Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar

University of Massachusetts Amherst

Stefan Müller, Elodie Winckel (Editors)

2023

Frankfurt/Main: University Library

Contents

Editor's note

The 30th International Conference on Head-Driven Phrase Structure Grammar (2023) took place as a hybrid conference at the University of Massachusetts Amherst and was organized by Gabriela Bîlbîie, Fabiola Henri, Manfred Sailer, and Elodie Winckel.

The conference featured 2 invited talks and 14 papers selected by the program committee (Anne Abeillé, Sascha Bargmann, Emily M. Bender, Felix Bildhauer, Olivier Bonami, Francis Bond, Gosse Bouma, Rui Chaves, Berthold Crysmann, Tony Davis, Frank van Eynde, Thomas Hoffmann, Anke Holler, Jonathan Ginzburg, Jong-Bok Kim, Jean-Pierre Koenig, Yusuke Kubota, Andy Lücking, Jakob Maché, Antonio Machicao Y Priemer, Nurit Melnik, Luis Morgado da Costa, Stefan Müller, Tsuneko Nakazawa, Joanna Nykiel, Petya Osenova, David Yoshikazu Oshima, Rainer Osswald, Gerald Penn, Frank Richter, Manfred Sailer, Sanghoun Song, Stephen Wechsler, Elodie Winckel, PC chair, Shuichi Yatabe, Eun-Jung Yoo, Olga Zamaraeva). There was a workshop on Experimental and Corpus-based Approaches to Ellipsis with one invited speaker and ten regular papers. This workshop had a separate program committee (Anne Abeillé, Gabriela Bîlbîie, PC chair, Katy Carlson, Rui Chaves, Berthold Crysmann, Jonathan Ginzburg, Andrew Kehler, Jong-Bok Kim, Jean-Pierre Koenig, Yusuke Kubota, Robin Lemke, Philip Miller, François Mouret, Joanna Nykiel, PC chair, Dan Parker, Jeffrey Runner, Shuichi Yatabe).

We want to thank the program committees for putting these nice programs together.

As in the past years the contributions to the conference proceedings are based on the five page abstract that was reviewed by the respective program committee, but there is no additional reviewing of the longer contribution to the proceedings. To ensure easy access and fast publication we have chosen an electronic format. The proceedings are published by the University Library of Goethe-Universität, Frankfurt am Main.

The proceedings include all the papers of the conference except the ones by Emily M. Bender, Paola Fritz-Huechante, Joanna Nykiel & Jong-Bok Kim, Jieun Oh & Stefan Müller, and Frank Van Eynde. Some of these papers and the workshop contributions will be published in journals.

Welsh clausal *i* and the hierarchical lexicon

Robert D. Borsley
University of Essex and Bangor University

Proceedings of the 30th International Conference on
Head-Driven Phrase Structure Grammar

University of Massachusetts Amherst

Stefan Müller, Elodie Winckel (Editors)

2023

Frankfurt/Main: University Library

pages 5–25

Keywords: Lexicon, Prepositions, Subordinate clauses

Borsley, Robert D. 2023. Welsh clausal *i* and the hierarchical lexicon. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 5–25. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2023.1.

Abstract

A number of types of Welsh subordinate clause are introduced by what looks like the preposition *i* 'to', 'for'. Earlier research has shown that there are three different lexemes here. It is not unusual for a language to have homophonous lexemes, but these lexemes share a variety of properties, and also share properties with the preposition *i*. The similarities and the differences among these lexemes can be captured if they are grouped together as four different realisations of a single 'super-lexeme' within the hierarchical lexicon.

1. Introduction

A number of types of Welsh subordinate clause are introduced by what looks like the preposition *i* 'to', 'for'. It appears with subjectless infinitives in some control sentences such as (1):

(1) Mae Heledd yn awyddus [i weld Rhiannon].
be.PRES.3SG Heledd PRED eager to see.INF Rhiannon
'Heledd is eager to see Rhiannon.'

This is shown to be a control sentence by the ungrammaticality of (2) with a dummy subject in the main clause:

(2) *Mae hi 'n awyddus [i fwrw glaw].
be.PRES.3SG she PRED eager to strike.INF rain
*'It's eager to rain.'

It also appears with subjectless infinitives in some raising sentences such as (3), which is shown to be a raising sentence by the grammaticality of (4):

(3) Mae Heledd yn mynd [i weld Rhiannon].
be.PRES.3SG Heledd PROG go.INF to see.INF Rhiannon
'Heledd is going to see Rhiannon.'
(4) Mae hi 'n mynd [i fwrw glaw].
be.PRES.3SG she PROG go.INF to strike.INF rain
'It's going to rain.'

---

* An earlier version of this paper was presented at the 27th Welsh linguistics seminar, in Bangor, in June 2023. I am grateful to Maggie Tallerman, Frank Van Eynde, Berthold Crysmann, and Jakob Maché for helpful comments, and to Howard Edwards, Peredur Webb-Davies, and Bob Morris Jones for help with the Welsh data. I alone am responsible for what appears here.

Finally, it appears with full clauses with an overt subject reminiscent of English *for-to* clauses such as (5):

(5) Disgwyliodd Heledd [i Sioned weld Rhiannon].
expect.PAST.3SG Heledd to Sioned see.INF Rhiannon
'Heledd expected Sioned to see Rhiannon.'

I will refer to such clauses as *i*-clauses. An obvious question here is: how many *i* lexemes are there in this area? There is evidence that there are three different lexemes (although (1)-(4) involve the same lexeme), but I will show that they share a number of properties, and also share properties with the preposition. I will go on to show that the HPSG hierarchical lexicon allows both the similarities and the differences in this area to be captured.

The discussion is organized as follows: In section 2, I show, drawing especially on Tallerman (1998), that there are three *i* lexemes introducing subordinate clauses. Then, in section 3, I argue that all are complementizers and heads of phrases. In section 4, I show how they share properties with each other and with the preposition *i*. In section 5, I develop an analysis exploiting the hierarchical lexicon of HPSG, and in section 6, I highlight the possibility of similar analyses for some other lexemes. Finally, in section 8, I offer some concluding remarks.

2. How many *i* lexemes are there?

In an early discussion of examples of the kind that we are concerned with here (Borsley 1986), I assumed that *i* is a complementizer, and argued that such an analysis was problematic for the then current Government Binding Theory (GB). The argument was fairly simple.

For GB, subjectless infinitives in a control sentence have a PRO subject and subjectless infinitives in a raising sentence have an NP trace subject. GB assumptions require that (a) PRO must be ungoverned, (b) NP trace must be governed but not case marked, and (c) an overt NP must be case-marked, normally by some governor. It seems to follow that *i* must (a) not govern, (b) govern but not case mark, and (c) govern and case mark. This looks like a problem. Of course, there would be no problem if there were three different *i* lexemes, but it would be unsatisfactory to adopt this position if the only motivation was the maintenance GB assumptions. Tallerman (1998) argues that there are in fact three different *i* lexemes although not in the way GB assumptions require.

There seem to be no reason to think that control and raising complements involve different *i* lexemes, but Tallerman provides evidence that the *i* of subjectless infinitives and the *i* of *i*-clauses are distinct lexemes. She shows that predicates which can take both a full clause introduced by *i* and a subjectless infinitive do not necessarily have *i* with the subjectless infinitive.

Instead, they may be introduced by zero or an element homophonous with the preposition *o* 'from':

(6) a. Disgwyliodd Heledd [i Sioned weld Rhiannon].
expect.PAST.3SG Heledd to Sioned see.INF Rhiannon
'Heledd expected Sioned to see Rhiannon.'

b. Disgwyliodd Heledd [weld Rhiannon].
expect.PAST.3SG Heledd see.INF Rhiannon
'Heledd expected to see Rhiannon.'

(7) a. Roedd hi 'n siŵr [iddi hi glywed
be.IMPF.3SG she PRED sure to.3SGF she hear.INF
y gwcw].
the cuckoo
'She was sure she heard the cuckoo.'

b. Roedd hi 'n siŵr [o gyrraedd yn
be.IMPF.3SG she PRED sure from arrive.INF PRED
hwyr].
late
'She was sure to arrive late.'

This suggests that there are two distinct lexemes here.

Tallerman (1998) also shows, building on Harlow (1993), that while some *i*-clauses are non-finite clauses and rather like English *for-to* clauses, others are finite. *I*-clauses with *disgwylio* 'expect' and many other verbs are clearly non-finite. They are negated by the negative verb *peidio* like subjectless infinitives. It is mutated as *beidio* in both cases.[1]

(8) Disgwyliodd Heledd [i Sioned beidio â gweld
expect.PAST.3SG Heledd to Sioned NEG with see.INF
Rhiannon].
Rhiannon
'Heledd expected Sioned not to see Rhiannon.'

(9) Disgwyliodd Heledd [beidio â gweld Rhiannon].
expect.PAST.3SG Heledd NEG with see.INF Rhiannon
'Heledd expected not to see Rhiannon.'

But other *i*-clauses appear with verbs which normally take a finite clause such as *meddwl* 'think'. A rather surprising fact about Welsh is that past tense forms of verbs are generally not acceptable in positive complement clauses (Jones 2010: 171). Thus, (10) is quite problematic:

[1] Mutation is ubiquitous in Welsh. I will pass over unimportant instances of mutation without comment, but I will discuss some important instances below.

(10) %Meddyliodd Heledd [aeth Sioned adre'].
think.PAST.3SG Heledd go.PAST.3SG Sioned home
'Heledd thought that Sioned had gone home.'

In colloquial Welsh, a perfect clause involving *bod* 'be' and *wedi* appears instead (Jones 2010: 172):

(11) Meddyliodd Heledd [bod Sioned wedi mynd
think.PAST.3SG Heledd be.INF Sioned PERF go.INF
adre'].
home
'Heledd thought that Sioned had gone home.'

Despite appearances, this is a type of finite clause, as Awbery (1976: 41-43), Tallerman (1998) and Bonami, Borsley & Tallerman (2016) show. In literary Welsh, an *i*-clause appears:[2]

(12) Meddyliodd Heledd [i Sioned fynd adre'].
think.PAST.3SG Heledd to Sioned go.INF home
'Heledd thought that Sioned had gone home.'

The interpretation suggests that this clause is actually finite, and so does the fact that it is in a context where a finite clause is expected. The fact that a negative counterpart of this clause is the ordinary finite clause in (13) points to the same conclusion:[3]

(13) Meddyliodd Heledd [aeth Sioned ddim adre'].
think.PAST.3SG Heledd go.PAST.3SG Sioned NEG home
'Heledd thought that Sioned had not gone home.'

Anaphora also suggests that there are non-finite and finite *i*-clauses. In a non-finite *i*-clause, a pronoun cannot be bound by an NP in the main clause NP, but a reflexive can:

[2] For some discussion of the relation between literary Welsh and other varieties, see Borsley, Tallerman & Willis (2007: section 1.3).

[3] Past tense verbs are also acceptable in interrogative complement clauses, as (i) illustrates (Jones 2010: 174-5):

(i) Dw i 'n gofyn (a) dda'th Mair
be.PRES.1SG I PROG ask.INF Q come.PAST.3SG Mair
ddoe.
yesterday
'I'm asking whether Mair came yesterday.'

(14) a. Dymunai Aled$_i$ iddo fo$_{j/*i}$ fynd.
want.PAST.3SG Aled to.3SGM he go.INF
'Aled wanted him to go.'

b. Dymunai Aled iddo ei hun ddarllen
want.PAST.3SG Aled to.3SGM 3SG REFL read.INF
y llyfr.
the book
'Aled wanted himself to read the book.' (Tallerman 1998: 92)

In contrast, in a finite *i*-clause, a pronoun can be bound by an NP in the main clause or can be free, but a reflexive cannot be:

(15) a. Dywedodd Aled$_i$ iddo fo$_{i/j}$ fynd.
say.PAST.3SG Aled to.3SGM he go.INF
'Aled said he's gone.'

b. *Dywedodd Aled iddo ei hun fynd.
say.PAST.3SG Aled to.3SGM 3SG REFL go.INF
*'Aled said that himself went.' (Tallerman 1998: 90)

Coordination also distinguishes between non-finite and finite *i*-clauses. A non-finite *i*-clause cannot coordinate with a normal finite clause, as noted by Sadler (1988: 40):

(16) ??Disgwyliodd Emrys [i Mair fynd i Gaerdydd]
expect.PAST.3SG Emrys to Mair go.INF to Cardiff
ac [y byddai Siôn yn mynd i
and PRT be.COND.3SG Siôn PROG go.INF to
Abertawe].
Swansea
'Emrys expected Mair to go to Cardiff and that Siôn would be going to Swansea.' (Tallerman 1998: 92)

In contrast, a finite *i*-clause can readily coordinate with a normal finite clause:

(17) Meddyliodd Aled [i Alys fynd adre'] ac
think.PAST.3SG Aled to Alys go.INF home and
[y byddai Mair yn mynd yn fuan].
PRT be.COND.3SG Mair PROG go.INF PRED soon
'Aled thought that Alys had gone home and that Mair would be going soon.' (Tallerman 1998: 79)

The preceding discussion focused on *i*-clauses as verbal complements. It seems that both types of *i*-clause may also appear as prepositional

complements, but the facts are complex. The following looks as if it contains a finite *i*-clause:

(18) Ges i air 'dag e cyn [iddo fe
get.PAST.1SG I word with he before to.3SGM he
fynd].
go.INF
'I had a word with him before he went.'

However, this appears to be a non-finite clause deriving a past time interpretation from the context, rather like the non-finite constituent in the following (and its English translation):

(19) Ges i air 'dag e cyn mynd.
get.PAST.1SG I word with he before go.INF
'I had a word with him before going/I went.'

Here is a similar example where the context gives rise to a future interpretation:[4]

(20) Dw i 'n moyn cael gair 'dag e
be.PRES.1SG I PROG want.INF get.INF word with he
cyn [iddo fe fynd].
before to.3SGM he go.INF
'I want to have a word with him before he goes.'

This is like the non-finite constituent in the following:

(21) Dw i 'n moyn cael gair 'dag e
be.PRES.1SG I PROG want.INF get.INF word with he
cyn mynd.
before go.INF
'I want to have a word with him before going/I go.'

But the following naturally occurring example shows an *i*-clause after *oherwydd* 'because', which normally takes a finite clause as its complement:

[4] Bonami, Borsley and Tallerman (2016) call finite *i*-clauses pseudo-non-finite clauses because they look like non-finite clauses but are really finite clauses. From this perspective, examples like (18) and (20) could be called pseudo-pseudo-non-finite clauses.

(22) Dirywiodd y wladwriaeth Carthaginaidd oherwydd
deteriorate.PAST.3SG the state Carthaginian because
[i 'r Rhufeiniaid eu trechu yn y Rhyfeloedd
to the Romans 3PL conquer.INF in the wars
Pwnig].
Punic
'The Carthaginian state deteriorated because the Romans conquered them in the Punic Wars.' https://cy.wikipedia.org/wiki/Algeria

This appears to be a genuine finite *i*-clause as a prepositional complement. Like examples with a finite *i*-clause as a verbal complement, it is quite literary, and a more colloquial example would have *bod* and *wedi*.

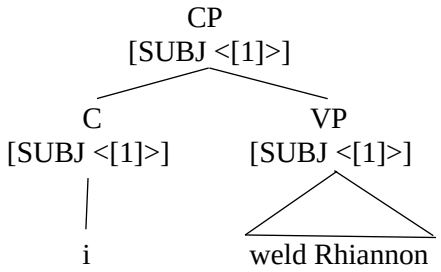
It seems, then, we have quite strong evidence that that there are two *i* lexemes in *i*-clauses: one non-finite, and one finite and past tense. I conclude that there are three clausal *i* lexemes altogether.

3. Three complementizers

What exactly are the three clausal i-lexemes? I will argue that they are all complementizers (essentially as in Borsley 1986) and heads taking complements.

In assuming that the *i* of subjectless infinitives is a complementizer and a head, I am essentially following Tallerman (1998). As is standard in HPSG, I assume that a subjectless infinitive is a VP. I assume, then, that this element is a complementizer taking a non-finite VP complement, and that it has the same value for SUBJ as its complement. In other words, it is a raising predicate, and apart from the fact that it is a complementizer is rather like English *to*. This means structures like the following:

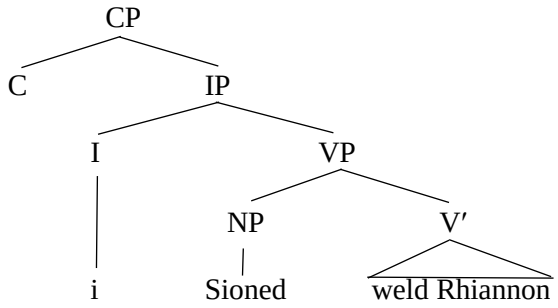
(23)



Not all HPSG work assumes that complementizers are heads taking a complement. Pollard and Sag (1994: 44-46) and others have proposed that they are markers combining with a clausal head of some kind. This looks like a possible alternative here. I will suggest, however, that it not plausible for the *i* lexemes in *i*-clauses.

Tallerman (1998) in fact assumes that *i* of *i*-clauses is not a complementizer. Assuming a fairly orthodox Chomskyan view of clause structure, in which there is a distinction between C(omplementizer) and I(nflection), she proposes that this *i* is in the I position. Thus, she has structures like the following:

(24)



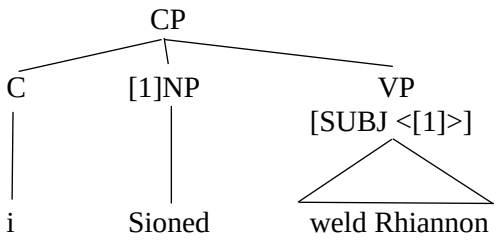
Tallerman sees the *i* of *i*-clauses as similar to finite verbs, which she assumes are in I because they can be preceded by certain particles, e.g. the affirmative particles *mi* in North Wales or *fe* in South Wales, which she assumes are in C:

(25) Mi/Fe welodd Sioned Rhiannon.
PART see.PAST.3SG Sioned Rhiannon
'Sionedd saw Rhiannon.'

However, there is evidence in Willis (1998: 70-71) and Borsley and Jones (2005: 57) that preverbal particles form a constituent with the following verb. This suggests that both are in C, and this is explicitly assumed by Willis working within a Chomskyan framework. Thus, the argument for this analysis seems quite weak even within Chomskyan assumptions. Outside those assumptions there is no reason to think that *i* occupies a different position in *i*-clauses and subjectless infinitives. I will assume, then, that the *i* lexeme in *i*-clauses is a complementizer.

I will also assume following Borsley (1999) that these elements are omplementizers taking two complements: an NP and a VP, where the NP is the subject of the VP. This is essentially the analysis that Sag (1997) proposes for English *for-to* clauses. It means structure like the following:

(26)



I think there is an objection here to an analysis in which complementizers are markers. As we will see in the next section, the *i* of *i*-clauses agrees with a following pronominal subject. As shown in Borsley (2009, 2022), agreement in Welsh generally involves a head and an immediately following complement. On the analysis in (26), *i*-clauses are just another example of this pattern. On a marker analysis, they would be something rather different. Markers combine with a single sister. Hence, on such an analysis, NP and VP would have to form a constituent and the agreement would involve a non-head and an element which is not its sister but a daughter of its sister. It seems preferable to maintain the assumption that agreement in Welsh involves a head and an immediately following complement, and the analysis in (26) allows one to do this.[5]

I conclude then there are three complementizers, one taking a single complement, a VP, and two taking two complements, an NP and VP. I will assume that *verb* and *complementizer* are subtypes of a type *verbal* as in Sag (1997: 457). This makes it unsurprising that there are positions in which both verb-headed and complementizer-headed constituents appear, and especially that finite *i*-clauses appear in the same positions as clauses headed by a finite verb.

4. **Similarities between the four *i* lexemes**

It is obviously not unusual for a language to have homophonous lexemes. Commonly, they have no other shared properties. In English, the preposition *to* and the infinitive marker *to* seem to have no other shared properties. The following Welsh examples illustrate a similar situation:

(27) a. Mae Heledd yn **Neiniolen**. (Deiniolen)
be.PRES.3SG Heledd in Deiniolen
'Heledd is in Deiniolen.'

b. Mae Heledd yn dawnsio.
be.PRES.3SG Heledd PROG dance.INF
'Heledd is dancing.'

These feature the preposition *yn* 'in' and the homophonous progressive marker. In (27a), the preposition triggers the alternation known as nasal mutation. Thus, the place name *Deiniolen* appears as *Neiniolen*. (Here and subsequently, I put important mutated words in bold and give the basic form

[5] It has been suggested to me that *i* could be a weak head in the sense of Tseng (2002), a head which derives many of its properties from its complement(s). But the various forms of *i* have little in common with their complements. The first of the forms is non-finite like its complement, and the second is non-finite like its second complement, but the third form has essentially no properties in common with its complements. Hence, I don't see any reason to think that we have weak heads here.

of the word in brackets.) In (27b), there is no mutation with the progressive marker. Historically, these are the same element (Sims-Williams 2015), but in the contemporary language they seem to be just two separate lexemes. However, the three complementizers that we are concerned with here are quite different. They have a variety of properties in common other than just their phonological form, and they all share properties with the preposition *i*.

Unlike the two *yn* lexemes just considered, which have different mutation properties, the four *i* lexemes have the same mutation property: they all trigger soft mutation on the following constituent. This is an NP in (28), (30) and (31), and a VP in (29):

(28) i **Fangor** (Bangor)
to Bangor

(29) Mae Heledd yn awyddus [i **weld**
be.PRES.3SG Heledd PRED eager to see.INF
Rhiannon]. (gweld)
Rhiannon
'Heledd is eager to see Rhiannon.'

(30) Disgwyliodd Heledd [i **ddau** dyn weld
expect.PAST.3SG Heledd to two man see.INF
Rhiannon]. (dau)
Rhiannon
'Heledd expected two men to see Rhiannon.'

(31) Meddyliodd Heledd [i **ddau** dyn fynd
think.PAST.3SG Heledd to two man go.INF
adre']. (dau)
home
'Heledd thought that two men went home.'

(The mutation of *weld* in (30) is triggered not by *i*, but by the preceding subject *ddau dyn*.)

The preposition *i* and the *i* of non-finite and finite *i*-clauses are also similar in showing agreement with a following third person pronoun:

(32) iddo fo / iddi hi / iddyn nhw
to.3SGM he to.3SGF she to.3PL they
'to him/her/them'

(33) Disgwyliodd Heledd [iddo fo / iddi hi /
expect.PAST.3SG Heledd to.3SGM he to.3SGF she
iddyn nhw weld Rhiannon].
to.3PL they see.INF Rhiannon
'Heledd expected him/her/them to see Rhiannon.'

(34) Meddyliodd Heledd [iddo fo / iddi hi /
think.PAST.3SG Heledd to.3SGM he to.3SGF she
iddyn nhw fynd adre'].
to.3PL they go.INF home
'Heledd thought that he/she/they had gone home.'

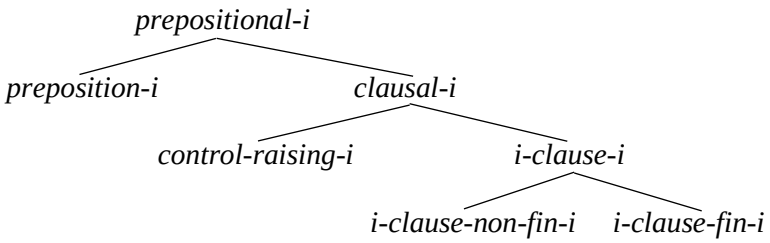
The preposition is unusual among prepositions in only showing agreement with a third person pronoun and not with all pronouns, and the complementizers have the same property. The *i* of subjectless infinitives does not show agreement, but it does not have the opportunity to because it is never immediately followed by a pronoun. Thus, we can say that all four lexemes have the same agreement potential, and more generally that they have the same morphological properties.

The four lexemes have different syntactic properties, but the three complementizers have in common the fact that they are complementizers, and the two *i*-clause complementizers have the same complement selection properties. Thus, it seems that there are four distinct *i* lexemes, but that they show a range of similarities. A satisfactory analysis needs to capture both the similarities and the differences in this area.

5. A hierarchical lexicon analysis

Standard HPSG assumptions about the lexicon stemming from Flickinger (1987) allow a fairly simple approach to situations like this. They allow the four lexemes to be analysed as four realisations of a 'super-lexeme' and all the shared properties to be specified just once. We can propose the type hierarchy in (35) for this part of the lexicon:

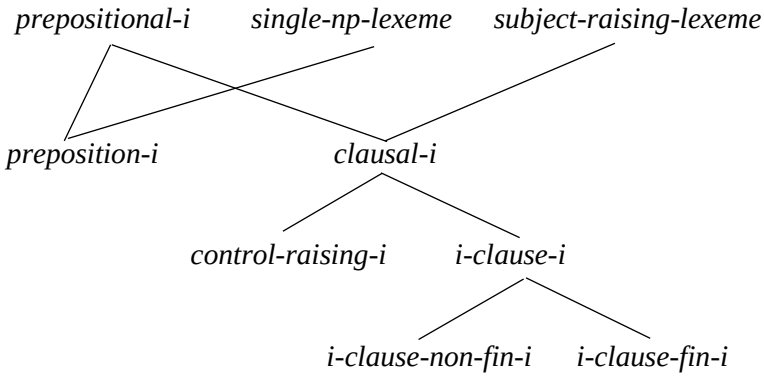
(35)



Note that *prepositional-i* and *preposition-i* are quite different types, and that I am using the type *control-raising-i* for the *i* of subjectless infinitives. This type hierarchy groups together all four lexemes as instances of *prepositional-i*, the three complementizers as instances as *clausal-i*, and the two *i*-clause complementizers as instances of *i-clause-i*. This hierarchy provides a basis for capturing the similarities and the differences in this area. However, it needs to be extended to take account of the fact that there is nothing unusual about the complement selection properties of the four lexemes. The preposition *i* is like

many prepositions in taking a NP complement. The *i* of subjectless infinitives is a raising predicate, as noted earlier. Finally, the two complementizers that appear in *i*-clauses are essentially raising predicates with both their arguments as complements.[6] These complement selection properties should be largely inherited from various argument selection types. I will assume two such types *single-np-lexeme* and *subject-raising-lexeme* and make *preposition-i* a subtype of the former, and *clausal-i* a subtype of the latter. This gives the following extended type hierarchy:

(36)



We begin with *prepositional-i* and the properties that are shared by all four lexemes. We have seen that all have the same morphological properties. I assume that these properties are a reflection of two features. First, following Borsley (2009, 2022), I assume that agreement in Welsh is the realization of a feature AGR, whose value is the index of a following pronoun with its PERSON, NUMBER, and GENDER features, or *none* when there is no following pronoun. I assume that the mutation-triggering property of a lexeme reflects a feature MUT(ATION)-TR(IGGER) with the values *soft*, *nasal*, and *aspirate* for the three kinds of mutation that occur in Welsh, or *none*. (Only the first is important here.) With these assumptions, we can attribute the phonological and morphological properties of the four lexemes to the following constraint on *prepositional-i* (where the MARKING feature allows heads to select a constituent headed by one of these lexemes):

(37) *prepositional-i* $\Rightarrow$ $\begin{bmatrix} \text{MARKING } i \\ \text{AGR } index \vee none \\ \text{MUT} - \text{TR } soft \end{bmatrix}$

[6] Following Borsley (1989), I assume that finite verbs also have all their arguments as complements. Thus, the *i*-clause complementizers are like finite raising verbs in their complement selection properties.

What about the form of the four lexemes? The grammar just needs to impose the following pairings of AGR value and form for *prepositional-i*:

AGR value	Form
3nd, sing, masc	*iddo*
3nd, sing, fem	*iddi*
3rd, plur	*iddyn*
Any value	*i*

AGR value-form pairings for *prepositional-i*

Following Bonami, Borsley and Tallerman (2016), I assume that more specific constraints take precedence over more general ones and hence that a general constraint does not apply if a more specific constraint requires something different. This means that the basic form *i* will not appear with a third person pronoun, but will appear in all other circumstances, i.e. with a first or second person pronoun or a non-pronominal NP.

We can turn now to the two immediate subtypes of *prepositional-i*. Here, we can propose the following simple constraints:

(38) *preposition-i* $\Rightarrow$ $\begin{bmatrix}\text{HEAD } prep \\ \text{SUBJ } <>\end{bmatrix}$

(39) *clausal-i* $\Rightarrow$ [HEAD *comp*]

The former will inherit properties from *single-np-lexeme* and the latter from *subject-raising-lexeme*. I assume these are subject to the following simple constraints:

(40) *single-np-lexeme* $\Rightarrow$ [ARG-ST <NP>]
(41) *subject-raising-lexeme* $\Rightarrow$ [ARG-ST <[1]NP, VP[*inf*, SUBJ <[1]>]>]

Numerous lexemes will inherit properties from these two types. I also assume the Argument Realization Principle in (42):

(42) *word* $\Rightarrow$ $\begin{bmatrix}\text{SUBJ [1]} \\ \text{COMPS [2]} \\ \text{ARG-ST [1]} \oplus \text{[2]}\end{bmatrix}$

As a subtype of *single-np-lexeme*, preposition *i* will have a single NP in its ARG-ST list. The SUBJ <> restriction in (38), interacting with the Argument Realization Principle, will ensure that this NP appears in its COMPS list. As a

subtype of *subject-raising-lexeme*, *clausal-i* will have the ARG-ST list specified by (41).

The two subtypes of *clausal-i*, *control-raising-i* and *i-clause-i*, will be subject to the following constraints:

(43) *control-raising-i* $\Rightarrow$ $\begin{bmatrix}\text{HEAD [VFORM } inf\text{]} \\ \text{SUBJ} < [] >\end{bmatrix}$

(44) *i-clause-i* $\Rightarrow$ [SUBJ <>]

Both types inherit a two member ARG-ST list from *subject-raising-lexeme*. The constraint on *control-raising-i* ensures that only the second member appears in its COMPS list. The constraint on *i-clause-i* ensures that both members appear in its COMPS list.

Finally, for the two subtypes of *i-clause-i*, we can propose the following quite simple constraints:

(45) *i-clause-non-fin-i* $\Rightarrow$ [HEAD [VFORM inf]]

(46) *i-clause-fin-i* $\Rightarrow$ $\left[\text{HEAD } \begin{bmatrix}\text{VFORM } fin \\ \text{TENSE } past\end{bmatrix}\right]$

(45) requires *i-clause-non-fin-i* to be non-finite, and (46) requires *i-clause-fin-i* to be finite and past tense. The past tense requirement ensures that finite *i*-clauses have the sort of interpretation that one would expect to be expressed by a complement clause with a past tense verb. The constraint in (46) could be extended to include the information that finite *i*-clauses are literary. It could be reformulated as follows:

(47) *i-clause-fin-i* $\Rightarrow$ $\begin{bmatrix}\text{SS|CAT|HEAD [VFORM } fin, \text{TENSE } past\text{]} \\ \text{REGISTER } literary\end{bmatrix}$

Within this analysis the four *i* lexemes have a variety of properties inherited from the various supertypes. Here are fairly full syntactic categories for each:

(48) *preposition-i*:

$$\left[\begin{array}{l}\text{HEAD } \begin{bmatrix}prep \\ \text{MARKING } i \\ \text{AGR } index \vee none \\ \text{MUT} - \text{TR } soft\end{bmatrix} \\ \text{SUBJ} <> \\ \text{COMPS} < \text{NP} >\end{array}\right]$$

(49) *control-raising-i*:

$$\begin{bmatrix} \text{HEAD} \begin{bmatrix} comp \\ \text{MARKING } i \\ \text{AGR } index \vee none \\ \text{MUT} - \text{TR } soft \\ \text{VFORM } inf \end{bmatrix} \\ \text{SUBJ} < [1] > \\ \text{COMPS} < \text{VP}[inf, \text{SUBJ} < [1] >] > \end{bmatrix}$$

(50) *i-clause-non-fin-i*:

$$\begin{bmatrix} \text{HEAD} \begin{bmatrix} comp \\ \text{MARKING } i \\ \text{AGR } index \vee none \\ \text{MUT} - \text{TR } soft \\ \text{VFORM } inf \end{bmatrix} \\ \text{SUBJ} <> \\ \text{COMPS} < [1]\text{NP}, \text{VP}[inf, \text{SUBJ} < [1] >] > \end{bmatrix}$$

(51) *i-clause-fin-i*:

$$\begin{bmatrix} \text{HEAD} \begin{bmatrix} comp \\ \text{MARKING } i \\ \text{AGR } index \vee none \\ \text{MUT} - \text{TR } soft \\ \text{VFORM } fin \\ \text{TENSE } past \end{bmatrix} \\ \text{SUBJ} <> \\ \text{COMPS} < [1]\text{NP}, \text{VP}[inf, \text{SUBJ} < [1\} >] > \end{bmatrix}$$

All these categories are [AGR *index* ∨ *none*]. Assuming the analysis of agreement developed in Borsley (2009, 2022), the value of AGR is an index when there is a following pronoun and otherwise *none*. The preposition and the two *i*-clause complementizers may be followed by a pronoun, but control-raising *i* is never followed by a pronoun. Thus, on the proposed analysis, it has an agreement potential which is never realised.

With these categories, the examples that we are concerned with here are all fairly ordinary head-complement phrases, two with one complement, and two with two. In each case, the head assigns soft mutation, and in each case, it will agree with an immediately following pronoun (but, as we have emphasized, control-raising *i* will never be immediately followed by a pronoun). (49) and (50) both head a non-finite clause, but (51) crucially heads a finite and past tense clause.

But what about the fact that a positive past tense verb is generally ungrammatical in a complement clause? One possibility is an analysis of the kind outlined in Bonami, Borsley & Tallerman (2016), in which finite *i* is literally a positive past tense form of the associated verb. However, as noted above, finite *i* is generally confined to the literary language, and in more colloquial Welsh a perfect clause involving *bod* 'be' and the particle *wedi* appears. I will assume, then, that there is a constraint ruling out a past tense verb in a positive complement clause, and that different varieties have different ways of expressing the meanings which cannot be expressed by a past tense verb, finite *i* fulfilling this role in the literary language.

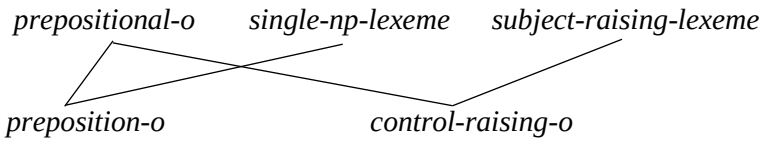
6. Some other super lexemes

There are some other cases in Welsh of homophonous lexemes which should probably be analysed as alternative realizations of a single super lexeme. I assume the element *o* in (7b) is another complementizer homophonous with a preposition. This element triggers soft mutation (the unmutated form of the following verb is *cyrraedd*). In this, it just like the proposition:

(52) Dw i wedi dôd o
be.PRES.1SG I PERF come.INF from
Gaernarfon. (Caernarfon)
Caernarfon
'I have come from Caernarfon.'

