aspects of the framework to be used here, *Lexical Resource Semantics* (Richter & Sailer, 2004), in Section 5 and extend it to our multi-dimensional semantics. In Section 6, we will apply this framework to the definiteness systems of Esperanto, Papiamentu, and Yiddish. We will end the paper with a conclusion.

Esperanto (Eo) is a constructed language that was created by Ludwik Leyzer Zamenhof (1859–1917) and first published in Warsaw as Zamenhof (1887). It has been shown, for example in van Oostendorp (1993), that Esperanto has all

[†]We would like to thank the program committee of HeadLex16 for inviting Manfred Sailer to the conference. We are grateful for the questions and the feedback we received during the conference and from the proceedings reviewers. Part of our material was also discussed in Sailer's class *Semantic Phenomena of Creole Languages*, Frankfurt a.M., summer term 2016. The presentation of the Papiamentu data profited considerably from the feedback of the participants, Florian Bauernschmitt, Lajana Knapp, David Krüger, Gabi Schmid-Hönisch, and Elisabeth Wolf. We are grateful to to Philippa Cook for discussion and to Maria Paunel for correcting our English.

properties of a natural language from a linguistic perspective. Therefore, we will treat it just as any other natural language that can and should be described by the means of formal linguistic theory.

There are two important reference grammars for Eo: Kalocsay & Waringhien (1985) and Wennergen (2016), both of which provide good information on the use of the definite article. Wennergen (2016) is strongly corpus based and uses primarily naturally occurring example sentences. In addition to these resources, the discussion of Eo will be based on data derived from the corpus *Tekstaro de Esperanto* (http://www.tekstaro.com/) and from web pages written in Eo.

Eo has an uninflected definite article, *la*, which is roughly used in the same way as the English *the*. Contrary to English, however, Eo has no indefinite article. A contracted form of the definite article, *l'*, exists, which can be used following a preposition that ends in a vowel, such as *de l'* 'of the' or *pri l'* 'about the'. These contracted forms are mainly used in poetry and not in spoken or prose Eo (Wennergen, 2016, p. 102). They will be ignored in the following.

Papiamentu (Pap), the second language to be discussed here, is a creole language spoken on the so-called ABC islands, Aruba (where the language is called "Papiamento"), Bonaire, and Curaçao. According to Kouwenberg (2013), its main lexifier languages are (Afro)Portuguese, Spanish, and Dutch. Pap is among the official languages of all three ABC islands. There is a vast linguistic literature on Pap, and it is highly present on the internet. Our data discussion is based on a reference grammar (van Putte & van Putte-de Windt, 2014) and a textbook (van Putte & van Putte-de Windt, 1992) as well as on data from web pages in Pap. In addition, we will rely on the discussion of the Pap article system in Kester & Schmitt (2007).

Pap has an uninflected definite article, *e* and an indefinite article, *un*. There exists a contracted form of the definite article when following the preposition *di* 'of': *dje* 'of-the'. We could not find a systematic discussion of when the contracted form is used and will, therefore, ignore it in this paper.

Finally, Yiddish (Yid) is an Indo-European language belonging to the Germanic branch. It is closely related to the High German dialects but has been in strong language contact with its surrounding languages. We will base our discussion on two reference grammars, Mark (1978) and Katz (1987). In addition, we use the *Corpus of Modern Yiddish* (CMY), which is available at http://web-corpora.net/YNC.

Yid has both a definite and an indefinite article. The definite article shows gender, number, and case agreement. In addition to the full forms of the articles, there are contracted forms with some prepositions, such as *tsum* 'to-the.MASC.SG'. The distribution of the contracted forms seems to be similar to what has been observed for standard German, for example in Schwarz (2009). There are also article-less uses with some preposition-noun combinations, such as *in gas* 'in the street' (Katz 1987: 80–81). We will not discuss the article-less forms in this paper.

After these preliminaries on the three languages, we can, now, look at how they use their definite articles in three semantically definite contexts.

2 Definite marking in Eo, Pap, and Yid

In this section, we will look at whether the three languages under consideration use a definite article in contexts in which we can assume a semantic definiteness. In particular, we will look at proper names, unique nouns, and anaphoric definites. We go through these three contexts one by one.

2.1 Proper names

We will look at proper names in their primary use only, i.e., as referring to an individual that bears the particular name (von Heusinger, 2010). In all three languages, no article is used in the standard case. This is shown in (1)–(3).

(1) En 1873 li transloĝis al Varsovio kun la tuta familio
in 1873 he moved to Warsaw with the entire family (Eo, tekstaro)

(2) I Korsou ta un isla chiki, ...
and Curaçao is an island small
'And Curaçao is a small island ...' (Pap)[1]

(3) khaym kumt bald.
Khaim comes soon
'Chaim is coming soon.' (Mark, 1978: 120)

While our three languages do not use an article with simple proper names, an article occurs when the noun is modified. This is illustrated for Yid in (4).[2]

(4) *(der) royter khaim kumt bald.
the red Khaim comes soon
'The red Chaim comes soon.' (Mark, 1978: 120)

2.2 Unique nouns

Unique nouns have exactly one individual satisfying their descriptive content. Löbner (2011: 284) lists *sun, pope, US president, weather*, etc. as nouns that are typically used as unique nouns.

The definiteness marking of unique nouns is particularly interesting for us as we see a great degree of variation here. As shown in (5), Eo requires a definite marking with unique nouns. In Pap, no article can be used, see (6). An article is used in Yid, but, where possible, this will be the contracted form of the article, illustrated in (7). If no contracted form exists in a given constellation, the full form is used, see (8).

[1] http://www.aav.cw/2014/02/07/spich-di-director-di-aav-na-okashon-di-nan-evento-di-anja-nobo-2014/, last consulted: 24.7.2016.

[2] Other languages use the definite article with proper names obligatorily (such as Modern Greek) or optionally (such as the varieties of Standard German spoken by the present authors).

(5) *(La) suno subite sin montris el la nuboj, ...
the sun suddenly itself showed from the clouds
'Suddenly the sun showed itself out of the clouds ...' (Eo, tekstaro)

(6) (*E) Solo ta kima sin miserikòrdia.
(the) sun PRES burn without mercy
'The sun is burning without mercy' (Pap, Kester & Schmitt, 2007: 113)

(7) un zey hobn geshikt tsum meylekh
and they have sent to-the king
'and they sent to the king' (Yid, CMY)

(8) az der meylekh hot gehert di zakh, ...
as the king has heard the affair ...
'as the king has heard about the affair, ...' (Yid, CMY)

2.3 Anaphoric definites

When a definite NP refers to a previously introduced referent, we speak of an anaphoric definite. In Eo, the definite article is used in these cases, see (9). Similarly, Pap uses its definite article in such contexts, see (10). The antecedent is dotted in our examples.

(9) Mi havas grandan domon. La domo havas du etaĝojn.
I have big house the house has two floors
'I have a big house. The house has two floors.' (Eo, Wennergen, 2016: 80)

(10) Mi a kumpra un bolo. 'I bought a cake'

*(E) bolo a wòrdu kome den 10 minüt.
the cake PART been eat in 10 minutes

'The cake was eaten in 10 minutes.' (Pap, Kester & Schmitt, 2007: 119)

We find a definite article with anaphoric definites in Yid as well, as shown in (11). However, the contracted form does not occur in these contexts, even in cases where a preposition-article contraction, *funem*, would be possible, see (12).

(11) hot im gefunen a man, ...; un der man hot im gefregt, azoy tsu zogn:
has him found a man and the man has him asked so to say
'A man found him and the man asked him to say: ...' (Yid, CMY)

(12) un a man iz gegangen .... 'And a man left ...'

un der nomen fun dem man iz gewen elimelekh ...
and the name of the man is been Elimelekh

'And the name of the man was Elimelekh ...' (Yid, CMY)

	proper name	unique noun	anaphoric definite
Esperanto	–	la	la
Papiamentu	–	–	e
Yiddish	–	der, di, dos, ... tsum, ... (P-Det)	der, di, dos, ...

Table 1: Definiteness marking in Esperanto, Papiamentu, and Yiddish

The distribution of the marking for the three languages in the three contexts is given in Table 1. It will be the task of the rest of this paper to provide a semantic characterization of the three contexts and lexical entries for the determiners and nouns that, in combination, produce the patterns found in the data.

3 Previous approaches

It is impossible to even summarize the most important contributions to the linguistic analysis of definites in the present paper—see Am-David (2014, 2016) for a systematic presentation of the state of the art. What we will do instead is to provide the basic ingredients of a semantics of definites as used in the literature and to point out some problematic aspects. We will start with the ι-operator (Section 3.1), then turn to the semantics of our three contexts (Section 3.2).

3.1 Definite meaning: The ι-operator

The most popular formalization of semantic definiteness uses the ι-operator. An ι-expression has the form $(\iota x : \phi)$ and refers to the single individual, a, that makes ϕ true, given that there is such a unique individual. If there is no such unique individual, the denotation of the ι-expression is undefined. This is stated more formally in (13).

(13) a. Syntax: For each type τ, for each variable x of type τ and for each formula ϕ, $(\iota x : \phi)$ is an expression of type τ.

b. Semantics: $[\![(\iota x : \phi)]\!]$

(i) is only defined if there is exactly one individual a such that $[\![\phi]\!]^{g[x \mapsto \mathsf{a}]} = 1$

(ii) when defined, then $[\![(\iota x : \phi)]\!] = \mathsf{a}$.

It is important to see that the definition in (13) combines three ingredients of meaning: First, the denotational aspect that the ι-expression refers to an individual. Second, the requirement that there exists such an individual and, third, the requirement that there is exactly one such individual. We will call the last two the *existence* requirement, and the *uniqueness* requirement. We will argue that it is more adequate to treat these three meaning components separately.

This separation is done in (14), where we redefine the semantics of ι-expressions. An ι-expression denotes an individual satisfying the description ϕ, (14-a), the existence requirement is expressed in (14-b), the uniqueness requirement in (14-c). When both existence and uniqueness are satisfied, (14) yields the same semantics as (13). We follow Coppock & Beaver (2015) in formulating uniqueness as "not more than one" instead of "exactly one" to keep it independent of existence.

(14) The semantics of $[\![(\iota x : \phi)]\!]$

a. Denotation: $[\![(\iota x : \phi)]\!] \in \{\mathsf{a} \mid [\![\phi]\!]^{g[x\mapsto \mathsf{a}]} = 1\}$

b. Existence: there exists some individual a satisfying ϕ, i.e. $\left\{\mathsf{a} \mid [\![\phi]\!]^{g[x\mapsto \mathsf{a}]} = 1\right\} \neq \emptyset$

c. Uniqueness: If there is an individual a satisfying ϕ, there is exactly one such individual, i.e. $\left|\left\{\mathsf{a} \mid [\![\phi]\!]^{g[x\mapsto \mathsf{a}]} = 1\right\}\right| \leq 1$

We will now argue that the existence requirement should be treated as a presupposition and the uniqueness requirement as a conventional implicature (CI).

Both presuppositions and CIs are types of projective meaning (Tonhauser et al., 2013), as they both can project in so-called *S-family contexts* (Gazdar, 1979), such as negation, *if*-clauses, and *yes/no*-questions. This means that if the trigger of the inference is embedded in such a context, the inference can be valid outside the effect of the operator that constitutes the context.

We will apply two criteria to distinguish between presuppositions and CIs that have been introduced already in Karttunen & Peters (1979) and are also discussed in Potts (2005, 2007). First, CIs obligatorily project in S-family contexts, whereas presuppositions can either project or be accommodated in the scope of the relevant operator. Second, if a presupposition projects, it needs to be satisfied for a sentence to be interpretable. A CI, on the other hand, has a truth value that is independent of that of the rest of the sentence.

In the standard definition of ι-expressions as in (13), both existence and uniqueness are expressed as definedness conditions, i.e., as presuppositions. Consequently, we are led to expect that the existence and the uniqueness requirements behave the same.

Let us turn to the first diagnostics. Horn & Abbot (2013) and Coppock & Beaver (2015) argue that existence need not project out of negation, but uniqueness always does. This is shown with the contrast in (15), from Horn & Abbot (2013: 341). In (15-a), the existence of a king of France is cancelled. In (15-b), the uniqueness of a consul of Illocutia is challenged. The example in (15-a) is rather natural, whereas (15-b) is not. This shows that existence and uniqueness project differently. Thus, approaches that glue them together, such as Elbourne (2005) and Schwarz (2009), are problematic.

(15) a. The king of France isn't bald—(because) there isn't any.

b. #The consul of Illocutia isn't bald—(because) there are two of them.

When looking at the formulations of denotation, existence, and uniqueness in (14), it becomes evident that the denotation of the expression, (14-a), is only defined if the existence requirement, (14-b), is met. Thus, existence is a pre-condition for the expression to be interpretable. The uniqueness requirement in (14-c), however, is independent. If there is no element that satisfies ϕ, the uniqueness requirement (in the sense of "at most one") is satisfied. On the other hand, if there are two or more elements that satisfy ϕ, there would be a denotation according to our definition, but the uniqueness requirement would be violated. Thus, the first two meaning components of ι-expressions in (14) are interdependent, whereas the third one is independent.[3]

3.2 The semantics of our three contexts

While we propose to reconsider ι-expressions, we largely base our analysis of the three definite contexts on existing proposals. In this subsection, we will start with unique nouns, then look at proper names, and, finally, at anaphoric definites.

We will relate our approach primarily to two recent proposals for the semantics of definites, Elbourne (2005) and Schwarz (2009). Just as these, we will adapt a version of situation semantics proposed in Kratzer (1989), which views situations as partial worlds. We will assume that there is a special variable, s_0, that is used for the situation under consideration—analogously to w_0 used in Montague grammar. All predicates have a situation argument, which we will use as the last argument of a predicate. So, instead of being a 1-place predicate, **student** is a 2-place predicate that is true of an individual a and a situation s if and only if a is a student in s.

Unique nouns Elbourne (2005) and Schwarz (2009) stress that the uniqueness of a unique noun is to be understood with respect to a referent situation, i.e., it is always situational uniqueness. In the case of so-called "globally unique nouns" such as *sun*, the situation is just taken to be considerably large. So, for the NP *the sun*, Schwarz assumes the semantics given in (16), i.e., a function that maps any reference situation (s_r) to the unique sun in that situation.

(16) the sun: $\lambda s_r.\iota x : \mathbf{sun}(x, s_r)$

We see a conceptual problem with uniqueness with respect to a particular reference situation. Instead, we think that what is at stake here is uniqueness in the common ground. This means that the communication partners know that uniqueness holds in all relevant situations.[4]

[3] See Horn & Abbot (2013) for more arguments in favor of the CI status of the uniqueness requirement. Coppock & Beaver (2015) take the predicative use of definite NPs as their starting point. They base their analysis on a distinction between the existence and the uniqueness requirement, but treat the uniqueness requirement as a presupposition, in contrast to us, who regard it as a CI.

[4] In the case of global uniques, the relevant situations would be all typical situations. This allows us to account for the global uniqueness of nouns like *pope* even if there had been untypical situations with two popes or situations with no pope.

Proper names Elbourne (2005) contains a recent approaches to the semantics of proper names. He follows Burge (1973) and proposes the analysis sketched in (17).

(17) Moseo/Moisés/moyshe (Eo/Pap/Yid): $(\iota y : (\textbf{moses}(y, s_0) \wedge y = x))$

Here, a name-predicate, **moses**, is used that is true of an individual if and only if that individual bears the indicated name. A proper name is, then, assumed to refer to an individual y that bears the name in a given situation, $\textbf{moses}(y, s_0)$, and that is co-referential with the value of some free variable, $y = x$. The value of x is either contextually instantiated or bound by some quantifier. Since Elbourne (2005) assumes a definition of ι-expressions as in (13), the name-bearing and the coreference are presuppositions.[5]

We will follow the basic insights of these approaches that proper names refer to an individual and that there is a naming predication that is a presupposition. We will deviate from them by our treatment of uniqueness as a CI. We will also assume that proper names are, basically, like unique nouns: if we use a proper name, we assume that there is a single bearer of this name in the common ground, i.e., in all situations that are relevant for the current conversation.

Anaphoric definites Elbourne (2005) assimilates anaphoric definites to unique nouns. He uses minimal situations that are established by the context that contain the antecedent of an anaphoric definite. In such a minimal situation, the anaphoric definite is, then, a situational unique. Schwarz (2009) treats anaphoric definites parallel to pronouns, emphasizing the anaphoric aspect. As a consequence, he largely ignores the descriptive content of anaphoric definites. We will use a variant of Elbourne's approach here: in the case of an anaphoric definite, uniqueness is established within a particular situation, s_0, the situation currently considered.[6]

We have shown in this section that we disagree with how uniqueness is often treated in the analysis of definite noun phrases. First, we argued that it needs to be separated from the existence requirement. Second, we argued that, depending on the type of definite context, the uniqueness requirement holds for all relevant situations (for unique nouns and proper names) or for a single, prominent situation (for anaphoric definites).

[5]The basic assumptions of this approach are also shared by Maier (2009). In Maier's analysis, a proper name introduces a discourse referent and a naming presupposition that is applied to that discourse referent. A DRT-style mechanism of presupposition accommodation allows for non-referring interpretations of proper names.

[6]So-called *bishop sentences* as in (i) are a potential problem to uniqueness-based theories of anaphoric definites. In such examples, there are two bishops given in the discourse, so the minimal situation needs to be defined carefully. Elbourne argues in Elbourne (2005) and subsequent papers that this is possible.

(i) If a bishop meets a(nother) bishop, the bishop blesses the (other) bishop.

4 Formalization of definite meaning

In this section, we will present a semantic analysis of definiteness that expresses our assumptions on uniqueness. We proceed in two steps. First, we clarify the relation among the three meaning components of a definite description as sketched above in (14). Second, we look at the various uniqueness inferences that we will need, in order to distinguish different kinds of definites contexts.

It became clear in Section 3.1 that the existence requirement of a definite noun phrase has the properties of a presupposition and the uniqueness requirement those of a CI. Rather than using a version of an ι-expression, we will split the three meaning component into three dimensions, see (18). The denotation of such an expression is an individual variable x. The presuppositional dimension contains an existential quantification binding this variable and the descriptive content ϕ. The CI content expresses the uniqueness requirement.