This suggests the type hierarchy in (53) and the constraints in (54)-(56):

(53)



(54) *prepositional-o* $\Rightarrow$ $\begin{bmatrix} \text{MARKING } o \\ \text{AGR } index \vee none \\ \text{MUT} - \text{TR } soft \end{bmatrix}$

(55) *preposition-o* $\Rightarrow$ $\begin{bmatrix} \text{HEAD } prep \\ \text{SUBJ } <> \end{bmatrix}$

(56) *control-raising-o* $\Rightarrow$ $\begin{bmatrix} \text{HEAD } \begin{bmatrix} comp \\ \text{VFORM } inf \end{bmatrix} \\ \text{SUBJ } < [] > \end{bmatrix}$

This is essentially a simplified version of the analysis for *i*.

There are at least two other cases for which an analysis of this kind seems appropriate. Welsh has a number of aspectual particles which are homophonous with a preposition. The most common, progressive *yn* and perfect *wedi*, which is homophonous with a preposition meaning 'after', seem to share no other properties with the preposition, but two others are different. The preposition *ar* 'on' and the homophonous aspect marker of imminence assign soft mutation (Jones 2010: 336-9).

(57) a. Mae 'r wylan ar **graig**. (craig)
be.PRES.3SG the seagull on rock
'The seagull is on a rock.'

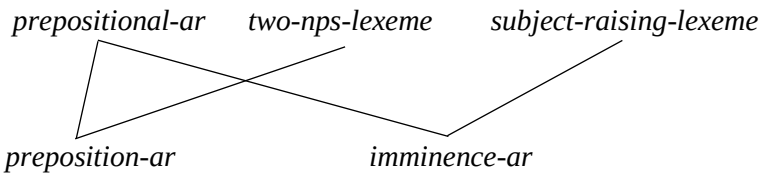
b. Mae o ar **ganu**. (canu)
be.PRES.3SG he on sing.INF
'He's about to sing.'

The preposition *ar* is predicative and has an object and a subject. This suggests that it inherits properties from a type *two-nps-lexeme* subject to the following constraint:

(58) *two-nps-lexeme* ⇒ [ARG-ST <NP, NP>]

This allows us to propose the following type hierarchy and constraimts:

(59)



(60) *prepositional-ar* ⇒ $\begin{bmatrix} \text{MARKING } ar \\ \text{AGR } index \vee none \\ \text{MUTR } soft \end{bmatrix}$

(61) *preposition-ar* ⇒ $\begin{bmatrix} \text{HEAD } prep \\ \text{SUBJ } < [] > \end{bmatrix}$

(62) *imminence-ar* ⇒ $\begin{bmatrix} \text{HEAD } imminence \\ \text{SUBJ } < [] > \end{bmatrix}$

We have a similar situation with the preposition *heb* 'without' and the homophonous negative perfect aspect marker (Jones 2010: 333-6). Both assign soft mutation:

(63) a. Dw i heb **gar** yr wythnos 'ma. (car)
be.PRES.1SG I without car the week here
'I'm without a car this week.'

b. Maen nhw heb **gyrraedd** eto. (cyraedd)
be.PRES.3PL they without arrive.INF yet
'They haven't arrived yet.'

This suggests an analysis like that proposed for *ar*.

7. Conclusions

I have argued in preceding pages that clausal *i* is three different lexemes (essentially as Tallerman 1998 showed), but that they are related lexemes with shared properties. I have also shown that the preposition *i* is a further related lexeme sharing various properties. I have shown that it is not too difficult to capture the similarities and differences among the four lexemes with the hierarchical lexicon of HPSG. With an appropriate type hierarchy the shared properties can be all be specified just once. There are a number of other cases in Welsh where an analysis of this kind may be appropriate.

A similar treatment is probably appropriate for a variety of phenomena in a variety of languages. In Borsley (2019), I analyze the Welsh predicational copula and identity copula as two realisations of a super-lexeme (without using the term), and Alotaibi and Borsley (2020) argue for a similar approach to the copula in Modern Standard Arabic. Also relevant here is recent unpublished work by Jacob Maché, who proposes an analysis rather like this for Germanic 'need' verbs, and also discusses how the type hierarchy it involves could emerge diachronically. It looks, then, as if the type of analysis developed here has considerable potential.[7]

---

[7] Naturally there are other matters that could be explored here. For example, the constructions we have been discussing may appear in relative clauses and other unbounded dependency clauses. The following, from (Borsley, Tallerman and Willis 207: 134), illustrate:

(i) Dw i 'n chwilio am rywbeth [i ('w)
be.PRES.1SG I PROG search for something to 3SGM
ddarllen ___].
read.INF
'I'm looking for something to read.'

(ii) Mae e wedi canu ar bob albwm [i ni
be.PRES.1SG he PERF sing.INF on every album to us
ei wneud erioed].
3SGM do.INF ever
'He's sung on every album we've ever done.'

But this is perhaps more a topic for research on Welsh unbounded dependencies.

REFERENCES

Alotaibi, Ahmad. & Robert D. Borsley 2020. The copula in Modern Standard Arabic. In Anne Abeillé & Olivier Bonami (eds.), *Constraint-based Syntax and Semantics: Papers in Honor of Danièle Godard*, 15-35. Stanford: CSLI Publications.

Awbery, Gwenllian 1976. *The Syntax of Welsh: A Transformational Study of the Passive*. Cambridge: Cambridge University Press.

Bonami, Olivier, Robert D. Borsley & Maggie O. Tallerman. 2016. On pseudo-non-finite clauses in Welsh. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller, *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, 104–124. Stanford, CA: CSLI Publications.

Borsley, Robert D. 1986. Prepositional complementizers in Welsh. *Journal of Linguistics* 22, 67–84.

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, 263–300.

Borsley, Robert D. 2009. On the superficiality of Welsh agreement. *Natural Language & Linguistic Theory* 27, 225–265.

Borsley, Robert D. 2019. The complexities of the Welsh copula. In Stefan Müller & Petya Osenova (eds.), *Proceedings of the 26th International Conference on Head-driven Phrase Structure Grammar*, 5–25. Stanford, CA: CSLI Publications.

Borsley, Robert D. 2022. On the structure of Welsh noun phrases. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 29th International Conference on Head-Driven Phrase Structure Grammar*, Online (Nagoya/Tokyo), 27–47. Frankfurt/Main: University Library.

Borsley, Robert D. & Bob M. 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.

Flickinger, Daniel P. 1987. *Lexical Rules in the Hierarchical Lexicon*. Stanford, CA: Stanford University dissertation.

Harlow, Stephen 1993. Finiteness and Welsh clause structure. In Hans-Georg Obenauer & Anne Zribi-Hertz (eds), *Structure de la Phrase et Thėorie du Liage*, 93-119. Paris: Presses Universitaires de Vincennes.

Jones, Bob Morris. 2010. *Tense and Aspect in Informal Welsh*. Berlin & New York: Mouton de Gruyter.

Pollard, Carl & Ivan A. Sag 1994. *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Sadler, Louisa 1988. *Welsh Syntax: A Government-Binding Approach*. London: Croom Helm.
Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33, 431–484.
Sims-Williams, Patrick 2015. Four types of Welsh YN. *Transactions of the Philological Society* 113:3, 286–304.
Tallerman, Maggie O. 1998. The uniform case-licensing of subjects in Welsh. *The Linguistic Review* 15, 69–133.
Tseng, Jesse 2002. Remarks on Marking. In Frank van Eynde, Lars Hellan & Dorothee Beermann (eds), *HPSG 2001: Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, 267-283. CSLI Publications.
Willis, David W. E. 1998. *Syntactic Change in Welsh: A Study of the Loss of Verb-Second*. Oxford: Clarendon Press.

Morphotactic competition in Murrinh-Patha: Rule composition and rule interaction in Information-based Morphology

Berthold Crysmann

Université Paris Cité, Laboratoire de linguistique formelle, CNRS

Proceedings of the 30th International Conference on
Head-Driven Phrase Structure Grammar

University of Massachusetts Amherst

Stefan Müller, Elodie Winckel (Editors)

2023

Frankfurt/Main: University Library

pages 26–43

Keywords: Morphotactics, Murrinh-Patha, Positional competition, Pāṇinian competition, Templatic morphology

Crysmann, Berthold. 2023. Morphotactic competition in Murrinh-Patha: Rule composition and rule interaction in Information-based Morphology. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 26–43. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2023.2.


Abstract

Murrinh-Patha, a polysynthetic Non-Pama-Nyungan language of Australia features competition of subject and object agreement markers for a particular position (i.e. slot 2), meaning that certain subject agreement markers are realised in this position, unless already occupied by overt object agreement markers. In their typology of variable morphotactics, Crysmann & Bonami (2016) cite the case of Murrinh-Patha as an instance of misaligned, conditioned placement. I shall propose a formal account of this positional competition in Murrinh-Patha within Information-based Morphology. To this end, I shall generalise the "pivot" features previously proposed for placement relative to the stem (Italian; Crysmann & Bonami, 2016) or the edge (Soranî Kurdish; Bonami & Crysmann, 2013; Salehi & Koenig, 2023) and show how this will facilitate the treatment of conditioned placement in Murrinh-Patha.

1 Introduction

In this paper[1], I shall discuss cases of positional competition between different exponents of subject and object agreement in Murrinh-Patha, a polysynthetic Non-Pama-Nyungan language of Australia. The data discussed here are taken from Nordlinger (2010, 2015) and Nordlinger & Mansfield (2021). The language features competition of subject and object agreement markers for a particular position (i.e. slot 2), meaning that certain subject agreement markers are realised in this position, unless already occupied by overt object agreement markers. According to Nordlinger (2010), this competition provides evidence for a templatic organisation of the language's morphology.

In their typology of variable morphotactics, Crysmann & Bonami (2016) cite the case of Murrinh-Patha as an instance of misaligned, conditioned placement, which they schematically represent as shown in Figure 1.

I shall propose a formal account of this positional competition in Murrinh-Patha, making explicit how "if available" can be implemented within Information-based Morphology. To this end, I shall generalise the "pivot" features previously proposed for placement relative to the stem (Italian; Crysmann & Bonami, 2016) or the edge (Sorani Kurdish; Bonami & Crysmann, 2013; Salehi & Koenig, 2023) and show how this will facilitate the treatment of conditioned placement in Murrinh-Patha.

[1]I would like to thank the audience at HPSG 2023 for their comments and discussion, in particular Jean-Pierre Koenig and Jakob Maché. A previous version of this paper had been presented at the International Symposium of Morphology (ISMo) 2021. Furthermore, I am also indebted to Olivier Bonami, Gilles Boyé and Baptiste Loreau Unger for discussion of several aspects of the issues discussed here. A great many thanks also go to Rachel Nordlinger for answering my questions on the intricate data discussed here. The research reported here has partially benefited from a public grant overseen by the French National Research Agency (ANR) as part of the program "Investissements d'Avenir" (reference: ANR-10-LABX-0083). It contributes to the IdEx Université Paris Cité (reference: ANR-18-IDEX-0001).

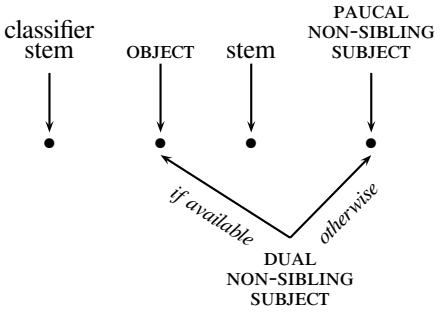


Figure 1: Schematic representation of Murrinh-Patha misaligned/conditioned placement (Crysmann & Bonami, 2016)

2 Murrinh-Patha

Verbs in Murrinh-Patha minimally consist of a lexical stem (open class) and a classifier stem (CS) from a set of 38 classifier stem paradigms. Together, these two stems express basic lexical meaning. While the lexical stem (in slot 5) is uninflected, the classifier stem (in slot 1) differentiates TAM as well as subject agreement.

1	2	3	4	5	6	7	8	9
CS.SUBJ.TAM	SUBJ NUM/OBJ	RR	IBP	**LEX-STEM**	TAM	ADV	SUBJ/OBJ NUM	ADV

Figure 2: Murrinh-Patha position classes (Nordlinger, 2015)

In addition to inflection by means of the classifier stem (CS), Murrinh-Patha verbs are inflected with a number of discrete markers, organised into a positional template, as shown in Figure 2. Of particular interest for this paper are slots 2 and 8, where exponents of subject and object agreement can be found.

			1		2	3
			INCL	EXCL		
SG			N/A	-ngi	-nhi	∅
DU	NSIB	M	-nhi	-nganku+nintha	-nanku+nintha	-nku+nintha
		F	-nhi	-nganku+ngintha	-nanku+ngintha	-nku+ngintha
	SIB		-nhi	-nganku	-nanku	-nku
PC	NSIB	M	-nhi+neme	-nganku+neme	-nanku+neme	-nku+neme
		F	-nhi+ngime	-nganku+ngime	-nanku+ngime	-nku+ngime
	SIB		-nhi	-ngan	-nan	-n
PL			-nhi	-ngan	-nan	-n

Table 1: Object agreement markers (Nordlinger, 2015, 505)

Agreement marking operates along up to four inflectional dimensions (illustrated by the paradigm of object agreement markers in Table 1): the language dis-

tinguishes four numbers (singular, dual, paucal, plural) and three persons, including a distinction between inclusive and exclusive for first person non-singular cells. Additionally, Murrinh-Patha marks a rather unique category of non-sibling in the dual and the paucal. Exponents of this category are differentiated for gender, which is otherwise not marked in the verb. Furthermore, the paucal is only distinguished for non-siblings. With siblings, paucal and plural are non-distinct. Another peculiarity of the non-sibling marker pertains to its morphotactics: while all other exponents of object agreement surface in slot two, the dual and paucal non-sibling markers are realised discontinuously in slot 8 (in the case of direct object agreement).

Subject agreement (cf. Table 2) is quite similar to object agreement, despite the difference in exponence: while object agreement is realised by discrete markers in slots 2 (person/number) and 8 (non-sibling number/gender), subject agreement is realised fusionally as part of the classifier stem (slot 1) plus discrete markers for non-sibling (slot 2/8) and for the non-future dual/paucal marker *ka* (slot 2). Recall that classifier stems cumulate expression of subject agreement with expression of TAM and lexical identity. See Mansfield (2020) for a detailed analysis of the classifier stem system. Another difference pertains to dual non-sibling marking: with direct object markers, the person/number exponent (slot 2) is syncretic with the person/number exponent of the sibling dual, whereas for subjects the classifier stem is syncretic with the singular.[2]

			1		2	3
			INCL	EXCL		
SG			N/A	bam	dam	bam
DU	NSIB	M	thubam	(ngu)bam+nintha	(nu)dam+nintha	(pu)bam+nintha
		F	thubam	(ngu)bam+ngintha	(nu)dam+ngintha	(pu)bam+ngintha
	SIB		thubam	ngubam+ka	nubam+ka	pubam+ka
PC	NSIB	M	thubam+neme	ngubam+ka+neme	nubam+ka+neme	pubam+ka+neme
		F	thubam+ngime	ngubam+ka+ngime	nubam+ka+ngime	pubam+ka+ngime
	SIB		thubam	ngubam	nubam	pubam
PL			thubam	ngubam	nubam	pubam

Table 2: Subject agreement (non-future sub-paradigm for classifier stem 'SEE(13)') (Nordlinger, 2015, 504)

As discussed above (cf. also Figure 2), the positions for the affixal markers of subject agreement overlap with those for object marking, so the central question is how conflict is actually resolved. Murrinh-Patha witnesses two strategies: displacement of the subject marker, and omission.

The first case of positional competition relates to the subject non-sibling markers *nintha/ngintha*. When marking subject agreement, these markers surface in slot

[2]See below for sensitivity of stem selection to the position of the non-sibling markers.

2, if available, i.e. before the lexical stem.[3]

(1) bam- **ngintha**- ngkardu
3SGS.SEE(13).NFUT DU.F see
'They (dual non-sibling) saw him/her.' (Nordlinger, 2010, 334)

However, if object agreement is overtly realised (any cell other than 3rd singular), slot 2 receives the object person/number marker and the subject non-sibling dual marker must surface in slot 8 instead, i.e. after the lexical stem, cf. (2).

(2) [pu]bam- ngi- ngkardu **-ngintha**
3SG/DUS.SEE(13).NFUT 1SGO see DU.F
'They (dual non-sibling) saw me.' (Nordlinger, 2010, 334)[4]

Given the fact that subject and object non-sibling markers are syncretic, and that object non-sibling markers are also realised in slot 8, non-sibling marking may end up ambiguous as to whether it refers to the subject or the object, cf. the examples from Nordlinger (2015) below.

(3) ma- nanku- rdarri- purl -nu **-ngintha**
1SGS.HANDS(8).FUT 2DU/PCO back wash FUT DU.F
'I will wash your (female dual non-sibling) backs.' *or*
'We (two exclusive female non-sibling) will wash your (dual sibling) backs.'
(Nordlinger, 2015, 506)

In (3), *ngintha* may either refer to the object, leaving subject agreement solely marked by the singular classifier stem, yielding singular. Alternatively, singular stem and dual non-sibling marker jointly express first person exclusive female non-sibling dual, leaving the object marker in slot 2 to express sibling dual.

What is important about realisation of the subject dual non-sibling markers is that realisation in slot 8 is only ever licit when slot 2 is blocked by another exponent. If slot 2 is free, subject *ngintha/nintha* must surface there.

The second case relates to the dual/paucal number marker *ka* which appears in slot 2 in the non-future, as shown in (4a,b) from Nordlinger (2010). Note that in the non-future, as opposed to other TAM categories, the dual and plural stems are syncretic.

(4) a. pubam- **ka-** ngkardu
3DUS.SEE(13).NFUT DU/PC.NFUT see
'They (dual sibling) saw him/her.'

[3]The paucal non-sibling marker *-neme/-ngime* are always realised in slot 8.

[4]The original example in Nordlinger (2010) provides a singular stem. However, as stated in Nordlinger & Mansfield (2021, Table 3), use of this stem instead of the dual stem is marginal, unless the non-sibling is found adjacent in slot 2. See also the discussion at the end of this section.

b. pubam- **ka-** ngkardu -ngime
3DUS.SEE(13).NFUT DU/PC.NFUT see PC.NON-SIB.F
'They (paucal, female, non-sibling) saw him/her.'

c. pubam- ngi- ngkardu
3DU/PLS.SEE(13).NFUT 1SGO see
'They (two siblings/plural) saw me.'

d. pubam- ngkardu
3PLS.SEE(13).NFUT see
'They (plural) saw him/her.' Nordlinger (2010, 333)

Again, in the case of overt object marking (4c), subject marking in slot 2 becomes unavailable. In contrast to the dual non-sibling markers, there is no alternate realisation for *ka*, even if a suitable position (like slot 8) happens to be unoccupied. Instead *ka* is simply dropped, possibly leading to ambiguity between dual and plural, as shown in (4c). Note that without a competitor in slot 2, only a non-dual interpretation is possible (4d).

The last morphotactic complication I shall discuss pertains to the choice of classifier stem for dual non-sibling: if the dual non-sibling marker is found in slot 2, the singular classifier stem is used, however, when the non-sibling marker is displaced by competition with an object marker, the dual stem must be used instead.

(5) a. ba- ngintha- ngkardu -nu
1SGS.SEE(13).FUT DU.F see FUT
'We two (non-siblings) will see it/him/her.' (Nordlinger & Mansfield, 2021, 8)

b. nguba- nhi- ngkardu -nu -ngintha
1DUS.SEE(13).FUT 2SGO see FUT DU.F
'We two (non-siblings) will see you (sg.).' (Nordlinger & Mansfield, 2021, 8)

Taking stock of the discussion of empirical patterns, we have found three challenges in the morphotactics of Murrinh-Patha agreement morphology, all of which revolve around slot 2, the templatic position right-adjacent to the classifier stem, and which is the only position available to object person/number markers, a fact responsible for a good deal of competition.

1. Subject non-sibling dual markers obligatorily surface in this position, unless already occupied. The alternate realisation is slot 8.

2. Classifier stems display allomorphic variation depending on slot 2.

3. The subject agreement marker *ka* (dual sibling, paucal nonsibling) obligatorily surfaces in slot 2, if possible, but is dropped otherwise.

As argued by Nordlinger (2010) and Nordlinger & Mansfield (2021), the high degree of overlapping exponence, involving discontinuous surface positions provides evidence against a morpheme-based view, favouring instead a templatic realisational perspective.

In the next section, I shall present an analysis in IbM, a formal theory of the morphological template.

3 IbM

The analysis I am going to propose will be cast within Information-based Morphology (=IbM; (Crysmann & Bonami, 2016)), a theory of inflectional morphology closely inspired by HPSG (Pollard & Sag, 1987, 1994). From its inception, IbM was developed to address cases of variable morphotactics, such as the ones addressed in this paper, essentially advocating a neo-templatic view of affix order. The framework has since been applied to a number of complex morphotactic systems, including Oneida verbal morphology (Diaz et al., 2019), dependent multiple exponence in Batsbi (Crysmann, 2021a), and morphotactic competition in Yimas (Crysmann, 2020).

As discussed in detail in Crysmann (2021b), IbM assumes a set of realisation rules, organised in a Koenig/Jurafsky-style type hierarchy (Koenig & Jurafsky, 1994; Koenig, 1999): this means that in addition to vertical abstraction (=underspecification), dimensions permit horizontal abstraction by means of cross-classification of rule types in different dimensions. Rules are minimally pairings of morphosyntactic properties to be expressed (MUD) and the list of morphs (MPH) that serve as exponents.

In order to ensure that rules of exponence are actually applied (completeness) and do not over-apply (coherence), IbM imposes a very general well-formedness constraint that dictates that the set of rules being applied must "consume" the entire morphosyntactic property set (MS): in essence, non trivial set union of the MUD values to yield the entire MS set ensures completeness and coherence. Similarly, the sequence union or "shuffle" of the rules' morph contributions MPH must yield the word's morphs list MPS, respecting the order implied by the position class (PC) indices (see Bonami & Crysmann, 2013, for details). The particular choice of non-trivial ($\uplus$) over ordinary ($\cup$) set union ensures that no rule can be applied twice, which may otherwise result in unwarranted repetition of morphs.

(6)
$$word \rightarrow \begin{bmatrix} \textsc{mps} & \boxed{e_1} \bigcirc \ldots \bigcirc \boxed{e_n} \\ \textsc{ms} & \boxed{0}\ (\boxed{m_1} \uplus \cdots \uplus \boxed{m_n}) \\ \textsc{rr} & \left\{ \begin{bmatrix} \textsc{mph} & \boxed{e_1} \\ \textsc{mud} & \boxed{m_1} \\ \textsc{ms} & \boxed{0} \end{bmatrix}, \ldots, \begin{bmatrix} \textsc{mph} & \boxed{e_n} \\ \textsc{mud} & \boxed{m_n} \\ \textsc{ms} & \boxed{0} \end{bmatrix} \right\} \end{bmatrix}$$

Furthermore, the well-formedness constraint exposes the entire MS set to every

rule, such that rules can be (allomorphically) conditioned on properties they do not express themselves.

However, there is an asymmetry between form features and function features, in the context of rules: for morphosyntactic function, rules have access to both local (MUD) and global properties (MS). For form, however, there is only access to local properties (MPH). Incidentally, the early work on IbM (Crysmann & Bonami, 2016; Bonami & Crysmann, 2013) already made use of "pivot" features in order to capture placement relative to the edge, or to a designated element, such as the stem.

Entirely analogous to the MUD/MS distinction, we can easily expose the global morphotactic structure of the word (MPS) to the individual rules, for conditioning:

(7)
$$word \rightarrow \begin{bmatrix} \textsc{mps} & \boxed{1}\ (\boxed{e_1} \bigcirc \ldots \bigcirc \boxed{e_n}) \\ \textsc{ms} & \boxed{0}\ (\boxed{m_1} \cup \cdots \cup \boxed{m_n}) \\ \textsc{rr} & \left\{ \begin{bmatrix} \textsc{mph} & \boxed{e_1} \\ \textsc{mps} & \boxed{1} \\ \textsc{mud} & \boxed{m_1} \\ \textsc{ms} & \boxed{0} \end{bmatrix}, \ldots, \begin{bmatrix} \textsc{mph} & \boxed{e_n} \\ \textsc{mps} & \boxed{1} \\ \textsc{mud} & \boxed{m_n} \\ \textsc{ms} & \boxed{0} \end{bmatrix} \right\} \end{bmatrix}$$

This provides a general mechanism for morphotactic conditioning: in addition to referring to the edge (8) or the stem (9), it will be possible to insist that some other morphotactic position be filled.

(8) Second position placement
$$\begin{bmatrix} \textsc{mph} & \langle \boxed{1} \rangle \\ \textsc{mps} & \left\langle \left[\textsc{pc} \quad \boxed{i} \right], \boxed{1} \left[\textsc{pc} \quad \boxed{i}\!+\!1 \right], \ldots \right\rangle \end{bmatrix}$$

(9) Stem-relative placement
$$\begin{bmatrix} \textsc{mph} & \langle \boxed{1} \rangle \\ \textsc{mps} & \left\langle \ldots, \begin{bmatrix} \textit{stem} \\ \textsc{pc} \quad \boxed{i} \end{bmatrix}, \ldots, \boxed{1} \left[\textsc{pc} \quad \boxed{i}\!+\!2 \right], \ldots \right\rangle \end{bmatrix}$$

This possibility will be explored in the following analysis, crucially making reference to slot 2, the locus of morphotactic competition in Murrinh-Patha.

4 Analysis

In the first section, we have seen several morphological dependencies that hold in the Murrinh-Patha verbal template, pertaining to both exponence and morphotactics.

First, marking of non-siblings is jointly achieved by a classifier stem or object marker expressing person/number and a non-sibling marker expressing number/gender. Second, placement of dual non-sibling gender markers is morphotactically dependent on position 2 being filled, either by an object marker, or by the

dual non-sibling gender marker itself. Third, placement of the dual non-sibling gender marker has an impact on the choice of classifier stem (singular vs. dual stem). Fourth, the dual/paucal marker *ka* is in positional competition with the object markers.

4.1 Non-sibling marking

In a non-revised version of IbM, each of these dependencies would have been captured by cross-classifying underspecified rule descriptions to yield rather complex rules that simultaneously talk about up to three morphotactic positions. However, the system of Murrinh-Patha non-sibling marking is quite self-contained, so it will be worthwhile experimenting with potential ways to reduce complexity by separating the treatment of the gender markers from that of the classifier stems (and object markers, respectively).

To this end, I shall propose a slightly refined representation of agreement information, that systematically separates gender/number and person/number information. As shown by the type hierarchies in Figure 3, person/number features are appropriate of the general *agr* type, its subtypes distinguishing between subject and object agreement. Values for sibling-status distinguish between *sib* and *nsib*, with only the latter having GEND(ER) as an appropriate feature.

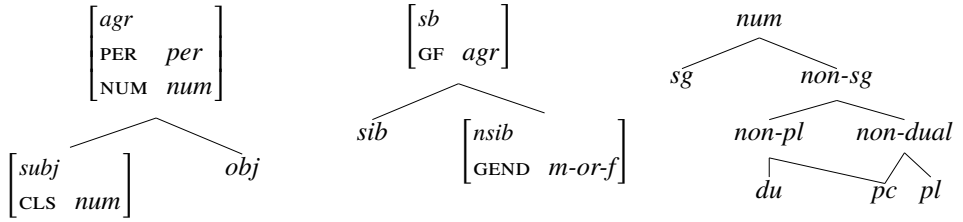


Figure 3: Agreement values

In order to link person/number and gender information, I shall propose to use the GF feature that embeds person/number agreement within sibling/nonsibling agreement[5] A sample MS representation for non-sibling dual 3rd person feminine subject is given in Figure 4:

Once a suitable MS representation is in place, like the one given in Figures 3 and 4, the rules for non-sibling gender marking are essentially quite straightforward, as shown in Figure 5. Note that the reentrancy between GF and person/number agreement makes it possible to refer to NUM information quite directly.

At the top, we find a most general statement about the morphotactics of the entire class of non-sibling gender markers (PC $2 \vee 8$), which is in turn narrowed down

[5]Other formalisations are of course conceivable. Minimally, ending grammatical function as type information would suffice. Maximally, one may even consider cyclic feature structures, such that sibling information embeds grammatical function and number and person/number information embeds sibling status.

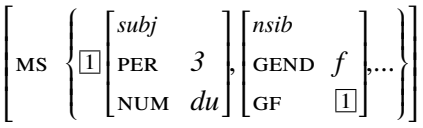


Figure 4: Sample MS representation for non-sibling dual 3rd person feminine subject

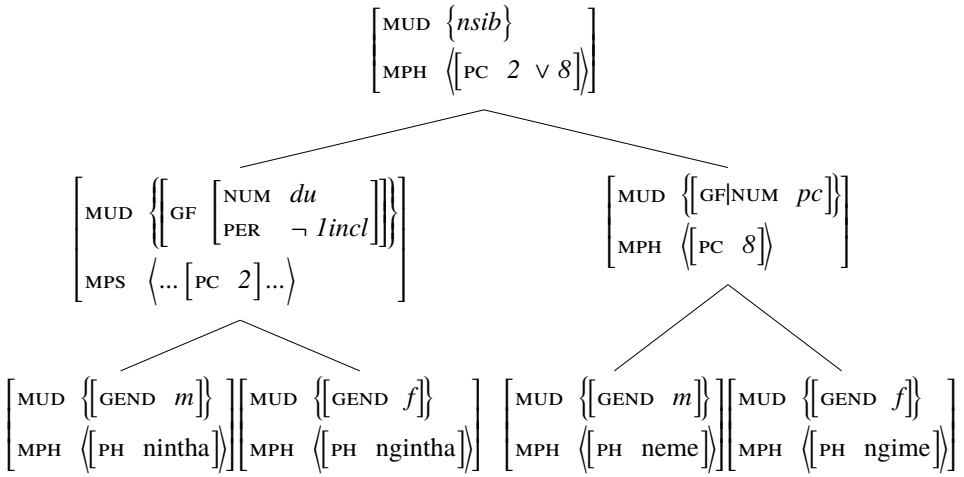


Figure 5: Non-sibling marking

for paucal and dual by its two immediate subtypes. While paucal markers are always in slot 8, dual markers retain the positional flexibility, yet require slot 2 to be filled, by way of the global morphotactic feature MPS. If there is some marker in slot 2, the dual marker will surface in slot 8, given that no two morphs can be assigned to the same positional index within a well-formed word. If, however there is not, placement of *nintha/ngintha* in slot 2 will be the only way to satisfy the constraint on the global morphs list MPS.[6] Note further that the rules are underspecified for grammatical function such that rule application can serve to narrow down to non-sibling referents the interpretation for either subject or object function.

A final remark is due regarding 1st person inclusive. As given by the paradigms in Tables 1 and 2, there is no overt marking of non-sibling dual in these cells, neither for object agreement nor for subject agreement. Nordlinger (2015) observes that the first person inclusive is characterised by a reduced paradigm, only distinguishing paucal non-sibling from all other cells. In the present analysis, the conspicuous absence of overt non-sibling marking is captured by a constraint regarding the person on the non-sibling dual rule types in Figure 5, thereby leaving zero exponence as the only option. Note that the rule type that is used for this is the same that serves

[6]Note that the analysis proposed here differs quite crucially from that in Stump (2022): here, placement of the non-sibling marker is directly conditioned on morphotactics, i.e. on slot 2 being filled. Stump, by contrasts conditions on the absence of object agreement properties, which he achieves by mapping third singular object agreement to the same representation as intransitives.

systematic zero exponence for dual sibling.[7]

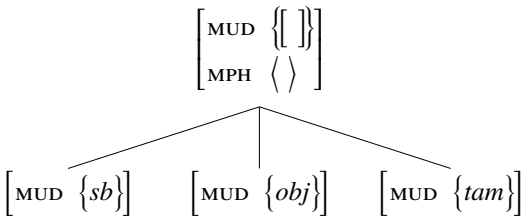


Figure 6: Rule types for zero exponence

4.2 Person/number marking

I shall now turn to the admittedly more complex hierarchy of rule types for person/number marking of core functions given in Figure 7. This hierarchy is organised into three orthogonal dimensions. As it is standard for IbM, fully expanded rules are obtained from this hierarchy by intersecting each leaf type from one dimension with each leaf type of every other dimension (Koenig & Jurafsky, 1994). Despite the complexity of the hierarchy, most of the properties postulated for the exponence rule types should be rather straightforward. E.g. the rule types in the STEM and SLOT-2 dimensions, which account for the bulk of rules in these dimensions, mostly pair the relevant morphosyntactic property with an exponent and its positional index.

Having laid out the overall shape of the hierarchy, I shall now zoom in to pairwise combinations of dimensions, focusing first on issues of stem selection, and then on properties expressed in the pivotal slot 2.

4.2.1 Classifier stems

The most straightforward dimension is STEMS: essentially, the rule types capture the expression of subject person number marking by specific forms of a classifier stem. Morphotactically, classifier stems are restricted to PC 1. Choice of stem form is further conditioned on TAM properties and, of course, lexical specification of the classifier stem (CS). Note that, with respect to subject marking, rule types in this dimension only restrict person in a direct fashion, while number is specified as a stem class property (CLS). This is necessary, given that stem selection and number do not always match up, thereby displaying the kind of morphomic properties we observed with singular stems being used for non-sibling duals.

Accordingly, the MORPHOMIX dimension associates morphomic CLS properties with actual number (NUM): while the general rule type just equates the two,

[7] As is standard in IbM, the denotation of the very general rules of zero exponence will be narrowed down under Paninian competition with more specific rules. See Crysmann (2021b) for in-depth discussion.

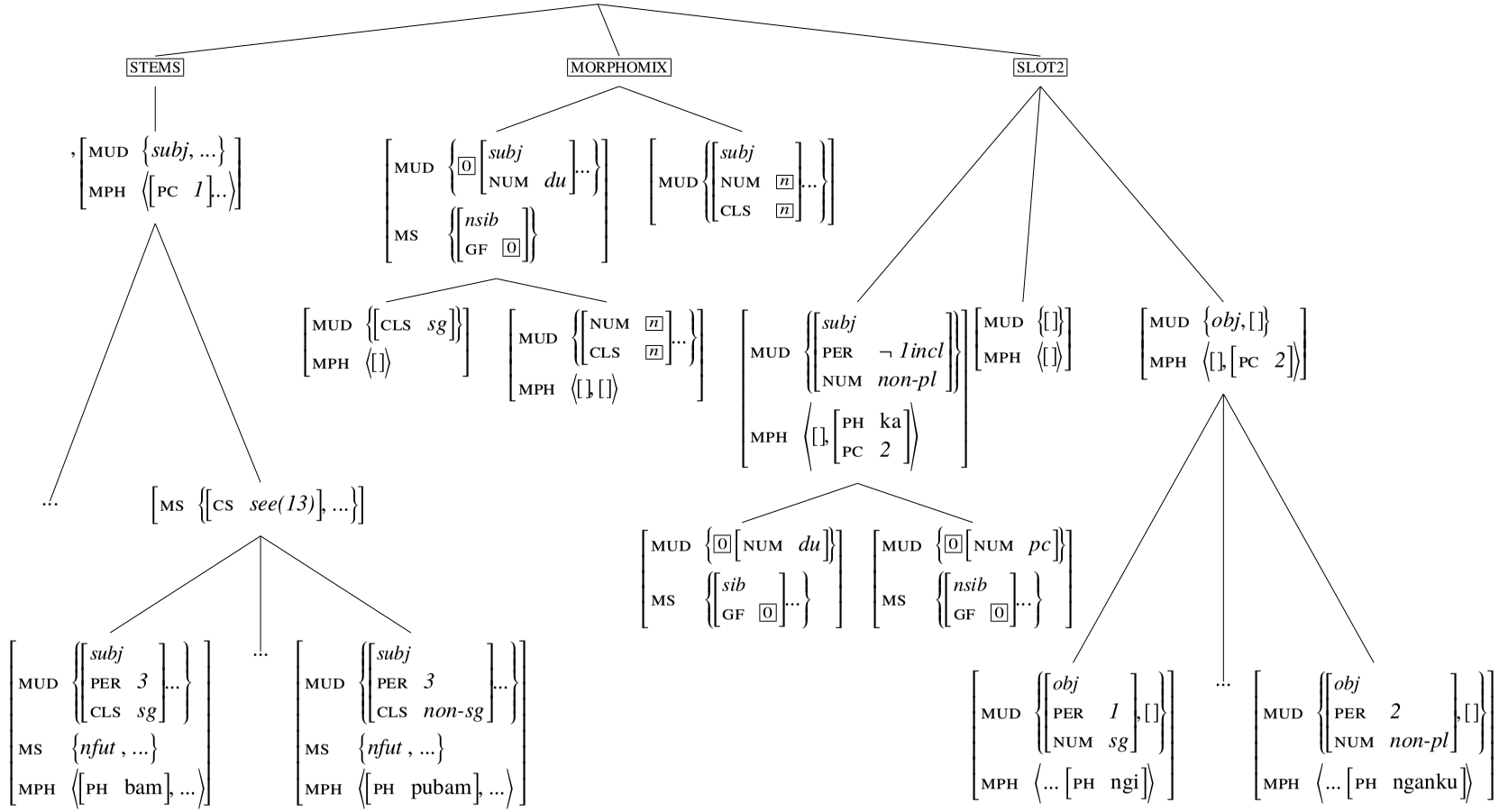


Figure 7: Person/number marking

there are specific rule types for non-sibling dual stems which also capture the morphotactic side effects: the rule type for choosing a singular stem is constrained to be mono-morphic, which means it will never unify with any rule types introducing another marker (in slot 2). As a consequence, slot 2 will be left free to host the non-sibling gender marker (cf. Figure 5). Conversely, the rule type for the more regular dual stem is bi-morphic, so it will only unify with rule types that are also bi-morphic, such as the ones for object markers. Since the object markers occupy slot 2, this means that the non-sibling gender marker will only be able to surface in slot 8. These two rule types thus account for the interaction between morphotactics and stem choice shown in (5).[8]

4.2.2 Slot 2

The SLOT-2 dimension on the right finally provides constraints on exponents in slot 2. On the very right, we find rules of exponence for object agreement. Crucially, these rules introduce a marker in slot 2, for every cell in the paradigm, except for third singular, which has zero exponence, cf. Figure 6.

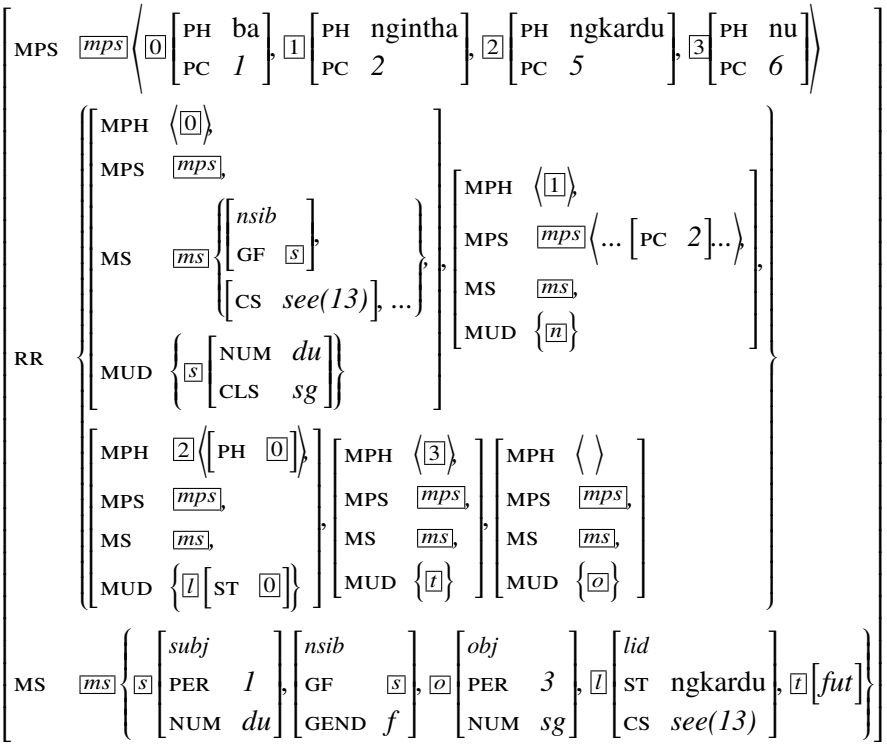
On the left are the exponence rule for the *ka* marker, where the supertype fixes shape and position and the two subtypes specify the feature combinations being expressed, i.e. the specific cases of paucal non-sibling and dual sibling. Finally, in the centre of this dimension, we find a rule type that serves as a target for any stem rule used without any of the aforementioned markers. Most crucially, it constrains the open MUD set and MPH list to each be of size 1.

4.3 Sample analyses

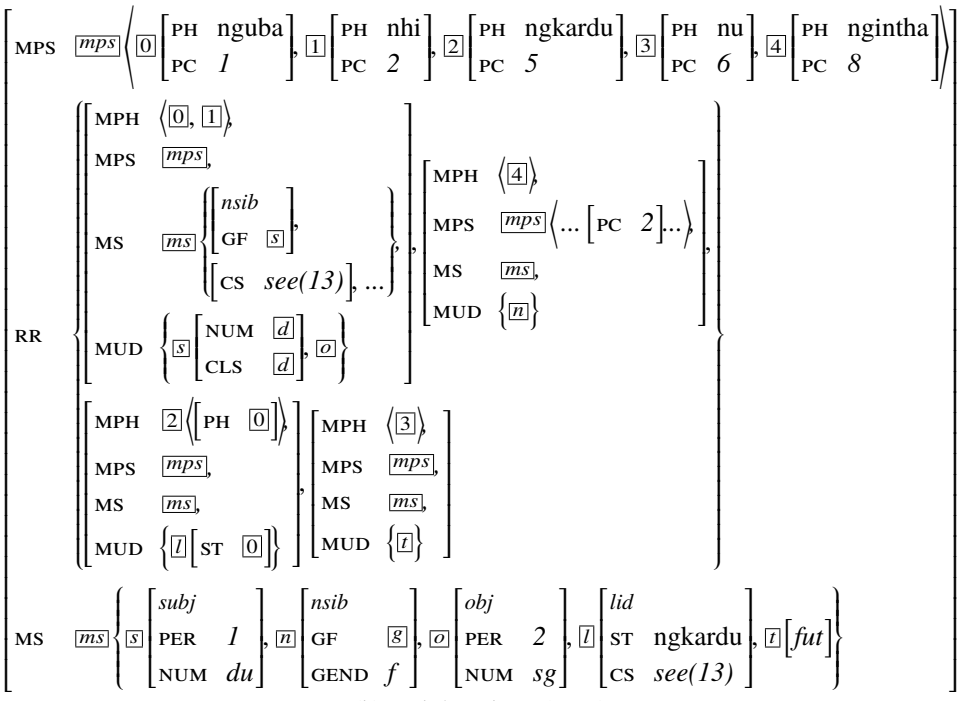
Now we have all the ingredients, we can see how they play together to derive the empirical patterns. Two morphotactic patterns are of concern here: the placement alternation of the non-sibling gender markers and the presence vs. absence of the *ka* marker.

Non-sibling dual The morphotactics of non-sibling marking are almost entirely contained within the rule hierarchy of the gender markers (Figure 5): the paucal markers, with their fixed position in slot 8 are trivial, but the mobile dual markers are dependent on a marker in slot 2, which we capture using the MPS feature. This latter condition for dual non-sibling can be met by any of the exponents introduced by a rule type from the SLOT-2 dimension in Figure 7. Yet, if none of these markers is present, the positionally flexible dual non-sibling marker itself will be the only one that can satisfy the requirement. These two situations correlate with stem selection: an object marker in slot 2 satisfies the morphotactic requirement for a dual non-sibling marker to surface in slot 8 and, by being part of a bi-morphic person-number marking rule, it selects the dual stem. Conversely, if no object marker is

[8] If one wants to rule in the marginal acceptability of a singular stem with a dual non-sibling marker in slot 8, all it takes is to remove the constraint that MPH to be mono-morphic.



(a) with 3sg object (zero exponence)



(b) with 2sg object (overt)

Figure 8: Derivations for non-sibling dual

present, only the mono-morphic dual non-sibling person/number rule can be selected, introducing the singular stem. And, as already stated above, slot 2 can and must be filled by the non-sibling gender marker in this case.

Figure 8 provides sample derivations of the contrasts given in (5) above.[9] I.e. Figure 8a provides the analysis for dual non-sibling subject acting on a third singular object, whereas Figure 8b provides the one with a second singular object, the crucial difference being that third singular object agreement has zero exponence, whereas second singular is expressed overtly in slot 2.

Recall from our discussion of Figure 7 above that we distinguished two rules for (non-sibling) dual classifier stem: one choosing the expected dual classifier stem, but requiring to combine with an object marker, and the exceptional mono-morphic rule, pairing dual number with a singular stem. Since there is no specific morphous rule for third singular objects, but only a zero exponence default, we get the exceptional singular classifier, as shown in Figure 8a. Furthermore, since zero exponence does not contribute any morphs (see the empty MPH list for object agreement in Figures 8a and 6), the non-sibling marker will be the only exponent that can fill its own requirement that position 2 be filled.

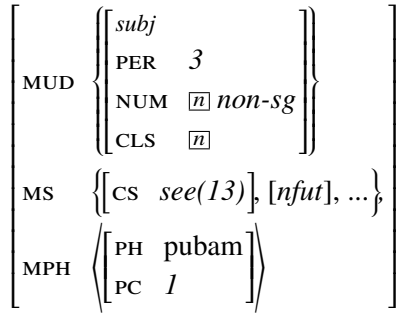
Conversely, second singular object agreement has a specific rule type, which will preempt default zero exponence by way of Panini's principle. Since rule types for object agreement in Figure 7 obligatorily combine with rule types for subject agreement, combination with mono-morphic rules is not viable. Therefore, we get a rule type that combines a first person dual classifier stem with a second singular object exponent. Since the exponent of object agreement occupies position 2, the global morphotactic requirement of the non-sibling marker is fulfilled. However, since position 2 is occupied now, the non-sibling marker must go into position 8.

Sibling dual *ka* The other morphotactically interesting case pertains to the *ka* marker. If no object marker is present, position 2 is available and the marker is obligatory in the sibling dual and the non-sibling paucal.

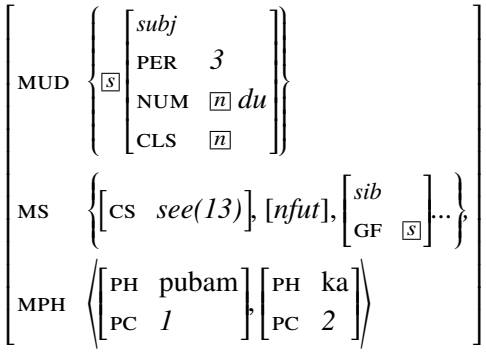
As can be verified from Figure 7, the combination of a non-singular stem rule type with any of the *ka*-rule types will be more specific, with respect to MUD and MS properties, than the mono-morphic classifier stem rule: as a result, Paninian competition will select the *ka*-inflected classifier stem over the bare one. The competing rules, both derived by cross-classification of leaf types in Figure 7 are given in Figure 9a,b. Thus, since the MUD and MS descriptions of the rule for the *ka*-marked classifier stem in Figure 9b are more specific than those for the bare classifier stem in Figure 9a, application of the more general bare rule is preempted by Paninian competition in exactly the cases described by the narrower *ka*-marked rule.

With a direct object marker in slot 2, the situation changes: since object markers equally combine with classifier stems into complex rules with equally complex

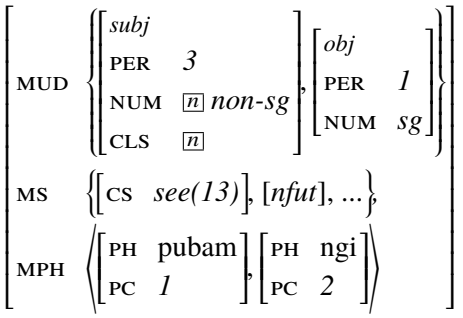
[9]The boxed coreference tags between the words' MS set and MPS list with the MUD, MPH, MS and MPS values of the rules (on RR) follow from the general well-formedness principle in (7). To ease readability, I use boxed letters for the morphosyntactic property set (corresponding to the property being realised), and boxed numbers for morphs (in consecutive order of appearance).



(a) Bare non-singular classifier stem (cf. (4d))



(b) *ka*-marked dual classifier stem (cf. (4a))



(c) Non-singular classifier stem with object agreement (cf. (4c))

Figure 9: Dual/paucal ka and rule competition

MUD values specifying both subject and object properties, they are not preempted by the *ka*-rule types via Paninian competition. As can be easily verified by comparing Figures 9b,c, neither rule's MUD and MS descriptions are more general or more specific than the other.

To summarise, rule combination by cross-classification achieves the correct behaviour with respect to Panini's principle here.

5 Conclusion

We have discussed complex morphotactic dependencies in Murrinh-Patha and shown how these can be modelled in IbM, a formal neo-templatic approach to morphology built on multiple inheritance hierarchies of type feature structures. The analysis of Murrinh-Patha has prompted me to revise the way relative placement can be addressed within IbM: in essence, specialised pivot features, as used in earlier work (Crysmann & Bonami, 2016; Bonami & Crysmann, 2013) have been generalised into a distinction between rule-local contributions of morphs and constraints on the word's global morphs list, a distinction that mirrors the one between properties a rule expresses (MUD value) and conditioning on the word's entire morphosyntactic

property set MS. The intricacy of morphotactic interactions in Murrinh-Patha have also helped to highlight that two different cases of morphotactic competition may require different answers: independent rules in case of pure morphotactic dependency on some slot being filled and complex rules built by cross-classification, to capture cases where morphotactic dependency interacts with Paninian competition.

References

Bonami, Olivier & Berthold Crysmann. 2013. Morphotactics in an information-based model of realisational morphology. In Stefan Müller (ed.), *Proceedings of the 20th conference on head-driven phrase structure grammar, berlin, 26–29 august*, 27–47.

Crysmann, Berthold. 2020. Morphotactic dependencies in Yimas. *Lingue e linguaggio* 19(1). 91–130.

Crysmann, Berthold. 2021a. Deconstructing exuberant exponence in Batsbi. In Berthold Crysmann & Manfred Sailer (eds.), *One-to-many relations in morphology, syntax and semantics* Empirically Oriented Theoretical Morphology and Syntax, 51–83. Berlin: Language Science Press.

Crysmann, Berthold. 2021b. Morphology. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-driven Phrase Structure Grammar: The handbook* (Empirically Oriented Theoretical Morphology and Syntax 9), chap. 21, 947–999. Berlin: Language Science Press.

Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics in Information-based Morphology. *Journal of Linguistics* 52(2). 311–374. doi: 10.1017/S0022226715000018.

Diaz, Thomas, Jean-Pierre Koenig & Karin Michelson. 2019. Oneida prepronominal prefixes in Information-based Morphology. *Morphology*, First Online, 1–43.

Koenig, Jean-Pierre. 1999. *Lexical relations*. Stanford: CSLI publications.

Koenig, Jean-Pierre & Daniel Jurafsky. 1994. Type underspecification and online type construction in the lexicon. In *West coast conference on formal linguistics*, vol. 13, 270–285. Stanford: CSLI Publications/SLA.

Mansfield, John. 2020. *Murrinhpatha mrphology and phonology*. Berlin: Mouton de Gruyter.

Nordlinger, Rachel. 2010. Verbal morphology in Murrinh-Patha: Evidence for templates. *Morphology* 20. 321–341.

Nordlinger, Rachel. 2015. Inflection in Murrinh-Patha. In Matthew Baerman (ed.), *Handbook of inflection*, Oxford: Oxford University Press.

Nordlinger, Rachel & John Mansfield. 2021. Positional dependency in murrinhpatha: expanding the typology of non-canonical morphotactics. *Linguistic Vanguard* 7(1).

Pollard, Carl & Ivan Sag. 1987. *Information–based syntax and semantics*, vol. 1. Stanford: CSLI.

Pollard, Carl & Ivan Sag. 1994. *Head–Driven Phrase Structure Grammar*. Stanford: CSLI and University of Chicago Press.

Salehi, Ali & Jean-Pierre Koenig. 2023. Argument realization and argument referencing in Soranî Kurdish. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 100–120. Frankfurt/Main: University Library.

Stump, Gregory T. 2022. *Morphotactics: A rule-combining approach*. Cambridge University Press.

Second position without movement: Enclitic particles in Passamaquoddy

Philip S. LeSourd
Indiana University

Proceedings of the 30th International Conference on
Head-Driven Phrase Structure Grammar

University of Massachusetts Amherst

Stefan Müller, Elodie Winckel (Editors)

2023

Frankfurt/Main: University Library

pages 44–57

LeSourd, Philip S. 2023. Second position without movement: Enclitic particles in Passamaquoddy. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 44–57. Frankfurt/Main: University Library.
DOI: 10.21248/hpsg.2023.3.

Abstract

Maliseet-Passamaquoddy (Algonquian, New Brunswick and Maine, MP) employs a set of enclitic particles to express tense, aspect, and various adverbial notions. These occupy second position in a clause: they follow either the first word in the clause or the first constituent. Johnson and Rosen (2015) propose an analysis of clitic placement in Menominee (Algonquian, Wisconsin) that takes clitics to occupy a functional head in the left periphery, postulating movement of one item into a specifier position to the left of this functional head, thus leaving the clitics in second position. Here I propose an alternative account for MP in the framework of Sign-Based Phrase Structure Grammar (Sag 2012) that makes no use of functional heads and postulates no movement operations. Instead, clitic positions are determined by a small number of maximally simple constructional statements.

1 Second-position enclitics in Maliseet-Passamaquoddy

Like many other Algonquian languages, Maliseet-Passamaquoddy (MP, New Brunswick and Maine) employs a set of enclitic particles that are stationed in second position in a clause. These may follow the first word of the clause (1a): second word placement. Less often, they follow the first phrase in the clause (1b): second daughter placement.[1]

(1) (a) [$_{\text{AdvP}}$ Kàt=**ona** qìn] cipok-eltù-wi-yol
not=also really intense-be.much-NEG-IN.PL

[1] *Notation:* c = /č/, q = /k^{w}/, o = /ə/, ’ = word-initial /h/ before C. Acute and grave accents mark distinctively high- and low-pitched stressed syllables, respectively. An equals sign (=) marks a clitic boundary. An m-dash (–) marks the boundary between a preverb and the verb or preverb-verb complex that it modifies.

pskihq-ís-ol.
grass-DIM-IN.PL[2]
'There is also not really a whole lot of grass.' (Maliseet)

(b) [$_{AdvP}$ Kàt qìn]=**yaq**=**ona** nokom-okil-ù.
not really=REPORT=also fairly-be.size-(3)-NEG
'And he was not really very big, they say.' (Maliseet)

Transformational analyses of second-position enclitics in languages such as Serbian and Croatian (see Diesing and Zec 2017 for a recent study) typically suppose that clitics occupy a functional head in the left periphery of the clause and that either a word or a phrase is then moved into the specifier of this functional head. An analysis along these lines is proposed for several Algonquian languages, with a focus on Menominee (Wisconsin), by Johnson and Rosen (2015). More on their approach shortly.

I will instead pursue a constraint-based analysis formulated within Sign-Based Construction Grammar (SBCG, Sag 2012). Three constructions are needed to account for the distribution of enclitics:

[2] *Abbreviations used in glosses:* 1 first person; 3 third person; AN, an. animate; AOR aorist; COND conditional; CONJ conjunct; DIM diminutive; DIR direct; EMPH emphatic; exc. exclusive; IN inanimate; MPL multi-plural (the subject of the verb refers to three or more individuals); N suffix *-(o)n(e)-* (several functions); NEG negative; OBV obviative; PL, pl. plural; PROX proximate; REPORT reportative; SG singular; UNC uncertain. Glosses are given in parentheses for morphemes that have no surface segmental shape.

1. A ***clitic-second-word-construction*** that states that a clitic or clitic group may follow the first (prosodic) word in a clause.
2. A ***clitic-second-daughter-construction*** that states that a clitic or clitic group may follow the first constituent in a clause.
3. A ***clitic-compacting-construction*** that forms clitic groups consisting of one or more enclitics.

I formalize these constructions below. But first, some preliminaries require attention.

2 The enclitics

The second-position particles of MP are given in (2). They are accented in utterance-final position, unaccented otherwise.

(2)

=àl 'uncertain'	*=lú* 'but, however'
=éhta 'indeed, in fact'	*=ná* 'also, as for X'
=kàhk 'but, however, certainly'	*=òc* ~ *=hc* 'future'
=kàl 'probably'	*=òp* ~ *=hp* 'would'
=kéte 'for example, moreover, thus'	*=tàhk* 'lo and behold'
=yáka 'afterward, furthermore'	*=yàq* 'they say, it is said'
=hk (idiomatic)	

Note that the items in (2) are semantically diverse: they include future and conditional markers, a reportative particle, a mirative marker, several adverbials, and particles indicating contrast and emphasis.

In addition, two conjunctions, *kenùk* 'but, however' and *cèl* 'and, moreover', may either introduce a clause or appear in second position, where they pattern like the items in (2).

3 Locating the left edge of the clause

Clitics are stationed with respect the left edge of the clause. But there is more than one such edge in some clauses. One or more phrases may be left-adjoined to the clause, with the result that there is sometimes more than one clausal boundary that can serve as a site for clitic placement.

An example with an adjoined AdvP is given in (3).

(3) [S **Malom=ehta=cel** [S yùt=**yaq**
finally=indeed=moreover this.IN=REPORT
mulahkepolásu ktáhkomiq
hole.be.trampled-(3) land
eli–kis-ká-hti-t.]]
thus–past-dance-PROX.PL-3AN
'Moreover, they danced there for so long in the end, they say, that a depression was trampled into the ground.' (Maliseet)

4 Discontinuous constituents: two analyses

Second-position particles freely occur between the words of constituents: an AdvP in (1a), an NP in (4).

(4) [NP Yùkk**=yaq=olu** kótok-ik kukéc-ok]
these=REPORT=but other-PROX.PL game.warden-PROX.PL
etuci–palitahas-ultí-hti-t nemiy-á-hti-t
to.extent–be.pleased-MPL-PROX.PL-3AN see-DIR-PROX.PL-3AN
w-itapé-wa-l…
3-friend-PROX.PL-OBV.SG
'But, they say, these other game wardens were very happy when they saw their friend…' (Passamaquoddy)

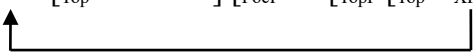
Alternatively, we might suppose that enclitics do not appear WITHIN constituents in such cases. Rather, the constituents they appear to interrupt might be DISCONTINUOUS.

MP does, in fact, permit the discontinuous expression of a wide variety of constituents, independently of second-position phenomena. In (5), for example, the demonstrative ***nòt*** 'that (an.)' is separated from the noun ***'puwìn*** 'corpse, body' that it modifies by an adverb and the verb of the clause.

(5) Mahkiyew-òss **[NPa nòt]** àpc mete-htéhsi-t
soon-DIM that.PROX again heard-fall-3AN
[NPb 'puwìn].
corpse
'After a little while **[the body]** was heard to fall again.'
(Passamaquoddy)

Johnson and Rosen (2015) attribute all discontinuity in the expression of Algonquian phrases to movement, including cases in which a clitic is stationed between segments of a phrase. For Menominee, they assume that a second-position clitic occupies a functional head, typically the head of Topic Phrase or Focus Phrase, at the left periphery of the clause. One word or a single constituent may be moved into the specifier of this head to satisfy the enclitic's requirement for a host. This puts the clitic into second position. If a segment of a constituent is left behind, a discontinuous constituent is the result. Note that they allow TopP and FocP to be iterated, as in (6a). The operations in question proceed as shown in (6b).

(6) a. [TopP … [FocP … TopP …]]]
(Johnson and Rosen 2015:142)
b. [$_{TopP}$ XPa [$_{Top°}$ = Clitic] [$_{FocP}$ … [$_{TopP}$ [$_{Top°}$ t_{XPa} XPb] …]]]



A Menominee example under Johnson and Rosen's analysis is shown in (7):

(7) [$_{TopP}$ [$_{D}$ **Ayom**] [$_{Top°}$ [$_{\&}$ **=taeh**]] [$_{FocP}$ [$_{Foc°}$ Ø] [$_{TopP}$ [$_{NP}$ [$_{D}$ t_{ayom}]
this.AN =and
owōhnema [$_{Top°}$ Ø] [$_{\&P}$ [$_{\&}$ t_{taeh}] [$_{TP}$'s osēqtahnacen
father AOR prepare.3/3OBV.CONJ
onīcianaesan 's maek-mesāhkataewāēnet]]]]…
his.child.OBV AOR while.fast.3OBV.CONJ
'And as this father prepared for his child's fast…' (Menominee, Johnson and Rosen 2015:145, simplified)

On this analysis, the enclitic conjunction *=taeh* 'and' initially occupies the position of head of &P, low in the clausal spine. It undergoes raising to become the head of TopP in the left periphery. ***Ayom owōhnema*** 'this father' is raised from subject position (not shown) to specifier position in an inner TopP, then its determiner ***ayom*** 'this (an.)' is raised again to specifier postion in a second, higher TopP, providing a host for the enclitic conjunction.

5 Against movement

Several problems arise if we try to adopt Johnson and Rosen's approach for MP. I will review just one here: second-position clitics may be stationed in MP in two locations with respect to the same phrase. Consider (8) in this connection.

(8) [$_{NP}$ Yùkt**=olu** wasís-ok]**=yaq**
these.PROX=but child-PROX.PL=REPORT
'totoli–tokom-á-wa-l.
(3)-ongoing–hit-DIR-PROX.PL-OBV.SG
'But the children, they say, were hitting him.' (Maliseet)

The reportative enclitic ***=yaq*** has been positioned after the clause-initial NP in this example by second-daughter placement. Thus, this NP must be intact; IT CANNOT BE DISCONTINUOUS.

It follows that ***=olu*** 'but' truly interrupts the bracketed NP in (8). It is not attached to the first segment of a discontinuous NP, one that is located in the Specifier of a functional projection that is headed by ***=olu***. There is accordingly no reason to suppose that movement has taken place in the derivation of (8): ***yùkt*** 'these' is not a separate constituent that has moved away from ***wasísok*** 'children' so as to constitute a host for ***=olu***. An analysis in Johnson and Rosen's terms is excluded.

6 A non-movement analysis: background

I propose instead an analysis that makes use neither of abstract underlying forms nor of movement. We can account for clitic placement in MP if we adopt a set of three word-order constructions, adapting the mechanisms of Wetta's (2011, 2014) analysis of verb-second phenomena, which is stated in the framework of Sign-Based Construction Grammar (SBCG, Sag 2012).

I extend SBCG to include the Linearization Theory of Reape 1994. Following Reape, I assume that each *sign* is specified for a feature DOMAIN (DOM or D), which is specified in turn for a list of DOMAIN ITEMS: the members of the domain. These are the sign's constituents.

I further assume (with Wetta) that each domain item is assigned one of two values of the feature LIN (for linearization): *fixed* (*fix*) or *flexible* (*flex)*. The second value is assigned by default: a sign is specified [LIN *flex*] unless some rule or principle states otherwise. Constructional statements may specify where a [LIN *fix*] element occurs in a structure. This is what our word-order constructions will do.

7 Putting these tools to work

A preliminary example: in (9), a single enclitic follows the first constituent in a clause.

(9) Nekòm=**ona** tol-ahsuwásu.
s/he=also ongoing-plan-(3)
'She also is making plans.' (Passamaquoddy)

Suppose for the moment that all of the enclitics of MP are lexically specified as [LIN *fix*], while all other syntactic expressions are specified as [LIN *flex*], by default. Further suppose that the grammar includes a constructional rule that states that one [LIN *fix*] element may follow a single [LIN *flex*] constituent at the beginning of a clause. This is the *clitic-second-daughter-construction* (*clitic-2D-cxt*), (10).

(10) *clitic-2D-cxt* ⇒

$$\left[\begin{array}{l}\text{MTR}\ \ [\text{SYN}\ [\text{CAT}\ \text{S}]] \\ \text{DTRS} < [\text{D} <[\text{LIN}\ \textit{flex}]>] \oplus [\text{D} <[\text{LIN}\ \textit{fix}]>] \oplus [\text{D} <[\text{LIN}\ \textit{flex}]>]_{o}>\end{array}\right]$$

This rule states that the mother (MTR) of the construction (of category S, a clause) consists of a concatenation (⊕) of domain items (D). The first of these is specified as [LIN *flex*]: it can be a constituent of any kind other than an enclitic. But the second is specified as [LIN *fix*]: it MUST be an enclitic. Any number of non-enclitic items may follow the enclitic within the clause. Thus, (10) is a template for a clause in which a clitic occupies the position following a single initial constituent.

Example (9) is analyzed by the construction in (10) as shown in (11).

(11) [DOM < [LIN *flex*] >] [DOM < [LIN *fix*] >] [DOM < [LIN *flex*] >]
[S Nekòm **=ona** tol-ahsuwásu.]
s/he =also ongoing-plan-(3)
'She also is making plans.'

The initial one-word phrase ***nekòm*** 'she' matches the initial [LIN *flex*] domain item specified in the construction. The enclitic ***=ona*** 'also' matches the specified [LIN *fix*] domain item. The verb ***tolahsuwásu*** 'she is making plans' is additional non-clitic material that the construction permits. Since this arrangement of material is sanctioned, the sentence as a whole is sanctioned.

8 Compaction: allowing for clitic groups

So far, we have allowed only for a single enclitic to appear in a clause. But combinations of two or more enclitics routinely appear together in second position, as illustrated in (12).

(12) Yùkt**=kahk=al=lu** tamà l-apàsu-w-ok.
these.AN=EMPH=UNC=but somewhere thus-pl.walk-3-PROX.PL
'But these (people) must surely be going somewhere.' (Maliseet)

Compaction (Kathol 2000:100; Wetta 2011:59) is a mechanism for forming a single domain item from a set of constituent domain items. For our analysis of MP clitics, we will use compaction to create clitic groups: these are single domain items that have one or more clitics as their constituents.

I postulate a *clitic-compacting-construction* (*clitic-comp-cxt*), as shown in (13).

(13) *clitic-comp-cxt* ⇒
$$\begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{DOM} & \left\langle \begin{bmatrix} \text{LIN} & [\,\textit{fixed}\,] \\ \text{FORM} & < \Phi\,(\text{L}) > \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{DTRS} & < \text{L: list}\,([\textit{clitic} +]) > \end{bmatrix}$$

Rule (13) states that the mother of the compacting construction (the compacted set of domain items) is itself a single domain item (a clitic group), that this is specified as [LIN *fixed*], and that it has as its constituents a set of (one or more) clitics that appear in the order specified by the function Φ.

The order of enclitics in a clitic group is relatively free in MP. In some closely related languages (such as Western Abenaki, LeSourd 2015:311–312), it is more nearly fixed. The feature [clitic+] that is employed here is simply shorthand for whatever property of the items in question causes them to require a host. (It should be noted that [clitic+] is not equivalent to a requirement that an item should appear in second position. The emphatic enclitic =*ote* equally requires a host, but may occur in any position in a clause.)

This formulation of compaction has a welcome consequence. We may now drop the assumption that clitics are lexically specified as [LIN *fix*]. It is the clitic group AS A WHOLE that is specified as [LIN *fix*] —and this assignment is made by the *clitic-compacting-construction* (13). Of course, the clitic group may consist of a single enclitic. But no lexical specifications for the feature LIN are required.

The *clitic-second-daughter-construction* (10), repeated below, now has the effect that the ENTIRE CLITIC GROUP occurs as a unit after the first constituent in a clause.

(10) *clitic-2D-cxt* ⇒

$$\begin{bmatrix} \text{MTR} \ [\text{SYN} \ [\text{CAT S}]] \\ \\ \text{DTRS} < [\text{D} <[\text{LIN } \textit{flex}]>] \oplus [\text{D} <[\text{LIN } \textit{fix}]>] \oplus [\text{D} <[\text{LIN } \textit{flex}]>]_{\circ} > \end{bmatrix}$$

9 The clitic-second-word-construction

Second-word clitic placement is considerably more common than second-daughter placement. Let us see how this mode of clitic placement may be formalized.

The evidence is not overwhelming, but second-word placement appears to be conditioned by prosody: the enclitic is stationed after the first prosodic word (ω) in the clause, as shown in (14).

(14) *clitic-2W-cxt* ⇒

$$\left[\begin{array}{l}\text{MTR}\ \ [\text{SYN}\ [\text{CAT}\ \text{S}]] \\ \text{DTRS} < [\text{DOM} < [\text{PHON} < \omega >] >] \oplus [\text{DOM} < [\text{LIN}\ \textit{fix}] >] \\ \qquad\qquad \oplus\ [\text{D} < [\text{LIN}\ \textit{flex}] >]_{\text{o}} >]\end{array}\right]$$

This rule states that one [LIN *fix*] item (a clitic group) may follow a clause-initial domain item that is specified as consisting of a single prosodic word **ω**. The clitic group may be followed by any number of non-clitic constituents.

Evidence for this formulation of the *clitic-2D-cxt* comes from two idiomatic expressions that are based on adverbial particles that include enclitics that are not part of an ordinary clitic group, as shown in (15).

(15) a. *tàn* 'such, how'
mèc 'still, yet'
b. *tàn=op=al* 'however'
mèc=op=al 'please; would it be possible?'

The conditional clitic ***=op*** may be repeated after the expressions in (15b), doubling the occurrence of this clitic that forms part of the idiom. Examples are given below in (16). Only in such cases are clitics ever repeated within a clitic sequence.

This situation makes sense if the second set of enclitics are stationed not with respect to the apparent adverbial base, but rather with respect to a base consisting of the adverb plus the inner enclitics. This is to say that the outer clitics do not follow the first SYNTACTIC word in the clause—the adverb. Rather, they follow the first PROSODIC word—formed by adding the inner enclitics to the adverb.

(16) *Clitic placement after the first phonological word in a clause*
a. [ω [ω Tàn=op=al]**=op=olu**]
how=COND=UNC=COND=but
't-oli–kisi–'sotuw-á-ni-ya
3-thus–able–understand-DIR-N-PROX.PL
kecciya-lí-c-il skicinúw-ol?
pure-OBV-3AN-OBV.SG Indian-OBV.SG

‘But how could they determine what a full-blooded Indian is?’ (Passamaquoddy)

b. [ω [ω Mèc=op=al]=**op**] nt-api–wikuwamkóm-a-n
still=COND=UNC=COND 1-go–visit-DIR-N
n-uhkomoss-òn?
1-grandmother-1PL
‘Could we please go see our (exc.) grandmother? (Passamaquoddy)

Diesing and Zec (2017) reach a similar conclusion in their analysis of Serbian: there is a phonological component to the placement of second-position enclitics in the language.