(18) The three-dimensional semantics of definite descriptions

a. Asserted content (AC): x

b. Presupposition (Presup): $\exists x(\phi \wedge \ldots)$

c. Conventional implicature (CI): $(\exists x\phi) \supset (\exists! x\phi)$

By distinguishing the three meaning dimensions, we can account for the differences in how the various meaning components project. The asserted content does not project at all. When a projective meaning stops being projected, it is integrated into the asserted content (see Section 5). Presuppositions can project in S-family contexts, but they need not do so. CIs obligatorily project in these contexts.

The interaction of the three meaning dimensions is shown in (19). The asserted content of the definite NP *the ambassador*, x, occurs as the second argument of the predicate **meet**. The existence presupposition is accommodated in the scope of the negation. The accommodation mechanism needs to be formulated in such a way that the variable x is bound by the existential quantifier contributed by the presupposition. Finally, the CI content is added conjunctively at the sentence level.

(19) Alex hasn't met the ambassador (because there is no ambassador).

$$\neg(\exists x(\mathbf{amb}(x, s_0) \wedge \mathbf{meet}(\mathbf{alex}, x, s_0))) \\ \wedge ((\exists x\, \mathbf{amb}(x, s_0)) \supset (\exists! x\, \mathbf{amb}(x, s_0)))$$

In (18-c) we only gave a very schematic uniqueness inference. Looking at languages with more than one definite article, Am-David (2016) arrives at three different uniqueness inferences, which are stated informally in (20): (A) uniqueness in the situation under consideration, (B) uniqueness in the common ground, and (C) uniqueness in the current universe of discourse.

(20) Three uniqueness inferences for *the N*:

a. A (Maximality): If there is an object satisfying $[\![N]\!]$ in the current situation, there is exactly one such maximal object.[7,8]

b. B (Common ground uniqueness): In every situation s in the common ground, if an object satisfying $[\![N]\!]$ exists in s, then there is exactly one such object in s.

c. C (Anaphoricity): If there is an object satisfying $[\![N]\!]$ in the current situation s, then there is exactly one such object in s that is part of the current universe of discourse.

For the three definite contexts considered in this paper, we only need to be concerned with the inferences (B) and (C): while proper names and unique nouns satisfy (B), anaphoric definites satisfy (C).

We can now provide the semantic analysis for the three definite contexts. We assume that the semantics of the relevant nouns is the same in all three languages discussed in this paper, irrespective of the way in which they mark definiteness. We will start with unique nouns. The semantics of the noun *sun* in our three languages is given in (21). We use the generic quantifier, **Gn**, to quantify over all relevant situations from the common ground.

(21) *suno/solo/zun* 'sun' (Eo/Pap/Yid):

a. AC: x

b. Presup: $\exists x(\mathbf{sun}(x, s_0) \wedge \ldots)$
(There is a sun in the current situation.)

c. CI: $(\ldots \wedge \mathbf{Gn}\ s\ (\exists x(\mathbf{sun}(x, s)) \supset \exists! x(\mathbf{sun}(x, s))))$
(Every situation in the common ground that contains a sun, contains exactly one sun.)

The next type of definite NP are proper names. Again, the same semantics is assumed for our three languages. Proper names are like unique nouns: They denote an individual. There is a presupposition that an individual with that name exists in the current situation. There is a CI that typically, we have only one person with that name in a given situation. This is formalized in (22).

(22) *Moseo/Moisés/moyshe* 'Moses' (Eo/Pap/Yid):

a. AC: x
(The name refers directly to an individual)

b. Presup: $\exists x(\mathbf{moses}(x, s_0) \wedge \ldots)$
(There is a person that is called Moses in the current situation)

c. CI: $(\ldots \wedge \mathbf{Gn}\ s(\exists x(\mathbf{moses}(x, s)) \supset \exists! x(\mathbf{moses}(x, s))))$
(Every situation in the common ground that contains a person called Moses contains exactly one such person.)

[7] The mentioning of a *maximal* individual in (20-a) is intended to include plural definites, as done in Sharvy (1980). In the present paper, we stick to singular definites and ignore this aspect.

[8] The *if*-part of (A) and (C) is not given in Am-David (2016) but seems to be adequate if we want to keep existence and uniqueness separated as much as possible.

In order to treat anaphoric definites, we need to introduce a new predicate, **d-acc** (**discourse-accessible**). This predicate holds of an individual a and the currently considered situation s if and only if a has been mentioned in the current discourse. The use of this predicate is illustrated in (23).

(23) *la studento/e studiante/der student* 'the student' (Eo/Pap/Yid):

a. AC: x

b. Presup: $\exists x(\textbf{stud}(x, s_0) \wedge \ldots)$
(There is a student in the current situation)

c. CI:
$(\ldots \wedge (\exists x(\textbf{stud}(x, s_0)) \supset (\exists! x(\textbf{stud}(x, s_0) \wedge x = y \wedge \textbf{d-acc}(y, s_0)))))$
(If there is a student in the current situation, there is a unique such student that is identical with some y which is accessible within the current discourse.)

The asserted content and the presuppostion in (23) are analogous to those for the other two contexts. The major difference lies in the CI content. Anaphoric definites have a uniqueness CI with respect to the currently considered situation, s_0, only. If there is an individual that satisfies the descriptive content (ϕ) of the NP, then there is exactly one individual that satisfies this description and that is coreferential with an individual y that has been mentioned in the discourse.

5 Framework

Now that the semantics of definite NPs has been introduced, we can turn to the integration into a formal theory of the syntax-semantics interface. We will assume HPSG (Pollard & Sag, 1994), enhanced with *Lexical Resource Semantics* (LRS, Richter & Sailer, 2004), a framework of underspecified semantics.

The basic idea of LRS is that words and phrases constrain the semantic representation of their utterance. These constraints specify what must occur in the representation and where. The constraints are given in the PARTS list of a sign.

For illustration, we provide an AVM for the NP *every student* in (24). The combinatorial semantic features are collected in the LRS attribute. The PARTS list is given in a short form, i.e., as a compact formula that contains meta-variables, represented by lower case Greek letters, α in (24).

(24) every student:

$$\begin{bmatrix} \text{PHON}\ \langle \textit{every, student}\rangle \\ \text{SYNS} \mid \text{LOC} \begin{bmatrix} \text{CAT} & [\text{HEAD}\ \textit{noun}] \\ \text{CONTEXT} & \begin{bmatrix}\text{SPEAKER} & \ldots \\ \text{HEARER} & \ldots\end{bmatrix}\end{bmatrix} \\ \text{LRS} \begin{bmatrix} \text{EXCONT} & \forall x(\textbf{student}(x, s_0) \supset \alpha) \\ \text{PARTS} & \langle \forall x(\textbf{student}(x, s_0) \supset \alpha)\rangle \end{bmatrix}\end{bmatrix}$$

The EX(TERNAL-)CONT(ENT) value of a sentence expresses the sentence's truth conditions. The *External Content Principle* requires that logical form of the sentence consists exactly of the material that is introduced by the lexical constraints. Technically, it is specified that all and only the elements of the PARTS list constitute the EXCONT value of a sentence, i.e. its logical form.

According to Richter & Sailer (2004), the EXCONT of a nominal element binds the discourse referent associated with the NP, x, in our example. For this reason, the quantifier $\forall x(\ldots)$ appears on both, the PARTS and the EXCONT values.[9]

So far, LRS publications did not discuss presuppositions and CIs, but there are proposals to build on: Bonami & Godard (2007) provide an encoding of CIs for evaluative adverbs using *Minimal Recursion Semantics* (Copestake et al., 2005). Hasegawa & Koenig (2011) propose a structured-meaning analysis for focus within LRS. Pollard & Sag (1994: 334) already mention that projective meaning could be treated in HPSG by applying a percolation mechanism, just as they use a SLASH percolation for extraction and Cooper storage mechanism for scope. We will adopt this suggestion.

We introduce two new list-valued features inside the LRS value of a sign: PRESUP(POSITIONS) for the presuppositions and CI for the CI content. The elements of PRESUP and CI also occur on the PARTS list. We assume special percolation principles for the new features, which are given in a preliminary form in (25) and (26). The percolation and retrieval principle for PRESUP states that presuppositions percolate unless they are retrieved in the scope of some appropriate operator.

(25) Percolation and retrieval for PRESUP:
In each phrase: All elements from the daughters' PRESUP lists are on the mother's PRESUP list unless the phrase is a clause and they appear in the clause's EXCONT value. In the latter case, they occur in the scope of some appropriate semantic operator.

CIs can only be retrieved in utterance-like contexts. Consequently, the percolation and retrieval mechanism in (26) says that elements of the CI list percolate unless they are retrieved at some utterance-like phrase (including indirect speech).

(26) Percolation and retrieval for CI:
In each phrase: All elements from the daughters' CI lists are on the mother's CI list unless the phrase is a matrix or an embedded utterance and they appear in the phrase's EXCONT value. In the latter case, they must occur in the immediate scope of some speech-act operator.

For illustration, consider the sentence in (27). We indicate the reading of the sentence, where we ignore the semantics of the proper names and only focus on the NP *the ambassador*. We will derive the reading in which the existence presupposition is accommodated inside the embedded clause, and the uniqueness CI

[9]The difference between PARTS and EXCONT will become more visible in later examples, where it will be clear that the PARTS value is a list, whereas the EXCONT value is always a single formula.

projects all the way up. In (28), we provide the three-dimensional encoding of the semantics of the definite NP *the ambassador*.

(27) [Context: There is no ambassador.]
Chris believes that Alex has met the ambassador.
$$\mathbf{believe}(\mathbf{chris}, \lambda s_1 \exists x(\mathbf{amb}(x, s_1) \wedge \mathbf{meet}(\mathbf{alex}, x, s_1)), s_0)$$
$$\wedge \mathbf{Gn}\, s(\exists x\, \mathbf{amb}(x, s) \supset \exists! x\, \mathbf{amb}(x, s)))$$

(28) The semantics of *the ambassador*:
$$\begin{bmatrix} \text{LRS} & \begin{bmatrix} \text{PARTS} & \langle x, \boxed{1}, \boxed{2} \rangle \\ \text{PRESUP} & \langle \boxed{1}\, \exists x(\mathbf{amb}(x, s_1) \wedge \alpha) \rangle \\ \text{CI} & \langle \boxed{2}\, (\beta \wedge \mathbf{Gn}\, s(\exists x\, \mathbf{amb}(x, s) \supset \exists! x\, \mathbf{amb}(x, s))) \rangle \end{bmatrix} \end{bmatrix}$$

The semantics of the embedded clause is shown in (29). Here, the presupposition is retrieved. The PRESUP list is empty and the presupposition appears in the clause's EXCONT value.

(29) The semantics of *Alex met the ambassador*:
$$\begin{bmatrix} \text{LRS} & \begin{bmatrix} \text{EXCONT} & \exists x(\mathbf{amb}(x, s_1) \wedge \mathbf{meet}(\mathbf{alex}, x, s_1)) \\ \text{PARTS} & \langle x, \boxed{1}, \boxed{2}, \mathbf{meet}(\mathbf{alex}, x, s_0) \rangle \\ \text{PRESUP} & \langle \rangle \\ \text{CI} & \langle \boxed{2}\, (\beta \wedge \mathbf{Gn}\, s(\exists x\, \mathbf{amb}(x, s) \supset \exists! x\, \mathbf{amb}(x, s))) \rangle \end{bmatrix} \end{bmatrix}$$

The semantics of the overall sentence is given in (30). Now, the CI list is also empty. The CI is added to the overall EXCONT. This overall EXCONT consists exactly of the elements of the PARTS list.

(30) The semantics of *Chris believes that Alex has met the ambassador*:
$$\begin{bmatrix} \text{LRS} & \begin{bmatrix} \text{EXCONT} & \mathbf{believe}(\mathbf{chris}, \lambda s_1 \exists x(\mathbf{amb}(x, s_1) \wedge \mathbf{meet}(\mathbf{alex}, x, s_1)), s_0) \\ & \wedge \mathbf{Gn}\, s(\exists x\, \mathbf{amb}(x, s) \supset \exists! x\, \mathbf{amb}(x, s))) \\ \text{PARTS} & \langle x, \boxed{1}, \boxed{2}, \mathbf{meet}(\mathbf{alex}, x, s_1), \mathbf{believe}(\mathbf{chris}, \lambda s_1 \gamma, s_0) \rangle \\ \text{PRESUP} & \langle \rangle \\ \text{CI} & \langle \rangle \end{bmatrix} \end{bmatrix}$$

In this section, we have shown how our three-dimensional analysis of semantic definiteness can be integrated into LRS. Since percolating feature values is one of the major analytic techniques of HPSG and LRS, we have a technique at hands to encode projective meaning. It is important for the rest of the paper that we assume the same semantics for the nouns in the three languages discussed in Section 2. The language-specific differences will, then, be attributed to differences in the semantics of the articles and to syntactic differences.

6 Analysis of definite marking in Eo, Pap, and Yid

We can, now, look at the analysis of definite NPs in Eo, Pap, and Yid. We will provide lexical entries for nouns of various types and determiners and, in the case of Yid, of preposition-determiner combinations. Since we focus on the occurrence of an article, we will only mention the SPR value among the valence features.

We saw in Section 2 that proper names normally do not take an article in Eo, Pap, or Yid. Consequently, we specifiy the SPR list as only optionally containing a determiner.[10] We already gave the semantics of proper names in (22), which re-appears in the LRS value in the lexical entry (31).

(31) Relevant parts of the lexical entry of the name *Moseo/Moisés/moyshe*:

$$\begin{bmatrix} \text{PHON} \left\langle \textit{Moseo/Moisés/moyshe} \right\rangle \\ \text{SYNS} \mid \text{LOC} \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{SPR} \left\langle (\text{Det}) \right\rangle \end{bmatrix} \end{bmatrix} \\ \text{LRS} \begin{bmatrix} \text{PARTS} & \langle x, \boxed{1}, \boxed{2} \rangle \\ \text{PRESUP} & \langle \boxed{1} \exists x(\textbf{moses}(x, s_0) \wedge \alpha) \rangle \\ \text{CI} & \langle \boxed{2} (\beta \wedge \textbf{Gn}\, s\, (\exists x(\textbf{moses}(x, s)) \supset \exists! x(\textbf{moses}(x, s)))) \rangle \end{bmatrix} \end{bmatrix}$$

Unique nouns look like proper names with respect to their semantic representation. Therefore, in Pap, the lexical entry of a unique noun such as *solo* 'sun' would look very similar to the one in (31). In Eo and Yid, unique nouns require a determiner. In (32), we sketch the lexical entry for the noun *suno/zun* 'sun'.

(32) Relevant parts of the lexical entry of *suno/zun* 'sun' (Eo/Yid):

$$\begin{bmatrix} \text{PHON} \left\langle \textit{suno/zun} \right\rangle \\ \text{SYNS} \mid \text{LOC} \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{SPR} \left\langle \text{Det} \right\rangle \end{bmatrix} \end{bmatrix} \\ \text{LRS} \begin{bmatrix} \text{PARTS} & \langle x, \boxed{1}, \boxed{2} \rangle \\ \text{PRESUP} & \langle \boxed{1} (\exists x(\textbf{sun}(x, s_0) \wedge \alpha)) \rangle \\ \text{CI} & \langle \boxed{2} (\beta \wedge \textbf{Gn}\, s\, (\exists x(\textbf{sun}(x, s) \supset \exists! x(\textbf{sun}(x, s)))))) \rangle \end{bmatrix} \end{bmatrix}$$

An ordinary singular count noun such as *studento/studiante/student* 'student' has a lexical entry as in (33): A determiner is required and the PRESUP and CI lists are empty.

(33) Relevant parts of the lexical entry of a singular count noun (Eo, Pap, Yid):

$$\begin{bmatrix} \text{PHON} \left\langle \textit{studento/studiante/student} \right\rangle \\ \text{SYNS} \mid \text{LOC} \begin{bmatrix} \text{VAL} \begin{bmatrix} \text{SPR} \left\langle \text{Det} \right\rangle \end{bmatrix} \end{bmatrix} \\ \text{LRS} \begin{bmatrix} \text{PARTS} & \langle x, \textbf{student}(x, s_0) \rangle \\ \text{PRESUP} & \langle \rangle \\ \text{CI} & \langle \rangle \end{bmatrix} \end{bmatrix}$$

We can now consider the lexical entries of the definite articles. We showed in

[10] All lexical entries in this paper are considerably simplified. In particular, we would need to allow for additional semantic material in the definite description, as in (4). We would write, for example, $\exists x(\gamma[\textbf{moses}(x, s_0)] \wedge \alpha)$ in the PRESUP value. There, γ is a meta-variable over formulæ and $\gamma[\textbf{moses}(x, s_0)]$ describes a formula γ that contains $\textbf{moses}(x, s_0)$ and possibly other material. There could, then, be a condition that the SPR list is empty if and only if $\gamma \equiv \textbf{moses}(x, s_0)$.

Section 2 that the article *la* is used in all definite contexts in Eo except for proper names. The lexical entry of this word is sketched in (34).[11]

(34) Relevant parts of the lexical entry of the Eo definite article *la*:

$$\begin{bmatrix} \text{PHON}\ \langle \textit{la} \rangle & \\ \text{SYNS} \mid \text{LOC} \begin{bmatrix}\text{HEAD} & \textit{det}\end{bmatrix} & \\ \text{LRS} \begin{bmatrix} \text{PARTS} & \langle x, \boxed{1}, \boxed{2} \rangle \\ \text{PRESUP} & \langle \boxed{1}\, \exists x(\phi \wedge \alpha) \rangle \\ \text{CI} & \langle \boxed{2} (\beta \wedge (\exists x \phi \supset \exists! x \phi)) \rangle \end{bmatrix} & \end{bmatrix}$$

When the words from (33) and (34) combine to form the NP *la studento*, the general LRS combinatorics ensure that their referential variables are identified, x here, and that the descriptive content of the head noun, $\mathbf{student}(x, s_0)$, appears in ϕ, i.e., as part of the existence and uniqueness requirements. No situation is mentioned in (34), but, since the noun *studento* uses the situation under consideration, s_0, existence and uniqueness will be required for s_0 as well.