10 Combining constructions

We have seen that enclitics may occur both after the first word and after the first constituent in the same clause, as in example (8), repeated here.

(8) [NP Yùkt=**olu** wasís-ok]=**yaq**
these.PROX=but child-PROX.PL=REPORT
’totoli–tokom-á-wa-l.
(3)-ongoing=hit-DIR-PROX.PL-OBV.SG
‘But the children, they say, were hitting him.’ (Maliseet)

That this situation should be possible is in fact PREDICTED by the analysis stated here: both of our clitic-placement constructions may be instantiated in the same clause.

On one parse, the Clitic Second Word Construction (14) analyzes ***yùkt*** ‘these’ as the host for a second-position enclitic, here ***=olu***. At the same time, the Clitic Second Daughter Construction (10) analyzes the entire initial NP ***yùkt wasísok*** ‘these children’ as a clitic host, here for ***=yaq*** ‘reportative’. The two analyses are fully compatible, and both are licensed in the same structure.

11 Conclusions

The proposed analysis accounts for the distribution of second-position enclitics in Maliseet-Passamaquoddy with a minimum number of constructional statements:

1) The ***clitic-second-word-construction*** (14): A clitic group may follow the first prosodic word in a clause.
2) The ***clitic-second-daughter-construction*** (10): A clitic group may follow the first constituent in a clause.
3) The ***clitic-compacting-construction*** (13): A single domain item (a clitic group) may be formed from a (possibly singleton) set of enclitics. Clitic groups formed in this way are specified as [LIN *fixed*], the only items in the language with this property.

This account of second-position phenomena in MP is as spare as an account can be, since it corresponds directly to the observed facts: second-position particles may follow the first word of a clause, or they may follow the first constituent in the clause, and enclitics may occur in clitic groups. The analysis makes no appeal to the properties or distribution of functional heads. It makes no appeal to movement operations of any kind.

It is worth noting as well that the theoretical devices that I have adapted from Wetta's (2011, 2014) work were not developed for the analysis of clitics, but for verb-second phenomena. Thus, my analysis of Maliseet-Passamaquoddy enclitics is appropriately seen as offering support for a larger research program that takes word-order constructions to play a central role in syntactic analysis.

Acknowledgments

I owe a great debt to many Maliseet and Passamaquoddy people who have helped me to understand the material discussed in this article. I must especially express my thanks to Maliseet speakers Victor Atwin of Kingsclear, NB, and Peter Paul of Woodstock, NB, and to Passamaquoddy speakers Margaret Apt and David Francis, Sr., of Pleasant Point, ME, as well as Anna Harnois, Estelle Neptune, and Wayne

Newell of Indian Township, ME. With the exception of Margaret Apt, all of these language experts have now passed away.

References

Diesing, Molly, and Draga Zec. 2017. Getting in the first word: Prosody and predicate initial sentences in Serbian. *Glossa: A Journal of General Linguistics* 2:1–25.

Johnson, Meredith, and Bryan Rosen. 2015. The syntax of discontinuous noun phrases in Algonquian languages: Left branch extractions and focus movements. *In* Monica Macaulay and J. Randolph Valentine, eds., *Papers of the Forty-Third Algonquian Conference*, 135–153. Albany, NY: SUNY Press.

Kathol, Andreas. 2000. Linear Syntax. Oxford: Oxford University Press.

LeSourd, Philip S. 2015. Enclitic particles in Western Abenaki: Form and function. *International Journal of American Linguistics* 81: 301–335.

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 (CSLI Lecture Notes 46), 151–198. Stanford, CA: CSLI Publications.

Sag, Ivan A. 2012. Sign-Based Construction Grammar: An informal synopsis. *In* Hans C. Boas and Ivan A. Sag, eds., *Sign-Based Construction Grammar,* 61–197. Stanford, CA: CSLI Publications.

Wetta, Andrew C. 2011. A construction-based cross-linguistic analysis of V2 word order. *In* Stefan Müller, ed., *Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar*, 248–268. Stanford, CA: CSLI Publications.

Wetta, Andrew C. 2014. Construction-based approaches to flexible word order. Ph.D. dissertation, University at Buffalo, State University of New York.

How things become red in Mandarin Chinese? A case study of deadjectival change of state predicates

Yanru Lu
Humboldt-Universität zu Berlin

Chenyuan Deng
Humboldt-Universität zu Berlin

Jian Ma
Humboldt-Universität zu Berlin

Proceedings of the 30th International Conference on
Head-Driven Phrase Structure Grammar

University of Massachusetts Amherst

Stefan Müller, Elodie Winckel (Editors)

2023

Frankfurt/Main: University Library

pages 58–78

Keywords: change of state predicate, deadjectival verb, Resultative Verb Compound, Mandarin Chinese, HPSG

Lu, Yanru, Chenyuan Deng & Jian Ma. 2023. How things become red in Mandarin Chinese? A case study of deadjectival change of state predicates. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 58–78. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2023.4.

Abstract

This paper provides an HPSG analysis for the morphosyntax and the semantics of deadjectival change of state (CoS) verbs in Mandarin Chinese. We first show that adjectives are a distinct word class from verbs in Mandarin Chinese and argue for the derivation of CoS verbs from property concept adjectives. We then model this derivation with a lexical rule. Finally, since CoS verbs can be combined with another verb to form a resultative verb compound (RVC) to express caused CoS, we also propose a lexical rule to account for RVCs.

1 Introduction

Languages may systematically derive change of state (CoS) and caused CoS from property concept state (PC, cf. Dixon 1982; Levin 1993, Koontz-Garboden 2005, Tham 2013, Beavers et al. 2017, among others), e.g., *loose* → *loosen* in English (examples from Koontz-Garboden 2005: 83):

(1) a. The knot loosened. (non-causative CoS)
b. Sandy loosened the knot. (causative CoS)
c. The knot is loose. (state)

In (1), both causative and non-causative CoS verbs *loosen* are derived from the state predicate *loose*, which is categorized as an adjective in English. Tham (2013) proposes a paradigm for this derivation from PC states to non-causative CoS and to caused CoS:

(2) Paradigm of Derivation from State to CoS (Tham 2013: 652, simplified[1])

	State Adj/Verb/Noun	CoS Verb	Caused CoS Verb
Property concept state-based	.....	.....	›

Note that the basic state words differ in distinct categories among languages, e.g., adjective in English as in (1c), or verb in Tongan (Koontz-Garboden 2005), or noun in Warlpiri (Wetzer 1992), etc. Additionally, since not all languages present a complete derivation from state to CoS and to caused CoS as English does, Tham displays the arrow with a dashed line.

Mandarin Chinese shows a similar derivation from PC state to CoS verb and, however, not directly to caused CoS verb in the same form, but to resultative verb compound (RVC) by combining with a new verb. We take *hóng* 'red', a common PC describing a kind of color, as an example for PC state in Mandarin Chinese. Unlike in English, it is unacceptable to use a basic state lexeme directly as a transitive verb

[1]Tham (2013: 652) also presents a (caused) CoS based deverbal derivation, i.e. a derivation from caused CoS verb to non-causative CoS verb and to state. This is not the focus of our study and is not represented in the paradigm.

in Mandarin Chinese (cf. (1b) and (3c)). In most cases, caused CoS can only be realized by RVC instead (Tham 2013: 653–654), as in (3d).[2]

(3) a. Mén hěn hóng. (State)
door very red
'The door is (very) red.'

b. Mén hóng-le. (CoS)
door red-PFV
'The door reddened.'

c. * Zhāngsān hóng-le mén. (Caused CoS)
Zhangsan red-PFV door
Intended: 'Zhangsan reddened the door.'

d. Zhāngsān shuā-hóng-le mén. (Caused CoS, RVC)
Zhangsan brush-red-PFV door
'Zhangsan brushed the door red.'

Due to the lack of morphological marking in derivation in Mandarin Chinese, the issue now is whether the state word in this language is an adjective only expressing a state, or a verb which can denote a state and a CoS. That is, *hóng* in (3a) is an adjective and the CoS verb *hóng* in (3b) is derived from it, or it is an intransitive verb, just being able to convey a state or a CoS in different structures. Indeed, whether Mandarin Chinese has the adjective word class remains controversial in previous studies. To this question, we follow Tham's (2013) assumption, that Mandarin Chinese does possess adjective as an independent category from verb and derives CoS verbs systematically from PC adjectives, instead of resorting to pragmatic coercion as being proposed for Tongan by Koontz-Garboden (2007). Based on our data, we argue that the basic state *hóng* in (3a) is an adjective 'red' and *hóng* in (3b) is a deadjectival CoS verb 'redden'. Further arguments will be given in Section 2.1.

It is worth mentioning that the perfective marker *le* needs to be attached to both causative and non-causative CoS verbs, cf. (3b) and (3d). However, the CoS is not ascribed to the perfective marker *le*. We argue that the derived verb *hóng* 'redden' has already a CoS interpretation, consistent with Tham (2013: 663) but at odds with Smith (1997: 265), who analyzes the derived verb as state and treats CoS as a "dynamic, shifted interpretation".

Our study aims to provide an analysis of this derivation in Mandarin Chinese, i.e., from PC state to non-causative CoS, and to caused CoS RVC. Taking the 'red → redden' counterparts as an example, i.e., *hóng → hóng(-le) → V-hóng(-le)*, We provide a formal account for the morphosyntax and semantics of deadjectival CoS predicates in Mandarin Chinese in the framework of Head-driven Phrase Structure Grammar (HPSG, Pollard & Sag 1994, Sag 1997, Müller et al. 2021).

[2] Abbreviations used in glossing of examples in this paper: CL = classifier; COP = copula; DE = noun phrase marker *de*; LOC = locative; NEG = negation; PFV = perfective; PREP = preposition; Q = interrogative particle; VPRT = post-verbal particle.

The paper is structured as follows: in Section 2, we present the predicatival uses of PC adjectives, deadjectival CoS verbs and RVCs. Formal analyses for deadjectival non-causative CoS verbs as well as the perfective marker *le* are given in Section 3. We deal with the caused CoS expressed by RVCs in Section 4. In the last section, we draw the conclusions of this paper.

2 The phenomenon

As briefly introduced in the previous section, a PC word describing a state in Mandarin Chinese can be categorized as an adjective. A non-causative CoS verb can be derived from the PC adjective. This deadjectival CoS verb can then be used in RVCs to express caused CoS.

2.1 Adjective as a distinct word class in Mandarin Chinese

Firstly, a PC word such as *hóng* 'red' denotes a state when used as a predicate, which is categorized as adjective. Unlike English adjectives, its predicative use is restricted to the following situations (cf. Liu 2010: 1018–1019, Grano 2012: 516): with degree adverb (4), with negation (5), in a polar question (6) or with contrastive focus (7). Using a bare adjective as a predicate is not possible, as in (8).

(4) with degree adverb
Zhāngsān yǎnjīng hěn hóng.
Zhangsan eye very red
'Zhangsan's eyes are (very) red.'

(5) with negation
Zhāngsān yǎnjīng bù hóng.
Zhangsan eye NEG red
'Zhangsan's eyes are not red.'

(6) in polar question
Zhāngsān yǎnjīng hóng mā.
Zhangsan eye red Q
'Are Zhangsan's eyes red?'

(7) with contrastive focus
píngguǒ hóng, qīngcǎo lǜ.
Apple red grass green
'The apple is red, the grass is green.'

(8) * Zhāngsān yǎnjīng hóng.
Zhangsan eye red
Intended: 'Zhangsan's eyes are red.'

One might argue that this predicate should be a stative verb and there is no adjective as an independent category in Mandarin Chinese. From this view, PC words

in Mandarin Chinese are morphologically categorized as stative verbs by Thompson (2004). She gives three reasons for treating PC words in Mandarin Chinese as verbs (Thompson 2004: 1113): a) No copular verb to distinguish this from other verbs; b) Same ability to occur with adverbial elements of degrees,[3] cf. (4) and (9); c) No distinction in attributive modification, cf. (10) and (11).[4]

(9) stative verb with degree adverb
tā hěn ài wǒ.
3.SG very love 1.SG
'S/he loves me very much.'

(10) PC as attributive modifier
gāo de rén
tall DE person
'tall person' or 'person who is tall'

(11) stative verb as attributive modifier
kū de rén
cry DE person
'person who is crying'

The use of copula *shì* is highly limited in Mandarin Chinese. It is required only when the predicate is an NP, cf. (12) and (13) and their English translations. That is, no copula is needed for predicates except for NPs. Thus, a) is not a sufficient indication that adjectives are the same as verbs in Mandarin Chinese.

(12) NP as predicate
tā shì yī-míng xuéshēng.
3.SG COP one-CL student
'S/he is (a/one) student.'

(13) PP as predicate
tā zài gōngyuán.
3.SG PREP.LOC.in garden
'S/he is in the/a garden.'

Furthermore, despite their identical abilities as proposed in b), Thompson (2004) fails to recognize that these elements showing degrees or contrast are necessary for adjectives but optional for verbs, cf. (8) and (14). Besides, the intensifier *hěn* may not have the meaning of intensified degree when used with PC words. Thus, the intensive reading is optional in (4), while *hěn* in (9) necessarily expresses a high(-er)

[3]She also mentions some inchoative suffixes like *-qǐlái* and modals. They are, however, the uses of deadjectival CoS verbs.

[4]Examples (9) – (11) are from Thompson (2004: 1113). We gloss the noun phrase marker *de* (Sun 2015: 374) as DE. Note that *de* was glossed by Thompson as REL, marker of relative clause. She assumed that the prenominal attributive uses of "adjective" and verb with *de* are relative clauses.

degree of loving compared to (14). In previous literature, *hěn* is analyzed as, for instance, positive interpretation morpheme (cf. Liu 2010, Grano 2012), as "subjective standard" (Hu. Fang 2018), etc.

(14) stative verb without degree adverb
tā ài wǒ.
3.SG love 1.SG
'S/he loves me.'

Lastly, the attributive use of adjectives with *de* differs from stative verbs. As Tham (2013: 658–661) points out, an adjective may occur prenominally without *de* and build a compound with no meaning changing, but stative verbs may not, cf. (15) and (16).[5] That is, not all prenomial modifications with *de* should be necessarily interpreted as relative clauses (Tham 2013: 659, cf. also Paul 2010: 117–136).

(15) adjective with/without *de*
gāo (de) jiàgé.
high DE price
'a high price'

(16) verb with/without *de*

a. (qīngshàonián) xǐhuán *(de) diànyǐng.
youth like DE movie
'a movie liked (by young people)'

b. xǐhuán (diànyǐng) *(de) qīngshàonián.
like movie DE youth
'young people who like (movies)'

In sum, Mandarin Chinese adjectives and stative verbs are distinguishable from each other. We argue that a PC word denoting a state is categorized as adjective in Mandarin Chinese and forms an independent category from stative verbs.

2.2 Behaviors of deadjectival non-causative CoS verbs

Having established the distinction between PC adjectives and stative verbs in Mandarin Chinese in the previous section, this section suggests that the non-causative CoS verbs can be derived systematically from the PC adjectives by showing the latter have systematic verbal CoS counterparts, which behave the same way as basic CoS verbs.

(17) and (18) show the CoS use of the word *hóng* 'redden', which has the same form as its adjectival counterpart. Note that *Zhāngsān* in (18) is not the subject of the sentence or the causative agent of the event, but rather the topic and the experiencer, i.e., *yǎnjīng* 'eye' is the subject of the verb *hóng-le* 'redden-PFV' in both (17) and (18). Thus, (18) is intransitive and non-causative. Compared to (3c), the licensing

[5]These examples are from Tham (2013: 661). Note that *de* was glossed by Tham as ASSOC, marker for associative phrase.

of an extra NP *Zhāngsān* to be the topic in (18) is due to the fact that this NP is an experiencer, or rather, the true subject *yǎnjīng* 'eye' is his body part.

(17) Zhāngsān yǎnjīng hóng-le.
Zhangsan eye redden-PFV
'Zhangsan's eyes reddened.'

(18) Zhāngsān hóng-le yǎnjīng.
Zhangsan redden-PFV eye
'Zhangsan's eyes reddened.'

Deadjectival CoS verbs behave the same as basic CoS verbs but differently from PC adjectives in the following ways. First, adjectives can be modified by classic intensifiers such as *hěn* 'very' and *fēicháng* 'extremely' (19), but deadjectival (20) and basic CoS verbs (21) can only be intensified by using the *de hěn lìhai* 'to a serious extent' phrase (Tham 2013: 664–665).[6] Note that (20a) can only have a stative interpretation, while (20b) only a CoS one.

(19) shù-yè fēicháng hóng. (Tham 2013: 664)
tree-leaf very red
'The leaves are extremely red.'

(20) a. Sānmáo de tóufa hěn bái.
Sanmao DE hair very white
'Sanmao's hair is very white.'
'Sanmao's hair turned drastically white.'

b. Sānmáo de tóufa bái de hěn lìhai. (Tham 2013: 665)
Sanmao DE hair white VPRT very serious
'Sanmao's hair turned drastically white.'
'Sanmao's hair is very white.'

(21) a. * Sānmáo hěn zuì. (Tham 2013: 664)
Sanmao very drunk
Intended: 'Sanmao is very drunk'

b. Sānmáo zuì de hěn lìhai. (Tham 2013: 664)
Sanmao drunk VPRT very serious
'Sanmao is drunk to a serious extent.'

Similarly, adjectives and CoS verbs interact differently with negation (Tham 2013: 665–667, cf. also Guo 2018). Compare (22) and (23), when *lǎo* 'old/become old' is negated by *bù*, it can only have a stative interpretation. In contrast, when it is negated by *méi*, only the CoS meaning is possible. Accordingly, basic CoS verbs cannot be negated by *bù* but only by *méi* (24).

[6] Tham (2013) glosses the *de* in *de hěn lìhai* 'to a serious extent' as VPRT, i.e. a post-verbal particle. The particle occurs immediately to the right of the verb, and may be followed by adverbial modifiers or resultative complements (Tham 2013: 664). Note that it is written with a different character as the noun phrase marker *de* glossed as DE.

(22) tā kàn-shàngqu yì diǎn dōu bù lǎo (Lin 2003: 437)
he look-appear one little all NEG old
'He is not old at all in appearance.'

(23) tā kàn-shàngqu yì diǎn dōu méi lǎo (Lin 2003: 437)
he look-appear one little all NEG old
'He hasn't become old at all in appearance.'

(24) a. * Sānmáo bú zuì (Tham 2013: 666)
Sanmao NEG drunk

b. Sānmáo méi zuì (Tham 2013: 666)
Sanmao NEG drunk
'Sanmao didn't get drunk.'

To sum up, PC adjectives have systematic CoS counterparts, which behave differently from adjectives but in the same way as basic CoS verbs. The systematicity indicates that there is an underlying grammatical process (Tham 2013: 668, 671–672), and the CoS meaning does not arise from pragmatic coercion as Koontz-Garboden (2007) proposes for Tongan. The Principle of Monotonic Composition (Rappaport Hovav & Levin 1998, Koontz-Garboden 2005: 98–99), as formulated in (25), constrains that the word meaning is built up by adding pieces of meaning rather than subtracting. With our example, the meaning of verbal *hóng* 'redden' is built up by adding the BECOME operator to the meaning of adjectival *hóng* 'be red'. See detailed discussions on the semantics of deadjectival CoS verbs in Section 3. Based on this, we assume that CoS verbs are derived from their PC adjective counterparts.

(25) The Principle of Monotonic Composition:
Word meaning is constructed monotonically on the basis of event structure constants and operators. (Koontz-Garboden 2005: 98)

The event structure of (17) can be changed when combined with different temporal modifiers, cf. (26) and (27). With the time point adverbial *sān-tiān hòu* 'in three days', (26) describes a CoS of the eyes' color, while (27), with time period adverbial *sān-tiān* 'for three days', indicates a state of the eyes' being red (after becoming red) either at a certain time in the past or continuing to the present.

(26) inchoative
sān-tiān hòu, Zhāngsān yǎnjīng hóng-le.
three-day later Zhangsan eye redden-PFV
'Three days later, Zhangsan's eyes reddened.'

(27) stative
Zhāngsān yǎnjīng hóng-le sān-tiān.
Zhangsan eye redden-PFV three-day
'Zhangsan's eyes were red for three days.' or 'Zhangsan's eyes have been red for three days.'

Finally, a word on the two types of *le* in Mandarin Chinese is needed here. In our examples above (17, 18, 26, 27), the verb-final *le* is obligatory. This *le* marks the perfective aspect, as shown in (28). While we mark the *le* in (17) as a verb-final perfective marker, V. Pan (2019: 16–17) argues instead that this is a sentence-final *le*, and that the CoS meaning comes from this sentence-final particle. However, not all our data are in line with this analysis. For instance, (18) still expresses a CoS meaning without the presence of the sentence-final *le*. Further, the two types of *le* can co-occur in one sentence (29):[7] the verb-final *le* describes the perfectivity of the event; the sentence-final *le* does not change the stative reading as in (27). Therefore, we maintain our view that the CoS meanig comes from the deadjectival verb itself, and we agree with Soh (2009) and Fang (2018) that the sentence-final *le* does not indicate a CoS, but is rather used for the speaker to express the unexpectedness towards the event.

(28) Tā mà-le tāde háizi. (Soh 2009: 628)
he scold-PFV his child
'He has scolded his child (and this is the whole event).'

(29) Zhāngsān yǎnjīng hóng-le sān-tiān le.
Zhangsan eye redden-PFV three-day LE
'Zhangsan's eyes have been/were red for three days.'

2.3 Behaviors of caused CoS RVCs

A deadjectival CoS verb builds an RVC with another verb to indicate caused CoS, cf. (30a) and (30b). In (30a), *hóng* 'redden' can be deleted because *shuā* 'brush' can be used transitively and can by itself take *mén* 'door' as the object. The deletion is not possible in (30b), since *kū* 'cry' is an intransitive verb and cannot take *yǎnjīng* 'eye' as a further argument. This indicates that the object is the argument of *hóng* 'redden' rather than of the preceding verb (cf. ECM resultatives discussed in Wechsler & Noh 2001: 394–395, Müller 2002: 247–250 and Williams 2008: Sec. 6.1, among others).

(30) caused CoS

a. Zhāngsān shuā-(hóng)-le mén.
Zhangsan brush-redden-PFV door
'Zhangsan brushed the door red.'

b. Zhāngsān kū-*(hóng)-le yǎnjīng.
Zhangsan cry-redden-PFV eye
'Zhangsan's eyes reddened from crying.'

[7] We gloss the sentence-final *le* as LE to distinguish it from the perfective verb-final *le*.

3 Analysis for non-causative CoS verbs

As mentioned in Section 1, we follow Tham (2013) and assume that the Chinese CoS verbs are derived from PC adjectives. The examples in the previous sections indicate that deadjectival CoS verbs can either be used by themselves as monomorphemes, or can be combined with other morphemes to form RVCs. This section is dedicated to the formalization of the derivation from adjectives to monomorphemic CoS verbs.

Taking into account all the structures monomorphemic deadjectival CoS verbs appear in, as we exemplify in Section 2, our analysis needs to account for all of the following simplified cases of derivation from the adjectival *hóng* 'red' (31) to the verbal *hóng* 'redden': an inchoative use without any modifications (32a), a stative use with a time period adverbial (32b), an inchoative use with a time point adverbial (32c), and lastly, again the stative use but with two different types of *le* (32d).

(31) yǎnjīng hěn hóng.
eye very red
'Eyes are (very) red.'

(32) a. yǎnjīng hóng-le.
eye redden-PFV
'Eyes reddened.'

b. yǎnjīng hóng-le sān-tiān.
eye redden-PFV three-day
'Eyes were red for three days.'

c. sān-tiān hòu, yǎnjīng hóng-le.
three-day later eye redden-PFV
'Three days later, eyes reddened.'

d. yǎnjīng hóng-le sān-tiān le.
eye redden-PFV three-day LE
'Eyes were red for three days.'

Syntactically, the category change does not result in a difference in the argument structure: the argument of the adjective remains the argument of the derived intransitive verb, i.e., the subject *yǎnjīng* 'eye' in (32).

As for the semantic part: judging from the four sentences mentioned above, two types of *hóng* 'redden' seem to be needed here, one inchoative, which can be used in conjunction with a time point expression, and one stative, which co-occurs with a time period expression.

However, if we decompose the content of these sentences, we find that they have a common semantic core: [BECOME(e, RED(s, x)) $\wedge$ e $<_t$ s],[8] which means, inchoative *hóng* 'redden' actually contains a state of being red, as shown in (33a). It accounts for why *hóng* 'redden' allows a time period adverbial, see the differences between

[8] We treat e (event) and s (state) as subcategories of eventualities, in the sense of Bach (1986: 6).

(33a) and (33b). Correspondingly, the eyes in (32b) with an "originally" stative *hóng* 'redden' must turn red before they can stay red for three days in this state. In other words, although the verbal stative *hóng* 'redden' emphasizes the state, it must contain the content of becoming red as shown in (33b). (33a) and (33c) are almost the same, except that (33c) has an extra e_2 in it, which is set to precede the BECOME event e_1, and it is, in our data (32c), spaced three days apart from e_1, i.e. t=3d. The semantics of (33b) and (33d) are exactly the same, although (33d) has an extra sentence-final *le*.

Judging from the semantic representation in (33), we need only the inchoative *hóng* 'redden'. The stative meaning is only activated when *hóng* 'redden' combines with a time period adverbial and in this case, the time period adverbial will only be linked to the *red* relation.

(33) a. *hóng-le* (inchoative)
$\lambda x \lambda e \lambda s$[BECOME(e, RED(s, x)) $\wedge$ e $<_t$ s]

b. *hóng-le* three days (stative)
$\lambda x \lambda e \lambda s$[BECOME(e, RED(s, x)) $\wedge$ e $<_t$ s $\wedge$ for-three-days(s)]

c. three days later *hóng-le* (inchoative)
$\lambda x \lambda e_1 \lambda s \lambda e_2$[BECOME($e_1$, RED(s, x)) $\wedge$ $e_1 <_t$ s $\wedge$ $e_2 <_{t=3d}$ e_1]

d. *hóng-le* three days *le* (stative)
$\lambda x \lambda e \lambda s$[BECOME(e, RED(s, x)) $\wedge$ e $<_t$ s $\wedge$ for-three-days(s)]

Comparing (33b) and (33d), the sentence-final *le* does not affect the content of the sentence. The "unexpectedness" in (29) is a pragmatic effect and will not be accounted for in the current analysis.

The feature descriptions of the adjectival *hóng* 'red' and the verbal *hóng* 'redden' are proposed in (34) and (35) respectively.

(34) adjectival *hóng* 'red'

$$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{adj} \\ \text{PRD} & + \\ \text{SUBJ} & \boxed{1}\left\langle \text{NP}_{\boxed{2}} \right\rangle \\ \text{COMPS} & \langle\rangle \\ \text{ARG-ST} & \boxed{1} \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{IND} & \boxed{3} \end{bmatrix} \\ \text{RELS} & \left\langle \begin{bmatrix} \textit{red} & \\ \text{LBL} & \textit{handle} \\ \text{ARG0} & \boxed{3} \\ \text{ARG1} & \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix}$$

(35) verbal *hóng* ‘redden’

$$\begin{bmatrix}
\textsc{cat} & \begin{bmatrix} \textsc{head} & \textit{verb} \\ \textsc{subj} & \boxed{1}\left\langle \textsc{np}\,\boxed{2} \right\rangle \\ \textsc{comps} & \langle\rangle \\ \textsc{arg-st} & \boxed{1} \end{bmatrix} \\
\textsc{cont} & \left[\textsc{ind} \quad \boxed{3}\right] \\
\textsc{rels} & \left\langle \begin{bmatrix} \textit{become} & \\ \textsc{lbl} & \textit{handle} \\ \textsc{arg}0 & \boxed{3} \\ \textsc{arg}1 & \boxed{4} \end{bmatrix}, \begin{bmatrix} \textit{red} & \\ \textsc{lbl} & \boxed{4} \\ \textsc{arg}0 & \boxed{5} \\ \textsc{arg}1 & \boxed{2} \end{bmatrix}, \begin{bmatrix} \textit{precedence} & \\ \textsc{lbl} & \textit{handle} \\ \textsc{arg}1 & \boxed{3} \\ \textsc{arg}2 & \boxed{5} \end{bmatrix} \right\rangle
\end{bmatrix}$$

To capture the productivity of the pattern, we suggest the deadjectical CoS lexical rule in (36). The argument of the PC adjective ($\boxed{2}$) becomes the subject of the intransitive verb; in the semantic part of the CoS verb, an additional relation of *become* is added to the original adjective content. An underspecified *precedence* relation is introduced to account for the temporal difference between the two events.

(36) Deadjectival CoS lexical rule

$$\begin{bmatrix}
\textsc{cat} & \left[\textsc{head} \quad \textit{adj}\right] \\
\textsc{cont} & \left[\textsc{ind} \quad \boxed{1}\right] \\
\textsc{rels} & \boxed{2}\left\langle \begin{bmatrix} \textsc{lbl} & \boxed{3} \\ \textsc{arg}0 & \boxed{1} \end{bmatrix} \right\rangle
\end{bmatrix} \mapsto$$

$$\begin{bmatrix}
\textsc{cat} & \left[\textsc{head} \quad \textit{verb}\right] \\
\textsc{cont} & \left[\textsc{ind} \quad \boxed{4}\right] \\
\textsc{rels} & \left\langle \begin{bmatrix} \textit{become} & \\ \textsc{lbl} & \textit{handle} \\ \textsc{arg}0 & \boxed{4} \\ \textsc{arg}1 & \boxed{3} \end{bmatrix} \right\rangle \oplus \boxed{2} \oplus \left\langle \begin{bmatrix} \textit{precedence} & \\ \textsc{lbl} & \textit{handle} \\ \textsc{arg}1 & \boxed{4} \\ \textsc{arg}2 & \boxed{1} \end{bmatrix} \right\rangle
\end{bmatrix}$$

This pattern is also applicable to other deadjectival CoS predicates. As the two pairs of examples, adjectival and verbal *gāo* ‘high’ and *piányi* ‘cheap’, from Tham (2013: 657) and the *People’s Daily* subcorpus in the Beijing Language and Culture University (BLCU) Corpus Center (Xun et al. 2016) show in (37) and (38).

(37) a. bìngrén xuèyā hěn gāo. (Tham 2013: 657)
patient blood.pressure very high
‘The patient’s blood pressure is (very) high.’

b. bìngrén xuèyā gāo-le. (Tham 2013: 657)
patient blood.pressure high-PFV
'The patient's blood pressure has raisen.'

(38) a. bīnguǎn hěn piányi. (*People's Daily*, Dec. 27. 2002)
hotel very cheap
'Hotels are (very) cheap.'

b. kànbīng piányi-le. (*People's Daily*, Nov. 10. 2013)
see.a.docotor cheap-PFV
'It became cheaper to see a doctor.'

The combination of the verbal *hóng* 'redden' with the verb-final *le* 'PFV' can be realized using the perfective lexical rule proposed by Müller & Lipenkova (2013: 246), as shown in (39).

(39) Perfective lexical rule

$$\begin{bmatrix} \textsc{phon} & \boxed{1} \\ \textsc{synsem}\,|\,\textsc{loc} & \begin{bmatrix} \textsc{cat}\,|\,\textsc{head} & \textit{verb} \\ \textsc{cont}\,|\,\textsc{ind} & \boxed{3} \end{bmatrix} \\ \textsc{rels} & \boxed{2} \end{bmatrix} \mapsto \begin{bmatrix} \textsc{phon} & \boxed{1} \oplus \langle \text{le} \rangle \\ \textsc{rels} & \left\langle \begin{bmatrix} \textit{perfective-rel} \\ \textsc{arg} \quad \boxed{3} \end{bmatrix} \right\rangle \oplus \boxed{2} \end{bmatrix}$$

As it is implemented in the CoreGram project (Müller 2015), the recursive application of this lexical rule is prevented in the way that the input must be of type *simple_word*. The output has a daughter and is thus automatically of type *complex_word*. In this way, (39) cannot be applied to its output.

4 Analysis for caused CoS RVCs

Caused CoS are expressed mainly by resultative verb compounds (RVCs) (Tham 2013: 653), as shown in (30) in Section 2.3. Müller (2002: Ch. 5, 2018: 70) provides the lexical rule in (41) for German resultative predicates such as (40), similar to the lexical rule proposed by Wechsler & Noh (2001) for predicative resultatives in English and Korean.

(40) Er fischt den Teich leer.
he fishes the pond empty

(41) is a lexical rule that licenses for each intransitive verb another lexical item that takes a secondary predicate as complement and forms a predicate complex. It maps an intransitive verb to a verb that takes an X(P) predicate and the subject of this X(P) as arguments. The RELS list of the output contains the RELS of the input (3), a *cause* and a *become* relation. The *cause* relation relates the event of the input verb (2) to the *become* event (7). The argument of the *become* relation is the contribution of the X(P) (5).

(41) Lexical rule for resultatives (Müller 2018: 70)

$$\begin{bmatrix} \textsc{arg-st} & \left\langle \boxed{1}\ \mathrm{NP}\left[\textit{str}\right] \right\rangle \\ \textsc{cont} & \left[\textsc{ind}\ \boxed{2}\right] \\ \textsc{rels} & \boxed{3} \end{bmatrix} \mapsto \begin{bmatrix} \textsc{arg-st} & \left\langle \boxed{1}, \boxed{4}\ \mathrm{NP}\left[\textit{str}\right], \mathrm{X(P)}\left[\textsc{prd+}, \textsc{subj}\left\langle \boxed{4} \right\rangle\right]\!:\!\boxed{5} \right\rangle \\ \textsc{cont} & \left[\textsc{ind}\ \boxed{6}\ \textit{event}\right] \\ \textsc{rels} & \boxed{3} \oplus \left\langle \begin{bmatrix} \textit{cause} & \\ \textsc{arg}0 & \boxed{6} \\ \textsc{arg}1 & \boxed{2} \\ \textsc{arg}2 & \boxed{7} \end{bmatrix}, \begin{bmatrix} \textit{become} & \\ \textsc{arg}0 & \boxed{7} \\ \textsc{arg}1 & \boxed{5} \end{bmatrix} \right\rangle \end{bmatrix}$$

Similar to (41), we propose the lexical rule in (42) for RVCs in Mandarin Chinese. It takes the intransitive form of a verb as input, as indicated by an empty COMPS list. The output is a verb that takes another intransitive verb as well as its subject (3) as complements. The subject of the output verb is taken over directly from the input verb and therefore needs not to be represented in the lexical rule. The meaning of the output is that the event expressed by the input verb (1) causes the event expressed by the verbal argument (4). Notice that different from (41), the meaning of the output does not contain a *become* relation. This is because based on our proposal in (36), the deadjectival CoS verb, namely the verbal argument in the output of (42), already contains the *become* relation. Thus it does not need to be represented again in (42).