In Eo, the definite article is also required with unique nouns. When the definite article *la* combines with a unique noun like *suno* from (32), ϕ will be identified with $\mathbf{sun}(x, s_0)$. The noun and the article, then, have identical presuppositions. In LRS, various elements can make identical contributions to the overall logical form—a property used extensively in the LRS analysis of negative concord in Richter & Sailer (2004), for example. The NP *la suno* inherits the CIs of both the article and the noun. The noun requires uniqueness in all relevant situations, and the article requires uniqueness in the considered situation, s_0. Since the considered situation is clearly relevant for a conversation, the article's uniqueness requirement is subsumed under that of the noun. In this sense, the use of an article with a unique noun is semantically redundant.

We saw above that Pap uses its definite article *e* only with anaphoric definites. Unique nouns can be used with *e* if they are used anaphorically. The lexical entry of *e* is sketched in (35). It has the same existence presupposition as Eo *la* in (34), but it has the anaphoric uniqueness requirement as its CI. Consequently, a unique noun can only combine with the article in anaphoric definite contexts, which is why the example in (6) above is ungrammatical unless the unique noun is used explicitly as an anaphoric definite. When the noun *studiante* combines with *e*, we arrive at exactly the semantics given for anaphoric definites in (23).

(35) Relevant parts of the lexical entry of the P definite article *e*:

$$\begin{bmatrix} \text{PHON}\ \langle \textit{e} \rangle & \\ \text{SYNS} \mid \text{LOC} \begin{bmatrix}\text{HEAD} & \textit{det}\end{bmatrix} & \\ \text{LRS} \begin{bmatrix} \text{PARTS} & \langle x, \boxed{1}, \boxed{2} \rangle \\ \text{PRESUP} & \langle \boxed{1}\, \exists x(\phi \wedge \alpha) \rangle \\ \text{CI} & \langle \boxed{2} (\beta \wedge \exists x \phi \supset \exists! x(\phi \wedge x = y \wedge \mathbf{d\text{-}acc}(y, s_0)))) \rangle \end{bmatrix} & \end{bmatrix}$$

[11] For our purpose, it is not important whether it is just the head noun selecting a determiner or whether there is also selection of the noun by the determiner. Therefore, we will ignore all valence and other selectional attributes in our descriptions of determiners.

In Yid, we saw that there is a full form of the definite article but also, in some cases, a reduced form which is contracted with a preposition. It is a lexical idiosyncrasy of individual prepositions whether such a contracted form exists for some definite articles or not. The preposition *af* 'on', for example, has a contracted form with the article *dem* (masculine singular dative/accusative; neuter singular dative), but not with any other definite article. We saw above that the contracted form is only used with unique nouns, i.e., nouns that require a determiner (not proper names) and that require uniqueness for all relevant situations.

To model this distribution, we assume that the lexical entry of the full form of the definite article looks as for Eo *la* in (34). The contracted P-Det form *afn* has a lexical entry as in (36). This word is a preposition but selects for a noun that requires a determiner. The P-Det contraction has the presupposition and the CI of a unique noun. Consequently, it only combines with unique nouns. In the PARTS list in (36), we write "..." for whatever is the genuine lexical semantics of *af* 'on'.

(36) Relevant parts of the lexical entry of *afn* 'on-the' (Yid):

$$\begin{bmatrix} \text{PHON}\ \langle \textit{afn} \rangle \\ \text{SYNS} \mid \text{LOC} \begin{bmatrix} \text{HEAD}\ \textit{prep} \\ \text{VAL} \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \left\langle \text{N} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{NUM}\ \textit{sg} \\ \text{GEN}\ (\textit{masc}\ \textbf{or}\ \textit{neutr}) \end{bmatrix} \\ \text{VAL SPR}\ \langle \text{Det} \rangle \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \\ \text{LRS} \begin{bmatrix} \text{PARTS} & \langle x, \boxed{1}, \boxed{2}, \ldots \rangle \\ \text{PRESUP} & \langle \boxed{1}\ \exists x(\phi \wedge \alpha) \rangle \\ \text{CI} & \langle \boxed{2}\ (\beta \wedge \mathbf{Gn}\ s\ (\exists x\phi \supset \exists! x\phi)) \rangle \end{bmatrix} \end{bmatrix}$$

The lexical entry in (36) guarantees that the contracted form P-Det only occurs with unique nouns. We do not restrict the non-contracted form of the preposition, because it can occur with proper names, with feminine unique nouns, and with anaphoric definites.

With this characterization, both the contracted form *afn* and the form *af dem* are compatible with unique nouns. We find a clear preference for the contracted form with plain unique nouns. If a unique noun occurs with a post-nominal PP, the non-contracted form is preferred.[12] This suggests that (36) should also include a constraint on the syntactic or the semantic complexity of the nominal complement. In addition, a performance strategy might have to be evoked: Since the contracted form is more restricted, and, thus, less ambiguous, it is preferred over the non-contracted form whenever possible.

The basic assumption of the present analysis is that the semantics of definiteness is encoded in the same way in the three languages that we have been addressing. Consequently, the semantic specifications for proper names, unique nouns, and

[12]In the consulted Yid corpus, CMY, there is no occurrence of the contracted form *tsum meylekh fun* 'to.the king of', but 48 occurrences of the non-contracted *tsu dem meylekh fun* 'to the king of'.

ordinary nouns are the same in all three languages. However, the languages differ with respect to their inventory of articles: Eo and Yid have very general definite articles. Pap has an article that is restricted to anaphoric definites, and Yid has P-Det contractions that are restricted to unique nouns. In Eo and Yid, unique nouns require an article syntactically, even though they already carry definite meaning (i.e., an existence and uniqueness requirement). This can easily be captured in LRS, as LRS allows for identical semantic contributions.

7 Conclusion

In this paper, we have implemented the theory of definiteness from Am-David (2016) in HPSG and applied it to three languages, namely Esperanto, Papiamentu, and Yiddish. We have argued for a three-dimensional analysis of definiteness, which distinguishes between denotation, presupposition, and conventional implicature. All definite contexts that we considered make the same existential presupposition, but they differ with respect to their uniqueness requirements: while proper names and unique nouns require uniqueness in all relevant situations, anaphoric definites require uniqueness only in the currently considered situation.

The question of whether or not a determiner is required syntactically is orthogonal to this semantic difference. The definite article in Eo is neutral with respect to the type of definite. Therefore, it is the syntax of the head noun that determines whether an article is present or not. The definite article in Pap is semantically restricted to anaphoric definites. Therefore, its use is constrained both syntactically and semantically. In Yid, the contracted P-Det combinations are semantically constrained, whereas the full form is only constrained syntactically and by the contrast to the contracted form.

We see the following differences between our approach and previous studies on definites: First, the separation of existence and uniqueness allows for an empirically more adequate formal modelling of definiteness. It also helps us to identify the differences between various definite contexts, which we relate to differences in the uniqueness requirement. Second, while the literature often concentrates on either the determiner (Hawkins, 1991; Elbourne, 2005; Schwarz, 2009) or the noun (Löbner, 2011), we can locate definite meaning in both the article and the noun if required. Third, we do not need to assume phonologically empty determiners for a comparable analysis across several languages, as done, for example, in Kester & Schmitt (2007) for Pap, because we can associate definiteness directly with lexical nouns. Fourth, whereas Schwarz (2009) has extensively discussed the meaning of P-Det combinations in German, we have tried to sketch a more complete picture for the similar cases in Yid, also covering the combination of a full article with a unique noun.

Our HPSG implementation of the theory of definiteness relies on standard mechanisms of the theory: different types of projective meaning are encoded by individual feature percolation principles. This technique has been the basis of the HPSG analyses of unbounded dependencies, quantifier scope, and other phenomena (Pollard & Sag, 1994). Our analysis of the semantically redundant—but syntactically required—use of a definite article with a unique noun exploits another basic analytic device of HPSG: token identity. Identities have been used in HPSG for Binding Theory and agreement (Pollard & Sag, 1994) and, in LRS, for negative concord and other phenomena of semantic concord.

The present paper is, admittedly, programmatic and leaves room for elaboration and extensions. Natural extensions would be to include more languages from the typological discussion in Am-David (2016), but also to adapt our analysis to languages without a definite article, such as Polish. Hawkins (1991) saw the main difference between definite articles and demonstrative determiners in the fact that demonstratives require uniqueness with respect to the sensually perceivable domain, whereas definites work on the domains of discourse. Our approach provides a good starting point for an implementation of this idea.

References

Am-David, Assif. 2014. *A semantic typology of definiteness. The crosslinguistic distribution of definiteness markers*. Saarbrücken: Lambert.

Am-David, Assif. 2016. Semantic typology of multiple definite articles. Manuscript Goethe University Frankfurt a.M. Copy available from the author.

Bonami, Olivier & Danièle Godard. 2007. Parentheticals in underspecified semantics: The case of evaluative adverbs. *Journal of Research on Language and Computation* 5(4). 391–413. Special Issue on Semantic Compositionality.

Burge, Tyler. 1973. Reference and proper names. *The Journal of Philosophy* 70. 425–439.

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan A. Sag. 2005. Minimal Recursion Semantics: An introduction. *Journal of Research on Language and Computation* 3. 281–332.

Coppock, Elizabeth & David Beaver. 2015. Definiteness and determinacy. *Linguistics and Philosophy* 38. 377–435.

Elbourne, Paul. 2005. *Situations and individuals*. Cambridge, Mass.: MIT Press.

Gazdar, Gerald. 1979. *Pragmatics: Implicature, presupposition and logical form*. New York: Academic Press.

Hasegawa, Akio & Jean-Pierre Koenig. 2011. Focus particles, secondary meanings, and Lexical Resource Semantics: The case of Japanese *shika*. In Stefan Müller (ed.), *Proceedings of the 18th international conference on Head-*

Driven Phrase Structure Grammar, University of Washington, 81–101. http://cslipublications.stanford.edu/HPSG/2011/hasegawa-koenig.pdf.

Hawkins, John A. 1991. On (in)definite articles: implicatures and (un)grammaticality prediction. *Journal of Linguistics* 27(2). 405–442. doi:10.1017/S0022226700012731.

von Heusinger, Klaus. 2010. Zur Grammatik indefiniter Eigennamen. *Zeitschrift für germanistische Linguistik* 38(1). 88–120.

Horn, Laurence R. & Barbara Abbot. 2013. <the, a>: (in)definiteness and implicature. In William P. Kabasenche & Matthew H. Michael O'Rourke, Slater (eds.), *Reference and referring*, vol. 10 Topics in Contemporary Philosophy, 325–355. Cambridge, Mass.: MIT Press.

Kalocsay, Kálmán & Gaston Waringhien. 1985. *Plena analiza gramatiko de Esperanto*. Rotterdam: Universale Esperanto Asocio 5th edn. http://luisguillermo.com/PAG/Plena_Analiza_Gramatiko_(K._Kalocsay,_G._Waringhien).pdf.

Karttunen, Lauri & Stanley Peters. 1979. Conventional implicature. In C. Oh & D. Dinneen (eds.), *Presupposition*, vol. 11 Syntax and Semantics, 1–56. New York: Academic Press.

Katz, Dovid. 1987. *Grammar of the Yiddish language*. London: Duckworth. http://dovidkatz.net/dovid/PDFLinguistics/2-1987-Grammar-Yiddish.pdf.

Kester, Ellen-Petra & Cristina Schmitt. 2007. Papiamentu and Brasilian Portuguese: A comparative study of bare nominals. In Marlyse Baptista & Jacqueline Guéron (eds.), *Noun phrases in creole languages. A multi-faceted approach*, vol. 31 Creole Language Library, 107–143. Amsterdam, Philadelphia: Benjamins.

Kouwenberg, Silvia. 2013. Papiamentu structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of pidgin and creole language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/47.

Kratzer, Angelika. 1989. An investigation of the lumps of thought. *Linguistics and Philosophy* 12. 607–653.

Löbner, Sebastian. 2011. Concept types and determination. *Journal of Semantics* 28. 279–333. doi:10.1093/jos/ffq022.

Maier, Emar. 2009. Proper names and indexicals trigger rigid presuppositions. *Journal of Semantics* 26. 253–315. doi:10.1093/jos/ffp006.

Mark, Yudel. 1978. *Gramatik fun der yidisher klal-shprakh/ Grammar of Standard Yiddish*. New York: Congress for Jewish Culture.

Pollard, Carl & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago and London: University of Chicago Press.

Potts, Christopher. 2005. *The logic of conventional implicatures* Oxford Studies in Theoretical Linguistics. Oxford: Oxford University Press.

Potts, Christopher. 2007. Into the conventional-implicature dimension. *Philosophy Compass* 4(2). 665–679.

Richter, Frank & Manfred Sailer. 2004. Basic concepts of lexical resource semantics. In Arne Beckmann & Norbert Preining (eds.), *Esslli 2003 – course material i*, vol. 5 Collegium Logicum, 87–143. Vienna: Kurt Gödel Society Wien.

Schwarz, Florian. 2009. *Two types of definites in natural language*: University of Massachusetts, Amherst dissertation.

Sharvy, Richard. 1980. A more general theory of definite descriptions. *The Philosophical Review* 89. 607–623.

Tonhauser, Judith, David Beaver, Craige Roberts & Mandy Simons. 2013. Toward a taxonomy of projective content. *Language* 89(1). 66–109.

van Oostendorp, Mark. 1993. Aangeboren principes en kunstmatige taal. *Gramma/TTT, tijdschrift voor taalkunde* 163–178. Accessible via http://helikon.ubn.ru.nl.proxy.ub.uni-frankfurt.de/tijd/g/gram/.

van Putte, Florimon & Igma van Putte-de Windt. 1992. *Dòsplak'i boka. Papiaments beginnerscursus*. Zutphen: Walburg Pers.

van Putte, Florimon & Igma van Putte-de Windt. 2014. *Grammatica van het Papiaments. Vormen en communicatieve strategiën*. Zutphen: Walburg Pers.

Wennergen, Bertilo. 2016. *Plena manlibro de Esperanta gramatiko*. http://bertilow.com/pmeg/.

Zamenhof, Ludwik Lejzer [Dr. Esperanto]. 1887. *Meždunarodnyj jazyk*. Warsaw: Kelter. http://dfiles.eu/files/nlpeafkuh. Reproduced 2004, Moscow: Impeto.

Peripheral ellipsis and verb mismatch

Aoi Shiraïshi
LLF, University Paris Diderot

Anne Abeillé
LLF, University Paris Diderot and IUF

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 662–680

Keywords: French, English, syncretism, Right-node raising, ellipsis

Shiraïshi, Aoi & Anne Abeillé. 2016. Peripheral ellipsis and verb mismatch. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 662–680. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.34.

Abstract

Right-node raising is usually set apart from other elliptical constructions for imposing a strict identity condition between the omitted and the peripheral elements. Since Pullum & Zwicky (1986), it is assumed that only syncretic forms may resolve a feature conflict between the two conjuncts (*I certainly will and you already have set the record straight.*). We present an empirical study of RNR with final verb in English and French that shows that verb mismatch does occur in corpora with and without syncretic forms, i.e. that syncretism does not appear to play a role. We present an acceptability judgement task on French that confirms this hypothesis. We therefore propose a new HPSG analysis of RNR that is based on sharing LID features and not morphophonological forms.

1 Peripheral Ellipsis and Syncretism

Peripheral ellipsis, or Right-node raising, is an elliptical construction where the right-peripheral elements of two or more clauses are shared. The construction was first mentioned by Ross (1967) (1a). It has been documented for many languages (Haspelmath (2007)) including French (1b).

(1) a. John liked and Mary disliked *the book*.
b. J' ai eu à traiter et je traite encore *un certain nombre de*
I have had to treat and I treat still a certain number of
dossiers de ce type.
cases of this kind.
'I had to deal with and I still deal with a certain number of cases of this kind.' (Abeillé & Mouret (2010))

Peripheral ellipsis is known for imposing stricter identity conditions between the elided and the peripheral material than other elliptical constructions. Syntactic mismatches occur when the elided and the peripheral material do not have the same syntactic features. VP ellipsis (2a), gapping (2b) and pseudo-gapping (2c) allow for tense mismatch or agreement mismatch. The examples in (2) show a verb form mismatch between the missing element and the antecedent.

(2) a. I haven't done it yet, but I will ~~do it~~. (VP ellipsis)
b. I want to stay and Mary ~~wants~~ to leave. (Gapping)
c. We'll let you know if it deals with the heat and humidity as well as it did ~~deal with~~ the frigid slop. (Pseudogapping, Miller (2014))

†This work was supported by IUF, strand 2 of the LabEx Empirical Foundations of Linguistics (ANR-10-LABX-0083), and a doctoral grant from Sorbonne Paris Cité to A. Shiraïshi. We thank for their help R. Chaves, B. Crysmann, B. Hemforth, P. Miller, the Headlex reviewers, and the Headlex audience, esp. T. King, L. Sadler, S. Yatabe for their comments.

Previous theories do not assume the existence of syntactic mismatches in peripheral ellipsis (Beavers & Sag (2004), Chaves (2014)), be they verb form mismatch (3a) or agreement mismatch (3c).

(3) a. *I like ~~*playing guitar*~~ and I will *play guitar*.
b. I like playing guitar and I will.
c. *Paul saved ~~*himself*~~, but Mary didn't save *herself*.
d. Paul saved himself but Mary didn't.

Note that the VP ellipsis counter parts (3b)(3d) are grammatical. Einsenberg (1973)) observes that in German, the final verb form in coordinated subordinate clauses must be the same (4a), but that "a difference in person can be sometimes compensated for by the identity of the phonological shape of the verbs": in (4c) *kaufen* can be a 1st person or 3rd person plural.

(4) a. weil Hans Bier und Franz Milch trinkt...
because Hans beer and Franz milk drink.3SG
'because Hans drinks beer and Franz milk...'
b. *weil ich Bier und du Milch trinkst/trinke...
because I beer and you milk drink.2SG/1SG
c. weil wir das Haus und die Mullers den Garten kaufen...
because we the house and the Mullers the garden buy.1PL/3.PL
'because we are buying the house and the Mullers the garden...'