(42) Lexical rule for RVCs:

$$\begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{head} & \textit{verb} \\ \textsc{comps} & \langle\rangle \end{bmatrix} \\ \textsc{cont} & \left[\textsc{ind}\ \boxed{1}\right] \\ \textsc{rels} & \boxed{2} \end{bmatrix} \mapsto$$

$$\begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{comps} & \boxed{3} \oplus \left\langle \begin{bmatrix} \textsc{loc} & \begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{head} & \textit{verb} \\ \textsc{subj} & \boxed{3}\left\langle \mathrm{NP} \right\rangle \\ \textsc{comps} & \langle\rangle \end{bmatrix} \\ \textsc{cont} & \left[\textsc{ind}\ \boxed{4}\right] \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \\ \textsc{cont} & \left[\textsc{ind}\ \boxed{5}\ \textit{event}\right] \\ \textsc{rels} & \boxed{2} \oplus \left\langle \begin{bmatrix} \textit{cause} & \\ \textsc{arg}0 & \boxed{5} \\ \textsc{arg}1 & \boxed{1} \\ \textsc{arg}2 & \boxed{4} \end{bmatrix} \right\rangle \end{bmatrix}$$

Rather than assuming different rules for intransitive and transitive verbs, we propose that (42) with an intransitive verb as the input can cover all cases of RVCs,

because as in German (Müller 2002: Sec. 5.1.7), most Mandarin Chinese verbs can be used without an object (Lü 1987: 2, Yang 1999: 35), as shown in (43).[9]

(43) Lü (1987: 2)

a. tā yào xiān chī fàn hòu hē jiǔ.
he want first eat rice after drink alcohol
'He wants to eat rice first and drink alcohol after.'

b. tā yào xiān chī hòu hē.
he want first eat after drink
'He wants to eat first and drink after.'

We assume that the second predicate is a verb, rather than an adjective. The second predicate in an RVC expresses a CoS meaning (Shibagaki 2010: Sec. 5), i.e. in (30a), repeated here as (44), Zhangsan's brushing causes the door to become red, rather than to stay in the state of being red. In Mandarin Chinese, the CoS meaning is expressed by verbs, while adjectives only express stative meaning (Tham 2013: 655, 661–667).

(44) Zhāngsān shuā-hóng-le mén
Zhangsan brush-redden-PFV door
'Zhangsan brushed the door red.'

Furthermore, there are RVCs whose second predicate is a basic verb, such as *pǎo* 'run' in (45).

(45) Zhāngsān xià-pǎo-le Lǐsì.
Zhangsan scare-run-PFV Lisi.
'Zhangsan scared Lisi, which caused Lisi to run away.'

If we assume that some second predicates are adjectives while others are verbs, we would have to assume two different rules for RVCs, because an adjectival secondary predicate requires a *become* relation in the output of the resultative lexical rule (41), while a verbal one does not (42). Thus, it is simpler to assume all second elements of RVCs to be verbs.

The two verbs in an RVC are not in a coordinated relation. First, switching the positions of the two verbs will result in a change in the meaning of the whole construction. Second, when negated by *bù*, as in (46), the negation only scopes over the second verb but not the first. If the denotation of the first verb is not true, the whole proposition is false regardless of the truth condition of the second verb. It goes to show that the second verb is truth-conditionally subordinate to the first verb (Song et al. 2015).

[9]Note that this is different from the null object construction (e.g. Huang 1991, H. Pan 2019), where there is a contextually salient antecedent of the unrealized object, as illustrated in (i).

(i) Zhāngsān kànjiàn-le tāde māmā, Lǐsì yě kànjiàn-le.
Zhangsan see-PFV his mom Lisi also see-PFV
'Zhangsan saw his mom, Lisi also saw.'

(46) Zhāngsān xǐ-bù-gānjìng yīfu.
Zhangsan wash-NEG-clean clothes
'Zhangsan cannot wash the clothes clean.'

Third, behaviors in imperatives show that the first verb is the head, as the whole structure inherits the ability to form imperative (47a) from the first verb (47b), while the second verb cannot form imperative (47c–d). The same behavior can be observed for resultative verb constructions in Yorùbá and thus, Maché (2022: 71) also assumes the first verb to be the head.

(47) a. chuī-gān tóufa!
blow-dry hair
'Blow the hair dry!'

b. chuī tóufa!
blow hair
'Blow the hair!'

c. * gān tóufa!
dry hair
Intended: 'Dry the hair!'

d. * gān!
dry
Intended: 'Be dry!'/'Become dry!'

All in all, it is desirable to analyze the second verb in a subordinate position to the first verb, as opposed to a headless structure such as what Müller & Lipenkova (2009) propose for the Serial Verb Construction in Mandarin Chinese.

For the example in (44), there is first a lexical entry (48) for the intransitive form of *shuā* 'brush'.

(48) *shuā* 'brush' (intransitive form)

$$\begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{subj} & \langle \text{NP}_{\boxed{1}} \rangle \\ \textsc{comps} & \langle\rangle \end{bmatrix} \\ \textsc{cont} & \begin{bmatrix} \textsc{ind} & \boxed{2} \end{bmatrix} \\ \textsc{rels} & \begin{bmatrix} \textit{brush} & \\ \textsc{arg}0 & \boxed{2} \\ \textsc{arg}1 & \boxed{1} \\ \textsc{arg}2 & \square \end{bmatrix} \end{bmatrix}$$

The value of ARG2 is □, as it is not linked to any element in the valence representation. As Müller (2002: 214) argues, the value of ARG2 is underspecified and is determined by the context. For (44), it is hard to imagine a situation where the brushing of something else caused the door to become red. However, Müller (2002: 211–215) shows with examples such as (49) that the accusative NP *Weinkeller* 'wine cellar' is not the object selected by the main verb *trinken* 'drink'.

(49) Die Gäste tranken den Weinkeller leer. (Müller 2002: 212)
the guests drank the wine.cellar empty

A similar example can be constructed in Mandarin Chinese, as in (50).[10]

(50) kèrén hē-kōng-le jiǔjiào.
guest drink-empty-PFV wine.cellar
'The guests drank the wine cellar empty.'

Applying the lexical rule (42) to (48), we get the lexical item (51) for *shuā* 'brush' as being used in an RVC such as *shuā-hóng* 'brush-red'.

(51) *shuā* 'brush' (as used in *shuā-hóng* 'brush-red')

$$\left[\begin{array}{l}
\textsc{cat}\ \left[\begin{array}{l}
\textsc{subj}\ \left\langle \mathrm{NP}_{\boxed{1}} \right\rangle \\
\textsc{comps}\ \boxed{2} \oplus \left\langle \left[\begin{array}{l}\textsc{loc}\ \left[\begin{array}{ll}\textsc{cat} & \left[\begin{array}{ll}\textsc{head} & \textit{verb} \\ \textsc{subj} & \boxed{2}\left\langle \mathrm{NP} \right\rangle \\ \textsc{comps} & \langle\rangle\end{array}\right] \\ \textsc{cont} & \left[\textsc{ind}\ \boxed{3}\right]\end{array}\right]\end{array}\right] \right\rangle
\end{array}\right] \\
\textsc{cont}\ \left[\textsc{ind}\ \boxed{4}\right] \\
\textsc{rels}\ \left\langle \left[\begin{array}{ll}\textit{brush} & \\ \textsc{arg}0 & \boxed{5} \\ \textsc{arg}1 & \boxed{1} \\ \textsc{arg}2 & \square\end{array}\right], \left[\begin{array}{ll}\textit{cause} & \\ \textsc{arg}0 & \boxed{4} \\ \textsc{arg}1 & \boxed{5} \\ \textsc{arg}2 & \boxed{3}\end{array}\right] \right\rangle
\end{array}\right]$$

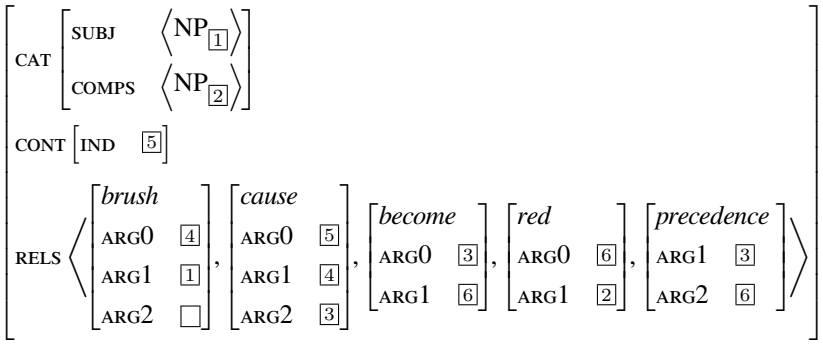
Combining (51) with the lexical item of *hóng* 'redden' as suggested in (35) yields the RVC *shuā-hóng* 'brush-red' as in (52). It means the subject's ($\boxed{1}$) brushing ($\boxed{4}$) causes the complement ($\boxed{2}$) to become ($\boxed{3}$) red ($\boxed{6}$), which correctly represents the meaning of *shuā-hóng* 'brush-red'.

[10]This analysis of an underspecified argument inferred from the context is not unique to resultative constructions, but also applies to, for instance, AcI (*Accusativum cum Infinitivum* 'accusative with infinitive') constructions, such as in (i). AcI verbs are those which embed an infinitive verb, whose subject appear in accusative (Müller 2002: 58).

(i) a. We saw them cross the river (# but we didn't see them).
b. I felt George get on the other end of the water bed (but, of course, I didn't actually feel George). (Kirsner & Thompson 1976: 209)

(i.a) seems to show that usually, when we perceive the event or the situation, we perceive the involved participants, too. However, based on examples such as (i.b), Kirsner & Thompson (1976) argue convincingly that the subject of the complement VP is not the direct object of the matrix verb, i.e. *George* is not the direct object of *feel*, rather, the event of George getting on the water bed is perceived globally. Thus, they conclude that the inference in (i.a) is made based on our world knowledge, rather than the grammatical structure of the sentence. De Geest (1970: 50–51) and Müller (2002: 63–64) argue for the same with Dutch and German examples, respectively.

(52) *shuā-hóng* ‘brush-red’



5 Conclusions

To summarize, Mandarin Chinese systematically derives change of state (CoS) verbs from property concept adjectives. These non-causative CoS verbs can then be combined with another verb to form resultative verb compounds (RVCs), which express caused CoS. We propose an HPSG account for deadjectival CoS verbs in Mandarin Chinese such as *hóng* ‘red/redden’. The derivation of non-causative CoS verbs can be analyzed as a lexical rule that changes an adjective to a verb and adds inchoativity to its meaning. RVCs expressing caused CoS can be handled by a lexical rule based on the proposal in Müller (2002: Ch. 5, 2018: 70). It maps an intransitive verb onto a verb which takes another intransitive verb and its subject as complements while the subject remains. The events denoted by the two verbs are in a causal relationship.

Acknowledgements

We would like to thank Stefan Müller, Antonio Machicao y Priemer, Giuseppe Varaschin, Jozina Vander Klok and the three anonymous reviewers, as well as Jakob Maché, Emily M. Bender, Jean-Pierre Koenig, and other participants in the 30th International Conference on Head-Driven Phrase Structure Grammar for comments on earlier versions of this paper.

References

Bach, Emmon. 1986. The algebra of events. *Linguistics and philosophy* 9(1). 5–16.

Beavers, John, Michael Everdell, Kyle Jerro, Henri Kauhanen, Andrew Koontz-Garboden, Elise LeBovidge & Stephen Nichols. 2017. Two types of states: A cross-linguistic study of change-of-state verb roots. In Patrick Farrell (ed.), *Proceedings of the linguistic society of america*, vol. 2, 1–15. Austin, TX: Linguistic Society of America. DOI: 10.3765/plsa.v2i0.4094.

de Geest, Wim. 1970. Infinitiefconstructies bij verba sentiendi. *Studia Neerlandica* 3. 33–59.

Dixon, Robert M.W. 1982. *Where have all the adjectives gone? And other essays in syntax and semantics* (Janua Linguarum: Series Maior 107). The Hague: Mouton. DOI: 10.1515/9783110822939.

Fang, Hongmei. 2018. Mirativity in Mandarin: The sentence-final particle *le* (了). *Open Linguistics* 4(1). 589–607. DOI: 10.1515/opli-2018-0029.

Fang, Huilin. 2018. Subjective standard-setting in gradable predicates: On the Mandarin *hen* structure. In *Proceedings of Sinn und Bedeutung*, vol. 21, 409–426.

Grano, Thomas. 2012. Mandarin *hen* and universal markedness in gradable adjectives. *Natural Language & Linguistic Theory* 30(2). 513–565. DOI: 10.1007/s11049-011-9161-1.

Guo, Rui. 2018. *Xiandai hanyu cilei yanjiu [A study of word classes in Modern Chinese]*. Beijing: Shangwu yinshuguan [The Commercial Press].

Huang, C.-T. James. 1991. Remarks on the status of the null object. In Robert Freidin (ed.), *Principles and parameters in comparative grammar*, 56–76. Cambridge, MA: MIT Press.

Kirsner, Robert S. & Sandra A. Thompson. 1976. The role of pragmatic inference in semantics: A study of sensory verb complements in English. *Glossa* 10(2). 200–240.

Koontz-Garboden, Andrew. 2005. On the typology of state/change of state alternations. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2005*, 83–117. Dordrecht: Springer. DOI: 10.1007/1-4020-4066-0_4.

Koontz-Garboden, Andrew. 2007. Aspectual coercion and the typology of change of state predicates. *Journal of Linguistics* 43(1). 115–152. DOI: 10.1017/S0022226706004464.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago press.

Lin, Jo-Wang. 2003. Aspectual selection and negation in Mandarin Chinese. *Linguistics* 41(3). 425–459. DOI: 10.1515/ling.2003.015.

Liu, Chen-Sheng Luther. 2010. The positive morpheme in Chinese and the adjectival structure. *Lingua* 120(4). 1010–1056. DOI: 10.1016/j.lingua.2009.06.001.

Lü, Shuxiang. 1987. Shuo "sheng" he "bai" [On "win" and "lose"]. *Zhongguo yuwen* 1. 1–5.

Maché, Jakob. 2022. Accounting for the variation in West Benue resultative constructions. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the International Conference on Head-Driven Phrase Structure Grammar*, 65–85. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2022.4.

Müller, Stefan. 2002. *Complex predicates: Verbal complexes, resultative constructions, and particle verbs in German* (Studies in Constraint-Based Lexicalism 13). Stanford, CA: CSLI Publications.

Müller, Stefan. 2015. The CoreGram project: Theoretical linguistics, theory development and verification. *Journal of Language Modelling* 3(1). 21–86. DOI: 10.15398/jlm.v3i1.91.

Müller, Stefan. 2018. *A lexicalist account of argument structure: Template-based phrasal LFG approaches and a lexical HPSG alternative* (Conceptual Foundations of Language Science 2). Berlin: Language Science Press. DOI: 10.5281/zenodo.1441351.

Müller, Stefan, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.). 2021. *Head Driven Phrase Structure Grammar: The handbook* (Empirically Oriented Theoretical Morphology and Syntax 9). Berlin: Language Science Press. DOI: 10.5281/zenodo.5543318.

Müller, Stefan & Janna Lipenkova. 2009. Serial verb constructions in Chinese: A HPSG account. In Stefan Müller (ed.), *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, Georg-August Universitat Göttingen, Germany*, 234–254. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2009.12.

Müller, Stefan & Janna Lipenkova. 2013. ChinGram: A TRALE implementation of an HPSG fragment of Mandarin Chinese. In *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27)*, 240–249. Taipei, Taiwan: Department of English, National Chengchi University.

Pan, Haihua. 2019. Null object constructions, VP-ellipsis, and sentence interpretation. In Haihua Pan & Jianhua Hu (eds.), *Interfaces in grammar*, 283–299. Amsterdam: John Benjamins.

Pan, Victor. 2019. *Architecture of the periphery in Chinese: Cartography and Minimalism*. London: Routledge. DOI: 10.4324/9781315158228.

Paul, Waltraud. 2010. Adjectives in Mandarin Chinese: The rehabilitation of a much ostracized category. In Patricia Cabredo Hofherr & Ora Matushansky (eds.), *Adjectives: Formal analyses in syntax and semantics* (Linguistics Today 153), 115–152. Amsterdam: John Benjamins. DOI: 10.1075/la.153.04pau.

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

Rappaport Hovav, Malka & Beth Levin. 1998. Building verb meanings. In Miriam Butt & Wilhelm Geuder (eds.), *The projection of arguments: Lexical and compositional factors*, 97–133. Stanford, CA: CSLI Publications.

Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33(2). 431–483. DOI: 10.1017/S002222679700652X.

Shibagaki, Ryosuke. 2010. Mandarin secondary predicates. *Taiwan Journal of Linguistics* 8(1). 57–93. DOI: 10.6519/TJL.2010.8(1).3.

Smith, Carlota. 1997. *The parameter of aspect*. Dordrecht: Springer. DOI: 10.1007/978-94-011-5606-6.

Soh, Hooi Ling. 2009. Speaker presupposition and Mandarin Chinese sentence-final-*le*: A unified analysis of the "change of state" and the "contrary to expectation" reading. *Natural Language & Linguistic Theory* 27(3). 623–657. DOI: 10.1007/s11049-009-9074-4.

Song, Sanghoun, Chen Bo, Francis Bond & Sio Joanna Ut-Seong. 2015. An HPSG-based Analysis of Resultative Compounds in Chinese. Paper presented at the 22nd International Conference on Head-Driven Phrase Structure Grammar.

https://bond-lab.github.io/events/2015-hpsg/pdf/vv-compounds.pdf (29 August, 2023).

Sun, Chaofen. 2015. The uses of DE as a noun phrase marker. In William S-Y. Wang & Chaofen Sun (eds.), *The Oxford Handbook of Chinese Linguistics*, 362–376. New York: Oxford University Press. DOI: 10.1093/oxfordhb/9780199856336.013.0027.

Tham, Shiao Wei. 2013. Change of state verbs and result state adjectives in Mandarin Chinese. *Journal of Linguistics* 49(3). 647–701. DOI: 10.1017/S0022226713000261.

Thompson, Sandra A. 2004. Property concepts. In Geert Booij, Christian Lehmann, Joachim Mugdan, Stavros Skopeteas & Wolfgang Kesselheim (eds.), *Morphology: An international handbook on inflection and word-formation*, vol. 2 (Handbooks of Linguistics and Communication Science [HSK] 17/2), 1111–1117. Berlin & New York: De Gruyter Mouton. DOI: 10.1515/9783110172782.2.14.1111.

Wechsler, Stephen & Bokyung Noh. 2001. On resultative predicates and clauses: Parallels between Korean and English. *Language Sciences* 23(4). 391–423. DOI: 10.1016/S0388-0001(00)00031-0.

Wetzer, Harrie. 1992. "Nouny" and "verby" adjectivals: A typology of predicative adjectival constructions. In Michel Kefer & Johan van der Auwera (eds.), *Meaning and grammar: Cross-linguistic perspectives* (Empirical Approaches to Language Typology [EALT] 10), 223–262. Berlin & New York: De Gruyter Mouton. DOI: 10.1515/9783110851656.223.

Williams, Alexander. 2008. Patients in Igbo and Mandarin. In Johannes Dölling, Tatjana Heyde-Zybatow & Martin Schäfer (eds.), *Event structures in linguistic form and interpretation* (Language, Context and Cognition 5), 3–30. Berlin: De Gruyter. DOI: 10.1515/9783110925449.3.

Xun, Endong, Gaoqi Rao, Xiaoyue Xiao & Jiaojiao Zang. 2016. Dashuju beijing xia BCC yuliaoku de yanzhi [The construction of the BCC corpus in the age of big data]. *Yuliaoku Yuyanxue [Corpus linguistics]* 3(1). 93–109.

Yang, Suying. 1999. Cong feibingedongci xianxiang jan yuyi yu jufajiegou zhijian de guanxi [The unaccusative phenomenon: A study on the relationship between syntax and semantics]. *Dangdai Yuyanxue [Contemporary Linguistics]* (1). 30–43.

Superlative *ever* in Dutch, French, German, and Spanish

Manfred Sailer
Goethe-University Frankfurt a.M.

Nicolas Lamoure
Goethe-University Frankfurt a.M.

Proceedings of the 30th International Conference on
Head-Driven Phrase Structure Grammar

University of Massachusetts Amherst

Stefan Müller, Elodie Winckel (Editors)

2023

Frankfurt/Main: University Library

pages 79–99

Keywords: borrowing, superlative, corpuslinguistics, Dutch, French, German, Spanish

Sailer, Manfred & Nicolas Lamoure. 2023. Superlative *ever* in Dutch, French, German, and Spanish. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 79–99. Frankfurt/Main: University Library.
DOI: 10.21248/hpsg.2023.5.

Abstract

The paper examines borrowed instances of what we call *emphatic superlative ever* (ES-*ever*) into two Germanic languages (Dutch and German) and two Romance languages (French and Spanish). We base our study on extensive corpus data. We model the data in three stages ranging from constructional borrowing (Stage-1: *el coolest job ever* 'the coolest job ever'), via diaconstructions (Stage-2: *la mejor canción ever* 'the best song ever'), up to lexical borrowing (Stage-3: *las portadas más photoshopeadas ever* 'the most photoshoped portals ever'). We extend an earlier approach to social meaning in HPSG to borrowing.
The data extracted for this study is available at: `https://osf.io/juewa/?view_only=215970c573d34b148815cc5653965697`

1 Introduction

The paper at hand examines borrowed instances of what we call emphatic superlative *ever* (ES-*ever*) into two Germanic languages (Dutch and German) and two Romance languages (French and Spanish), see (1), in order to deepen our understanding of borrowing, which in turn will feed a theoretic implementation into a HPSG model for borrowing.

(1) a. nl: de beste opmerking *ever* 'the best comment ever'
matrix language alternative: *ooit* 'ever'

b. de: bestes Bild *ever* 'best picture ever'
matrix language alternative: *aller Zeiten* 'of all times'

c. fr: la meilleure idée *ever* 'the best idea ever'
matrix language alternative: *de tous les temps* 'of all times'

d. es: la mejor foto *ever* 'the best picture ever'
matrix language alternative: *de todos los tiempos* 'of all times'

Based on data from an extensive corpus research via Sketch Engine (Kilgarriff et al. 2014), we will show that although the languages at hand exhibit ES-*ever* with varying frequencies, they do so in a homogenous way:

1. Only ES-*ever* is borrowed, other well-formed and well-attested uses in English (such as with negation or in questions) do not occur or only in all-English passages.

2. ES-*ever* respects the rules of the source language (English): (i) it requires licensing by either a morphological superlative form or a semantic/pragmatic superlative; (ii) it can only occur in an extraposed position.

[†]Aspects of this paper were presented at the *45th Annual Meeting of the German Linguistic Society*, March 2023. We would like to thank the reviewers and the audience of that event and of *HPSG 2023* for their comments, in particular Emily M. Bender, Nurit Melnik, and Elodie Winckel. All errors are ours.

3. The usage of ES-*ever* can be modelled in stages.

4. ES-*ever* has blended in with the grammatical rules of the respective matrix languages allowing to be combined with constructions not permissive in English (what we will take as indicative of the last stage)

These findings will serve as a starting point to an HPSG approach to borrowing, which we envision as a three stage model ranging from constructional borrowing (Stage-1) up to lexical borrowing (Stage-3).

We will proceed in the following fashion: We will present our corpus study in Section 2. Section 3 will give deeper insights into the syntactic structure of ES-*ever* in the respective borrowing languages. Section 4 will present our HPSG approach to borrowing. We end with a conclusion.

2 Corpus study

The main goal of our corpus study was to collect relevant data in order to determine the licensing conditions of ES-*ever* in the different matrix languages. In the following, we will outline our methodology and present some of our results.

2.1 Methodology

In order to acquire comparable results in all the languages scurtinized we used the Timestamped corpora which are available for all four languages. In order to optimize the quantity of results we opted for the largest version of the corpus for each language, i.e. Timestamped JSI web corpus 2014-2021 French, Spanish, German and Dutch respectively. Since the Timestamped corpora contain texts extracted from RSS feeds of News-websites, we hoped to limit the extraction of irrelevant hits such as e.g. fragments from all-English texts, gibberish and machine created texts. Creating appropriate queries turned out to be a challenging task: ES-*ever* constructions allow for the (recursive) embedding of constituents, e.g. PPs, relative clauses and so forth. However, such elementary properties of human languages can only be indirectly operationlized in CQL (Corpus Query Language, the language which Sketch Engine provides for conducting more advanced searches), as it is generally restricted to describing "flat" linear order of strings and is generally unaware of hierarchical structures. Consequently, any query can only be an approximation; ours are no exception. The expressions in Q1, Q2, and Q3 show our queries for French, which – except for references to language specific lexical items (i.e. *et* 'and') – are the same for all languages.

Q1
```
[tag="ADJ.*"] []{0,5}
[tag="NOM|NAM" & word!="than|for|4"]
[word="ever|EVER"] [word!="after|closer"] within <s/>
```

```
Q2 [tag="ADJ.*"] []{0,5}
   [tag="NOM|NAM" & word!="than|for|4"]
   []{0,5} [word!="et|ever|for|than|4"]
   [word="ever|EVER"] [word!="after|closer"] within <s/>

Q3 [tag="NOM|NAM" & word!="than|for|4"] []{0,5}
   [tag="ADJ.*"] [word="ever|EVER"] within <s/>
```

Q1 and Q2 represent the queries for pre-nominal adjectives (used in all four languages) and Q3 the query for post-nominal adjectives (used only in the Romance part of the sample). Q1 will find "simple" examples of ES-*ever* as in (2), which may contain additional pre-nominal material.

(2) de: der höflichste, freundlichste, teamwilligste Camper ever
the kindest, friendliest, most teamwilling camper ever
'the kindest, friendliest, most teamwilling camper ever'

Q1 translates to: "Find some adjectival form; followed by a span of zero to five unspecified word-forms; followed by a common or proper noun which may not have the form "than", "for", or "4"; followed by the word-form "ever" or "EVER"; not followed by the word-forms "after" or "closer". All within one sentence."

Q2 will find examples with material between a noun and *ever*, see (3).

(3) de: die wohl beste Werbung für das Hotel ihrer Eltern ever
the probably best advertisment for the hotel of.her parents ever
'the apparently best advertisement for her parents' hotel'

Q2 translates to: "Find a string which begins by some form of an adjective; followed by zero to five occurrences of some undefined wordform; followed by a noun which is either tagged as a noun or a proper name but has not the form "than", "for", or "4"; followed by zero to five occurrences of some undefined word; followed by the wordforms "ever" or "EVER", which may not be succeeded by the wordforms "after" or "closer". All within one sentence."

Q3 will find examples such as the following, featuring a post-nominal adjective and an optional post-nominal span of arbitrary words:

(4) fr: l'un des joueurs les plus fragiles ever
the one of.the players the most fragile ever
'one of the most fragile players ever'

Q3 translates to: "Find a string which begins by some common or proper noun, which may not have the wordform "than", "for", or "4"; followed by followed by zero to five occurrences of some undefined wordform; followed by some adjective; followed by the wordforms "ever" or "EVER". All within one sentence."

The reasoning behind the queries is the following:

- **Reference to an unspecified adjectival form**: We decided to not restrict the query to only include morphological superlatives, which would be supported by the tagging system, as not all languages attest synthetic superlatives for all adjectives and we wanted to evaluate if indeed all instances of ES-*ever* included a superlative form.

- **Reference to "than" "for" and "4", "after", "closer"**: in prior versions of this query we identified quite a lot of false hits, i.e. fixed expression, which included irrelevant instances of *for/4 ever*, *than ever*, *ever closer union* etc.

- **Reference to zero to five occurrences of some undefined word forms**: These two wild cards of varying length allow us to account for recursion, like additional pre- or post-nominal material such as PPs, additional adjectives, adverbs, relative clauses etc. We restricted the span to maximally five items, because enlarging the span (i) leads strings matching the query multiple times, and (ii) increases the amount of false hits.

After extracting all the hits found by the queries, we evaluated them by hand in order to sort out any false hits. Subsequently we collected all results for each language into one set via the identifiers of each hit. This allowed us to further analyze the results according to their geographical origin. Finally, we exported them to .csv files and tagged them according to the following criteria:[1]

- Pattern inside the noun phrase in terms of POS tags

- Type of superlative, i.e. analytic (*most beautiful*), synthetic (*best*) or inherent (*absolute*)

- positioning of the adjective, i.e. pre- or post-nominal

- Stage of nativization (see Section 2.2)

2.2 Results and discussion

After merging the results of each query by language, eliminating false hits by hand and removing duplicates, we received 369 instances of ES-*ever* for Dutch (0.23 hits per million tokens), 2,230 for German (0.26 hits per million tokens, hmt), 159 for French (0.02 hmt) and 120 for Spanish (0.01 hmt).[2]

Generally speaking, ES-*ever* is strikingly more frequent in the Germanic than in the Romance languages of our sample. We can further correlate the occurrences of ES-*ever* with their source country by looking at the top-level domain of the site an utterance is taken from. We show the variation for some

[1] An exception in this approach was German: Since we were left with 2,230 hits in German after excluding false positives, we decided to create a random sample of 300 items for evaluation.

[2] The data extracted for this study is available at: `https://osf.io/juewa/?view_only=215970c573d34b148815cc5653965697`

Country	*N*	hmt
Netherlands	266	0.25
Belgium	79	0.18
All countries	369	0.23

Country	*N*	hmt
Germany	1,661	0.26
Austria	142	0.27
Switzerland	187	0.29
All countries	2,230	0.26

Table 1: Frequencies in Dutch and German

Country	*N*	hmt
France	77	0.02
Canada	27	0.05
Belgium	3	< 0.01
Switzerland	1	< 0.01
All countries:	159	0.02

Country	*N*	hmt
Spain	44	< 0.01
Mexico	13	< 0.01
Chile	13	0.02
Peru	8	0.01
All countries:	120	0.01

Table 2: Frequencies in French and Spanish

countries in Table 1 for Dutch and German, and in Table 2 for French and Spanish.

Canada is English-French bilingual and shows the highest relative frequency of ES-*ever* in the French data, but this value is still much lower than for the Germanic languages. For Belgium, the relative frequency for both Dutch and French is lower than the overall results for these languages. For Switzerland, the relative frequency for German is the highest in the table, but for French, it is among the lowest. This shows that societal bilingualism does not explain the variation among the French-speaking countries. We tentatively conclude that the Germanic-Romance contrast is the prominent, consistent, determining factor of the frequency of ES-*ever* in our data.

As mentioned in the previous section we classified the extracted and hand-sorted findings according in four structurally distinct stages of nativization: Stage-0 contains fully English expressions, like (5). As there is no interaction between the grammar of the noun phrase in the matrix and the source language in Stage-0, we will ignore this stage in rest of this paper. Stage-1 refers to expressions which contain an uninflected English adjective followed by the noun, see (6). An expression that features a matrix language determiner, noun and adjective was sorted under Stage-2 (cf. (7)). Finally, expressions that are additionally incompatible with the English source language grammar were collected under Stage-3, see (8).

(5) de: Best party ever

(6) es: el worst deal ever
the worst deal ever ‘the worst deal ever’

Language	Stage-1		Stage-2		Stage-3	
Dutch	38	(10%)	276	(75%)	1	(< 1%)
German	3	(1%)	287	(96%)	7	(2%)
French	14	(9%)	123	(59%)	15	(9%)
Spanish	11	(1%)	71	(59%)	33	(28%)

Table 3: Distribution of the nativization stages

(7) fr: la plus belle fin de chanson ever
the most beautiful end of song ever
'the most beautiful song ending ever'

(8) nl: Gisteren de allermooiste babyshower gehad ever ...
yesterday the utmost beautiful babyshower had.PTCP ever
'Yesterday (we) had the utmost beautiful babyshower ever ...'

For the Romance languages, which allow for pre- and post-nominal adjectives we found that pre-nominal adjectives are more common, yet post-nominal adjectives are clearly possible: Our French sample contained 144 ES-*ever* instances featuring a pre-nominal adjective and 14 featuring a post-nominal adjective, as in (4) above. Our Spanish data revealed 88 ES-*ever* instances with a pre-nominal adjective and 32 with a post-nominal adjective. We classified all instances of post-nominal adjectives as belonging to Stage-3, as post-nominal uses of adjectives in English are rather rare, but see Section 3.

For every language we found examples from every stage. The numbers and percentages are given in Table 3. It shows that Stage-2 is the most frequently represented stage in all four languages. The higher percentage of Stage-3 cases in the Romance languages is primarily due to the use of ES-*ever* with post-nominal adjectives.

3 Syntax of emphatic superlative expressions

In this section, we will first look at the general syntax of adjectival modification in the languages under discussion. Then, we will show how the matrix language emphatic superlative expressions are integrated.

Pollard & Sag 1994 analyzed adjectival modification in English as an AP combining with a nominal category that is saturated for complements. Sadler & Arnold (1994) show that this is not adequate since pre-nominal adjectives are rather restricted in their complexity. For example, they allow degree particles but no complements (*the very proud (*of their kids) parents*). In contrast to this, post-nominal adjectives can show full complexity (*the parents very proud of their kids*). Consequently, Sadler & Arnold analyze only post-nominal adjectives as full APs

that combine with a complement-saturated nominal projection. Combinations with pre-nomial adjectives are treated as "small constructions," for which they propose that an A^0 category combines with an N^0 head.

Abeillé & Godard 2000 argue that the structures of French adjectival modification are analogous to those found in English. French also has severe restrictions on the syntactic complexity of pre-nominal adjectives, but none on post-nominal adjectives. However, only very few adjectives can occur pre-nominally at all. Abeillé & Godard use a head feature WEIGHT whose value is *light* for pre-nominal adjectives, i.e. for what Sadler & Arnold call "small constructions." Post-nominal adjectives have the WEIGHT value *non-light*. Machicao y Priemer & Winckel (2015) propose that Spanish can be analyzed like French, though with an even smaller set of pre-nominal adjectives.

In Dutch and German, all adjectives occur pre-nominally, allowing full APs in prenominal position, i.e., an analysis like that proposed in Pollard & Sag 1994 is unproblematic for Dutch and German, see (9).

(9) nl: Nederland is een met zichzelf tevreden natie, ...
The Netherlands is a with itself content nation
'The Netherlands is a nation content with itself.'

To sum up, the grammar of English adjective placement is like that of French and Spanish, even though, most adjectives occur in pre-nominal position which, therefore, make English look more like Dutch and German from the point of frequency. This means that in all five languages under discussion we find the same word order for the NP *the best book*. However, in English, French, and Spanish, *best book* is a "small construction," i.e. an $[A^0\ N^0]$ combination, whereas it is an [AP N′] combination in Dutch and German.

We can now turn to the syntax of emphatic superlative expressions from (1) in the languages looked at in this paper. Just like *ever* in English, they all occur post-nominally. Therefore, we can assume the same syntactic position for these expressions in all of our languages. In English, ES-*ever* has the same syntax as other NP-internally extraposed degree phrases. We will follow the extraposition analysis developed in Kay & Sag (2012) for the obligatorily extraposed clauses introduced by degree particles such as *so*. Their analysis is sketched in (10).

(10) en: [[so willing to help out][that they called early]]

The degree particle *so* selects the *that* clause via a list-valued feature EXTRA. In a phrase, the EXTRA values of all daughters are concatenated, unless it is an extraposition structure. There, the non-head daughter corresponds to the first element of the head daughter's EXTRA list, and the mother's EXTRA value is the rest of that list.