Pullum & Zwicky (1986) consider Right-Node Raising as a special case of factorable coordination. They confirm that syntactic conflicts between the conjuncts may be resolved by phonological identity. For example in English, the verb *are* is acceptable in (5b) because it has values consistent with both subjects.

(5) a. *Either they or I are/am/is going to have to go.
b. Either they or you are going to have to go.
c. *At present the project managers, but in the past the executive directors, set the research priorities. (Pullum & Zwicky (1986))

However, not all syncretic forms may be appropriate in such contexts. In (5c), the present and past forms of *set* are the same, but phonological identity is not sufficient, at least for some speakers. The tense feature in (5c) is meaningful and not syntactically imposed, whereas the person feature is triggered by agreement (5b). Pullum & Zwicky (1986) thus conclude that for a syntactic feature conflict to be resolved in a factorable coordination, the feature value must be "syntactically imposed on the factor", and "a phonological form is available [...] which is ambiguous between these values". Syncretic forms are thus an exception to the Principle of phonology-free syntax.

They have also been considered as a difficult challenge for unification-based grammars (Ingria (1990)) if underspecified features (PERS=2/3) have to be resolved when unification takes place: if *are* is resolved PERS=2 when combining with *you*, how can it be resolved PERS=3 and combine with *they* ? Both LFG (Dalrymple & Kaplan (2000), Dalrymple et al. (2009)) and HPSG (Levy & Pollard (2002), Sag (2003)) have provided formal solutions, assigning syncretic forms a specific value, for example PERS= $\{2, 3\}$, instead of an underspecified one.

However, we have found cases without phonological syncretism in English and French peripheral ellipsis. This is problematic for RNR analyses assuming strict identity condition as well as for analyses giving special status to syncretic forms.

This paper is structured as follows. Section 2 provides corpus data which show the existence of verbal sharing without phonological identity in English peripheral ellipsis. Section 3 presents both corpus data and an acceptability judgment test which shows that syncretic forms do not have a special status in French peripheral ellipsis. Section 4 suggests that mismatch resolution may be due to closest conjunct agreement, and compares peripheral ellipsis with lexical coordination, and section 5 presents our HPSG analysis.

2 Verbal Mismatch in English Peripheral Ellipsis

Peripheral ellipsis, or Right-node raising (RNR), is exemplified in (6).

(6) a. John likes ~~*bananas*~~ but Mary dislikes *bananas*.
b. She learns how to relax them to accept ~~*the entering object*~~ -instead of contracting them to repel- *the entering object*.
(Brown Corpus, Bîlbîie (2013))

As noted by Hudson (1976), Williams (1990), it is not restricted to coordination, as in (6b). It may also apply to non maximal constituents (7a) and even to wordparts (7b) (Chaves (2008)):

(7) a. It was a sweet ~~*dog*~~ and an intelligent *dog*. (SWB corpus)
b. These events took place in pre~~*war Germany*~~ or in post*war Germany*?

Chaves (2014) proposes a rule of backward deletion under phonological identity, with prosodic rather than syntactic constraints.

In case of peripheral verb ellipsis, Pullum & Zwicky (1986) observe the following contrast:

(8) a. *I certainly will ~~*clarify the situation*~~, and you already have *clarify/clarified the situation*.
b. I certainly will ~~*set the record straight with respect to the budget*~~ and you already have *set the record straight with respect to the budget*.

Although they note that *clarified* may be acceptable in (8a) for some speakers, they claim that only verbs such as *put*, *set*...with syncretic base and participle are fully grammatical in such environments. An informal Google search provides both examples with syncretic (9a) and non syncretic forms (9b):

(9) a. I encourage anyone who has ~~*come across my presence*~~ or who will *come across my presence* to never limit yourself. (thecashlayproject.com/post/4690385610)

b. Her publicist Max Clifford said: "I think she's going to be remembered as a young girl who has ~~*saved an awful lot of lives*~~, and who will, *save an awful lot of lives*". (news.bbc.co.uk/2/hi/entertainment/)

We conducted a corpus study, looking for coordinated clauses with a pronominal subject, or relative clauses, with conflicting auxiliaries, such as 'who has and who will', 'that will and that have'....In the COCA (520 million words), (Davies (2010)), we only found 3 relevant examples, all with non syncretic forms (10). In The English web corpus (19 billion words) (Baroni et al. (2009)), we only found 30 examples, again all with non syncretic forms (Table 1).

(10) a. The two teachers who have ~~*encouraged me the most*~~ and who continue to *encourage me the most* to follow my heart for photography are Mr. Thomas Collins and Mr. Andrew Shapiro. (PSA Youth Showcase, 2008, COCA)

b. We have ~~*persevered*~~ and we shall *persevere*, in no small measure because of the plucky brand of people true to these ideas. (USA Today Magazine, COCA)

This may be due to the fact that verbs with non syncretic base and participles outnumber by far those with syncretic forms.

Sequences	Number of occurrences (with syncretism)	Number of occurrences (without syncretism)
who have/has and/or who V + to inf	0	2
who have/has and who will + inf	0	14
who have/has or who will +inf	0	1
who have and who are - ing	0	5
who have or who are - ing	0	7
which have and which will	0	1
Total	0	30

Table 1: Verbal mismatch in RNR in English web 2013

Could cases such as (10) be analysed as cataphoric VP ellipsis as noted in Chaves (2014)? We assume that cataphoric VP ellipsis follows Langacker (1966)'s constraint on backward anaphora (11).

(11) a. $Paul_i$ came and he_i was angry.

b. *He_i came and $Paul_i$ was angry.

c. When he_i came, $Paul_i$ was angry.

Thus cataphoric VP ellipsis only occurs in a subordinate contexts (12a) and coordinations such as (12b) should be analyzed as peripheral ellipsis.

(12) a. If you can, you should come tomorrow.
b. If you're scared, you can, and you should, leave now.
(Frank Gallagher, John M Del Vecchio, *The Bremer Detail*, 2014)

3 Syntactic Mismatch in French Peripheral Ellipsis

3.1 Previous Work on French Peripheral Ellipsis

Syntactic mismatch has been reported for indefinite determiners (13a) and weak prepositions (13b) in French Peripheral ellipsis (Abeillé & Mouret (2010), Abeillé et al. (2015)).

(13) a. Il y a des langues qui ont
There CLIT are INDF.PL languages REL.SUBJ have
~~*une flexion casuelle*~~ et des langues qui n'
~~INDF.SG inflection case~~ and INDF.PL languages REL.SUBJ NEG
ont pas *de flexion casuelle*.
have NEG NPI inflection case.
'There are languages that have and languages that don't have case inflection.' (Abeillé & Mouret (2010))

b. Ce parti ne parvient pas ~~*à surmonter ses contradictions*~~,
This party NEG manages NEG ~~to overcome its contradictions~~,
voire ne souhaite pas, *surmonter ses contradictions*.
and.even NEG wishes NEG, overcome its contradictions.
'This party cannot manage, and may not even want to overcome its contradictions.' (Le Monde, Abeillé & Mouret (2010))

A positive verb form *ont* 'have' cannot take a direct object marked by *de*, which is only allowed by a negative verb form *n'ont pas* 'have not' in (13a). An infinitival complement after a verb like *parvenir* 'succeed' must be marked by *à*, which is missing in the periphery of (13b), since *souhaiter* 'wish' takes a bare infinitival complement. We thus conclude that RNR allows for the non identity of meaningless markers.

Voice mismatch has also been reported in French peripheral ellipsis (Abeillé et al. (2015), Shiraïshi et al. (2016)): in (14), the reflexive auxiliary *se sont* expects an active participle, whereas the passive auxiliary *ont été* requires a passive one.

(14) Ce pharmacien doit des explications à ceux qui se
This pharmacist owes INDF.PL explanations to those REL.SUBJ REFL
sont ~~*mobilisés pour lui*~~, ou qui ont été *mobilisés pour lui*.
AUX ~~mobilized for him~~ or REL.SUBJ have been mobilized for him.
'This pharmacist owes explanations to those who rallied to his cause, or who were rallied to it. ' (www.ipreunion.com, 2013)

Since past and passive participles are syncretic in French, as in English, this mismatch does not challenge the phonological identity constraint on meaningful elements.

Corpus data for determiner, preposition and voice mismatch have been conformed by acceptability judgement tasks (Abeillé et al. (2015)).

3.2 Verbal Mismatch in French Peripheral Ellipsis

For French, we have tested relative clause coordination with a shared verbal form, and found feature mismatch without phonological syncretism. In (15), from a spoken corpus, *qui ont* 'who have' expects a past participle *vu* whereas *qui vont* 'who will' expects an infinitival *voir*.

(15) ...une carte interactive de tous les sites de production à grande
...a map interactive of all the facilities of production at large
échelle qui ont *~~vu le jour~~*, ou qui vont *voir le jour*
scale REL.SUBJ have ~~seen the day~~, or REL.SUBJ will see the day
dans les mois qui viennent en France.
in the months REL.SUBJ come in France.
'...an interactive map of large scale production facilities that have, or that will see the day in the months to come in France.'
(France Inter, 2015/02/20)

Since we do not have large corpora annotated for ellipsis, we searched for coordinated relative clauses with conflicting auxiliaries in a web corpus (frtenten 2012, 10 billion words). We found 49 examples (table 2), out of which, only 27 (55 %) are syncretic forms.

(16) a. Parler de sujets scientifiques, des innovations qui
Talking of subject scientific, INDF.PL innovations REL.SUBJ
ont *~~impacté le quotidien du grand public~~* ou qui
have ~~impacted the dailylife of.DET.M.SG large public~~ or REL.SUBJ
vont *impacter le quotidien du grand public*.
will impact the dailylife of.DEF.M.SG large public.
'Talking about scientific topics, innovations that have or that will impact the daily life of the public.'
(http://www.cnrs.fr/centre-est)

b. Parmi les nominés, on retrouve les artistes qui ont
Among the nominees, we find the artists REL.SUBJ have
~~investi les scènes de France~~ ou qui vont *investir les scènes*
~~invested the scenes of France~~ or REL.SUBJ will invest the scenes
de France.
of France.
'Among the nominees, we find the artists who have or who will invest the French scenes.'
(www.etudiant-france.info)

(16a) shows verbal mismatch with phonological identity, between the past participle *impacté* [ɛ̃pakte] and the infinitive *impacter* [ɛ̃pakte] ‘impact’, (16b) shows verbal mismatch without identity, since the past participle *investi* [ɛ̃vɛsti] is different from the infinitive *investir* [ɛ̃vɛstir] ‘invest’. Both are from well written sites. In all cases, the feature conflict is resolved by the form required by the second conjunct.

One may wonder why the percentage of syncretic forms is much lower in English than in French: it may well be due to the lexical frequency of the relevant verbs: 90 % of French verbs belong to the 1st inflexion group (with -er infinitive and -é participle), while only about 20 English verbs have syncretic infinitival and participles (*come, cost, cut, hit, put, set*...).

Sequences	Number of occurrences (with syncretism)	Number of occurrences (without syncretism)
qui sont et/ou qui vont +inf (who are and /or who will + inf)	3	3
qui ont et/ou qui vont +inf (who have and/or who will +inf)	16	12
qui a et/ou qui va + inf (who have and/or who will +inf)	5	3
qui ont déjà et/ou qui vont + inf (who have already and/or who will + inf)	0	3
qui a déjà et/ou qui va +inf (who has already and/or who will +inf)	1	0
qui ont/a et/ou qui peuvent/peut +inf (who have/has and/or who can +inf)	1	1
qui ont et/ou qui doivent +inf (who have and/or who must +inf))	1	0
Total	27	22

Table 2: Verbal mismatch in RNR in frTenTen 2012

3.3 Testing French Data with an Experiment

We investigated whether verbal mismatch with or without phonological syncretism is acceptable in French peripheral ellipsis. We conducted an online acceptability judgment test, on the French RISC platform (http://www.risc.cnrs.fr/). 37 French native speakers from age 19 to 70, working in French Universities, rated the sentences from 0 to 10. We had 24 target items (12 items with syncretism, 12 items without), 13 control items and 24 fillers. The experimental materials are inspired from corpus examples. Target items have three conditions: ellipsis with mismatch (17a), (18a), no ellipsis (17b), (18b), ellipsis without mismatch (17c), (18c):

(17) a. Certaines agences immobilières ont déjà, ou vont bientôt
Certain agencies estate have already, or will soon
fermer leurs portes.
close their doors.
‘Some estate agencies have already, or will soon close their doors.’

b. Certaines agences immobilières ont déjà fermé leurs portes,
Certain agencies estate have already closed their doors,
ou vont bientôt les fermer.
or will soon them close.
'Some estate agencies have already closed their doors, or will soon close them.'

c. Certaines agences immobilières ont déjà, ou auront bientôt
Certain agencies estate have already, or will have,
fermé leurs portes.
closed their doors.
'Some estate agencies just have, or will have soon closed their doors.'

(18) a. Quelques électeurs vont bientôt, ou ont peut-être déjà *rejoint*
Some voters will soon, or have perhaps already joined
le centre.
the center.
'Some voters will soon, or may have already joined the center.'

b. Quelques électeurs vont bientôt rejoindre le centre, ou l' ont
Some voters will soon join the center, or it have
peut-être déjà rejoint.
perhaps already join.
'Some voters will soon join the center, or may have already joined it'

c. Quelques électeurs auront bientôt, ou ont peut-être déjà
Some voters will-have soon, or have perhaps already
rejoint le centre.
join the center.
'Some voters will soon have, or may have already joined the center.'

Control items have tree conditions: grammatical (19a), ungrammatical with wrong verb form (19b), ungrammatical with wrong preposition (19c):

(19) a. Certains commerçants ont déjà ouvert leurs magasins.
Certain shopkeepers have already opened their stores.
'Some shopkeepers have already opened their stores.'

b. *Certains commerçants ont déjà ouvrir leurs magasins.
Certain shopkeepers have already open their stores.

c. *Le syndic cherche de régler ce problème de fuite.
The trustee tries of address this problem of leakage.

Figure 1 presents the average judgments in each condition. The data was analyzed using linear mixed-effects models. Items with ellipsis and mismatch are judged slightly less acceptable than non elliptical items (mean rate 7.8) but much more acceptable than ungrammatical controls (mean rate 3.5). There is no significant difference between ellipsis with and without mismatch. Furthermore, there is no significant difference between syncretic (mean rate 7) and non syncretic (mean rate 6.9) mismatch.

The experiment suggests that syncretic forms do not have a special status in peripheral ellipsis in French and that peripheral ellipsis with verbal mismatch with and without syncretism should be integrated in the grammar.

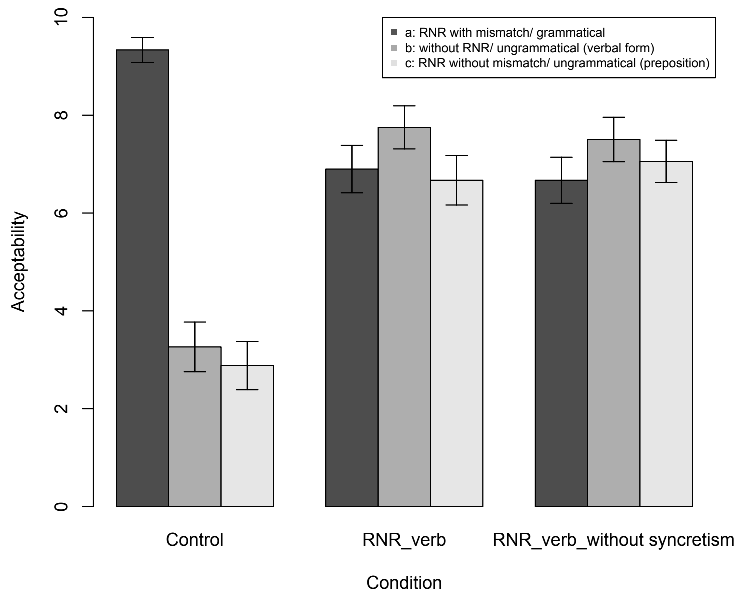


Figure 1: Results of the acceptability judgement (French). Error bars represent standard error.

4 Peripheral ellipsis and closest conjunct agreement

We hypothesize that closest conjunct agreement may interfere with peripheral ellipsis in our data. Closest Conjunct Agreement is the way some languages resolve feature conflict in nominal coordination (Corbett (1991), Sadler (1999), Yatabe (2004), Villavicencio et al. (2005)).

In French, Abeillé (2006) assumes that determiner coordination imposes identity on the shared elements, as in (20a), where *secrétaire* 'secretary' is a syncretic form for gender. But Shiraïshi (2014) has found numerous examples of determiner coordination with number or gender mismatch: in (20b), *travail* 'job' is the non syncretic plural of *travaux* 'jobs' and in (20c), *chanteuse* the non syncretic feminine of *chanteur* 'singer'.

(20) a. un ou une secrétaire
INDF.M or INDF.F secretary.M/F

b. ...pour rediriger le ou les travaux vers leur nouvelle
...to redirect DEF.SG or DEF.PL jobs to their new
destination.
destination.
'...to redirect the jobs to their new destination.'
(Gilles Lemaitre, *Backup exec pour Windows server: sauvegarde et restau*, 2007)

c. Il faut attendre que le, ou la chanteuse soit au
It must wait that DEF.M, or DEF.F singer is TO.DEF.SG
top.
top
'One must wait until the singer is at the top.'
(Bernard Tellez, *L'aube d'hiver de Barcelone*, 2010)

Looking at English auxiliary coordination, Mallinson & Blake (1981) observe that "proximity is an important factor in judgements of acceptability on agreement and government", with the following contrast (p. 202):

(21) a. ?Bill has and will underestimate the opposition.
b. Bill has and, if I'm not mistaken which I rarely am, probably always will underestimate the opposition.
c. Bill has and will upset the opposition.

They further note that "the control problem is neutralized by using a verb that has identical stem and past participle forms" as in (21c). They consider that the "desire to communicate" may win over grammatical constraints "particularly when the minor features of government and agreement are to some extent redundant" (p. 205).

We conducted a corpus search on lexical coordination of auxiliaries in English and French, using the same corpora and the same patterns as in sections 2 and 3. We found 79 examples of auxiliary coordination with verb mismatch in the COCA (table 3), compared to 3 with RNR. The percentage of syncretic forms (5 %) was higher than with RNR, but the non syncretic cases (95%) are too numerous to be considered as performance errors.

(22) a. Those demands have and will come into conflict with protection of river flows in national parks.
(Ebba Hierta, *Rivers at risk.*,1995)
b. all of them have or will become something other than Leninist...
(Journal of International Affairs, 1991)

(22) are cases of syncretism. Non syncretic cases occur both in spoken (23a) and well written genre (23b).

(23) a. And Jan is a very strong Christian woman, and so God has really— has and will carry her through this. (ABC, 1999/08/20)

b. the strength of the city's tourism industry and the recent rise in visitors has and will continue to play a major role. (USA Today, 2010/04/27)

In French, we found 238 cases of auxiliary coordination with verb mismatch (table 4). The percentage of syncretic forms (72%) was higher than for RNR, suggesting that distance may play a role.

(24) a. La réolution a ou va gagner.
The revolution has or will win.
'The revolution has or will win.' (europeecologie.eu, 2012/02/22)

b. De 2009 à 2012, de vraies révolutions ont et vont voir
From 2009 to 2012, INDF.PL real revolutions have and will see
le jour sur Internet chez BNP Paribas.
the day on Internet at BNP Paribas.
'From 2009 to 2012, real revolutions have and will see the light on the Internet at BNP Paribas.'
(epargnebourse.com, 2012/02/23)

(25) a. Sans oublier deux autres frontrunners que McCain a
Without forgetting two other frontrunners REL.OBJ McCain has
ou va rencontré/er
or will met/meet.
'Without forgetting two other frontrunners who McCain has or will meet.'
(Le Figaro, 2012/04/03)

b. Le Tea Party a et va redéfinir le paysage politique
The Tea Party has and will redefine the landscape political
américaine et forcer les Républicains à retrouver leur valeurs
american and force the Republican to regain their values
conservatrices.
conservative.
'The Tea Party has or will redefine the American political landscape or force the Republicans to regain their conservative values.'
(Le Figaro, 2012/02/24)

(24a) is an example of mismatch with syncretism (*gagné* / *gagner*) 'win' and (24b) without (*vu* / *voir*) 'see'. (25a) is an example with an innovative disjunctive spelling, showing that the writer is aware of the feature conflict. (25b) is an example with a conjunction of two shared verbal complements: a non syncretic one (*redéfini* / *redéfinir*) 'redefine' followed by a syncretic one (*forcé* / *forcer*) 'force'. The results are summarized in tables 3 and 4.

Sequences	Number of occurrences (with syncretism)	Number of occurrences (without syncretism)
will and/or have/has + psp	0	3
have/has and/or will + inf	4	54
shall and/or have/has + psp	0	0
have/has and/or shall + inf	0	0
is/are +-ing and/or have/has	0	2
have/has and/or is/are + -ing	0	16
Total	4	75

Table 3: Verbal mismatch in lexical coordination in COCA

Sequences	Number of occurrences (with syncretism)	Number of occurrences (without syncretism)
est et/ou va +inf (is and/or will +inf)	10	7
sont et/ou vont +inf (are and/or will +inf)	29	20
a et/ou va+inf (has and/or will+inf)	74	13
ont et/ou vont+inf (have and/or will+inf)	10	3
a et/ou peut+inf (has and/or can+inf)	2	1
ont et/ou peuvent +inf have and/or can +inf	2	1
a et/ou doit +inf (has and/or must +inf)	3	0
ont et/ou doivent +inf (have and/or must +inf)	2	0
Total	171	67

Table 4: Verbal mismatch in lexical coordination in frTenTen 2012

5 An HPSG Analysis of Peripheral Ellipsis

Previous analyses of peripheral ellipsis in terms of movement (Ross (1967)), multiple dominance (McCawley (1982)) or deletion do not lead one to expect verbal mismatch between the missing material and the shared material.

In HPSG, Yatabe (2001, 2012), Crysmann (2003), Beavers & Sag (2004) propose linearization-based analyses of RNR, and Chaves (2014) and Abeillé et al. (2015) a unary deletion rule. These analyses suppose phonological identity between the elided and the right peripheral elements. In LFG, Maxwell & Manning (1996) and Kuhn et al. (2010) propose a non-constituent coordination analysis, which does not take into account the possibility of mismatch either.

We follow Chaves (2014) in assuming that an RNR rule can target morphophonological units in Morphophonology (MP). According to Chaves (2014), morphophonological units can be deleted under identity of the FORM feature which lists morph form. However, the data in section 2 and 3 suggest that FORM mismatch is acceptable in English and French as shown in (26).

(26) a. ⟨We have⟩ ⟨~~persevered~~⟩ ⟨and we shall⟩ ⟨persevere⟩
b. ⟨ont déjà⟩ ⟨~~ouvert leur portes~~⟩ ⟨ou vont bientôt⟩ ⟨ouvrir leur portes⟩

The existence of RNR without phonological syncretism shows that lexeme identity plays an important role. In RNR without phonological syncretism, the past participle and the infinitive share the same lexeme. Homonyms cannot be shared as shown in (27): *bat* cannot be both the baseball instrument and the animal. *volé* cannot be shared as stolen and flown at the same time.

(27) a. #Robin swung and Leslie tamed *an unusual bat*.
(Levine & Hukari (2006))
b. #On a des avions qui ont et des accusés
We have INDF.PL planes REL.SUBJ have and INDF.PL defendants
qui n' ont pas *volé*.
REL.SUBJ NEG have NEG flown/stolen.

We assume that the feature MP includes PHON, VFORM and LID. Lexeme identity can be captured by the LID feature. The LID feature is used to individuate lexical items semantically: it includes a list of semantic frames that canonically specify the meaning of a lexeme (Sag (2012)). Identity between the infinitive and the past participle can be captured by LID. (28) shows the lexical entries for the infinitive *impacter* and the past participle *impacté* 'impact'. (29) shows the lexical entries for the infinitive *investir* and the past participle *investi* 'invest'.

(28) Infinitive *impacter* and past participle *impacté* with syncretism:

$$\begin{bmatrix} \textit{word} & \\ \text{MP} & \left\langle \begin{bmatrix} \text{PHON} & \left\langle \tilde{\varepsilon}\text{pakte} \right\rangle \\ \text{VFORM} & \textit{base} \\ \text{LID} & \left\langle \left[\text{impacter-fr}\right] \right\rangle \end{bmatrix} \right\rangle \end{bmatrix} \quad \begin{bmatrix} \textit{word} & \\ \text{MP} & \left\langle \begin{bmatrix} \text{PHON} & \left\langle \tilde{\varepsilon}\text{pakte} \right\rangle \\ \text{VFORM} & \textit{psp} \\ \text{LID} & \left\langle \left[\text{impacter-fr}\right] \right\rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

(29) Infinitive *investir* and past participle *investi* without syncretism:

$$\begin{bmatrix} \textit{word} & \\ \text{MP} & \left\langle \begin{bmatrix} \text{PHON} & \left\langle \tilde{\varepsilon}\text{v}\varepsilon\text{stiR} \right\rangle \\ \text{VFORM} & \textit{base} \\ \text{LID} & \left\langle \left[\text{investir-fr}\right] \right\rangle \end{bmatrix} \right\rangle \end{bmatrix} \quad \begin{bmatrix} \textit{word} & \\ \text{MP} & \left\langle \begin{bmatrix} \text{PHON} & \left\langle \tilde{\varepsilon}\text{v}\varepsilon\text{sti} \right\rangle \\ \text{VFORM} & \textit{psp} \\ \text{LID} & \left\langle \left[\text{investir-fr}\right] \right\rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

We also assume that the LID values of the past participle and the passive participle are the same and the mismatch between the active and the passive as shown in (14) can also be explained.

We posit a RNR unary deletion rule as follows. The rule states that the elements with the same LID value can be elided in the first conjunct. Thus the shared elements are always those expected by the second conjunct.

(30) *rnr - phrase*$\longrightarrow$

$$\begin{bmatrix} \text{MP} & \boxed{l_1}\oplus\boxed{r_1}\oplus\boxed{r_2}\oplus\boxed{r_3} \\ \text{SYNSEM} & \boxed{0} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{MP} & \left\langle \boxed{l_1}\oplus\boxed{l_2}\left\langle \begin{bmatrix}\text{PHON}\backslash\boxed{p_1}\\ \text{LID}\boxed{1}\end{bmatrix},\ldots,\begin{bmatrix}\text{PHON}\backslash\boxed{p_n}\\ \text{LID}\boxed{n}\end{bmatrix}\right\rangle \oplus \boxed{r_1}\oplus\boxed{r_2}\left\langle \begin{bmatrix}\text{PHON}\backslash\boxed{p_1}\\ \text{LID}\boxed{1}\end{bmatrix},\ldots,\begin{bmatrix}\text{PHON}\backslash\boxed{p_n}\\ \text{LID}\boxed{n}\end{bmatrix}\right\rangle\oplus\boxed{r_3}\right\rangle \\ \text{SYNSEM} & \boxed{0}\end{bmatrix}\right\rangle \end{bmatrix}$$

The MP list of the Daughter is divided into 5 sublists, which must obey prosodic constraints, which we ignore here. The first sublist $\boxed{l_1}$ is kept. The deleted list is $\boxed{l_2}$: it must comprise elements with the same LID as $\boxed{r_2}$. Note that the elements in $\boxed{l_2}$ are not preserved in the MP list of the mother. Thus the forms of the peripheral elements $\boxed{r_2}$ is always that required by the second conjunct. $\boxed{r_1}$ is the sublist before the shared elements and may comprise a coordinating conjunction.The extra $\boxed{r_3}$ list accounts for Right-node Wrapping as in (31) and can be empty. The rule in (30) works as shown in Figure 2.

(31) a. I've got friends in low places, where the whiskey drowns ~~*my blues*~~ and the beer chases *my blues* away. (Whitman (2009))

b. des églises qui se sont ~~*rattachées à Rome*~~ ou
INDF.PL churches REL.SUBJ REFL AUX ~~attached to Rome~~ or
qui ont été *rattachées à Rome* par la force
REL.SUBJ have been attached to Rome by the force
'churches who have or who have been attached to Rome by force '
(Abeillé et al. (2015))

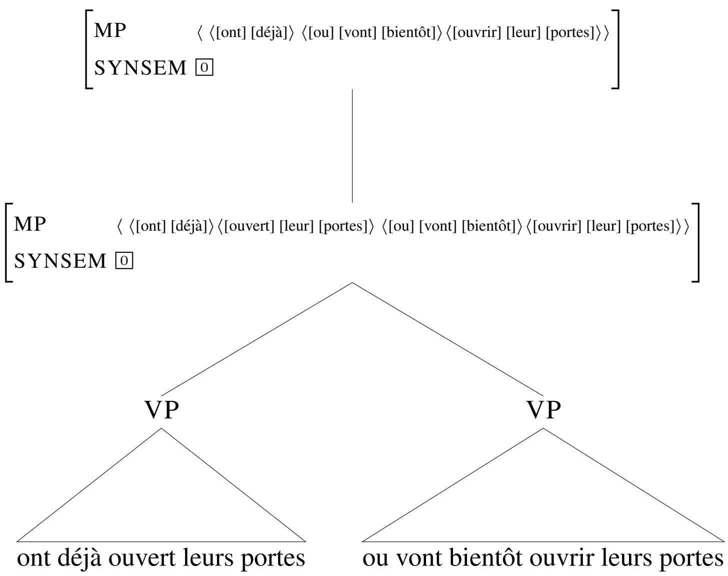


Figure 2: French RNR with non syncretic verb mismatch

6 Conclusion

In this paper, we have offered evidence that lexematic identity may be sufficient in peripheral ellipsis. Contrary to what has usually been assumed since Pullum & Zwicky (1986), corpus data indicates the existence of verbal mismatch without phonological syncretism in both French and English Right node raising. An acceptability judgment test suggests that verbal mismatch with and without phonological syncretism is as acceptable as without mismatch in French peripheral ellipsis. A possible explanation is that a principle of closest conjunct agreement overrides the identity constraint. We indeed find lexical coordination of auxiliaries with verb mismatch in both English and French. Following Chaves (2014), we adopt a deletion-based analysis of peripheral ellipsis. But we reformulate the constraint on the shared material in terms of LID feature identity.

References

Abeillé, Anne. 2006. In Defense of Lexical Coordination. In Olivier Bonami & Patricia Cabredo Hofherr (eds.), *Empirical Issues in Formal Syntax and Semantics*, 7–36. CSSP.

Abeillé, Anne, Berthold Crysmann & Aoi Shiraïshi. 2015. Syntactic Mismatch in French Peripheral Ellipsis. In Christopher Pinon (ed.), *Empirical Issues in Formal Syntax and Semantics*, Paris: CSSP.

Abeillé, Anne & François Mouret. 2010. Quelques contraintes sur les coordinations elliptiques en français. *Revue de sémantique et de pragmatique* 24. 177–207.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation* 43 (3). 209–226.

Beavers, John & Ivan A. Sag. 2004. Ellipsis and Apparent Non-Constituent Coordination. In Stefan Müller (ed.), *Proceedings of the HPSG-2004 Conference, Center for Computational Linguistics, Katholieke Universiteit Leuven*, 48–69. Stanford: CSLI Publications.

Bîlbîie, Gabriela. 2013. A quantitative study on Right Node Raising in the Penn Treebank. In *International congress of linguists*, Geneva.

Chaves, Rui. 2008. Linearization-based word-part ellipsis. *Linguistics and Philosophy* 31. 261–307.

Chaves, Rui. 2014. On the disunity of Right-Node Raising phenomena: Extraposition, Ellipsis, and Deletion. *Language* 90(4). 834–886.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Crysmann, Berthold. 2003. An Asymmetric Theory of Peripheral Sharing in HPSG: Conjunction Reduction and Coordination of Unlikes. In Gerhard Jäger, Paola Monachesi, Gerald Penn & Shuly Wintner (eds.), *Proceedings of formal grammar 2003*, 47–62.

Dalrymple, Mary & Ronald M. Kaplan. 2000. Feature Indeterminacy and Feature Resolution. *Language* 76(4). 759–798.

Dalrymple, Mary, Tracy Holloway King & Louisa Sadler. 2009. Indeterminacy by underspecification. *Journal of Linguistics* 45. 31–68.

Davies, Mark. 2010. The corpus of contemporary american english as the first reliable monitor corpus of english. *Literary and Linguistic Computing* 25(4). 447–464.

Einsenberg, Peter. 1973. A note on the identity of constituents. *Linguistic Inquiry* 4. 417–420.

Haspelmath, Martin. 2007. *Language Typology and Syntactic Description* chap. Coordination, 1–50. Cambridge: Cambridge University Press.

Hudson, Richard A. 1976. Conjunction Reduction, Gapping and Right-Node Raising. *Language* 52(3). 535–562.

Ingria, Robert. 1990. The limits of unification. In *Proceedings of the 28th Annual Meeting of the Association for computational linguistics*, 194–204. Pittsburgh, Pennsylvania, USA: Association for Computational Linguistics.

Kuhn, Jonas, Christian Rohrer & Sina Zarrieß. 2010. Right Node Raising in Parsing and Generation. In Miriam Butt & Tracy Holloway King (eds.), *The Proceedings of the LFG '10 Conference*, 311–331. Ottawa, Canada.

Levine, Robert D. & Thomas E Hukari. 2006. *The Unity of Unbounded Dependency Constructions*. Stanford University: CSLI Publications.

Levy, Roger & Carl J. Pollard. 2002. Coordination and Neutralization in HPSG. In Frank Van Eynde, Lars Hellan & Dorothee Beermann (eds.), *The Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, 221–234. Stanford: CSLI Publications.

Mallinson, Graham & Barry J. Blake. 1981. *Language typology: cross-linguistic studies in syntax*. Amsterdam: North-Holland Publishing Company.

Maxwell, John T. & Christopher D. Manning. 1996. A theory of non-constituent coordination based on finite-state rules. In Miriam Butt & Tracy Holloway King (eds.), *The Proceedings of the LFG '96 conference*, Rank Xerox, Grenoble.

McCawley, James D. 1982. Parentheticals and discontinuous constituent structure. *Linguistic Inquiry* 13. 91–106.

Miller, Philip. 2014. A corpus study of pseudogapping and its theoretical consequences. In Christopher Pinon (ed.), *Empirical Issues in Formal Syntax and Semantics*, Paris: CSSP.

Pullum, Geoffrey K. & Arnold M. Zwicky. 1986. Phonological resolution of syntactic feature conflict. *Language* 64. 751–773.

Ross, John Robert. 1967. *Constraints on variables in syntax*: MIT dissertation.

Sadler, Louisa. 1999. Non-Distributive Features and Coordination in Welsh. In Miriam Butt & Tracy Holloway King (eds.), *The Proceedings of the LFG '99 Conference*, CSLI Publications.

Sag, Ivan A. 2003. Coordination and Underspecification. In Jong-Bok Kim & Stephen Wechsler (eds.), *The Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar*, 267–291. Stanford: CSLI Publications.

Sag, Ivan A. 2012. Sign-Based Construction Grammar: An Informal Synopsis. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-based Construction Grammar* 193, 69–202. CSLI Publications.

Shiraïshi, Aoi. 2014. *Syntactic Mismatches in Right-Node Raising and Lexical Coordination in French*. University Paris Diderot MA thesis.

Shiraïshi, Aoi, Barbara Hemforth & Anne Abeillé. 2016. Voice mismatch in Right-Node Raising: Evidence from acceptability judgments and an eyetracking reading experiment. In *the 22nd AMLaP conference, Architectures and Mechanisms for Language Processing*, Bilbao.

Villavicencio, Aline, Louisa Sadler & Doug Arnold. 2005. An HPSG Account of Closest Conjunct Agreement in NP Coordination in Portuguese. In Stefan Müller (ed.), *The Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, Department of Informatics, University of Lisbon*, 427–447. CSLI Publications.