Just as degree particles can select extraposed clauses, we assume that elements with superlative semantics can select an emphatic superlative expression via the EXTRA list. These come in two groups. The first group consists of elements

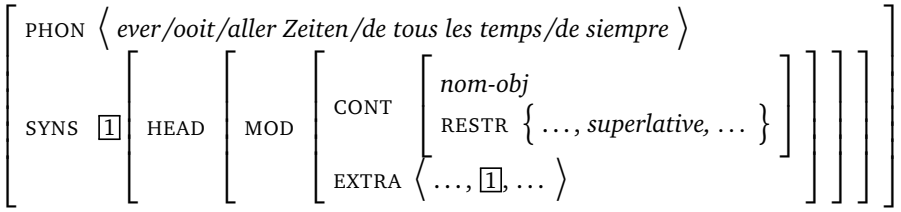


Figure 1: Sketch of an emphatic superlative expression

that introduce a morpho-syntactic superlative form: a superlative particle (like English *most* or French *plus* 'most'), a superlative morpheme attached to the positive form of an adjective by some lexical rule (like the morpheme *-st* in Dutch, English, and German), or a suppletive superlative form (like English *best* or Spanish *peor* 'worst'). Such morpho-syntactic superlatives can alternatively select a clausal comparative class expression, see (11).

(11) en: the best book ever/ of all times/ [that I have read in a long time]

The second group of expressions that can introduce an emphatic superlative expression are purely semantic/pragmatic superlatives. These include adjectives in their positive form if they have a superlative-like semantics such as *favorite* or Spanish *único* 'only', but also top-degree nouns such as *highlight* or the combination of a noun with a top-degree prefixoid like *top-* or German *Lieblings-N* 'favorite N'. For items of the second group, the occurrence of a clausal comparative class expression is less typical, though not ungrammatical, see (12)

(12) en: In the gallery, you'll find 15 of our favorite essays that we published this year. (Timestamped JSI English 2014–2021)

We will largely ignore the semantics and pragmatics of emphatic superlative expressions in this paper. Clearly, they indicate the comparison class of the superlative operator having the effect of a domain widening. Pragmatically, this has the effect of stressing the extraordinary degree to which the property in the scope of the superlative operator holds. This pragmatic effect may vary between different expressions.

We sketch a description of an emphatic superlative expression in Figure 1. Note that it modifies an element whose semantics is of the sort *nominal-object*, i.e., a noun or an adjective. The modified element must have a superlative operator on its RESTR list. Finally, the SYNSEM value of the emphatic superlative expression, 1, must be on the EXTRA list of the modified element.[3]

In Figure 2 we sketch the resulting syntactic structure for the English NP *the best book ever* and its French and Spanish translations. The structure for the Dutch and German equivalents (*het beste boek ooit/das beste Buch* [*aller Zeiten*])

[3]From here on, we will drastically simplify the AVMs, ignoring, for example the SYNSEM feature. Even when not displaying it, we commit to a full, standard HPSG feature geometry.

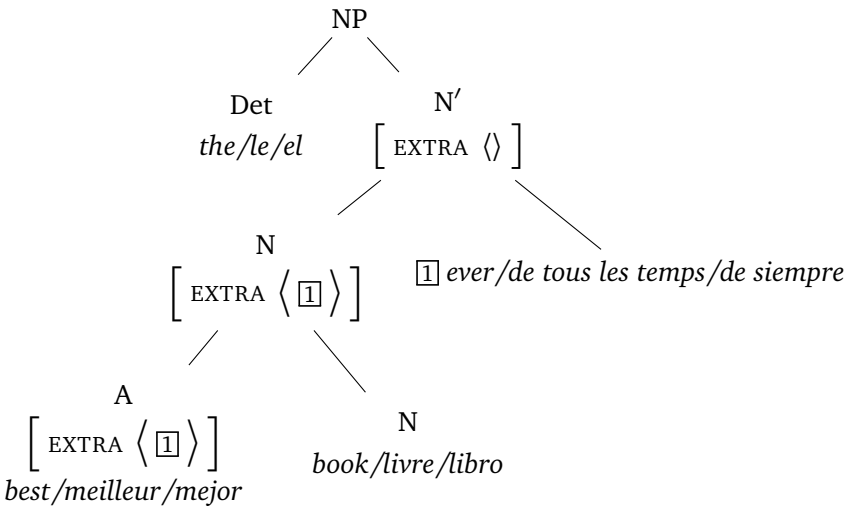


Figure 2: Sketch of the structure of an NP with emphatic superlative (en/fr/es)

looks the same, but the adjective *beste* would be an AP and the noun *boek/Buch* would be required to have an empty COMPS list.

4 A model of borrowing

In this section, we will turn to the main theoretical contribution of this paper, a development of a modelling of borrowing in HPSG. We will model the data presented in Section 2 as a three-stage process, which reflects the stages described in Section 2. We provide a Spanish example NP for each of these stages in (13).

(13) a. es: el coolest job *ever* 'the coolest job ever'
b. es: la mejor canción *ever* 'the best song ever'
c. es: las portadas más photoshopeadas *ever*

For the first stage, we postulate that a structure of the form "A N *ever*" is borrowed from English (constructional borrowing). In the second stage, we still only find the combination with pre-nominal adjectives, but matrix language lexical material is used, leaving *ever* as the only English item. Finally, *ever* is turned into a lexical borrowing and can occupy all positions found for emphatic superlative expressions in the matrix language. This means that ES-*ever* can occur with post-nominal adjectives in French and Spanish from that stage on.

In Section 4.1, we will sketch the two existing approaches whose ideas we (partly) incorporate in our own approach in Section 4.2. Finally, we will show how the three steps sketched in (13) are expressed in our model (Section 4.3).

4.1 Background

We will embed in our approach some fundamental concepts of *Diasystematic Construction Grammar* (DCxG, Höder 2012, 2018) and of the *Communicative Situations* approach CSA, Wiese 2021). Common to them is the assumption that the linguistic knowledge of multilectal and multilingual language users consists of a single repertoire, containing all elements of the apparently different linguistic system available to them. Each of these elements is, however, marked for the varieties to which they belong, or the communicative contexts in which they are appropriate. However, while DCxG is *usage based*, CSA is *competence based*. We will side with the latter.

Clearly, a single item can be appropriately used in a number of communicative contexts. In this case, DCxG assumes that items can be specific or unspecific for a particular variety, with unspecific items (so-called *diaconstructions*) being compatible with more than one. In this approach, the borrowing of ES-*ever* could be modelled in the following way: ES-*ever* starts off as specific for English and is, then, turned into a diaconstruction and can appear in structures which are similar between English and the matrix language. However, we saw that in French and Spanish, ES-*ever* occurs with post-nominal adjectives, a position barely found in English. Consequently, there must, eventually, be a language-specific variant of borrowed ES-*ever*. Nonetheless, we will assume something along the lines of a diaconstruction for Stage-2 as in (13b).

Wiese's (2021) CSA shares many basic assumptions with DCxG, but takes a competence-based stance and is less devoted to a particular framework for expressing linguistic generalizations, though she uses Jackendoff's *Parallal Architecture* framework, summarized for example in Jackendoff 2007. Wiese annotates each unit of linguistic knowledge for the *communicative situation* (ComSit) in which this unit is usually, and recurrently encountered. These ComSits can be of any degree of concreteness or abstractness. Wiese argues that the notion of a "named language" or a particular "named dialect" can be understood as very abstract socially constructed entities, and, consequently, as ComSits.

Wiese (2021: Section 2.4) provides an example of a lexical borrowing from English to German. She argues that the English word *chicken* occurs as a borrowing in German only as chicken meat and in the context of diners. I.e., the English word is integrated into German with a subpart of its English meanings and in special types of communicative situations. At the same time, the German word *chicken* is still explicitly marked as a borrowing from the (semantically and situationally more general) English word *chicken* (Wiese 2021: 14).

If we want to adopt this approach to the borrowing of ES-*ever*, we could say that English, more general *ever* is borrowed only in its emphatic superlative particle use into the discussed matrix languages. It is not clear to us if the borrowing of a particular meaning of *ever* can be as plausibly attributed to the communicative situations in which the word is used as in the case of *chicken*, where a particular food item is connected to the places in which it is usually

consumed. Nonetheless, we will, in fact, model the final borrowing stage, i.e. ES-*ever* as in (13c), in a way that is an HPSG-rendering of the CSA analysis just sketched. We will also make the link between the borrowing from "English" situations to matrix language situations even more explicit than Wiese 2021.

4.2 An HPSG-approach to borrowing in HPSG

In this section, we will provide the basic architecture for our modelling of the borrowing of ES-*ever*. So far, there has been no work on borrowing in HPSG to our knowledge. However, there has been work on social meaning and/or register.[4] Under the assumption that language users have just one all-comprising grammar with marking for communicative situations, borrowing from one language to another is nothing else but, as in Wiese's (2021) *chicken* example, making an element from ComSits typical for one language available in ComSits typical for another language. Within the CSA, there is, thus, no principled difference between borrowing among languages and among registers, or from an item acquiring a new social meaning.

In this paper, we will follow the pragmatic tradition of the modelling of sociolinguistic aspects of language, based on the architecture of CONTEXT proposed in Green (1994), and applied to diglossia in Paolillo (2000). Asadpour et al. 2022 provide a recent incarnation of this approach, modelling regional and register variation in the realization of relative clauses in English and Kurdish.[5] They assume that social meaning takes the form of statements as in (14).

(14) (X believes that) X and Y mutually believe that community Z normally believes that expression E signals ϕ. (Asadpour et al. 2022: 18)

Such statements have the formal status of *conventional implicatures* (Grice 1975, Potts 2005), or rather of *expressive/use-conditional meaning* (Potts 2007, Gutzmann 2013), i.e., we take it that social meaning has the following properties: its truth conditions are independent of the at-issue content; it relates to the current utterance situation (non-displaceability), usually it expresses something about the speaker (perspective dependency); it is hard to paraphrase explicitly (descriptive ineffability); it performs its meaning simply by being uttered (immediacy); and using several items with the same social meaning reinforces their effect rather than being perceived as redundant (repeatability).

In the pragmatic approach, statements as (14) are introduced as elements of the projective, non-at-issue content. Green (1994) follows Pollard & Sag (1994) in using the set BACKGROUND for this. Asadpour et al. (2022) assume

[4]We adopt the view that the *social meaning* of a linguistic entity is the knowledge of its typical/conventionalized association with particular communicative situations; a *register* is the subset of a language user's linguistic repertoire consistent with a particular communicative situation.

[5]For reasons of space, we cannot elaborate on the differences to other HPSG approaches such as Wilcock (1999), Bender (2007), and Machicao y Priemer et al. (2022).

that there can be different types of projective meaning, each of which having its own set- or list-valued attribute. Therefore, they use attributes PRESUPPOSITION and CONVENTIONAL-IMPLICATURE (CI) instead of a single BACKGROUND feature.[6] Social meaning statements are, then, treated as elements in a sign's CI value.

In the context of the present paper, we look at particular types of social meaning statements, namely at statements of the form in (15), i.e., that the social meaning associated with an expression is marked for a particular named language (the matrix language or the source language, here English).

(15) (X believes that) X and Y mutually believe that the speech community of the matrix language normally believes that expression *E* signals that X and Y are in a matrix-language communicative situation/ in an English communicative situation.

We will use abbreviated forms of social meaning statements in this paper which ignore the various embeddings of attitude predicates. An example of such an abbreviated form is given in (16). The AVM in (16) is part of the lexical entry of English *ever*. It expresses that the word *ever* is perceived as signaling communicative situations in which English is typically used. The relevant element in the CI set specifies the C(OMM)-SIT value as *en*. In an UTT(ERANCE) value it indicates which expression is marked for this communicative situation.[7]

(16)
$$\left[\begin{array}{l} \text{PHON } \boxed{1}\ \textit{ever} \\ \text{CTXT } \left[\text{CI } \left\langle \ldots, \left[\begin{array}{l}\text{C-SIT } \textit{en} \\ \text{UTT } \boxed{1}\end{array}\right], \ldots \right\rangle\right] \end{array}\right]$$

Such a marking for C-SIT is part of each linguistic expression. An utterance is fully English, for example, when all linguistic expressions in it have an element of the form $\left[\text{ C-SIT } \textit{en} \right]$ in their CI set. If we find elements with different C-SIT specifications, this does not lead to an ungrammatical utterance, but simply to one that is not purely monolingual.[8]

We will ignore the problematic difference between borrowing and codeswitching here, and, instead, assume that what we find in ES-*ever* is a development from what Muysken (2000: 72) calls *conventionalized code mixing* to an *established loan*. Muysken (2000) uses the term *borrowing* exclusively for lexical material that is, in clear cases, morphologically integrated into the matrix language. We will, instead, use the term *borrowing* for conventionalized or "listed" elements

[6] See Sailer & Am-David (2016) and Rizea & Sailer (2020) for other applications of this refined structure of the CONTEXT feature.

[7] We use the PHON value of an expression as value for UTT. Green 1994 even includes the entire sign, not just its PHON value. Similarly, a semantic type *u* for *utterance* is often assumed in the literature on quotes and meta-linguistic language use, such as in Potts 2005, for example.

[8] We assume a supertype *lang* for named languages, with subtypes *de*, *en*, *es*, *fr*, and *nl* for the languages discussed here.

of mixing of any level of linguistic complexity – here, at the word and the phrasal level. This reflects the rather constructional approach taken in our paper. We would use the term *codeswitching*, then, for the non-listed cases, i.e., for cases in which word- or phrase-level elements of different languages co-occur spontaneously – though not randomly, but rather in well-defined patterns, as elaborated in Muysken 2000.

We assume that a borrowed element is considered part of the matrix language, but that it is still connected to the corresponding element from the source language, just as in Wiese's (2021) analysis of the German use of *chicken*. We treat the information on borrowing as a kind of social meaning, i.e., communication participants can have mutual beliefs on what a borrowing is. This means that we will have additional elements in the CI values of borrowed expressions that note the borrowing property.

Borrowings need not satisfy the principles of grammar of the matrix language. To achieve this, they are marked as idiosyncratic signs by means of the specification $\left[\text{ COLL } \textit{irregular} \right]$, introduced in Richter & Sailer (2009: 307). This exempts them from the regular principles of grammar (such as the Head Feature Principle or the Immediate Dominance Principle).

We introduce a sort *borrowing* with at least two subsorts *lexical-borrowing* (*l-borrow*), and *constructional-borrowing* (*cx-borrow*). Each *borrowing* object has an attribute SOURCE encoding information on the source element, and an attribute TARGET with information on the target element.

In the case of lexical borrowing, these will be the lexical identifiers of the related items, i.e., the commonly used LID value or the LISTEME value of Soehn (2006).[9] For the *chicken* example, we would assume a LID value *chicken_en* for the English chicken and a LID value *chicken_de* for its borrowed version. The latter would contain a C-SIT specification as German, but also a *l-borrow* object indicating that *chicken_en* is the source, and *chicken_de* is the target.

For constructional borrowing, the source and the target are more complex. We assume that the SOURCE value is a list of signs. This captures the observation that borrowed constructions need not be structurally analyzable by the borrower. The TARGET value is the (possibly phonologically adapted) concatenation of the PHON values of the elements in the SOURCE list. The TARGET value is identical with the PHON of the overall phrase.

4.3 Modelling the borrowing steps

In this subsection, we will use the machinery introduced in Section 4.2 to model the borrowing stages for ES-*ever*, as illustrated in (13).

[9] Whereas LID is assumed to be a head feature, LISTEME is not. This allows Soehn (2006) to maintain the Head-Feature Principle within a default-free grammar even for idiomatic expressions, where the lexical identifier changes, but morpho-syntactic properties percolate regularly.

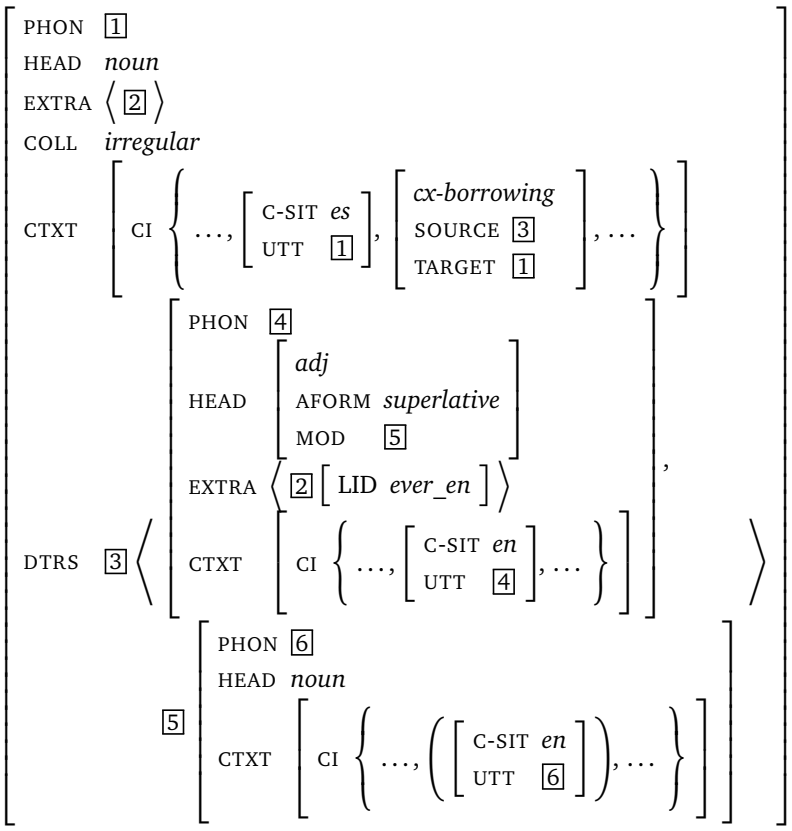


Figure 3: Description of the borrowing construction for Stage-1

Stage-1: Constructional borrowing (es: *el coolest job ever* 'the coolest job ever') In Stage-1, we find an English adjective with synthetic superlative combining with a noun and ES-*ever*, as in (13a). The noun is typically an English word that can also occur freely in Spanish utterances, whereas the adjectives can't. We model this stage by the borrowing construction in Figure 3.

The source, 3, consists of two daughters. The first daughter is a superlative adjective that is marked for English communicative situations and that modifies the second daughter (tag 5), and has the English ES-*ever* (2) on its EXTRA list. The second daughter is a noun that is also marked as English. The construction does not specify whether any standard construction of the matrix language grammar is used to combine the daughters. In fact, there need not be any, as the overall phrase is marked as *irregular*.

English ES-*ever* (2) occurs on the EXTRA list of the embedded adjective. It is explicitly inherited by the overall construction. Since Spanish has an NP-internal extraposition in its grammar, ES-*ever* can combine with the construction in Figure 3 by an ordinary construction of Spanish. Note, however, that within that construction, the extraposed element is the English word *ever*.

We sketch the resulting tree in Figure 4. The subscript on words and phrases

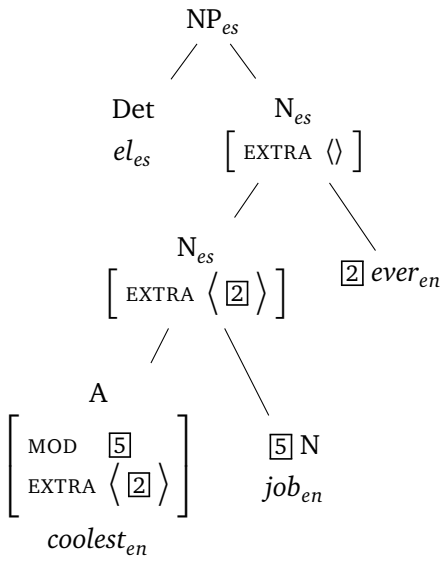


Figure 4: Structure of the Spanish NP *el coolest job ever* 'the coolest job ever'

indicates the C-SIT value associated with the phonology of those signs. The node dominating *coolest job* is licensed by the borrowing construction in Figure 3.

Stage-2: Diaconstruction (es: *la mejor canción ever* 'the best song ever') In this step, we will adopt the a DCxG-style modelling. While the borrowing construction in Figure 3 does not specify any immediate dominance schema, the word order A N is compatible with the matrix grammar: An A-N combination in French and Spanish, and AP-N combination in Dutch and German. In line with the ideas of DCxG, we assume that a more abstract, less idiosyncratic construction emerges, which is syntactically regular, allows for matrix language lexical elements, but has as its only idiosyncratic property the requirement that the adjective introduces the English word *ever*. We provide the constraint on this construction in Figure 5.

The construction in Figure 5 is marked as *regular* and as adequate for Spanish communicative situations. Consequently, all Spanish principles of grammar apply. It specifies two daughters: a superlative adjective, followed by a noun modified by that adjective. Within the Spanish grammar this restricts the set of possible adjectives. The adjective is, furthermore, required to have the English word *ever* on its EXTRA list. Since the phrase is regular, we need not specify that the EXTRA value percolates from the adjective to the mother.

Stage-3: Lexical borrowing (es: *las portadas más photoshopeadas ever* 'the most photoshoped portals ever') In Stage-2, the English word *ever* was the only idiosyncratic item in the construction. In Stage-3, this lexical item gets

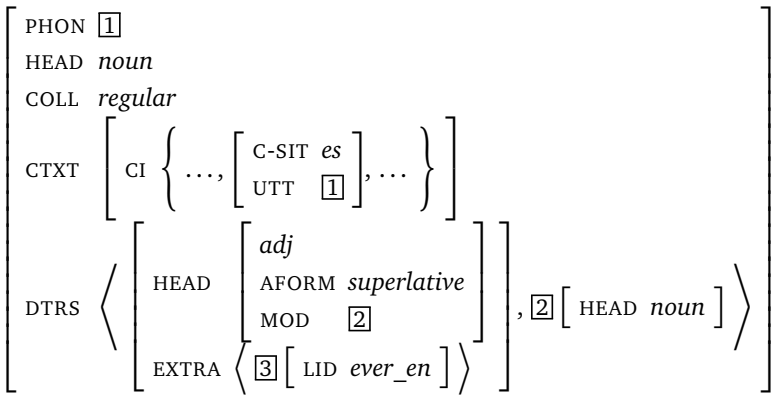


Figure 5: Description of the diaconstruction for Stage-2

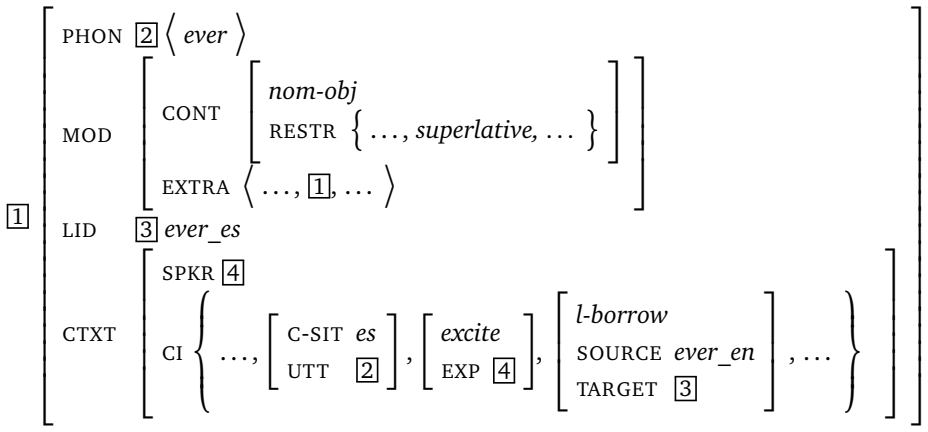


Figure 6: Lexical entry of matrix language ES-*ever* as a lexical borrowing

borrowed into the matrix language, but only with the particular meaning that it has in this construction. Figure 6 shows the lexical entry of borrowed *ever*.

Borrowed ES-*ever* is specified as modifying an item that has a nominal object in its content with a superlative semantics. It must be on the EXTRA list of the modified element (indicated with tag 1). Borrowed ES-*ever* has its proper LID value *ever_es* and is marked as appropriate for Spanish communicative situations in the CI value. The CI value also expresses the pragmatic effect of ES-*ever*, skteched with the *excite(ment)* object as some high excitement of the speaker. The word is marked as a borrowing whose source is the English word *ever*, and whose target is the LID value of the matrix language ES-*ever* (3).

At this stage, ES-*ever* can occur in any position in which matrix language emphatic superlative expressions are possible. As a consequence of our analysis, all words and phrases in an NP like the one in (13c) are marked for matrix language communicative situations. The lexical item *ever* is specified as being related to English *ever*, but it is a matrix language word and its semantics and pragmatics is much more specific than that of its English source item.

The three-stage approach developed here allows us to capture a number of observations. First, the fact that some English adjectives can occur in the matrix languages exclusively when accompanied by *ever*. These cases are licensed by constructional borrowing. Second, the tendency in French and Spanish for using *ever* with pre-nominal adjectives, contrary to the overall dominance of post-nominal adjectives. This is achieved by the diaconstruction in Stage-2. Third, even though the syntactic rules of adjectival modification of English are more similar to those of French and Spanish than to those of Dutch and German, the surface combinations of English look more like those of the other Germanic languages than those of the Romance languages. The diaconstructional perspective in Stage-2 captures this, as the surface sequence adjective-noun is interpreted differently depending on the matrix language. The Dutch/German interpretation as AP N makes the structure immediately available for all adjectives of the matrix language, whereas the French/Spanish interpretation as A N opens it only for a restricted set of adjecitves.

Finally, and perhaps most importantly, the lexical borrowing is a generalization step resulting from the earlier two stages. In those stages, the ordinary English word *ever* is used, independently of whether it has an emphatic superlative semantics by itself or gets this interpretation through the use as modifying a superlative. In Stage-3, the new, matrix language word is "extracted" from the diaconstruction and, consequently, has the meaning that is specific to an emphatic superlative marker. Thus the question of why *ever* is borrowed in its emphatic superlative reading receives a natural answer as the lexical borrowing is just a final stage in a sequence starting with a constructional borrowing.

5 Conclusion

We discussed the occurrence of emphatic superlative *ever* in Dutch, French, German, and Spanish. Besides strong similarities, our corpus data revealed qualitative and quantitative differences, primarily between the Romance and the Germanic languages. We provided an analysis of the data within the pragmatic approach to social meaning in HPSG outlined in Asadpour et al. 2022. Adopting insights and techniques from Diasystematic Construction Grammar, and the Communicative Situation Approach, we proposed a three-stage modelling of the borrowing process that accounts for the observation that while, by now, ES-*ever* should be considered a lexical borrowing, it is only in that specific use that English *ever* occurs in all four languages we considered.

It is important that often, a structure that is compatible with a certain stage can also be analyzed according to the next stage. In particular, the Dutch and German data from Stage-2 are all also compatible with a Stage-3 analysis, i.e., we cannot tell whether ES-*ever* is introduced constructionally or lexically. We consider this a strength of our approach, as it allows us to model the transition from one step to the next in terms of a re-analysis of existing data.

This paper shows that Asadpour et al.'s (2022) approach to dialect and register variation can be extended to model phenomena of code-mixing, in particular conventionalized code-mixing such as constructional and lexical borrowing. Placing our analysis in the broader picture of code-mixing, we can distinguish between structures in which all expressions belong to the same language (our Stage-3) – even if some are marked as borrowings – and structures in which expressions belong to different languages (our Stage-1 and Stage-2). Within our particular technical implementation, the latter type of structures might be called instances of *codeswitching*. Even though this is far beyond the scope of this paper, we hope that our contribution is a first step towards a formal modelling of different phenomena of language (and register) mixing and shifting in HPSG.

References

Abeillé, Anne & Danièle Godard. 2000. French word order and lexical weight. In Robert Borsley (ed.), *The nature and function of syntactic categories*, vol. 32 (Syntax and Semantics), 325–360. San Diego: Academic Press. DOI: 10.1163/9781849500098_012.

Asadpour, Hiwa, Shene Hassan & Manfred Sailer. 2022. Non-*wh* relatives in English and Kurdish: Constraints on grammar and use. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 29th International Conference on Head-Driven Phrase Structure Grammar, Online (Nagoya/Tokyo)*, 6–26. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2022.1.

Bender, Emily M. 2007. Socially meaningful syntactic variation in sign-based grammar. *English Language and Linguistics* 2(11). 347–381. DOI: 10.1017/S1360674307002286.

Green, Georgia M. 1994. The structure of CONTEXT: The representation of pragmatic restrictions in HPSG. *Studies in the Linguistic Sciences* 24(1/2). 215–232.

Grice, Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics*, vol. 3, 41–58. New York: Academic Press.

Gutzmann, Daniel. 2013. Expressives and beyond: An introduction to varieties of use-conditional meaning. In Daniel Gutzmann & Hans-Martin Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning*, 1–58. Leiden: Brill. DOI: 10.1163/9789004183988_002.

Höder, Steffen. 2012. Multilingual constructions: A diasystematic approach to common structures. In Kurt Braunmüller & Christoph Gabriel (eds.), *Multilingual individuals and multilingual societies*, 241–257. Amsterdam, Philadelphia: Benjamins. DOI: 10.1075/hsm.13.17hod.

Höder, Steffen. 2018. Grammar is community-specific: Background and basic concepts of Diasystematic Construction Grammar. In Hans C. Boas & Steffen Höder (eds.), *Constructions in contact: Constructional perspectives on contact phenomena in Germanic languages*, 37–70. Amsterdam: Benjamins. DOI: 10.1075/cal.24.02hod.

Jackendoff, Ray. 2007. A Parallel Architecture perspective on language processing. *Brain Research* 1146. 2–22. DOI: 10.1016/j.brainres.2006.08.111.

Kay, Paul & Ivan A. Sag. 2012. Cleaning up the Big Mess: Discontinuous dependencies and complex determiners. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-based construction grammar*, 229–256. Stanford: CSLI Publications.

Kilgarriff, Adam, Vít Baisa, Jan Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1. 7–36. DOI: 10.1007/s40607-014-0009-9.

Machicao y Priemer, Antonio, Stefan Müller, Roland Schäfer & Felix Bildhauer. 2022. Towards a treatment of register phenomena in HPSG. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 29th International Conference on Head-Driven Phrase Structure Grammar, Online (Nagoya/Tokyo)*, 86–101. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2022.5.

Machicao y Priemer, Antonio & Elodie Winckel. 2015. *Adjectives within Spanish and French NPs: Position and interpretation*. Handout.

Muysken, Pieter. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

Paolillo, John C. 2000. Formalizing formality: An analysis of register variation in Sinhala. *Journal of Linguistics* 36(2). 215–259. DOI: 10.1017/S0022226700008148.

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

Potts, Christopher. 2005. *The logic of conventional implicatures* (Oxford Studies in Theoretical Linguistics 7). Oxford: Oxford University Press.

Potts, Christopher. 2007. Into the conventional-implicature dimension. *Philosophy Compass* 4(2). 665–679.

Richter, Frank & Manfred Sailer. 2009. Phraseological clauses in constructional HPSG. In Stefan Müller (ed.), *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, Göttingen 2009*, 297–317. Stanford: CSLI Publications. DOI: 10.21248/hpsg.2009.15.

Rizea, Monica-Mihaela & Manfred Sailer. 2020. A constraint-based modeling of negative polarity items in result clause constructions in Romanian. *Lingvisticae Investigationes* 43(1). 129–168. DOI: 10.1075/li.00042.riz.

Sadler, Louise & Doug Arnold. 1994. Prenominal adjectives and the phrasal/lexical distinction. *Journal of Linguistics* 30(1). 187–226. DOI: 10.1017/S0022226700016224.

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. Standford: CSLI Publications. DOI: 10.21248/hpsg.2016.33.

Soehn, Jan-Philipp. 2006. *Über Bärendienste und erstaunte Bauklötze: Idiome ohne freie Lesart in der HPSG*. Ph.D. thesis, Friedrich-Schiller-Universität Jena. Frankfurt am Main: Peter Lang.

Wiese, Heike. 2021. Communicative situations as a basis for linguistic systems. *Working Papers in Urban Language & Literacies (WPULL)* 287.

Wilcock, Graham. 1999. Lexicalization of context. In Gert Webelhuth, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and constructional aspects of linguistic explanation*, 373–387. Stanford: CSLI Publications.

Argument realization and argument referencing in Soranî Kurdish

Ali Salehi
University at Buffalo

Jean-Pierre Koenig
University at Buffalo

Proceedings of the 30th International Conference on
Head-Driven Phrase Structure Grammar

University of Massachusetts Amherst

Stefan Müller, Elodie Winckel (Editors)

2023

Frankfurt/Main: University Library

pages 100–120

Keywords: Soranî Kurdish, Argument Indexing, Edge Inflection, Morphology

Salehi, Ali & Jean-Pierre Koenig. 2023. Argument realization and argument referencing in Soranî Kurdish. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 100–120. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2023.6.

Abstract

Soranî Kurdish can reference up to two arguments morphologically, a subject agreement marker and an incorporated object pronoun. One of the argument referencing morphs is verb-bound and occurs in a fixed position in the verb template (after the stem), while the other is a mobile morph that can occur either verb-internally (in second or last position) or verb externally. Either the subject agreement marker or the object incorporated pronoun can be verb bound or mobile morphs, depending on the tense and presence of an NP complement. Previous literature has analyzed mobile morphs as (VP) endoclitics. We argue that this is not the case as verb-external mobile morphs occur at the end of the last word of the least oblique NP complement and cannot attach to the last word of VP-internal PPs. We provide an edge-feature based analysis of verb-external mobile morphs and show that the same realizational rules account for the exponents of mobile morph features whether they occur verb-internally or verb-externally. We furthermore suggest that the dissociation between paradigm class (verb-bound or mobile morph) and syntactic status (subject or object; agreement marker vs. incorporated pronoun) challenges views that treat morphological structure as isomorphic to syntactic structure.

1 Introduction

Semantic arguments of predicators can be realized syntactically or be morphologically referenced on the head. Haspelmath (2013) uses the term *argument indexing* to cover both possibilities, a usage we follow, although our paper focuses on the second possibility, the use of affixes to index arguments. We use the term *morphological argument referencing* or *argument referencing* for short for the *morphological* indexing of semantic arguments. Soranî's argument referencing poses interesting challenges to the interface between morphology and syntax and has for this reason received quite a bit of attention (Samvelian 2007, Bonami & Samvelian 2008, Walter 2012, Bonami & Crysmann 2013, Crysmann 2021, Akkus et al. 2023). The issue that makes Soranî argument referencing of particular interest is that it lives a double life. Some argument indices are always expounded verb internally after the verb stem as run-of-the-mill verbal affixes but some argument indices either occur verb internally (in all but one case in second position within the verb form) or are affixed/cliticized on the last word of a constituent that precedes the verb within the VP. We refer to these two ways of expounding Soranî argument indices as verb-bound and mobile morphs, respectively (Bonami & Samvelian (2008) use the terms *Verbal Person Endings* and *Mobile Person Markers*). In previous HPSG literature, mobile morphs have been treated as second-position (endo-)clitics within the VP, in a way that is reminiscent of Pashto endoclitics (Dost 2007). In this paper, we argue mobile morphs are not second position (endo-)clitics. They do not necessarily occur at the end of the first VP constituent; they occur at the end of the least oblique NP complement. We further argue that a proper analysis of Soranî morphological argument referencing requires dissociating the syntactic status of the argument being referenced (which members of the ARG-ST list is being referenced) from the morphological status of the

inflectional feature being expounded (which paradigm the inflectional feature belongs to). We present an analysis of Soranî morphological argument referencing within Information-based Morphology (IbM) (Crysmann & Bonami 2016) and provide a set of realizational rules for mobile morphs that ensure they are second-position/last position affixes either within the verb form or after the last word of the NP complement to which they are affixed (making use of edge features in the latter case, see Miller & Halpern 1993).