Whitman, Philip Neal. 2009. *Right-node wrapping* 235–56. Stanford, CA: CSLI Publications.

Williams, Edwin. 1990. The ATB-theory of parasitic gaps-theory of parasitic gaps. *The Linguistic Review* 6. 265–279.

Yatabe, Shûichi. 2001. The Syntax and Semantics of Left-Node Raising in Japanese. In Dan Flickinger & Andreas Kathol (eds.), *The Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar*, 325–344. Stanford.

Yatabe, Shûichi. 2004. A Comprehensive Theory of Coordination of Unlikes. In Stefan Müller (ed.), *Proceedings of the HPSG-2004 Conference, Center for Computational Linguistics, Katholieke Universiteit Leuvenenter for Computational Linguistics, Katholieke Universiteit Leuven*, 335–355. CSLI Publications.

Yatabe, Shûichi. 2012. Comparison of the Ellipsis-Based Theory of Non-Constituent Coordination with its Alternatives. In Stefan Müller (ed.), *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar, Chungnam National University Daejeon*, 453–473.

Medial left-node raising in Japanese

Shûichi Yatabe
University of Tokyo

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

Stanford, CA: CSLI Publications

pages 681–701

Keywords: right-node raising, non-constituent coordination, linearization

Yatabe, Shûichi. 2016. Medial left-node raising in Japanese. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 681–701. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2016.35.

Abstract

In this paper, it is demonstrated that there is a phenomenon that can be viewed as a mirror image of medial right-node raising and thus might be designated as *medial left-node raising*, and it is argued that the properties of this phenomenon are consistent with the predictions of the HPSG-based theory of non-constituent coordination first proposed in Yatabe (2001) and modified in later works such as Yatabe (2015).

1 Introduction

In a canonical right-node raising (RNR) construction, a string is shared by multiple phrases, typically conjuncts, and that string is pronounced at the right edge of the rightmost of those phrases, as in (1). Here and elsewhere, expressions shared by multiple phrases in this type of construction are shown in boldface.

(1) This tall and that short **student** are a couple.
(from Shen (2016))

It has been noted in the literature that a string that is shared by multiple phrases in an RNR construction is sometimes pronounced at a location other than the right edge of the rightmost of the phrases that share it. The sentence in (2) illustrates this phenomenon, which will be referred to as *medial right-node raising* in what follows. In this example, the string *boyfriend*, which is shared by two NPs (viz. *a new boyfriend* and *that ex-boyfriend you used to date*), is pronounced within the second of those NPs, but is not at its right edge.

(2) Are you talking about a new or that ex-**boyfriend** you used to date?
(from Chaves (2014))

In this paper, it will be demonstrated that there is a phenomenon which can be viewed as a mirror image of medial RNR and thus might be designated as *medial left-node raising (LNR)*, and it will be argued that the properties of this phenomenon are consistent with the predictions of the HPSG-based theory of non-constituent coordination first proposed in Yatabe (2001) and modified in later works such as Yatabe (2015).

What is going to be dealt with in this paper is not merely a descriptive issue within Japanese linguistics but is of theoretical import. As we will see in section 5 below, facts regarding medial RNR have been shown to have the potential of ruling out some theories of non-constituent coordination, but it has been unclear whether medial RNR is truly a grammatical phenomenon or a result of some kind of performance error. On the assumption that LNR and RNR are mirror images of each other, the view that medial RNR is allowed by grammar and not merely a type

†I thank the anonymous reviewers of the HeadLex2016 conference for their invaluable comments on the extended abstract and/or the near-final version of this paper.

Table 1: The 4-point scale used in the questionnaires

rating	meaning of the rating
1	‘The sentence is perfectly natural (under the intended reading).’
2	‘The sentence is slightly unnatural (under the intended reading).’
3	‘The sentence is considerably unnatural (under the intended reading).’
4	‘The sentence is completely impossible (under the intended reading).’

of performance error predicts that medial LNR will be found in a language like Japanese, which, as we will see in section 2, allows canonical, non-medial LNR. It is less clear what is predicted by the view that medial RNR is a type of performance error, but this latter view is compatible with there being no medial LNR, since there is no reason to expect patterns of performance errors to have left-right symmetry. Thus, if medial LNR does not exist, it will be possible to argue on that basis for the latter view of medial RNR. One of the implications of what follows is that it is not possible to make such an argument.

2 Left-node raising in Japanese

It is shown in Yatabe (2001) that Japanese has what might be called left-node raising constructions, i.e. structures in which a string that is shared by multiple phrases, typically conjuncts, is pronounced only once at the left edge of the leftmost of those phrases. (3) is an example of this construction, and can be viewed as the result of applying LNR to (4). The compound verb *omoidas-* ‘to recall’ consists of a noun *omoi* ‘thought’ and a verb *das-* ‘to exude’, and what has been left-node-raised in (3) is its first half, which appears at the left edge of both disjuncts in (4).

(3) [[**Omoi**dasu ka] [dasanai ka]] ga mondai da. <12, 3, 1, 0>
[[recall-PRES or] [‘exude’-NEG-PRES or]] NOM problem COP
‘Whether you recall it or you don’t is the problem.’

(4) [[Omoidasu ka] [omoidasanai ka]] ga mondai da.
[[recall-PRES or] [recall-NEG-PRES or]] NOM problem COP

The figures shown in angle brackets after (3) and other examples below are the result of questionnaire studies in which the respondents were asked to judge the acceptability of given sentences on the scale of 1 to 4 described in Table 1. Each sentence was accompanied by a description of what the intended reading of that sentence was, when the 4-point scale presented to the respondents contained the parenthesized expression in Table 1, i.e. the phrase “under the intended reading”. The order of sentences was randomized for each respondent. The four figures shown after a sentence indicate the number of respondents who chose 1, 2, 3, and 4 respectively for that sentence. A sentence for which the mean acceptability rating was R is shown throughout this paper with no symbol if $1 \leq R < 2$, with ‘?’ if

$2 \leq R < 2.5$, with '??' if $2.5 \leq R < 3$, with '?*' if $3 \leq R < 3.5$, and with '*' if $3.5 \leq R \leq 4$.

Choice of the 4-point scale is justified by the finding reported in Weskott & Fanselow (2011) that *n*-point scale data are no less informative than data gathered by the magnitude estimation method. On the other hand, the way the questionnaire results are classified into the five categories of "no symbol", "?", "??", "?*", and "*" is unavoidably arbitrary to a certain extent, and is meant to be merely a useful expedient.

The two questionnaires whose results are reported in this section were conducted in order to test the factual claims made in Yatabe (2001). In the first of the two questionnaires, there were three experimental sentences and 29 fillers (for the purpose of this paper), and 16 respondents. In the second questinnaire, there were six experimental sentences and 37 fillers (for the purpose of this paper), and 19 respondents.

Although Japanese is a so-called pro-drop language in which more types of expressions are omissible than in a language like English, part of a compound verb is generally not omissible, even when it is recoverable from the context. This is shown by the contrast between (5b) and (5c), which are both to be interpreted as responses to the question in (5a).

(5) a. Omoidashita?
recall-PAST
'Have you recalled it?'

b. Iya, omoidasanai. <12, 2, 1, 1>
no recall-NEG-PRES
'No, I don't recall it.'

c. ?? Iya, dasanai. <3, 3, 4, 6>
no 'exude'-NEG-PRES
'(Same as (5b))'

This observation lends support to the view that (3) above cannot be explained away simply as a case of context-dependent omission of part of a word.

The examples in (6) and (7) below, whose syntactic structure parallels that of (3) above, show that what licenses (3) is a mechanism of some generality, not some idiosyncratic properties of the particular lexical items involved.

(6) [[[**Sô yû toki ni**] **atari**chirasu ka] [chirasanai ka]]
[[[such occasion DAT] throw tantrums-PRES or] ['sprinkle'-NEG-PRES or]]
de, zuibun inshô ga chigaimasu yo. <17, 0, 1, 1>
INST considerably impression NOM differ-POL.PRES I tell you
'The impression you leave will be considerably different, depending on whether you throw tantrums on such occasions or you don't, I tell you.'

(7) [[[**Dasareta** **tabemono o**] **tabe**kireru ka] [kirenai
[[[serve-PASS-PAST food ACC] eat up-can-PRES or] [‘cut’-can-NEG-PRES
ka]] ga wakaremichi desu. <11, 6, 1, 1>
or]] NOM crossroads COP.POL
‘Whether you can eat up the food that you’re served or you cannot is the deciding issue.’

The compound verb *atarichiras-* ‘to throw tantrums’ in (6) consists of two verb stems, *atar-* ‘to bump’ and *chiras-* ‘to sprinkle’, and what is left-node-raised in this sentence is the first part of that compound verb and a temporal adjunct that modifies the compound verb as a whole. Example (7) involves a compound verb *tabekir-* ‘to eat up’, which consists of two verb stems, *tabe-* ‘to eat’ and *kir-* ‘to cut’; what is left-node-raised in this sentence is the first part of that compound verb and the complement of the compound verb.

As shown by the following examples, ellipsis of the first part of the compound verbs, *atarichiras-* and *tabekir-*, which appears to be involved in (6) and (7) above, is not licensed by mere pragmatic recoverability. (8b) can be, while (8c) cannot be used as an answer to the question in (8a); likewise, (9b) can be, but (9c) cannot be used as an answer to the question in (9a).

(8) a. Atarichirashita no?
throw tantrums-PAST NML
‘Did you throw tantrums?’

b. Iya, atarichirasanakatta. <18, 1, 0, 0>
no throw tantrums-NEG-PAST
‘No, I didn’t throw tantrums.’

c. ?? Iya, chirasanakatta. <2, 3, 12, 2>
no ‘sprinkle’-NEG-PAST
‘(Same as (8b))’

(9) a. Tabekireta no?
eat up-can-PAST NML
‘Were you able to eat it up?’

b. Iya, tabekirenakatta. <19, 0, 0, 0>
no eat up-can-NEG-PAST
‘No, I couldn’t eat it up’

c. ?* Iya, kirenakatta. <0, 4, 11, 4>
no ‘cut’-can-NEG-PAST
‘(Same as (9b))’

The data presented in this section point to the conclusion that Japanese allows LNR of part of a compound.

3 Medial left-node raising

In (3), (6), and (7) above, the left-node-raised string, which is shown in boldface, is at the left edge of the first of the two conjuncts that share it, and is missing from the left edge of the second conjunct. They are all instances of canonical, i.e. non-medial, LNR. If LNR is a mirror image of RNR, medial LNR, that is to say a phenomenon corresponding to (2), must be possible too; more specifically, it is expected that a left-node-raised string can be pronounced at a non-initial position within the initial conjunct as well. At the same time, it is expected to be impossible for a left-node-raised string to be missing from a non-initial position within a non-initial conjunct, since a right-node-raised string cannot be missing from a non-final position within a non-final conjunct, as shown by the following example, which is a result of right-node-raising the head noun *boyfriend* out of the two conjuncts in *that tall boyfriend you used to date or a new boyfriend*:

(10) *that tall you used to date or a new **boyfriend**

In other words, in the case of RNR, the pronunciation site, i.e. the location at which the shared string is pronounced, can be medial while the ellipsis site, i.e. the location from which the shared string is missing, cannot be medial, and it is expected that the same is true in the case of LNR, if the latter is truly a mirror image of the former.[1]

It turns out that cases of medial LNR can be found on the internet. The following, found at http://q.hatena.ne.jp/1427552124, is one such example.

(11) Ima to natte wa [[mare ni **omoi**dasu ka] [dasanai ka]] no
now [[rarely recall-PRES or] [‘exude’-NEG-PRES or]] COP
kusare-kioku desu ga...
rotten memory COP.PRES although
‘Although it is now a rotten memory that I either rarely recall or do not recall at all...’

This is a case of medial LNR, because the expression *mare ni* ‘rarely’ at the beginning of the initial conjunct is semantically incompatible with the second conjunct but precedes the left-node-raised string *omoi*. This observation lends support to the first of our hypotheses, namely the hypothesis that the pronunciation site of an LNR construction can be medial.

However, corpus evidence is hard if not impossible to obtain regarding our second hypothesis, namely the hypothesis that the ellipsis site of an LNR construction cannot be medial. Thus, two questionnaire studies were conducted in order to test the two hypotheses simultaneously.

(12) and (13) are the experimental sentences in the first of these questionnaires, in which there were two experimental sentences and 14 fillers (for the purpose of

[1] It has been claimed by some authors that the ellipsis site of RNR *can* be medial. We will come back to this point in section 5.

this paper), and 28 respondents. Both sentences involve LNR of the first part of the compound verb *omoidas-* ‘to recall’.

(12) ? [[Sukoshi wa **omoi**dasu no ka], [dasanai no ka]], ga
[[at least a little recall-PRES NML or] [‘exude’-NEG-PRES NML or]] NOM
mondai da. <10, 10, 4, 4>
problem COP
‘Whether you recall it at least a little or you don’t is the problem.’

(13) ?? [[Sukoshi mo **omoi**dasanai no ka], [sukoshi wa dasu no
[[at all recall-NEG-PRES NML or] [at least a little ‘exude’-PRES NML
ka]], ga mondai da. <8, 6, 6, 8>
or]] NOM problem COP
‘Whether you don’t recall it at all or you do at least a little is the problem.’

Since the phrase *sukoshi wa* ‘at least a little’ at the beginning of (12) is a positive polarity item and is not semantically compatible with the second conjunct, which means ‘you don’t recall’, we know that the phrase unambiguously belongs to the first conjunct. The left-node-raised expression in this example, i.e. the string *omoi-*, which is missing from the left edge of the second conjunct, follows this phrase within the first conjunct. Therefore the fact that (12) was rated as only slightly unnatural indicates that Japanese allows medial LNR.

In (13), which is also an instance of medial LNR due to the presence of the phrase *sukoshi mo* ‘at all’ at its beginning, the left-node-raised string *omoi-* is missing from a non-initial position within the second conjunct. Thus, the fact that (13) was rated as considerably unnatural if not completely impossible tends to confirm the hypothesis that an expression cannot be left-node-raised from a non-initial position within a non-initial conjunct.

The contrast between (12) and (13) is subtle, but the one-sided Wilcoxon signed-rank test showed that the difference in acceptability between (12) and (13) was statistically significant ($Z = 2.27$, $p \leq 0.05$). Furthermore, as will be explained below, the subtlety of the contrast was in fact part of the prediction of the theory being tested, i.e. the theory that the phenomenon that we are examining can be regarded as the mirror image of right-node raising.

On the one hand, medial right-node raising is often slightly awkward, as shown in Yatabe (2015) using questionnaire results involving medial RNR in Japanese. Thus, medial left-node raising, exemplified by (12), was expected to be slightly awkward as well. The less than perfectly acceptable status of sentences involving medial LNR or RNR can be interpreted as the result of the necessarily degraded structural parallelism between the conjuncts in such sentences.

And on the other hand, the example showing that the first part of the compound *omoidas-* is normally not elidable, i.e. example (5c), was in the “??” range, so the example of impossible left-node raising, i.e. (13), was predicted to be in the “??”

range, too. Sentences like (5c), (8c), (9c), and (13) are unnatural but not completely impossible probably because it is marginally possible for the verbs *das-* 'to exude', *chiras-* 'to sprinkle', and *kir-* 'to cut' to metaphorically signify something analogous to what is expressed by the verbs *omoidas-* 'to recall', *atarichiras-* 'to throw tantrums', and *tabekir-* 'to eat up' respectively. Given such an interpretation, the bimodality of the responses to (13) becomes understandable in the following way; this sentence is acceptable to those speakers who feel that the verb *das-* can metaphorically signify something analogous to what is expressed by the verb *omoidas-*, and it is not acceptable to those speakers who feel that the verb *das-* cannot be interpreted in that way.

A second questionnaire was conducted to test the same hypotheses that the first questionnaire tested using different compound verbs and relying on a different set of respondents. This questionnaire had four experimental sentences and 12 fillers (for the purpose of this paper), and 27 respondents. (14) and (15) are one of the two experimental sentence pairs in this second questionnaire. They can both be interpreted as involving LNR of a temporal modifier and the first part of the compound verb *atarichiras-*.[2]

(14) [[[**Sô yû toki ni**] sukoshi wa **atari**chirasu no ka], [chirasanai no ka]] de, zuibun inshô ga chigaimasu yo. <14, 7, 4, 2>
[[[such occasion DAT] at least a little throw tantrums-PRES NML or] [‘sprinkle’-NEG-PRES NML or]] INST considerably impression NOM differ-POL.PRES I tell you
'The impression you leave would differ considerably, depending on whether you throw tantrums at least a little on such occasions or you don't.'

(15) ?? [[[**Sô yû toki ni**] **atari**chirasu no ka], [sukoshi mo chirasanai no ka]] de, zuibun inshô ga chigaimasu yo. <4, 8, 12, 3>
[[[such occasion DAT] throw tantrums-PRES NML or] [at all ‘sprinkle’-NEG-PRES NML or]] INST considerably impression NOM differ-POL.PRES I tell you
'The impression you leave would differ considerably, depending on whether you throw tantrums on such occasions or you don't at all.'

The high rating of (14) shows that medial LNR is possible, and the low rating of (15) indicates that LNR is not possible from a non-initial position within a non-initial conjunct. The difference in acceptability between (14) and (15) was statistically significant ($Z = 3.43$, $p \leq 0.05$).

[2]These sentences can also be interpreted as involving LNR of *atari-* alone. In other words, the temporal modifier *sô yû toki ni* in these sentences can be interpreted as belonging only to the first conjunct.

Sentences (16) and (17) are the other experimental sentence pair in the second questionnaire. They can both be interpreted as involving LNR of an accusative NP and the first part of the compound verb *tabekir-*.[3]

(16) ? [[[**Dasareta** **tabemono** **o**] dônika kônika **tabe**kireru
[[[serve-PASS-PAST food ACC] somehow or other eat up-can-PRES
ka] [kirenai ka]] ga wakaremichi desu. <4, 16, 5, 2>
or] [‘cut’-can-NEG-PRES or]] NOM crossroads COP.POL
‘Whether you can somehow or other eat up the food that you’re served or you cannot is the deciding issue.’