2 Morphological argument referencing in Soranî

Soranî, previously described as a split-ergative language (Thackston 2004, McCarus 2009, MacKenzie 1960), can morphologically reference up to two arguments. When referencing the subject argument in the present tense, a *verb-bound* post-stem affix is employed and a subject external NP can co-occur and agree with the verbal affix. Verb-bound morphs in sentences (1)-(2), for example, agree with the external subject NP[1].

(1) *minał-ekan/(ewan) e-řo-n*
kid-DEF.PL/they IPFV-GO.PRS-3PL
'The kids/they are leaving.'

(2) *(êma) nan e-xo-yn*
we food IPFV-eat.PRS-1PL
'We are eating food.'

In contrast to what is the case with subject NPs, object NPs are in complementary distribution with argument referencing *mobile morphs*. Thus, in example (3), the object is expressed as a lexical NP, while in example (4) from Thackston (2004), the object is morphologically referenced by the mobile morph *-t* in second position within the verb form (preceding the stem in this case). Importantly, an object NP cannot co-occur with the mobile morph. In other words, while the morphs that reference the subject in examples (1)-(2) are subject agreement markers, morphs that reference the object realize the relevant argument of the verb in the sense of Levin & Rappaport Hovav (2005). The complementary distribution between morphological argument referencing and syntactic argument realization suggests that object argument referencing affixes are incorporated pronouns in the terminology of Bresnan & Mchombo (1987) who discuss a similar distinction in Chichewa.

(3) (*min*) *kořˇ-eke e-bîn-im*
I boy-DEF.SG IPFV-see.PRS-1SG:A
'I (will) see the boy.'

[1] All examples in this work are from fieldwork data unless stated otherwise. Soranî speakers were all from Suleymanî in Iraq or Baneh in Iran.

(4) *e-t-bîn-ê*
IPFV-2SG:P-see.PRS-3SG:A
'S/he (will) see you.'

In the past tense, the correspondence between syntactic arguments and verb-bound vs. mobile morphs switches. Subject agreement markers are now second position mobile morphs while post-stem verb-bound morphs reference and realize the verb's object. Consider examples (5) and (6). The second-position mobile morphs *-man* and *-yan* are now subject agreement markers and can co-occur with a lexical NP while the object markers are now verb-bound morphs that occur after the stem and they cannot co-occur with external NPs.

(5) (*ême*) *e-man-kêşan-n*
we IPFV-1PL:A-pull.PST-3PL:P
'We were pulling them.'

(6) (*ewan*) *girt-yan-în*
they catch.PST-3PL:A-1PL:P
'They caught us.'

Although verb internal mobile morphs typically occur in second position, a third singular mobile morph occurs after the verb-bound morph rather than in the usual second position within the verb template, as shown in (7) and (8) (from Thackston 2004).

(7) *xward-in-î*
eat.PST-3PL-3SG
'S/he ate them'

(8) *dît-în-î*
see.PST-1PL-3SG
'S/he saw us.' (Thackston 2004)

Table 1 summarizes the phonology and position (PC) of verb-bound and mobile morphs (we omit a table of the Soranî verb template for reasons of space). Note that although this paper deals almost exclusively with the position of mobile and verb-bound morphs, the exponents of the two classes of morphs also differ segmentally. The distinction between verb-bound and mobile morphs is thus one of paradigm class.

The distinction between mobile morphs and verb-bound morphs is straightforward when there is no NP preceding the verb as in the examples we provided so far. The presence of external NPs complicates the distribution of mobile morphs. In example (9), the proto-agent argument is referenced by the verb-bound suffix *-n* in the present tense; however, in the past tense example (10), the proto-agent is referenced by the mobile morph *-yan* suffixed to the last word of the object NP.

(9) *minał-ekan hermê-kan beş e-ke-n*
kids-DEF.PL pear-DEF.PL share IPFV-do.PRS-3PL
'The kids share the pears'

	Verb Bound	PC	*Mobile Morph*	PC
1SG	-m	9	-(i)m	2ND
2SG	î(t)	9	(i)t	2ND
3SG	-ê(t)/a(t)/Ø	9	-î	LAST
1PL	-în	9	-man	2ND
2PL	-(i)n	9	-tan	2ND
3PL	-(i)n	9	-yan	2ND

Table 1: Morphological argument marking and position class in Soranî

(10) *heřmê-ek-an-yan beş kird*
pear-DEF-PL-3PL:A share do
'They shared the pears.' (Mohammadirad 2020)

The mobile morph in sentence (10) is suffixed to the direct object NP. The position of mobile morphs in sentences that contain complex predicates shows that mobile morphs can be suffixed to a complement NP even if it is not a direct object or does not correspond to a semantic argument. Mobile morphs simply attach to the least oblique NP complement.[2] Complex predicates in Soranî can be formed by combining a light verb and a so-called deverbal, something that is widely assumed to be a nominal. A mobile morph can correspond to the proto-patient argument and be suffixed to the deverbal (11) with the proto-agent expounded through a verb-bound morph on the light verb in the present tense or can be a subject agreement affix (12) in the past tense while the proto-patient is expounded morphologically on the light verb of the complex predicate. The deverbal in both sentences behaves as an NP.

(11) (*ewan*) *bang-man e-ke-n*
they call-1PL:P IPFV-do.PRS-3PL:A
'They (will call) are calling us.'

(12) (*ême*) *bang-man kird-in*
we call-1PL:A do.PST-3PL:P
'We called them.'

Importantly, when a complex predicate also combines with a direct object external NP, the subject agreement marker (in the past tense) is now affixed to the last word of that object NP. This phenomenon can be observed in example (13), where the subject agreement is suffixed to the object complement of the complex predicate *heřmekan* 'the pears.' The subject marker can still be affixed to the last word of the constituent headed by the deverbal even when the verb semantically selects for a

[2] B. Crysmann (p.c.) asked whether mobile morphs attach to NP complements or to second least oblique NP syntactic argument (the second NP member of the ARG-ST list). Examples that involve displaced NP arguments are critical to decide between these two alternatives. Unfortunately, the data we have been able to gather up to this point is not entirely conclusive. We continue to talk about least oblique NP complement for convenience's sake.

proto-patient argument, provided the object NP modifies the deverbal, as illustrated in example (14). The complex predicate in this example is semantically dyadic, corresponding to the English verb *look*. However, the realization of the second semantic argument, the entity being looked at, modifies the deverbal of the complex predicate. This modification is indicated by the presence of an ezafe morph (EZF) on the deverbal. As the constituent headed by the deverbal is now the only NP within the VP, the subject agreement marker is suffixed to the last word of the modifier of the deverbal. Overall, the contrast between examples such as (11)-(12) or (14), on the one hand, and (13), on the other, shows that verb external mobile morphs are suffixed to the last word of the *least oblique* complement NP, the deverbal when there is no other NP complement, the object NP, when there is one.

(13) *heřmê-ek-an-yan beş kird*
pear-DEF-PL-3PL:A share do
'They shared the pears.' (Mohammadirad 2020)

(14) [*seyr-î wêne-kan-yan*] *kird*
look-EZF photo-DEF.PL-3PL:A do.PST
'They looked at the photos.'

While complex predicates in Soranî are typically semantically dyadic, some complex predicates are semantically monadic and still exhibit a subject agreement marking pattern typically associated with dyadic predicates as examples (15) and (16) illustrate. In both cases, the complex predicate is semantically monadic, yet subject agreement is marked by a mobile morph on the deverbal: the subject agreement marker is suffixed to the right edge of the deverbal NP complement just as for verbs that have an object NP complement.

(15) *pyase-man kird*
walk-1PL do.PST
'We walked.'

(16) *yarî-yan kird*
play-3PL do.PST
'They played.'

The contrast between the verb form-internal subject agreement marker in (17) and the verb form-external subject agreement markers in (15) and (16) shows that it is the presence of an external NP complement, rather than semantic dyadicity, that governs the occurrence of the mobile morph outside of the verb form.

(17) *(ewan)/dar-ek-an kewt-in*
they/tree-DEF-PL fall.PST-3PL:S
'They/the trees fell.'

Stepping back, the dissociation between the morphological and syntactic status of Soranî argument referencing this section discusses poses significant challenges to

linguistic theories that assume an isomorphism between the linear order of morphs and constituency or grammatical function (e.g., some version of the Mirror Principle proposed by Baker 1985). Verb-bound and mobile morphs do not consistently correspond to subject agreement markers or the realization of the proto-patient argument, *contra* what one would expect if morphology mirrored syntactic structure.

We have up to now provided examples that suggest that verb external mobile morphs are suffixed to the least oblique NP complement. This contrasts with previous HPSG literature that proposed that mobile morphs are second position clitics (within the VP) when they occur verb externally. We turn to examples that support our generalization and invalidate previous descriptions of Soranî mobile morphs as endoclitics. As PP complements play a critical role in comparing the two descriptions of the distribution of mobile morphs, we first discuss general properties of Soranî adpositions. Traditionally, Soranî adpositions (prepositions) have been grouped into two classes, simplex and compounded adpositions. The inflectional features and morphological attachment of these adpositions have been described extensively (MacKenzie 1960, Thackston 2004, McCarus 2009). Critical for our purposes, each class of adpositions can be further divided into simple and absolute adpositions, as proposed first in MacKenzie (1960). Basically, the pronominal complement of simple adpositions are separate words while the pronominal argument of absolute adpositions is referenced as a morph suffixed to those adpositions.

Most relevant to the issue of the status of the verb-external occurrence of mobile morphs is the fact that the position of mobile morphs seems to depend on whether the adposition is semantically potent. When the adposition is not semantically potent, the mobile morph may be suffixed to the last word of the phrase that contains the adposition, as seen in (18), where the adposition *be* functions as a case marker. By contrast, the same adposition is semantically potent (encoding the notion of goal of motion) in (19) and the mobile morph is now on the last word of the object NP, not on the last word of the PP. The prepositions in (19)-(21) are similarly semantically potent and the mobile morph does not occur on the last word of the phrases that contain these semantically potent adpositions; rather, the mobile morph occurs verb internally, as if the verb had no complement.

(18) *be Ali-m gut*
to Ali-1SG say.PST
'I told Ali.'

(19) *bo qotabxane minał-ekan-man nard*
to school kid-DEF.PL-1PL:A send.PST
'We sent the kids to school'

(20) *le Kurdistan bar-man kird-in*
from Kurdistan load-1PL:A do.PST-3PL:P
'We loaded them from Kurdistan.'

(21) *berew dorge-ke mele-yan e-kird*
toward island-DEF.SG swim-3PL IPFV-do.PST
'They were swimming toward the island.'

Data from conjunction reduction (22) and VP-anaphors (23) show that semantically potent PPs are VP internal in Soranî. The mobile morph agreement marker for the subject *Ali* is suffixed to the last word of the direct object NP in the first conjunct in (22), but occurs verb internally in the second conjunct because the complement of the complex predicate is a PP, not an NP. The fact the combination of the PP and the complex predicate (but to the exclusion of the subject) can be combined with the combination of a direct object and a complex predicate shows that the PP is indeed VP internal. The same conclusion can be drawn from (23). The anaphor in the second conjunct is a VP anaphor, as the two conjuncts have different subjects and its antecedent is *legeł mindałekan qisey kird* suggesting that the PP *legeł mindałekan* 'with the kids' is VP internal.

(22) *Ali mindał-ekan-$\hat{\text{i}}_{\text{MM}}$ timaşa kird û legeł mîwan-ekan qise-y_{MM} kird*
Ali kid-DEF.PL-3SG watch do.PST and with guest-DEF.PL talk-3SG do.PST
'Ali watched the kids and talked to the guests.'

(23) *Ali legeł mindał-ekan qise-y_{MM} kird û min-iş herwa*
Ali with kid-DEF.PL talk-3SG:A do.PST and I-too so
'Ali talked with the kids and so did I.'

To conclude, the contrast between (18)-(19) and (20)-(21) indicates that mobile morphs do not always attach to the last word of the first complement of the VP when they occur verb-externally, as had been previously claimed (Samvelian 2007, Bonami & Samvelian 2008). Rather, they are suffixed to the least oblique NP complement, if there is one, or to a phrase that contains a semantically inert preposition (as in (18)). We analyze semantically inert prepositions as markers: the head of *be Alim* 'to Ali' in (18) is thus an NP, in line with our hypothesis that verb external mobile morphs suffix to the last word of the least oblique NP complement.

We round up our description of Soranî morphological argument referencing with two constructions, the applicative and possessor raising, that add an NP argument to the ARG-ST list and can affect argument referencing. Absolute adpositions can function as applicative markers and increase the number of direct syntactic arguments of the verb as shown in (24)-(25) and discussed in Karim & Salehi (2022). The number of direct arguments increases from 1 to 2 or 2 to 3 when *lê* is added . As the recipient argument in (24) is now expressed as a direct syntactic argument, it is referenced via a verb-bound suffix and the subject agreement marker becomes a mobile morph. Similarly, the addition of *lê* in (25) allows the proto-patient to be referenced by a mobile morph.

(24) *pirsyar-man lê kird-in*
question-1PL:A ABS.P do.PST-3PL:P
'We asked them a question.'

(25) *lê-t-e-ç-ê*
ABS.P-2SG:P-IPFV-GO-3SG
'S/he looks like you.'

More generally, adpositions can head a PP that realize a recipient argument or they can combine with a verb as absolute adpositions. In the latter case, the member of the ARG-ST that corresponds to the recipient of a typical semantically tryadic verb is an object NP (when the verb is in the active voice; we omit discussion of passives for reasons of space) and that object NP can be referenced morphologically when it bears a pronominal index. The contrast between these two alternative ways of expressing a recipient is illustrated in (26) vs. (27) and (28).

(26) *mamosta-ke bo ême kitêb-eke-y hênaw*
teacher-DEF.SG for we book-DEF.SG-3SG:A bring.PST
'The teacher brought a book for us.'

(27) *mamosta-ke kitêb-eke-y bo hênaw-în*
teacher-DEF.SG book-DEF.SG-3SG.A ABS.P bring-1PL
'The teacher brought the book for us.'

(28) *kitêb-ek-it bo e-hên-im*
book-DEF.SG-2SG:BEN for IPFV-bring.PST-1SG:A
'I will bring a book for you.'

All Soranî verbs with three NP dependents must include an absolute adposition that functions as an applicative marker, except for the verb *dan* 'give', as shown in (29), where the pronominal recipient is referenced by a verb-bound morph or as in (30) where it is referenced by a mobile morph.

(29) *kitêb-ek-an-im da-n-ê*
book-DEF-PL-1SG:A give.PST-3PL:R-ABS.P
'I gave them the books.'

(30) (*min*) *sêw-êk-it e-de-m-ê*
(1SG) apple-IND-2SG:R IPFV-give.PRS-1SG.A- ABS.P
'I will give you an apple' (Mackenzie 1961: Sul.2)

The verb *dan* 'give' is also exceptional in that when the postposition *-ê* is present, all three arguments of the verb can be referenced morphologically, as shown in (31).

(31) *da=m-in-in-ê*
give.PST=1SG.A-3PL.O-3PL.R-APPL
'I gave them to them.' (Karim & Salehi 2022)

We now turn to the possessor raising construction, which like applicative uses of absolute adpositions, adds an NP to the ARG-ST list and thus affects argument referencing. Soranî simple verb stems, complex predicates and applicative constructions

all allow for possessor raising when the verb is in the past tense under certain conditions. In such cases, the pronominal possessor of an NP is realized as a verb-bound morph after the verb stem and the subject agreement marker is a mobile morph suffixed to the least oblique member of the ARG-ST. Examples are provided in (32)-(34). While possessor raising can be observed with a variety of verbs, it is subject to certain constraints. For example, the verbs *birdin* 'to take' and *peřandin* 'to fly/to jump' allow possessor raising only for certain person/number combinations and some verbs do not allow possessor raising at all. Compare (32) to (34): the possessor of the book in sentence (32) is referenced on the verb with a verb-bound morph but the possessor of the kids in sentence (34) is referenced on the direct object itself, despite the main verb being identical in both sentences. The exact conditions under which possessor raising occurs are outside the scope of this paper.

(32) *ktêb-ek-it* *bird-im*
book-DEF.SG-2SG:A take.PST-1SG:POSS
'You took away my book.'

(33) *xew-it* *lê-peřand-im*
sleep-2SG:A APPL-jump-1SG:POSS
'You ruined my sleep.'

(34) *minal-ek-m-yan* *bird*
kid-DEF.PL-1SG:POSS-3PL:A take.PST
'they took my kids'

3 An IbM account of Soranî argument referencing

Having described how morphological argument referencing works in Soranî, we turn to our model of these facts. To model Soranî morphological argument referencing, we make use of Information-based Morphology (Crysmann 2021, Crysmann & Bonami 2016). In Information-based Morphology (IbM), an INFL feature that records all relevant information for inflection is included in the representation of each word, as shown in (35).

(35) $word \Rightarrow \left[\begin{array}{ll} \textsc{infl} & \left[\begin{array}{ll} \textsc{mph} & list(mph) \\ \textsc{rr} & \left[\begin{array}{ll} \textsc{mud} & set(msp) \\ \textsc{ms} & set(msp) \\ \textsc{mph} & list(mph) \end{array}\right] \\ \textsc{ms} & set(msp) \end{array}\right] \end{array}\right]$

The INFL feature includes three distinct features. The first feature, MPH, is a partially ordered list of the exponents that make up a word and corresponds to the "output" of the set of realizational rules the word instantiates. This list is critical in establishing the linear order of exponents that make up the phonological shape of the word. The fact that the list is partially ordered allows for some degree of flexibility

in the ordering of morphs within the word when a language allows for it. The second feature, RR, or set of realizational rules, is responsible for mapping morphosyntactic properties to exponents. Each realizational rule is made up of three properties. MUD (Morphosyntactic features Under Discussion) includes features that a rule expounds while MPH is the set of exponents of the MUD features. Finally, MS is the entire set of morphosyntactic features a word expounds and provides a context for the application of each realizational rule when appropriate. Each member of MPH includes both phonological information (recorded in the value of PH) along with position in the word template (recorded in the value of PC). The final feature of INFL is the MS feature, which, as just mentioned, includes all the morphosyntactic features a word expounds, including the form of the stem on which the word is based. In order to ensure that the word's morphosyntactic property set is present on each realizational rule and provides a possible context for the application of rules, structure sharing is employed between the MS value of each realizational rule and the MS value of the INFL feature.

To guarantee the application of realizational rules and the generation of well-formed words, Crysmann (2021) proposes a general principle for the well-formedness of words. This principle, stated in (36), ensures that the properties expounded by any rule are consistent with the word's morphosyntactic features and that the exponents for each rule appear in the list of morphs of the word: each morphosyntactic feature is expounded in one and only one rule and each exponent of a rule appears in the word form.

(36) *word* ⇒
$$\left[\begin{array}{l}\textsc{infl}\ \left[\begin{array}{ll}\textsc{mph} & \boxed{e_1} \bigcirc \ldots \bigcirc \boxed{e_n} \\ \textsc{rr} & \left\{\left[\begin{array}{ll}\textsc{mph} & \boxed{e_1}\\ \textsc{mud} & \boxed{m_1}\\ \textsc{ms} & \boxed{0}\end{array}\right], \ldots \left[\begin{array}{ll}\textsc{mph} & \boxed{e_n}\\ \textsc{mud} & \boxed{m_n}\\ \textsc{ms} & \boxed{0}\end{array}\right]\right\} \\ \textsc{ms} & \boxed{0}\left(\boxed{m_1} \uplus \ldots \uplus \boxed{m_n}\right)\end{array}\right]\end{array}\right]$$

To model Soranî's mobile morphs one important feature needs to be added since the position of mobile morphs in the verb template is not absolute but relative: mobiles morphs (except for third person) occur after the first morph, irrespective of the position in the verb template of that first morph. Crysmann (2021) proposes an analysis of mobile morphs that makes use of a "pivot feature" 1ST-PC to model their relative position within the verb template. This pivot feature is included in all members of the list of morphs (as shown in (37)) and can thus be referenced by rules that realize mobile morphs verb-internally such as the one he proposes in (39).

(37) *word* ⇒
$$\left[\textsc{infl}\ \left[\textsc{mph}\ \left\langle\left[\begin{array}{ll}\textsc{pc} & \boxed{1}\\ \textsc{1st-pc} & \boxed{1}\\ \textsc{stm-pc} & \boxed{s}\end{array}\right], \left[\begin{array}{ll}\textsc{1st-pc} & \boxed{1}\\ \textsc{stm-pc} & \boxed{s}\end{array}\right], \ldots, \left[\begin{array}{ll}\textsc{1st-pc} & \boxed{1}\\ \textsc{stm-pc} & \boxed{s}\end{array}\right]\right\rangle\right]\right]$$

The rule in (39) is intended to model the mobile morph *-man* in (38): it requires *-man* to occur one position after the first position in the verb template, i.e. in second position.

(38) *e-man-xward-in*
IPFV-1PL:A-eat.PST-3PL:P
'We didn't eat them.'

(39)
$$\begin{bmatrix} \textsc{mud} & \left\{ \begin{bmatrix} \textsc{pers} & 3 \\ \textsc{num} & \textit{pl} \end{bmatrix} \right\} \\ \textsc{mph} & \left\langle \begin{bmatrix} \textsc{ph} & \langle \text{man} \rangle \\ 1\textsc{st-pc} & \boxed{1} \\ \textsc{pc} & \boxed{1}+1 \end{bmatrix} \right\rangle \end{bmatrix}$$

The analysis presented in Crysmann (2021) is meant to cover verb-internal exponence of mobile morphs. In the rest of this section, we provide a more general account of Soranî argument referencing, an account that covers both verb-bound morphs and verb-external as well as verb-internal occurrence of mobile morphs. To provide a comprehensive analysis of Soranî morphological argument referencing, we need to first distinguish between two sorts of argument referencing features or *arg-ref* whose informational structure is provided in (40). *Vb-arg-ref* and *mm-arg-ref* (for *verb-bound* and *mobile morph* argument referencing features) are subsorts of *arg-ref* (see Bonami & Crysmann (2013) for a similar distinction between *verbal person ending* and *mobile person markers* morphosyntactic feature sorts). These two subsorts of *arg-ref* encode the fact that verb-bound and mobile morphs constitute two distinct paradigms. Realizational rules can then distinguish between the position of the exponents of these two subsorts (second morph or post-stem morph) as well as their phonology, except in the case of the first person singular form where the segmental make-up of the morph is the same for the two subsorts of argument referencing features (but their position remains distinct). The realizational rule for the first person affix will target the *arg-ref* sort rather than either of its subsort, ensuring that the PH, but not the position class, of first person affixes is shared between the relevant verb-bound and mobile morphs.

(40)
$$\begin{bmatrix} \textit{arg-ref} \\ \textsc{arg-ind} & \textit{nom-ind} \end{bmatrix}$$

Having introduced the feature that are targeted by argument referencing realizational rules, we now turn to how argument referencing is achieved. Conceptually, Soranî argument referencing involves three distinct sets of constraints. The first set relates indices of syntactic arguments (members of ARG-ST) with morphosyntactic features (members of MS or the scrapbook feature MM-FEAT, see below), what we call *argument referencing features* for ease of reference. This set of constraints ensures that argument referencing features are assigned the right sort, *vb-arg-ref* or *mm-arg-ref*, respectively, as well as the appropriate person and number. The rule we provided in (39) from Crysmann (2021) is thus amended as follows so that it only applies to argument referencing features of subsort *mm-arg-ref*.

(41)
$$\begin{bmatrix} \textsc{mud} & \left\{ \begin{bmatrix} \textit{mm-arg-ref} & \\ \textsc{arg-ind} & \begin{bmatrix} \textsc{pers} & 3 \\ \textsc{num} & \textit{pl} \end{bmatrix} \end{bmatrix} \right\} \\ \textsc{mph} & \left\langle \begin{bmatrix} \textsc{ph} & \langle \text{man} \rangle \\ 1\textsc{st-pc} & \boxed{1} \\ \textsc{pc} & \boxed{1}+1 \end{bmatrix} \right\rangle \end{bmatrix}$$

The second set of constraints—which is only relevant to verb-external mobile morphs—ensures that a *mm-arg-ref* feature is present on the last word of an NP complement if there is one, rather than being expounded on the verb. The third set of constraints are the realizational rules themselves, for both verb-bound and mobile morphs. We tackle each set of constraints in turn.

The first step in modeling Soranî argument referencing involves relating indices of syntactic arguments with argument referencing features. As discussed extensively in previous literature as well as in Section 2, subjects of present tense verbs are referenced by verb-bound affixes whether the verb is intransitive or transitive. In other words, whether the verb's ARG-ST includes one or two NPs, the argument referencing feature for the subject is of sort *vb-arg-ref*. The constraint in (42) models this generalization as follows. The ARG-ST includes an NP corresponding to the verb's subject followed by a (possibly empty) *list* of other syntactic arguments. The NP argument's index is structure-shared with the value of the ARG-IND feature of *vb-arg-ref* in the morphosyntactic property set. This structure-sharing ensures that the exponent of the argument referencing feature agrees with the subject NP. As the argument referencing feature (42) introduces is of sort *vb-arg-ref*, subject agreement affixes will be verb-bound morphs.

(42)
$$\begin{bmatrix} \textsc{head}|\textsc{tns} & \textit{pres} \\ \textsc{arg-st} & \left\langle \textsc{np}\boxed{1} \right\rangle \oplus \textit{list} \end{bmatrix} \Rightarrow \begin{bmatrix} \textsc{infl} & \begin{bmatrix} \textsc{ms} & \left\{ \begin{bmatrix} \textit{vb-arg-ref} \\ \textsc{arg-ind} \ \boxed{1} \end{bmatrix} \right\} \uplus \textit{set} \end{bmatrix} \end{bmatrix}$$

When the verb is in the past tense, the argument referencing feature of subjects is not uniformly of sort *vb-arg-ref*. It is of sort *vb-arg-ref* if the verb only has one NP syntactic argument (is intransitive); it is of sort *mm-arg-ref* if it has two or more NP syntactic arguments. The constraint in (43) ensures that the subject argument referencing feature of past tense verbs with only one NP in their ARG-ST list is of sort *vb-arg-ref*; the constraint in (44) ensures that the subject argument referencing feature of past tense verbs with two or more NPs in their ARG-ST list is of sort *mm-arg-ref*. Note that this last constraint makes use of the MM-FEAT feature whose value is a set of *mm-arg-ref* features. As we discuss shortly, this feature—a sort of scrapbook feature—proves useful when modeling the alternation between verb-internal and verb-external exponence of mobile morph features.

(43)
$$\begin{bmatrix} \textsc{head}|\textsc{tns} & \textit{past} \\ \textsc{arg-st} & \left\langle \textsc{np}\boxed{1} \right\rangle \oplus \textit{list}\left(\left[\textsc{cat} \ \neg \textit{noun} \right] \right) \end{bmatrix} \Rightarrow \begin{bmatrix} \textsc{infl} & \begin{bmatrix} \textsc{ms} & \left\{ \begin{bmatrix} \textit{vb-arg-ref} \\ \textsc{arg-ind} \ \boxed{1} \end{bmatrix} \right\} \uplus \textit{set} \end{bmatrix} \end{bmatrix}$$

(44)
$$\begin{bmatrix} \textsc{head}|\textsc{tns} & \textit{past} \\ \textsc{arg-st} & \left\langle \textsc{np}\boxed{1}, \textsc{np} \right\rangle \oplus \textit{list} \end{bmatrix} \Rightarrow \begin{bmatrix} \textsc{infl} & \begin{bmatrix} \textsc{mm-feat} & \left\{ \begin{bmatrix} \textit{mm-arg-ref} \\ \textsc{arg-ind} \ \boxed{1} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

The sorts of argument referencing features of pronominal direct objects is the mirror image of the sorts of subject argument referencing features. Object argument referencing features are of sort *mm-arg-ref* when the verb is in the present tense and of sort *vb-arg-ref* when the verb is in the past tense, as shown in the constraints in (45) and (46), respectively. Note that while there was no restriction as to the sort of *synsem* targeted by the constraints introducing argument referencing features for subjects (Soranî is a *pro*-drop language), object argument referencing features introduced by the constraints in (45) and (46) require the relevant ARG-ST member to be of sort *pron-aff*, i.e. to be a non-canonical *synsem* that is not structure-shared with valence lists, following the work of Miller & Sag (1997).

(45) $$\left[\begin{array}{ll}\textsc{head}|\textsc{tns} & \textit{pres} \\ \textsc{arg-st} & \left\langle \textsc{np}, \textsc{np}^{\textit{pron-aff}}\boxed{2}\right\rangle \oplus \textit{list}\end{array}\right] \Rightarrow \left[\textsc{infl}\ \left[\textsc{mm-feat}\ \left\{\left[\begin{array}{ll}\textit{mm-arg-ref} \\ \textsc{arg-ind}\ \boxed{2}\end{array}\right]\right\}\right]\right]$$

(46) $$\left[\begin{array}{ll}\textsc{head}|\textsc{tns} & \textit{past} \\ \textsc{arg-st} & \left\langle \textsc{np}, \textsc{np}^{\textit{pron-aff}}\boxed{2}\right\rangle \oplus \textit{list}\end{array}\right] \Rightarrow \left[\textsc{infl}\ \left[\textsc{ms}\ \left\{\left[\begin{array}{ll}\textit{vb-arg-ref} \\ \textsc{arg-ind}\ \boxed{2}\end{array}\right]\right\} \uplus \textit{set}\right]\right]$$

The set of constraints we just provided ensures that an argument referencing feature of the appropriate sort is part of the inflectional structure of the verb and is co-indexed with the relevant ARG-ST member. In most languages, all that would then be needed would be realizational rules that expound those argument referencing features verb internally. But, Soranî's mobile morphs can occur within the verb form (mostly, in second position in the verb template) or be suffixed at the right edge of the first NP complement. So, we need to account for the fact that inflectional features of sort *mm-arg-ref* are alternatively expounded verb-internally and verb-externally. A second set of constraints is needed to model this alternate exponence of *mm-arg-ref* features. As we want the same constraints relating ARG-ST members and inflectional features to apply whether a *mm-arg-ref* feature is expounded verb-internally or verb-externally, the scrapbook feature MM-FEAT we introduced in constraints (44) and (45) comes in handy, as this feature helps keep the introduction of an argument referencing feature independent of whether that feature is expounded verb-internally or verb-externally. The constraint in (47) says that when a verb has no NP complement, the mobile morph feature set (the value of MM-FEAT) is part of the verb's MS (and will thus be expounded within the verb form, as per the Well-formedness constraint in (36)).[3] The constraint in (47) suffices to ensure the verb-internal realization of mobile morph features when no NP complement is present.

(47) $$\left[\begin{array}{ll}\textsc{infl} & \left[\textsc{mm-feat}\ \boxed{1}\right] \\ \textsc{comps} & \textit{list}\left(\left[\textsc{cat}\ \neg\ \textit{noun}\right]\right)\end{array}\right] \Rightarrow \left[\textsc{infl}\ \left[\textsc{ms}\ \boxed{1} \uplus \textit{set}\right]\right]$$

[3]We assume here that the constraint applies to *in situ* NP complements and therefore references NP members of the COMPS list. As mentioned above, the data we have been able to gather so far is not entirely probative and further fieldwork is needed. If mobile morphs can be suffixed to displaced syntactic arguments, the constraint would reference ARG-ST members. Nothing substantial hinges on this issue.

The constraint in (48) introduces via structure-sharing the mobile morph feature set (the value of MM-FEAT) on the least oblique NP complement as an edge feature rather than on the MS of the verb, thus guaranteeing the feature will not be expounded verb internally. Standard approaches to edge features further guarantee that the mobile morph argument referencing feature set is part of the representation of the last word of the least oblique NP complement (the NP whose *synsem* is tagged 2 in (48)).

(48) $$\begin{bmatrix} \textsc{infl} & [\textsc{mm-feat}\ \boxed{1}] \\ \textsc{comps} & \langle \boxed{2}\textsc{np} \rangle \oplus \boxed{3} \end{bmatrix} \Rightarrow \left[\textsc{comps}\ \left\langle \boxed{2}[\textsc{edge}|\textsc{trig}|\textsc{right}\ \boxed{1} \uplus \mathit{set}] \right\rangle \oplus \boxed{3}\right]$$

Although many different approaches to edge features would fit the bill, we follow the approach developed in Miller & Halpern 1993 and Halpern (1995) (see also Tseng 2003) and its HPSG implementation in Crysmann (2010). In that approach, edge features "percolate down" from the root of the phrase they are introduced on to the leftmost/rightmost word they are expounded on: in our case, from the root of the least oblique NP to the last word of that NP. These approaches distinguish between edge trigger features, launching an edge inflection dependency, and edge marking features. Percolation of feature values is achieved through the Edge Feature Principle as follows: "The right (left) MARK feature of the right (left) daughter is the concatenation of the right (left) MARK and TRIG features of the mother." (Crysmann 2010: 278), as stated semi-formally in (49) (adapted from Crysmann 2010). The principle in (49) together with the constraint in (48) guarantees that the mobile morph is realized at the right edge of the noun phrase that corresponds to the first NP member of the COMPS list.

(49) $$\begin{bmatrix} \textsc{ss} & \left[\textsc{edge} & \begin{bmatrix} \textsc{mark}|\textsc{right} & \boxed{2} \\ \textsc{trig}|\textsc{right} & \boxed{1} \end{bmatrix}\right] \\ \textsc{dtrs} & \left\langle [\textsc{ss}|\textsc{edge}|\textsc{mark}|\textsc{right}\ \boxed{1} \uplus \boxed{2}] \right\rangle \oplus \mathit{list} \end{bmatrix}$$

The second set of constraints we have laid out makes sure that the mobile morph argument referencing feature is either part of the verb's set of inflectional features (member of its MS) or is part of the set of right edge features of the least oblique NP and percolates down to the last word of that NP. Before one can apply the third set of constraints (the realizational rules themselves for either verb-bound or mobile morphs), we need to make sure mobile morph argument referencing features on the set of right edge features are part of the MS of the last word of the least oblique NP complement. To that end, we posit the word-to-word construction in (50).

(50) $$\begin{bmatrix} \mathit{wd\text{-}to\text{-}wd\text{-}infl} & \\ \textsc{infl} & \left[\textsc{ms}\ \left\{ \boxed{3}, \begin{bmatrix} \mathit{mm\text{-}stem\text{-}lid} \\ \textsc{stem}\ \boxed{2} \end{bmatrix} \right\}\right] \\ \textsc{edge} & \left[\textsc{mark}|\textsc{right}\ \left\{\boxed{3}\right\}\right] \\ \textsc{dtrs} & \left\langle \begin{bmatrix} \mathit{word} \\ \textsc{ph}\ \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix}$$

A few comments on this construction are in order. First, since mobile morphs can attach to uninflected as well as inflected last words of NPs, the argument referencing

feature is only part of the value of the MS attribute of the mother, as the daughter may already be a fully inflected word. Second, the construction creates a new stem whose stem phonology is identical to that of the phonology of the word to which the mobile morph is suffixed. Third, the extended word created by the construction contains in its inflectional feature set (the value of its MS) the right edge feature (1 in (50)), i.e., the mobile morph argument referencing feature as well as the new stem feature (6 in (50)). That's it! Nothing else is needed aside from the *wd-to-wd-infl* construction to properly expound mobile morph argument referencing features verb-externally. The same rules used for expounding the features verb-internally apply when they are expounded verb-externally. So, the realizational rule in (41) applies whether *-man* is realized verb-externally as in (12) or verb-internally as in (38). The *wd-to-wd-infl* construction provided in (50) directly accounts for the expounding of the *mm-arg-ref* in the same position class as when it is expounded verb-internally: the mobile morph follows the first morph of the extended word just as it does when it follows the first morph of an inflected verb.