(17) ?* [[[**Dasareta** **tabemono** **o**] **tabe**kireru ka]
[[[serve-PASS-PAST food ACC] eat up-can-PRES or]
[dô shite mo kirenai ka]] ga wakaremichi desu.
[for the life of you ‘cut’-can-NEG-PRES or]] NOM crossroads COP.POL
<0, 4, 13, 10>
‘Whether you can eat up the food that you’re served or you cannot for the life of you is the deciding issue.’

The difference in acceptability between (16) and (17) was statistically significant ($Z = 4.23$, $p \leq 0.05$). Sentence (16), which was rated as slightly unnatural but acceptable, is an instance of medial LNR, due to the presence of the positive polarity item *dônika kônika* ‘somehow or other’, which unambiguously belongs to the first conjunct but precedes part of the left-node-raised string. (17), which was rated as considerably unnatural, shows, together with (13) and (15), that LNR is not possible from a non-initial position within a non-initial conjunct.

The result of this second questionnaire was exactly as predicted by the theory, as was the result of the first questionnaire. The examples of medial LNR, i.e. (14) and (16), were expected to be slightly awkward, and they were found to be slightly awkward. Of the two examples of impossible LNR, the first one, i.e. (15), which was expected to be in the same range as (8c), i.e. the “??” range, was found to be in the “??” range, and the second one, i.e. (17), which was expected to be in the same range as (9c), i.e. the “?*” range, was found to be in the “?*” range.

Thus, the two hypotheses stated at the outset of this section were both confirmed. In an LNR construction in Japanese, the pronunciation site can be medial, but the ellipsis site cannot be medial, just as in an RNR construction. In other words, Japanese allows not only canonical, non-medial LNR but also medial LNR, which is a mirror image of medial RNR. Given these findings, we now have one fewer potential reasons to believe that instances of medial RNR are results of some kind of performance error and are in fact ungrammatical.

[3]These sentences can also be interpreted as involving LNR of *tabe-* alone.

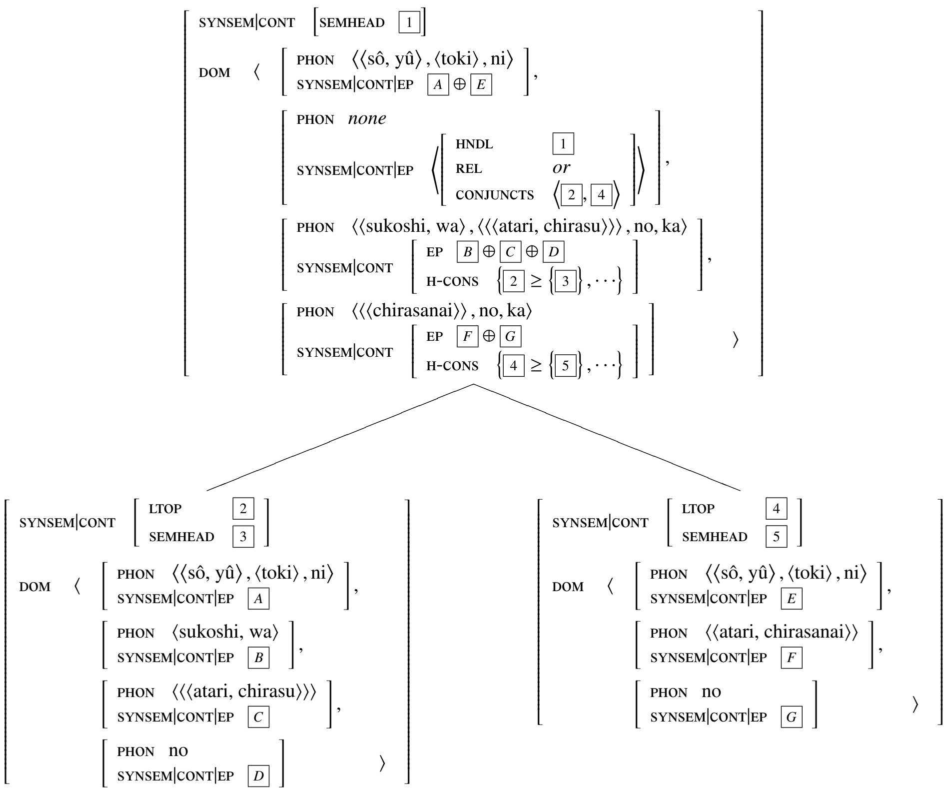


Figure 1: Part of the structure assigned to example (14) when the first half of the compound verb (namely *atari-*) and the temporal modifier (namely *sô yû toki ni*) are taken to have undergone phonological and syntactic LNR respectively

4 A linearization-based explanation

The HPSG-based theory of medial RNR and LNR proposed in Yatabe (2012) and slightly modified in Yatabe (2015) is fully compatible with the findings of this paper. According to this theory, there are two types of RNR and two types of LNR: a phonological kind of RNR and LNR that is merely prosodic ellipsis and a syntactic kind of RNR and LNR that involves merging of multiple domain objects that has the potential of affecting semantic interpretation. LNR of part of a compound must be phonological LNR, whereas LNR of things like a temporal modifier and an accusative NP may be either of the two types of LNR. Note that, pace Kubota & Levine (2015), there is nothing in this theory that is inconsistent with the long-known fact that RNR and LNR can affect semantic interpretation; Kubota and Levine's criticism of HPSG-based theories of non-constituent coordination is sound if read as a critique of the theory proposed in Beavers & Sag (2004), but not if read as an assessment of the theory under discussion, in which order domains are not mere phenogrammatical representations but principal carriers of semantic information (see Yatabe & Tam (2016)).

Figure 1 shows part of the structure assigned to example (14) in this theory when the temporal modifier *sô yû toki ni* is assumed to have undergone the syntactic type of LNR. The figure depicts the local subtree where two conjuncts, namely *sô yû toki ni sukoshi wa atarichirasu no* and *sô yû toki ni atarichirasanai no*, are conjoined by two instances of the coordinator *ka* to become a larger phrase *sô yû toki ni sukoshi wa atarichirasu no ka, chirasanai no ka*. Each node is associated with the the SYNSEM feature and the DOM feature. The value of the DOM feature is an order domain, which is a list of domain objects, each of which has the PHON feature and the SYNSEM feature. A coordinator like *ka* is assumed to be introduced into a syntactic structure by a linearization-related mechanism, and does not appear as a node in the syntactic tree (see Yatabe (2012)).

The first domain object in the order domain of the mother (pronounced "sô yû toki ni"), which represents the expression that has undergone the syntactic type of LNR, is the result of extracting the leftmost domain object from the order domain of each conjunct and merging those two domain objects, whose PHON values are identical with each other but whose SYNSEM values are not identical with each other because the two occurrences of this temporal adjunct modify different expressions. The second domain object is there to represent the meaning of disjunction, and has no phonological content. The third domain object (pronounced "sukoshi wa atarichirasu no ka") is the result of (i) compacting (i.e. turning into a single domain object) the first daughter with its leftmost domain object (which has undergone syntactic LNR) removed, and then (ii) adding *ka* as the last element of the PHON value of the newly created domain object. And the fourth domain object (pronounced "chirasanai no ka") is the result of (i) applying phonological LNR to (i.e. eliding) the string *atari* at the left edge of the domain object "atari chirasanai" in the order domain of the second daughter (which became the leftmost domain object in that order domain when the domain object "sô yû toki ni" was syntactically left-node-raised out of it), (ii) compacting the second conjunct thus altered, and then (iii) adding *ka* as the last element of the PHON value of the newly created domain object.

Sentence (14) satisfies the constraints on medial LNR that are stated in Yatabe (2012), irrespective of whether the temporal modifier *sô yû toki ni* is taken to have been (i) syntactically left-node-raised as in Figure 1, (ii) phonologically left-node-raised as in Figure 2, or (iii) part of the first conjunct alone all along rather than part of the left-node-raised string. According to Yatabe (2012), medial LNR is allowed only if all the left-node-raised expressions can be made to line up at the left edge of the order domain of the initial conjunct by removing one or more domain objects. The left-node-raised expressions in the example do line up at the left edge of the order domain of the initial conjunct if one domain object (namely the one to be pronounced "sukoshi wa") is removed, in the first two of the three scenarios above, and if two domain objects (namely "sô yû toki ni" and "sukoshi wa") are removed, in the third scenario.

Impossible cases of LNR and RNR in which the ellipsis site is medial are correctly ruled out by a generalized version of the Persistence Constraint. The Persis-

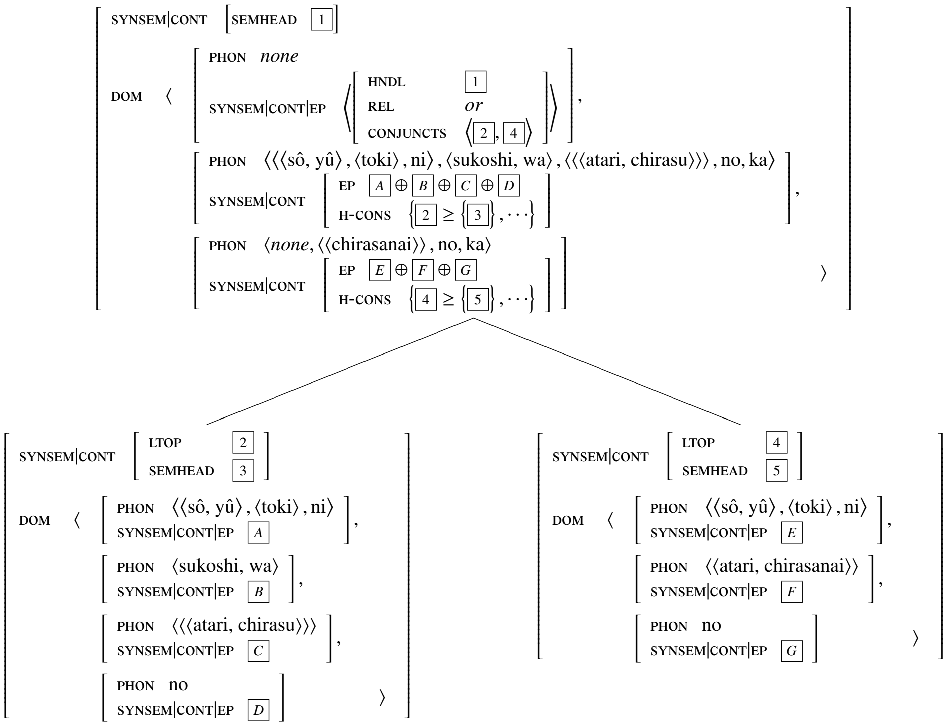


Figure 2: Part of the structure assigned to example (14) when the first half of the compound verb (namely *atari-*) and the temporal modifier (namely *sô yû toki ni*) are both taken to have undergone phonological LNR

tence Constraint as it is formulated in Kathol (1995) is shown in (18).

(18) Any ordering relation that holds between domain objects α and β in one order domain must also hold between α and β in all other order domains that α and β are members of.

This constraint can be generalized in the following way, using the term *string* to refer to any contiguous part of the PHON value of a domain object.

(19) Any ordering relation that holds between strings α and β in one order domain must also hold between α and β in all other order domains that α and β are both contained in.

What this constraint says is that the order of strings can never be reversed once it is fixed inside some order domain. Here is how this generalized version of the Persistence Constraint rules out example (10) (**that tall you used to date or a new boyfriend*), which is a result of combining (20) and (21) and right-node-raising the noun *boyfriend*.

(20) that tall boyfriend you used to date

(21) a new boyfriend

In the order domain corresponding to (20), the string *boyfriend* precedes the string *you used to date*. However, after the two phrases (20) and (21) are combined into one order domain corresponding to (10), the string *boyfriend* comes to *follow* the string *you used to date*, thus violating the generalized version of the Persistence Constraint. It is easy to see that this constraint is violated whenever the ellipsis site is medial in an RNR or an LNR construction.

In contrast, canonical RNR and LNR and medial RNR and LNR such that the pronunciation site is medial but the ellipsis site is not are consistent with the generalized version of the Persistence Constraint. Take, for example, the coordinate structure inside (2), namely *a new or that ex-boyfriend you used to date*, which is a result of combining (22) and (23) and right-node-raising *boyfriend*.

(22) a new boyfriend

(23) that ex-boyfriend you used to date

The string *a new* precedes the string *boyfriend* throughout, that is, both in the order domain of the first conjunct and in the order domain of the coordinate structure as a whole. Similarly, *that ex-* precedes *boyfriend*, and *boyfriend* precedes *you used to date* throughout, that is, both in the order domain of the second conjunct and in the order domain of the coordinate structure as a whole. Thus, there are no two strings whose order is reversed in violation of the generalized version of the Persistence Constraint. The constraint is likewise satisfied in all the other acceptable examples that have been discussed in this paper.

5 Comparison with other theories

In contrast to the theory advocated here, theories of medial RNR proposed within the framework of Categorial Grammar (CG), such as those described in Whitman (2009), Kubota (2014), and Warstadt (2015), arguably cannot be applied to the data presented in section 3. In these theories, a right-node-raised or left-node-raised string is assumed to be located outside the relevant coordinate structure. Thus, if they are to be applied to (14), for example, it will be necessary to assume that the coordinate structure here is of the form (*sô yû toki ni*) *sukoshi wa chirasu no ka chirasanai no ka*, and that the string *atari* is infixed into it when the left-node-raised string and the coordinate structure are combined. This is an unnatural assumption, and when such an assumption is made, the low acceptability of example (15) becomes a mystery, because in this analysis a degree modifier like *sukoshi mo* and *sukoshi wa* must be allowed to combine with an incomplete verb like *chirasanai* and *chirasu* to form a grammatical and hence conjoinable unit. Thus, these theories, which are shown in Yatabe (2015) to be unable to account for the full range of facts involving medial RNR,[4] have trouble dealing with medial LNR as well.

The theory proposed in Maxwell & Manning (1996) is another theory of non-constituent coordination that is unable to deal with medial RNR and LNR in an appropriate fashion. This theory is based on LFG and utilizes what the authors refer to as finite-state rules, which license phrases such as NPs and VPs that are missing their left edge and/or the right edge. Phrases that are missing the same type of string at their left and/or right edge are allowed to be coordinated with each other, and the resulting structure involving non-constituent coordination is then combined with the kinds of strings that each of the non-constituent conjuncts is missing at its left and/or the right edge. As the authors note, this way of licensing non-constituent coordination naturally does not allow medial RNR or LNR. The theory employs the HPSG-style SLASH mechanism as well because the finite-state rules alone cannot generate all cases of canonical, non-medial RNR, but this additional mechanism still does not allow the theory to generate any instance of medial RNR or LNR. It is not clear in what way the SLASH mechanism is intended to be incorporated into the LFG setup, but if the standard type of SLASH mechanism is employed, then an additional problem arises, since such a theory allows the ellipsis site to be medial, while not allowing the pronunciation site to be medial.

The properties of medial RNR and medial LNR that we have been discussing are problematic for the theory proposed in Chaves (2014) as well. Chaves argues that there are three distinct categories of grammatical phenomena that have all been referred to as right-node raising: (i) phenomena involving VP ellipsis or N′ ellipsis, (ii) phenomena involving across-the-board extraposition, which could affect semantic interpretation, and (iii) phenomena involving prosodic ellipsis, which does not affect semantic interpretation. The first of these three categories clearly should be distinguished from the rest, and will be ignored in the remainder of the discussion. At first blush, there seems to be little difference between this theory and the theory proposed in Yatabe (2001) and modified in Yatabe (2012); the latter theory also distinguishes two types of RNR, as noted in the previous section, and treats one type of RNR (the syntactic type of RNR) using the same mechanism that it uses to deal with extraposition. However, there turn out to be important differences between the two theories.

First, prosodic ellipsis that is postulated in Chaves's theory is allowed to delete a string that is not at the edge of any phrase. In other words, his theory makes the wrong prediction that not only the pronunciation site but also the ellipsis site can be medial. More specifically, the schema presented in Chaves (2014, (128)) states, in effect, that a sequence of morphophonological units *X* can be omitted if (i) it precedes another sequence of morphophonological units *Y* that has the same morph form as *X* and (ii) there are one or more morphophonological units before *X* and also between *X* and *Y*. Let us see what prediction this theory makes concerning the sentence shown in (24).

[4]In Yatabe (2015), it is shown that the theories of medial RNR proposed in Whitman (2009), Kubota (2014), and Warstadt (2015) are all incapable of dealing with an example like (2).

(24) In this country, the second Tuesday of every month, which is a weekday, is an election day, and as a result of that, the second Tuesday of every month is important.

A parenthetical like *which is a weekday* and a long subject like *the second Tuesday of every month* are likely to be phrased as separate intonational phrases (see Gussenhoven (2004, p. 287)), so (24) is likely to be pronounced as in (25), where each bracketed string is an intonational phrase.

(25) [In this country], [the second Tuesday of every month], [which is a weekday], [is an election day], [and as a result of that], [the second Tuesday of every month] [is important].

Consequently, Chaves's theory predicts incorrectly that the first occurrence of the phrase *the second Tuesday of every month* in (24) can be omitted, as in (26).

(26) *In this country, which is a weekday, is an election day, and as a result of that, **the second Tuesday of every month** is important.

Second, Chaves's theory treats the potentially meaning-changing type of RNR, which he claims is across-the-board extraposition, using the SLASH mechanism, whereas in my theory extraposition and syntactic RNR (as well as syntactic LNR) are viewed as involving dislocation of domain objects. As a result, in Chaves's theory, this second type of RNR is also predicted to allow the ellipsis site to be medial. This prediction itself is not necessarily problematic; it may not lead to overgeneration if it is assumed, for example, that only complements and right-adjoined adjuncts are extraposable in English.[5] However, the theory makes analogous predictions when applied to left-node raising, and those predictions are clearly incorrect. Consider the following example, taken from Kubota & Levine (2015).

(27) I **said different things to** Robin on Thursday and Leslie on Friday.

This sentence has a reading in which it means "What I said to Robin on Thursday was different from what I said to Leslie on Friday." If we are to generate that reading by applying the mirror-image version of Chaves's theory to LNR, we need to assume that (27) involves SLASH-mediated dislocation of a verb (*said*) and of its object (*different things*) and prosodic ellipsis (or, less plausibly, SLASH-mediated dislocation) of the preposition *to*. Given the way Chaves's theory deals with the semantics of phrases like *different things*, it is not possible to view this sentence as involving SLASH-mediated dislocation of a partial VP of the form *said different things* or *said different things to*; the phrase *different things* must be treated as having been dislocated as an independent unit. Thus, unless some additional constraint

[5]Postal (1998, p. 194–195) argues that the prediction in question is in fact correct. His argument, however, is inconclusive, because it is based on the assumption that Heavy NP Shift cannot apply to a prepositional object, an assumption that is disputed in Rochemont & Culicover (1990, p. 191).

is added to it, the theory predicts that the sentence in (28), which involves SLASH-mediated dislocation of *different things* (but not SLASH-mediated dislocation of *said* or prosodic ellipsis of *to*), is also allowed by the grammar.

(28) *I **different things** said to Robin on Thursday and said to Leslie on Friday.

Shown below is the structure that can be assigned to this example according to the theory under discussion.

(29) *I [different things]$_i$ [[said t_i to Robin on Thursday] and [said t_i to Leslie on Friday]].

In the structure in (29), the two gaps corresponding to the dislocated grammatical object are in medial positions within the VPs, but that should not be a problem according to the theory. Thus, it turns out that Chaves's theory of RNR cannot be applied to LNR in a simple fashion.

There is a further problem with the theory that this second, potentially meaning-changing type of RNR involves filler-gap dependency mediated by the SLASH mechanism. Since what has been dislocated out of a phrase using the SLASH mechanism is necessarily pronounced outside that phrase, this theory predicts that the pronunciation site of meaning-changing RNR cannot be medial whereas the pronunciation site of meaning-preserving RNR can be medial. This prediction turns out to be also incorrect. It has been shown by Kubota (2014) and Warstadt (2015) that the pronunciation site of meaning-changing RNR can be medial, just like the pronunciation site of meaning-preserving RNR. (30) is the example that Warstadt uses to establish this point.[6]

(30) Carl Philip Emmanuel Bach secretly hid or donated **every manuscript in his father's collection** to the library. (Many of the former type remain lost, while the latter are well preserved.)

In this example, the right-node-raised expression, i.e. *every manuscript in his father's collection*, is in a medial position within the second disjunct, and the interpretation of the sentence, in which the universal quantifier takes wide scope over disjunction, is different from the only possible interpretation of the corresponding sentence in which the quantifier has not been right-node-raised, viz. *Carl Philip Emmanuel Bach secretly hid every manuscript in his father's collection or donated every manuscript in his father's collection to the library*.

6 A revision of the linearization-based theory

As it turns out, the theory proposed in Yatabe (2012) also incorrectly predicts that the pronunciation site of meaning-changing RNR cannot be medial, and hence

[6]Not all native speakers of English find this example perfectly acceptable under the intended reading, although here I will assume that it is grammatical under that reading.

does not fare any better than Chaves's theory in this regard. This is because the theory contains a stipulation to the effect that medial RNR and LNR can only be the phonological type of RNR and LNR respectively. It was noted in Yatabe (2015) that this stipulation needs to be excised from the theory, but it was not stated exactly how that could be done.

The problematic stipulation is contained in the following statement, which was given in (29) of Yatabe (2012).

(31)
$$\begin{aligned}
&\text{pnr_dom}\left(\left\langle \boxed{1}, \cdots, \boxed{n} \right\rangle, \boxed{D_0}, \boxed{f}\right) \equiv \\
&\left(\boxed{A_L} \neq \langle\rangle \quad \vee \quad \boxed{A_R} \neq \langle\rangle \quad \vee \quad \boxed{B_L} \neq \langle\rangle \quad \vee \quad \boxed{B_R} \neq \langle\rangle\right) \\
&\wedge \quad \text{syn_pnr}\left(\left\langle \boxed{1} \right\rangle \bigcirc \cdots \bigcirc \left\langle \boxed{n} \right\rangle, \boxed{H}, \left\langle \boxed{l_1}, \cdots, \boxed{l_n} \right\rangle, \left\langle \boxed{r_1}, \cdots, \boxed{r_n} \right\rangle\right) \\
&\wedge \quad \text{phon_pnr}\left(\boxed{H}, \boxed{G}, \boxed{B_L}, \boxed{B_R}\right) \\
&\wedge \quad \text{totally_compact_each}\left(\boxed{G}, \boxed{F}\right) \\
&\wedge \quad \text{add_conjunction}\left(\boxed{F}, \boxed{E}, \boxed{f}\right) \\
&\wedge \quad \text{fuse_each}\left(\left\langle \boxed{l_1}, \cdots, \boxed{l_n} \right\rangle, \boxed{A_L}, \boxed{f}\right) \\
&\wedge \quad \text{fuse_each}\left(\left\langle \boxed{r_1}, \cdots, \boxed{r_n} \right\rangle, \boxed{A_R}, \boxed{f}\right) \\
&\wedge \quad \boxed{D_0} : \boxed{A_L} \oplus \boxed{E} \oplus \boxed{A_R}
\end{aligned}$$

This statement defines a relation that may or may not hold among the list of daughter signs ($\left\langle \boxed{1}, \cdots, \boxed{n} \right\rangle$), the order domain of the mother ($\boxed{D_0}$), and a phonologically empty domain object which is produced by the phrase-structure schema for coordinate structures and which carries the meaning of conjunction or disjunction ($\boxed{f}$). A local subtree in a syntactic phrase-structure tree has the option of being licensed by virtue of satisfying this constraint as opposed to other, more usual types of constraint, which license structures that do not involve any RNR or LNR.

Let us use the word *peripheral-node raising (PNR)* as a cover term for RNR and LNR. When a given local subtree is licensed by virtue of satisfying the constraint in (31), the licensed structure is going to involve some type of PNR. In (31), $\boxed{A_L}$, $\boxed{A_R}$, $\boxed{B_L}$, and $\boxed{B_R}$ denote syntactically left-node-raised material, syntactically right-node-raised material, phonologically left-node-raised material, and phonologically right-node-raised material, respectively. Notice that line 2 of (31) requires that at least one of the four tags should denote something other than an empty list.

What (31) as a whole dictates can be stated in ordinary English in the following way. Syntactic PNR deletes a list of domain objects at the right (or left, respectively) edge of each daughter (line 3 of (31)), fuses those lists of domain objects item by item to create a possibly modified list of domain objects (lines 7 and 8 of (31)), and places the resulting list of domain objects at the right (or left, respectively) edge of the order domain of the mother (line 9 of (31)). Phonological PNR simply deletes some phonological material at the right (or left, respectively) edge of non-final (or non-initial, respectively) daughters, on condition that the same phonological material is contained in the final (or initial, respectively) daughter (line 4 of (31)). Those parts of each daughter node that do not undergo syntactic or phonological PNR are totally compacted and become a single domain object (line 5 of (31)), and the newly created domain objects, each corresponding to one of the daughter nodes, are placed in the order domain of the mother (line 9 of (31)), after possibly having a coordinator added to their PHON values (line 6 of (31)).

The reader is referred to Yatabe (2012) for the definitions of the relations *syn_pnr*, *phon_pnr*, *totally_compact_each*, *add_conjunction*, and *fuse_each*, which are used in (31) above.

The last line of (31) states that the syntactically left-node-raised expressions, i.e. $\boxed{A_L}$, have to be placed at the left edge of the order domain of the mother, and that the syntactically right-node-raised expressions, i.e. $\boxed{A_R}$, have to be placed at the right edge of the order domain of the mother. This is what produces the incorrect prediction that medial RNR and LNR can only be of the phonological type. Thus, (31) needs to be replaced by (32), which is identical to (31) except in lines 3, 7, 8, and 9.

(32)
$$\begin{array}{ll}
\multicolumn{2}{l}{\mathrm{pnr_dom}\left(\left\langle \boxed{1}, \cdots, \boxed{n} \right\rangle, \boxed{D_0}, \boxed{f}\right) \equiv} \\
\multicolumn{2}{l}{\left(\boxed{A_L} \neq \langle\rangle \quad \vee \quad \boxed{A_R} \neq \langle\rangle \quad \vee \quad \boxed{B_L} \neq \langle\rangle \quad \vee \quad \boxed{B_R} \neq \langle\rangle\right)} \\
\wedge & \mathrm{syn_pnr}\left(\left\langle \boxed{1} \right\rangle \bigcirc \cdots \bigcirc \left\langle \boxed{n} \right\rangle, \boxed{H}, \left\langle \boxed{U_1} \oplus \boxed{T_1} \right\rangle \oplus \boxed{S_L}, \boxed{S_R} \oplus \left\langle \boxed{T_n} \oplus \boxed{U_n} \right\rangle\right) \\
\wedge & \mathrm{phon_pnr}\left(\boxed{H}, \boxed{G}, \boxed{B_L}, \boxed{B_R}\right) \\
\wedge & \mathrm{totally_compact_each}\left(\boxed{G}, \boxed{F}\right) \\
\wedge & \mathrm{add_conjunction}\left(\boxed{F}, \boxed{E}, \boxed{f}\right) \\
\wedge & \mathrm{fuse_each}\left(\left\langle \boxed{T_1} \right\rangle \oplus \boxed{S_L}, \boxed{A_L}, \boxed{f}\right) \\
\wedge & \mathrm{fuse_each}\left(\boxed{S_R} \oplus \left\langle \boxed{T_n} \right\rangle, \boxed{A_R}, \boxed{f}\right) \\
\wedge & \boxed{D_0} : \boxed{U_1} \oplus \boxed{A_L} \oplus \boxed{E} \oplus \boxed{A_R} \oplus \boxed{U_n}
\end{array}$$

The last line of the new definition states that the sequence of syntactically left-node-raised expressions, i.e. $\boxed{A_L}$, may be preceded by some domain objects coming from the left edge of the order domain of the first daughter, i.e. $\boxed{U_1}$, and that the sequence of syntactically right-node-raised expressions, i.e. $\boxed{A_R}$, may be followed by some domain objects coming from the right edge of the order domain of the last daughter, i.e. $\boxed{U_n}$. This allows a sentence like (30), as desired.

The proposed analysis predicts correctly that medial RNR in Japanese cannot be meaning-changing, unlike medial RNR in English. The following three sentences, taken from Yatabe (2015), illustrate this property of medial RNR in Japanese.

(33) [Kyôko wa] [Pari de] [muji no] [masshiro na], soshite [Jirô wa] [aoi
[Kyoko TOP] [Paris in] [plain COP] [pure white COP] and [Jiro TOP] [blue
moyô no] **o-sara o** [Honkon de], **sorezore kônyû shita no**
pattern COP] plate ACC [Hong Kong in] individually bought NML
desu. <11, 6, 1, 1>
COP

‘Kyoko bought pure white plates without patterns in Paris, and Jiro bought plates with blue patterns in Hong Kong, the two of them acting individually.’

(34) ?[Kyôko wa] [Pari de] [muji no] [masshiro na], soshite [Jirô wa]
[Kyoko TOP] [Paris in] [plain COP] [pure white COP] and [Jiro TOP]
[Honkon de] [aoi moyô no], [**gôkei jû-mai ijô no**] **o-sara o**
[Hong Kong in] [blue pattern COP] [in total ten or more] plate ACC
sorezore kônyû shita no desu. <4, 6, 5, 4>
individually bought NML COP
‘Kyoko bought pure white plates without patterns in Paris, and Jiro bought plates with blue patterns in Hong Kong, buying ten or more plates in total between them and the two of them acting individually.’

(35) *[Kyôko wa] [Pari de] [muji no] [masshiro na], soshite [Jirô wa]
[Kyoko TOP] [Paris in] [plain COP] [pure white COP] and [Jiro TOP]
[aoi moyô no], [**gôkei jû-mai ijô no**] **o-sara o** [Honkon de],
[blue pattern COP] [in total ten or more] plate ACC [Hong Kong in]
sorezore kônyû shita no desu. <0, 2, 4, 13>
individually bought NML COP
‘(Same as (34))’

Sentence (33) is an instance of meaning-preserving medial RNR,[7] and sentence (34) is an instance of meaning-changing non-medial RNR. Both sentences are more or less acceptable. In contrast, sentence (35) is an instance of meaning-changing medial RNR, and it is not acceptable. The constraint in (32) correctly captures this pattern. In sentence (35) and other cases of meaning-changing medial RNR in Japanese, the string that belongs only to the last conjunct (the phrase *Honkon de* in the case of (35)) but follows a syntactically right-node-raised expression (the phrase *gôkei jû-mai ijô no o-sara o* in the case of (35)) is in turn followed by another syntactically right-node-raised expression (the phrase *sorezore kônyû shita no desu* in the case of (35)). This is not permitted by the constraint in (32); according to the constraint, when a syntactically right-node-raised expression is followed by a sequence of expressions that belong only to the last conjunct, that sequence of expressions cannot in turn be followed by another syntactically right-node-raised expression.[8] Unlike (35), sentence (33) is possible because the medial RNR of the phrase *o-sara o* in this example is meaning-preserving and hence can be of the phonological type.

The effects of the constraint stated in (32) overlap extensively with those of

[7]More specifically, sentence (33) involves meaning-preserving medial RNR of the phrase *o-sara o* and meaning-changing non-medial RNR of the phrase *sorezore kônyû shita no desu*.

[8]Unlike the theory that is sketched in Yatabe (2015, section 3), the theory proposed here does not simultaneously account for the fact (noted in Yatabe (2007)) that conjuncts are scope islands in Japanese but not in English.

(19). There is probably a better, more succinct way to state this constraint that reduces that overlap. Investigation of that possibility will be left for future occasion.

7 Summary

In this paper, it has been shown that there is a phenomenon that is essentially a mirror image of medial right-node raising and thus might be designated as medial left-node raising. It has also been shown that, in both RNR and LNR, the ellipsis site cannot be medial while the pronunciation site (i.e. the surface position of the left-node-raised or right-node-raised string) can be medial. It was observed that these findings are consistent with the linearization-based theory proposed in Yatabe (2012) and modified in later works and are inconsistent with theories that are based on Categorial Grammar mechanisms and with theories that employ the SLASH mechanism to deal with RNR and LNR. In the penultimate section, a modification of the linearization-based theory was proposed that eliminates the incorrect prediction that medial RNR can only be of the phonological type.

References

Beavers, John & Ivan A. Sag. 2004. Coordinate ellipsis and apparent non-constituent coordination. In Stefan Müller (ed.), *Proceedings of the 11th international conference on Head-Driven Phrase Structure Grammar*, 48–69. Stanford: CSLI.

Chaves, Rui Pedro. 2014. On the disunity of right-node raising phenomena: Extraposition, ellipsis, and deletion. *Language* 90(4). 834–886.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Kathol, Andreas. 1995. *Linearization-based German syntax*. Columbus: Ohio State University dissertation.

Kubota, Yusuke. 2014. Medial right-node raising and multi-modal categorial grammar. Ms. Posted at http://ling.auf.net/lingbuzz/002302.

Kubota, Yusuke & Robert Levine. 2015. Against ellipsis: Arguments for the direct licensing of 'noncanonical' coordinations. *Linguistics and Philosophy* 38. 521–576.

Maxwell III, John T. & Christopher D. Manning. 1996. A theory of non-constituent coordination based on finite-state rules. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '96 conference*, Stanford: CSLI. Available at http://cslipublications.stanford.edu/LFG/.

Postal, Paul M. 1998. *Three investigations of extraction*. Cambridge, Mass.: MIT Press.

Rochemont, Michael S. & Peter W. Culicover. 1990. *English focus constructions and the theory of grammar*. Cambridge: Cambridge University Press.

Shen, Zheng. 2016. No clash constraint in nominal RNR number agreement. In Sunghye Cho (ed.), *University of Pennsylvania working papers in linguistics*, vol. 22.1, 245–253. Available at http://repository.upenn.edu/pwpl/vol22/iss1/27.

Warstadt, Alexander. 2015. Right-Node Wrapping: A combinatory account. In Yusuke Kubota & Robert Levine (eds.), *Proceedings for ESSLLI 2015 workshop 'Empirical advances in Categorial Grammar' (CG 2015)*, 183–210. Available at http://www.u.tsukuba.ac.jp/˜kubota.yusuke.fn/cg2015.html.

Weskott, Thomas & Gisbert Fanselow. 2011. On the informativity of different measures of linguistic acceptability. *Language* 87. 249–273.

Whitman, Neal. 2009. Right-node wrapping: Multimodal categorial grammar and the 'friends in low places' coordination. In Erhard Hinrichs & John Nerbonne (eds.), *Theory and evidence in semantics*, 235–256. Stanford: CSLI.

Yatabe, Shûichi. 2001. The syntax and semantics of left-node raising in Japanese. In Dan Flickinger & Andreas Kathol (eds.), *Proceedings of the 7th international conference on Head-Driven Phrase Structure Grammar*, 325–344. Stanford: CSLI.

Yatabe, Shûichi. 2007. Evidence for the linearization-based theory of semantic composition. In Stefan Müller (ed.), *Proceedings of the 14th international conference on Head-Driven Phrase Structure Grammar*, 323–343. Stanford: CSLI.

Yatabe, Shûichi. 2012. Comparison of the ellipsis-based theory of non-constituent coordination with its alternatives. In Stefan Müller (ed.), *Proceedings of the 19th international conference on Head-Driven Phrase Structure Grammar*, 454–474. Stanford: CSLI.

Yatabe, Shûichi. 2015. The semantic inertness of medial right-node raising in Japanese. In *Nihon gengogakkai dai-151-kai taikai yokô-shû (Proceedings of the 151st meeting of the Linguistic Society of Japan)*, 318–323. Kyoto: Linguistic Society of Japan. Available at http://ling.auf.net/lingbuzz/002778.

Yatabe, Shûichi & Wai Lok Tam. 2016. In defense of an HPSG-based theory of non-constituent coordination: A reply to Kubota and Levine. Ms. Posted at http://ling.auf.net/lingbuzz/003152.