Interestingly, the introduction via (50) of the mobile morph argument referencing feature onto the MS of last word of the least oblique NP complement also accounts for cases where the mobile morph is not the second morph in a word. Until now, we have focused on cases where the mobile morph is the second morph in an inflected word. But this is not always the case, as (53) and (54) show (to be compared to the more usual pattern in (51) and (52)). As noted by Samvelian (2007) and others, third singular mobile morphs occur *after* verb-bound morphs; in fact, they occur in the final position in the verb template, i.e. as the last suffix of the verb.

(51) *dît-man-Ø*
see.PST-1PL:A-3SG:P
'We saw him/her.'

(52) *dît-man-in*
see.PST-1PL:A-3PL:P
'We saw them.'

(53) *dît-în-î*
see.PST-1PL:P-3SG:A
'He saw us.'

(54) *xward-in-î*
eat.PST-3PL:P-3SG:A
'He ate them.'

The realizational rule in (55) models such verb-final realization of mobile morphs. As the position class is again relative rather than absolute (last morph in the inflected word), we make use of the feature LAST-PC to ensure *-î* is the last morph. This rule also applies when the mobile morph occurs verb-externally: as the MS of the mother node of any construct based on the construction in (50) contains only two inflectional

features, the stem feature and the mobile morph argument referencing feature, the last position (for third singular mobile morphs) and second position (for all other mobile morphs) are the same position.

(55)
$$\left[\begin{array}{ll} \textsc{mud} & \left\{\left[\begin{array}{ll} \textit{mm-arg-ref} & \\ \textsc{arg-ind} & \left[\begin{array}{ll}\textsc{pers} & 3\\ \textsc{num} & \textit{sg}\end{array}\right]\end{array}\right]\right\} \\ \textsc{mph} & \left\langle\left[\begin{array}{ll}\textsc{ph} & \langle\hat{1}\rangle\\ \textsc{last-pc} & \boxed{1}\\ \textsc{pc} & \boxed{1}\end{array}\right]\right\rangle \end{array}\right]$$

We have now discussed three sets of constraints required for modeling morphological argument referencing in Soranî. The first set of constraints identifies indices of NP members of ARG-ST to argument referencing inflectional features and assign the appropriate paradigm sort to these inflectional features. The second set of constraints ensures the mobile morph argument referencing feature is either part of the verb's MS (when the mobile morph occurs verb-internally) or on the MS of the last word of the least oblique NP (when the mobile morph occurs verb-externally). The third set of constraints are the realizational rules themselves that provide segmental and positional information for combinations of person and number that are appropriate for the verb-bound and mobile morph paradigms. The last aspect of Soranî argument referencing to cover is the neutralization of paradigm sort for the phonological, but not positional, exponence of first person indices, and the variable morphotactics of mobile morphs (last position for exponents of third singular indices, second position in all other cases). In both cases, our model makes use of the dissociation between phonology and morphotactics proposed in Crysmann & Bonami (2016) and the notion of Online Type Construction first proposed in Koenig & Jurafsky (1994) and Koenig (1999), as Figure 1 illustrates.

The MORPHOTACTICS dimension specifies positional information about exponents of argument referencing features. Verb-bound morphs always occur in position 9 in the verb template; mobile morphs occur in last or second position depending on whether the argument referencing feature is third singular or not. The PHON dimension specifies segmental information about exponents of argument referencing features. The segmental exponence of first person singular is the same for verb-bound and mobile morphs and the rule *1sg-phon* therefore mentions the supersort *arg-ref*. It applies whether the argument referencing feature is of subsort *vb-arg-ref* or *mm-arg-ref*. Exponents of other person/number combinations differ segmentally between verb-bound and mobile morphs, as illustrated in (41) above for the first plural mobile morph *-man*.

4 Final remarks

At first glance, the morphological argument referencing system of Soranî appears to resemble the second position (endo-)clitic system of Pashto and this is indeed how it has been analyzed in previous HPSG work. We showed in this paper that Soranî's

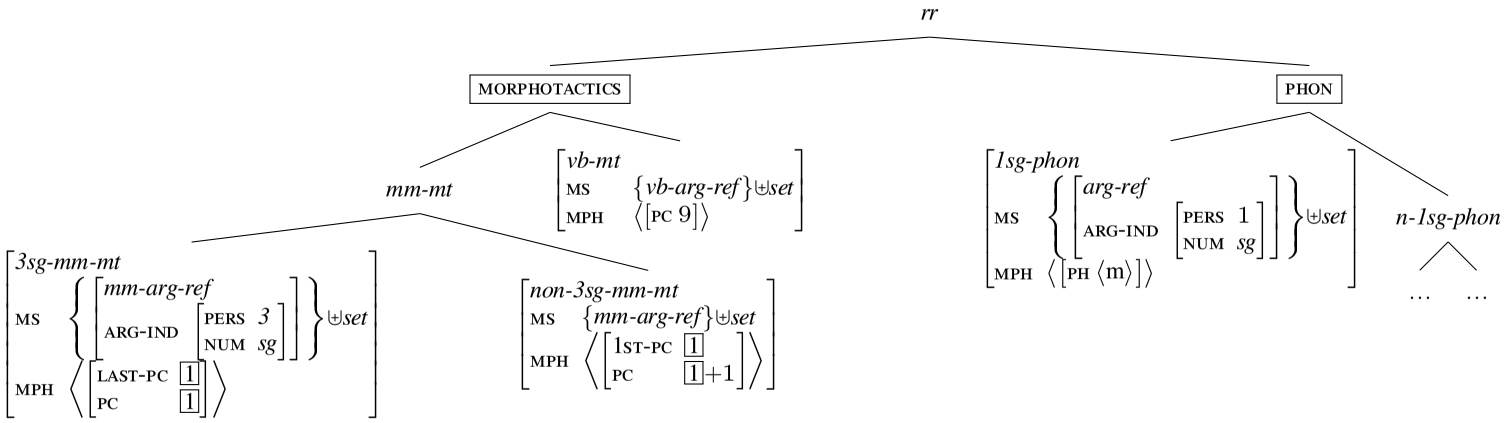


Figure 1: A partial hierarchy of argument referencing realizational rules for Soranî argument referencing

verb-external argument referencing morphs are not second position clitics. Rather, they function as edge affixes on the least oblique NP complement. Mobile morphs are, we claim, best modeled as second position inflectional affixes (leaving aside the case of the third person singular morph), using the relative morph placement analysis proposed by Crysmann (2021: 983), as well as the approach to edge inflection outlined in Miller & Halpern (1993) and Crysmann (2010).

More broadly, Soranî morphological argument referencing demonstrates a unique dissociation between syntax and morphology, which has implications for the architecture of grammars. Specifically, both subject and object argument referencing morphs can be either *verb-bound* or *mobile*, and both *verb-bound* and *mobile* morphs can correspond to either agreement markers or so-called incorporated pronouns. Such dissociations between syntactic status and morphological status are difficult to reconcile with theories that assume an isomorphism between morphological linear order and constituency/grammatical function, such as the mirror principle proposed by Baker (1985), or the view that morphological expounding feeds off functional syntactic terminals as proposed by Embick (2015) (but see Akkus et al. (2023) for an analysis of Soranî argument referencing within an approach of the kind advocated in Embick (2015) that may overcome the difficulties the Soranî data presents). In frameworks that maintain a distinction between inflectional features (members of the MS set) and syntactic information (order on the ARG-ST list or presence on both the ARG-ST list and valence lists), such as HPSG, such dissociations are easily modeled.

References

Akkus, Faruk, David Embick & Mohammed Salih. 2023. Case and the syntax of argument indexation an analysis of Soranî Kurdish. University of Massachusetts Amherst and University of Pennsylvania.

Baker, Mark. 1985. The mirror principle and morphosyntactic explanation. *Linguistic Inquiry* 16(3). 373–415.

Bonami, Olivier & Berthold Crysmann. 2013. Morphotactics in an information-based model of realisational morphology. In Stefan Müller (ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar, Freie Universität Berlin*, 27–47. Stanford, CA: CSLI Publications.

Bonami, Olivier & Pollet Samvelian. 2008. *Soranî Kurdish person markers and the typology of agreement*. Talk presented at the 13th International Morphology Meeting, Vienna.

Bresnan, Joan & Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chichewa. *Language* 63(4). 741–782.

Crysmann, Berthold. 2010. Discontinuous negation in Hausa. In Stefan Müller (ed.), *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar, Université Paris Diderot, Paris 7, France*, 269–287. Stanford, CA: CSLI Publications.

Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics and Information-based Morphology. *Journal of Linguistics* 52(2). 311–374.

Crysmann, Bertholt. 2021. Morphology. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head Driven Phrase Structure Grammar*, 1st edn., chap. 21, 219–244. Berlin: Language Science Press. DOI: 10.5281/zenodo.5543318.

Dost, Ascander. 2007. A domain-based approach to second position clitics in Pashto. In Frederick Hoyt, Nikki Seifert, Alexandra Teodorescu & Jessica White (eds.), *Texas Linguistic Society IX: the morphosyntax of underrepresented languages*, 89–110. Stanford, CA: CSLI Publications.

Embick, David. 2015. *The morpheme: a theoretical introduction* (Interface Explorations 31). Berlin: de Gruyter Mouton.

Halpern, Aaron. 1995. *On the placement and morphology of clitics* (Dissertation in Linguistics). Stanford, CA: CSLI Publications.

Haspelmath, Martin. 2013. Argument indexing: a conceptual framework for the syntactic status of bound forms. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: studies in memory of Anna Siewierska*, 197–227. Berlin: De Gruyter Mouton.

Karim, Shuan & Ali Salehi. 2022. An applicative analysis of soranî "absolute prepositions". In Sara Pacchiarotti & Fernando Zuniga (eds.), *Applicative Morhphology: Neglected Syntactic and Non-syntactic Functions*, 1st edn., chap. 10, 263–298. Berlin, Boston: De Gruyter Mouton. DOI: 10.1515/9783110777949.

Koenig, Jean-Pierre. 1999. *Lexical relations* (Stanford Monographs in Linguistics). Stanford: CSLI Publications.

Koenig, Jean-Pierre & Dan Jurafsky. 1994. Type underspecification and on-line type construction in the lexicon. In.

Levin, Beth & Malka Rappaport Hovav. 2005. *Argument realization* (Research Surveys in Linguistics). Cambridge, UK: Cambridge University Press.

MacKenzie, D N. 1960. *The Origins of Kurdish*. Oxford University Press. 2–16.

Mackenzie, David N. 1961. *Kurdish Dialect Studies*. Oxford: Oxford University Press.

McCarus, Ernest N. 2009. Kurdish. In Gernot Windfuhr (ed.), *The iranian languages*, 1st edn., chap. 10, 587–633. New York: Routledge.

Miller, Philip & Aaron Halpern. 1993. English possessives and the syntax of morphological features. In *Proceedings of the 3rd meeting of the formal linguistics society of Mid-America*, 219–234. Evanston, Northwestern University.

Miller, Philip & Ivan Sag. 1997. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory* 15(3). 573–639.

Mohammadirad, Masoud. 2020. *Pronominal clitics in Western Iranian languages: Description, mapping, and typological implications*. Paris: Université Sorbonne Nouvelle Paris 3. (Doctoral dissertation).

Samvelian, P. 2007. What Sorani Kurdish absolute prepositions tell us about cliticization. In *Texas linguistic society ix: the morphosyntax of underrepresented languages*, 263–283.

Thackston, Wheeler M. 2004. *Sorani Kurdish: a reference grammar with selective readings*. Cambridge, MA: Harvard University.

Tseng, Jesse L. 2003. Edge features and French liaison. In Jong-Bok Kim & Stephen Wechsler (eds.), *On-line proceedings of HPSG 2002*, 313–333. Stanford, CA: CSLI Publications.

Walter, Géraldine. 2012. Fitting into morphological structure: accounting for Sorani Kurdish endoclitics. In Angela Ralli, Geert Booij, Sergio Scalise & Athanasios Karasimos (eds.), *On-line proceedings of the eighth Mediterranean morphology meeting*, 299–321.

Generation of MRS Abstract Predicates from Paninian USR

Sukhada Sukhada
IIT (BHU) Varanasi

Sirisipalli Veera Hymavathi
IIT (BHU), Varanasi

Soma Paul
IIIT Hyderabad

Proceedings of the 30th International Conference on
Head-Driven Phrase Structure Grammar

University of Massachusetts Amherst

Stefan Müller, Elodie Winckel (Editors)

2023

Frankfurt/Main: University Library

pages 121–141

Keywords: Universal Semantic Representation, Minimal Recursion Semantics, Abstract Predicates, Transfer Grammar, Indian Grammatical Tradition, Panini, vivakṣā, and kāraka.

Sukhada, Sukhada, Sirisipalli Veera Hymavathi & Soma Paul. 2023. Generation of MRS Abstract Predicates from Paninian USR. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 121–141. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2023.7.

Abstract

Semantic Representations become useful resources for various multilingual NLP applications such as Machine Translation, Multilingual Generation, cross Lingual QA, to name a few. No Semantic Representation, to our knowledge, adopts *vivakṣā* (*Speaker's intention)* as a guiding principle for the representation. This motivates us to develop a new Semantic Representation system – **Universal Semantic Representation (USR)** – following Indian Grammatical Tradition (IGT) and Paninian grammar. Since USR is designed to be language-independent, we have currently taken up the task of generating English, Hindi, Tamil and Bangla from the USR. For English generation, the USR is mapped to ERG meaning representation (Flickinger, D. 1999) which is couched in Minimal Recursion Semantics (MRS). We use an off-the-shelf ACE generator that uses ERG as a resource-grammar for generating English. While designing the transfer module from USR to ERG-based MRS, we came across various Abstract Predicates (APs) in MRS representation as described in ErgSemantics_Basic (Flickinger et al., 2014). These APs are used to represent the semantic contribution of grammatical constructions or more specialized lexical entries such as compounding or the comparative use of more and so on. This paper presents the strategy for postulating the APs from the information given in USR and then reports the implementation of the transfer module keeping the focus on the postulation of APs. We get around 95% accuracy in postulating APs from USRs.

1. Introduction

One major advantage of Semantic Representations (SemRep) is the potential cross-linguistic *universality* (Abend and Rapparport, 2017) that these SemReps can ideally represent. Languages differ in terms of their form but they have often been assumed to be much closer in terms of their semantic content (Bar-Hillel, 1960; Fodor, 1975) and SemRep can capture that content. Thus Semantic Representations become useful resources for various multilingual NLP applications such as Machine Translation (MT) (Hajič 2002), Multilingual Generation (Cabezudo et al., 2019), cross Lingual QA, to name a few.

Generally, all SemReps abstract away from grammatical and syntactic idiosyncrasies inherent in natural languages (Boguslavsky et al., 2021). As is evident in Semantic Role Labeling (Gildea and Jurafsky, 2002), FrameNet (Baker et al., 1998), Propbank (Kingsbury and Palmer, 2002), Abstract Meaning Representation (Banarescu et al., 2013), the fundamental component of the content conveyed by SemReps of texts is argument structure – who did what to whom, where, when and why, i.e., events, their participants and the

relations between them (Abend and Rapparport, 2017). However, in communication, speakers express through an utterance how (s)he views the situation which the mere argument structure of events can never capture. Thus what is expressed in communication is vivakṣā: the intention of the speaker about the meaning to be conveyed by the words. IGT views discourse composition as the manifestation of the speaker's vivakṣā. Example (1.a) and (1.b) explain how vivakṣā determines the syntactic expressions:

(1) a. **umā ko** kala rāta cāṃda <u>dikhā</u>
umā k4a yesterday-r6 night-k7t moon-k1 see(intr)-past
'umā happened to see the moon yesterday night'

b. ***umā ne*** *kala rāta cāṃda* <u>*dekhā*</u>
umā k1 yesterday-r6 night-k7t moon-k2 see(tr)-past
'*umā* saw the moon yesterday night'

The activity of 'seeing' licenses an animate *seer* and a *seen entity*. That is the *semantic frame* for the verb. However, in communication, it is not the semantic frame of a chosen event alone that depicts the situation. Two other important factors also play a major role: (a) how the speaker conceptually cognizes the situation? (b) which linguistic expressions are available to translate that cognition into languages? For example, in (1), does the speaker want to express Uma's agency or does (s)he want to foreground the moon's appearance over the seer's agency? This is termed the speaker's *vivakṣā*. Depending on that, the speaker would choose the most appropriate linguistic expressions to convey his/her thoughts. For example, the speaker views the situation very differently when (s)he says (1.a) '*Uma happened to see the moon yesterday night'* vis-à-vis (1.b) *'Uma saw the moon yesterday night'*. In Hindi, two different verb roots are used and the post-position on the seer also indicates different kāraka relations. In (1.a), Uma is an experiencer, while in (1.b), the volitionality of Uma is maintained.

To our knowledge, no SemRep adopts *Speaker's intention* as a guiding principle for the representation. This motivates us to develop a new Semantic Representation system – **USR** – following IGT (Sukhada et al., 2023) and Paninian grammar (Zdeněk Žabokrtský et al., 2020). The application task chosen is Multilingual Natural Language Generation. Since USR is designed to be language-independent, we have currently taken up the task of generating English, Hindi, Tamil and Bangla from the USR.

For English generation, the USR is mapped to ERG meaning representation (Flickinger, D. 1999) which is couched in MRS (Copestake et al., 2005). We use an off-the-shelf ACE generator that uses ERG as a resource-grammar for generating English. Since both USR and MRS are semantics-based representations, we assume that the USR-MRS transfer would be straightforward. While designing the transfer module from USR to ERG-based MRS, we came across various APs in MRS representation as described in ErgSemantics_Basic (Flickinger et al., 2014). These APs are used to represent

the semantic contribution of grammatical constructions or more specialized lexical entries such as compounding or the comparative use of more and so on.

This paper presents the strategy for postulating the APs from the information given in USR and then reports the implementation of the transfer module keeping the focus on the postulation of APs.

Section 2 introduces the new SemRep USR briefly. Section 3 presents the similarities and differences between USR and MRS to motivate the significance of writing a transfer grammar. Section 4 describes the APs as postulated in ERG meaning representation. Section 5 discusses in detail the implementation of Transfer Grammar for APs in CLIPS. The experiment, results and error analysis for the task *generation of APs* is reported in Section 6. Finally Section 7 concludes the paper.

2. A Brief Introduction to USR

USR attempts to design a structured representation for the speaker's vivakṣā. IGT views language as a holistic phenomenon (Sukhada et al., 2023). Words are not derived as isolated units in Paninian grammar, but as units that are semantically connected with other words in the sentence (Raster, 2015). Sentences are connected across the discourse. This is explicitly recognized by the Paninian rule (A 2.1.1): *samarthaḥ padavidhiḥ*. Keeping in mind Natural Language Generation as the targeted application, the lexico-semantic and relational information is specified in the USR at various layers so that proper word forms, relations among words and finally relations across sentences can be generated systematically.

USR is a csv-formatted multilayered information packaging system that encapsulates (a) lexico-conceptual, (b) syntactico-semantic relational and (c) discourse level information (Garg et al., 2023). The uniqueness of this representation is that information on each layer is distinctly yet interactively maintained through attribute value matrix and co-referencing as shown in sentence (2). The USR for the semantics of sentence (2) is given in Table 1:

(2) hari ne apane guru jī ko garama dūdha aura miṭhāī dekara
hari erg his teacher respect dat hot milk and sweet offering
ābhāra vyakta kiyā
gratitude express do.pst
'hari expressed gratitude to his respected teacher by offering hot milk and sweet'

Astract layer	Con(cept)	hari	apanā /self	guru_1/ teacher_5	garama_1/hot_1	dūdha_1/milk_1	miṭhāī_1/sweet_1	de_17/ of-fer_1	ābhāra_1/ grati-tude_1	vyakta+kara_1/ ex-press_1-yA_1/past_1
Lexico-Con-ceptual	Sem(antic)_Cat(egory)	per-son male		anim						
	Morpho-Sem(antic)						plural			
	Speaker-view			re-spect						
	Index	1	2	3	4	5	6	7	8	9
Syntactico-Se-mantic	Dep(endency)	9:k1 (doer/agent)	3:r6 (genitive)	7:k4 (recipient)	5:mod	7:k2 (object)	7:k2	9:rpk (sequence of events)	9:k2 (theme)	0:main
	Con-struction	conjunction:[5,6]								
Dis-course	Coreference & Con-nective		1:coref							
	Sent Type	affirmative								

Table 1: USR for sentence (2)

In Table 1, the **Concept row** represents *unique* concepts (not words) that refer to entities, events, quality, quantity and other properties of an entity or event. For sentence (2), the concepts are *Hari, apanā/self*, *guru_1/teacher_5*,

garama_1/hot_1, dūdha_1/milk_1, de_9/offer_1, miṭhāī_1/sweet_1, ābhāra_1 /gratitude_1, vyakta+kara_1/express_1. TAM (Tense-aspect-modality) is also treated as a concept and therefore the TAM 'past' has been specified on the main verb. We have not considered *aura(Conjunction and), jī(respect marker)* as concepts in the concept row. The speaker intends to show *respect* to *his* teacher. Therefore the information *'respect'* has been specified in the **Speaker's view row** under the concept *guru_1/teacher_5*. During generation, the corresponding word *jī* in Hindi, *bābu* in Bangla, *gāru* in Telugu and the modifier 'respected' in English will be generated. Currently, the Semantic category row contains named entity information (C S & Lalitha Devi, WILDRE 2020), animacy and inherent gender information for the concepts. The Morpho-semantic row captures semantic information such as number, comparison and causation which can be marked in languages morphologically. Relations among concepts are specified in terms of dependency relation in the Dependency row. The Construction row conveys non-dependency relational information. For example, in a conjoined construction, all entities involved enjoy equal status. The discourse level information such as inter-sentential connectivity, co-referencing are represented in the Discourse row. Finally, the Sentence type is also specified.

3. Motivation for Transfer Grammar Module

As stated earlier, the main application task planned for USRs is multilingual generation. Since for English generation, the open-source ACE generator is available and the input it takes is a kind of semantic representation in MRS format, we examined if a transfer grammar module can be developed for converting USR to MRS. The advantage is that we would not be required to develop an English generator from scratch. Moreover, apart from English, already large-scale MRS-based grammar is available for a few other languages such as German, Japanese, and Korean. Thus in the future, those languages can also be generated from USRs via MRS. This section describes the similarities and differences found during examining the USR and MRS representations.

3.1 Similarities between USR and MRS representation

The motivating factor for writing a transfer module from USR to MRS is that there are many similarities between the two representations such as the following:

- Both USR and MRS are semantic representations that abstract way the syntactic idiosyncrasies of languages
- The finite verb is the head/root of the representation
- TAM (tense-aspect-modality) is represented as features on the verb
- GNP (gender-number-person) information are attested on the nouns.

- Adjectives are treated as a stand-alone concept even where they are derived from nouns or verbs.
- Verb-argument structure is specified in the representation
- No canonical representation. For example, active voice and passive voice sentences are represented as different semantic representations.

Thus principally, the conversion between USR and MRS can be direct. However, there are differences observed between the two representations, especially in terms of the postulation of APs in MRS, that necessitate a constraint-based transfer grammar module. The abstract predicate mapping that is the focus of the paper highlights the dissimilarity between the two frameworks as shown in Section 4. However this paper discusses differences related to APs alone.

4. Abstract Predicates in MRS

The predicate symbols in ErgSemantics have been divided into two classes: *surface* predicates and *AP*s. In non-lexical contexts, APs come into play, whether to represent ordinals such as "*first*" with "/*ord/*" or to denote negative constructions using "/*neg/*." ERG has around 108 APs. They can be classified into the following broad categories:

i. Quantifier
ii. Abstraction
    a. Degrees of Comparison
    b. Pronoun
    c. Named Entity
    d. Time and Place
    e. Question
    f. Number
iii. Construction
iv. Other

Table 2 shows the list of the APs handled so far:

Quantifier					
/def_explicit_q/		/udef_q/	/proper_q/		/every_q/
/def_implicit_q/		/which_q/	/pronoun_q/		
Abstraction					
Degrees of Comparison	**Pronoun**	**Named Entity**	**Time & Place**	**Question**	**Number**
/comp/	/pron/	/named/	/loc_nonsp/	/measure/	/card/
/comp_equal/		/dofw/	/place_n/	/thing/	/ord/
/comp_less/		/mofy/	/time_n/	/reason/	
/superl/		/yofc/		/property/	

/abstr_deg/		/season/		/manner/	
				/person/	
Construction					
Compounds	**Passive**	**Non-Finite**	**Negation**	**Possession**	**Reciprocal Pronoun**
/compound/	/parg_d/	/subbord/ /nominalization/	/neg/	/poss/	/recip_pro/
Others					
/unspec_manner/					

Table 2: List of the APs handled so far

This paper attempts to identify where and how information encoded in USR enables to postulate the aforementioned APs. In most of the cases, semantic information encoded in USR is used to determine the APs while there are few cases where we are currently using mainly entries of Concept row to postulate APs. Table 3 to Table 6 specify which information from USR is being used to predict the right AP of different categories.

i. Quantification

As described in the ErgSemantics_Basic document, the ERG assumes that all instance variables (of type *x*) are bound by a generalized quantifier. Such an assumption is not taken in USR. Table 3 indicates the information that we are using from USR to postulate *_q APs. In column 3 of Table 3 to Table 6, the convention ('<' is used for *binds*, '|' for *when*, ':' *in* * row) is used.

MRS quantifier	**Context**	**Rules from USR information**	**Example**
/def_explicit_q/ (1)	Possessive nouns & pronouns	1 < noun, | *r6* for **i** : Dep row	Ram**'s** book
/def_implicit_q/ (2)	Spatial & temporal adverbs	2 < here & /place/ | *yahāṃ* : Con row 2 < there & /place/ | *vahāṃ* : Con row 2 < now & /time/ | *aba* : Con row 2 < today & /time/ | *āja* : Con row 2 < tomorrow & /time/ | *kala* : Con row *2* < /poss/, /person/ & /which_q/ | *kim*, *:* Con row + *r6* for **i** : Dep row	The boy lives **here.** I am going **there.** He will come **now.** The meeting is **today.** She will catch the train **tomorrow.** **Whose**, house is this?
/every_q/ (3)	Universal	3 < /person/ | *saba* : Con	Rama calls **eve-**

	quantifier	row	**rybody** in the school..
/proper_q/ (4)	Proper noun	4 < /named/ | *per :* Sem_Cat row 4 < /named/ | *place :* Sem_Cat row 4 < /dofw/ | *dow :* Sem_Cat row 4 < /mofy/ | *mofy:* Sem_Cat row *4* < /yofc/ | *yofc :* Sem_Cat row	**Sanju** is good. **India** is a sub-continent. Babies eat fruits on **Monday.** **January** is the first month. He will come to India in **2024.**
/pronoun_q/ (5)	Personal pro-noun	5 < /pron/ | *speaker :* Con row + +*sg/pl :* Morpho-Sem_row 5 < /pron/ | *addressee :* Con row +*sg/pl :* Morpho-Sem_row 5 < /pron/ | *3rd person/wyax:* Con row + *coref :* dis-course_row	**We** are going to a party. **You** are a good person. **He** is smart.
/which_q/ (6)	Interrogative pronoun	6 < /person/ | *kim*$_i$: Con row + *k1* for **i** : Dep row + anim : Sem_Cat row 6 < /time/ + /loc_nonsp/ | *kim*$_i$: Con row + *k7t* for **i** : Dep row 6 < /place/ + /loc_nonsp/ | *kim*$_i$: Con row +*k7p*/*k2p* for **i** : Dep row 6 < /thing/ | *kim*$_i$: Con row + *k2* for **i** : Dep row - animacy : Sem_Cat row 6 < /reason/ | *kim*$_i$: Con row + rh for **i** : Dep row 6 < /manner/ & /unspec_manner/ | *kim*$_i$: Con row + *krvn*: Dep row	**Who**$_i$ filled the bottle? **When**$_i$ will you come? **Where**$_i$ are you going? **What**$_i$ are you buying? **Why**$_i$ are you sad? **How** did you finish the work?
/abstr_deg/ (7)	Interrogative Degree	7 < /measure/ & /which_q/ | *kim*$_i$: Con row + *degree* re-lation for **i** : Dep row + *in-terrogative* : Sent_Type row	**How**$_i$ **happy** was Sita?

Table 3: MRS quantifiers from the USR information utilized

ii. Abstraction

This category consists of cases where MRS representation goes one level more abstract than the surface predicates to capture certain generalization in the representation, example */comper_equal/* for the similarity between two different entities. Similarly APs for the named entities, adverbs of time, and numerals as well as the information specified in the USR for these predicates have also been listed in Table 4:

MRS Predicates	**Context**	**Rules from USR information**	**Example**
/comp/ (8)	Comparative degree more	ARG1 of 8 is adj$_i$ & ARG2 is noun$_j$ | *compermore* of **i** : Mor_Sem_row + *rv* for **j** : *Dep row*	Sanju is **more** intelligent$_i$ than Rahul$_j$.
/comp_less/ (9)	Comparative degree less	ARG1 of 9 is adj$_i$ & ARG2 is noun$_j$ | *comperless* of **i** : Mor_Sem_row + *rv* for **j** : *Dep row*	Mohan is **less** intelligent$_i$ than Rama$_j$.
/comp_equal/ (10)	Similarity	ARG1 of 10 is adj$_i$ & ARG2 is noun$_j$ | *ru* relation for **j** : Dep row	Sita is **as beautiful**$_i$ **as** Radha$_j$.
/superl/ (11)	Superlative degree	ARG1 of 11 is adj$_i$ | *superl* of **i** : Morpho_Sem_row + *k1s* for **j** : Dep row	The sun is the **biggest**$_i$ star$_j$.
/pron/ (12)	Personal pronouns	12 | *speaker:* Con row + *sg/pl:* Mor-Sem_row 12 | *addressee:* Con row +*sg/pl:* Mor-Sem_row 12 | *wyax:* Con row + *coref :* discourse_row	**I** bought a diary. **You** are smart. **They** will go to Banaras.
/named/ (13)	Proper noun	13 | *per :* Sem_Cat row 13 | *place :* Sem_Cat row	**Rama** ate an apple. **Rama** lives in Ayodhya.
/dofw/ (14)	Name of the days of week	14 | *dow :* Sem_Cat row	**Sunday** is a holiday.
/mofy/ (15)	Name of the months of year	15 | *mofy:* Sem_Cat row	**December** is the final month of the year.

/yofc/ (16)	Year of centuries	16 | *yofc:* Sem_Cat row	What will happen in **2025**?
/season/ (17)	Name of the seasons	17 | *season* : Con row	Christmas is celebrated in **winter**.
/loc_nonsp/ (18)	Spatial & temporal entities	18 | *yahāṃ* : Con row 18 | *vahāṃ* : Con row 18 | *aba* : Con row 18 | *āja* : Con row 18 | *kala* : Con row	He lives **here.** I will be **there** in five minutes. The teacher will teach **now.** He is happy **today.** **Tomorrow** is a holiday.
/place_n/ (19)	Spatial entities	19 | kim_i : Con row + *k7p*/*k2p* for i : Dep row 19 | *yahāṃ* : Con row 19 | *vahāṃ* : Con row 19 | kim_i : Con row + *k5* for i : Dep row	**Where**$_i$ do you live? Kids are **here**. Your bicycle is **there.** **Where**$_i$ did you come from?
/time_n/ (20)	temporal adverbs	20 | *aba* : Con row 20 | *āja* : Con row 20 | *kala* : Con row 20 | *kim* : Con row + *k7t* for **i** : Dep row	She is reading the book **now.** He plays the guitar **today.** We will buy groceries **tomorrow.** **When**$_i$ are you leaving?
/measure/ (21)	Abstract Measuring	ARG1 of 21 is adj_i & ARG2 is which_q | kim_j : Con row + *degree* relation for **j** : Dep row + *interrogative* : Sent_Type row	**How**$_j$ sad_i was Sita?
/thing/ (22)	Wh - word "*What"*	22 | kim_i : Con row+ *k2* of **i** : Dep row	**What** are you doing?

/property/ (23)	How are you?	23 | *kim*$_i$: Con row + *k1s for* **i** : Dep row + *interrogative* : Sent_Type row + *animacy* : Sem_Cat row	**How**$_i$ **are you?**
/person/ (24)	wh_words with animacy	24 | *kim*$_i$: Con row + *k1* for **i** : Dep row + *animacy* : Sem_Cat row 24 | *kim*$_i$: Con row + *k1s* for **i** : Dep row + *animacy* : Sem_Cat row 24 | *kim*$_i$: Con row + *k2/k2g/k4 for* **i** : Dep row + *animacy* : Sem_Cat row 24 | *kim*$_i$: Con row + *r6* for **i** : Dep row 24 | *kim*$_i$: Con row + *k5* for **i** : Dep row + *animacy* : Sem_Cat row	**Who**$_i$ finished the work? **Who**$_i$ is Rama? **Whom**$_i$ did Rama meet? **Whose**$_i$ car is that? **Who**$_i$ is Mohan afraid of?
/reason/ (25)	Why word	25 | *kim*$_i$: Con row + *rh* for **i** : Dep row + *interrogative* : Sent_Type row	**Why**$_i$ are you crying?
/manner/ (26)	Interrogative Manner	26 | *kim*$_i$: Con row + *krvn* for **i** : Dep row + *interrogative* : Sent_Type row	**How**$_i$ did you come?

Table 4**:** Generic APs with Examples

iii. Construction

This category includes what we commonly call as construction, form-meaning pairs. For different constructions different kinds of information from USR is being utilized for the mapping.

MRS predicates	**Context**	**Rules from USR information**	**Example**
/compound/ (27)	Compound words & English Honor-	27 | noun$_j$+noun$_i$: Con row	He laid the **foundation**$_j$

	ific words	27 | noun$_i$: Con row; *respect* : Speaker_View row of noun$_i$	**stone**$_i$ for Rama's office. **Ms$_j$. Sita**$_i$ joined the course.
/parg_d/ (28)	(i) Passive sentences & (ii) rbks	(i) 28 | *Passive TAM* : Con row + *passive* : Sent_Type row (ii) 28 | verb$_i$: Con row + rbks for **i** : Dep row + *affirmative* : Sent_Type row	Ravana$_j$ was **killed**$_i$ **by** Rama. The fruit **eaten**$_i$ **by** Rama$_j$ was sweet.
/subord/ (29)	Subordinate clauses	(i) 29 | verb$_i$: Con row + *rpk* for **i**: Dep row (ii) 29 | verb$_i$: Con row + *krvn* for **i** : Dep row	Having been **going**$_i$ to the school, Rama **ate** food. Mohan walks **limping**$_i$ **.**
/nominalization/ (30)	Nominalized	30 | verb$_i$: Con row + k1 of : Dep row	**Chasing**$_i$ the cat is old.
/neg/ (31)	Negation	31 | *neg* : Dep row	I am **not** going$_i$ to the function.
/poss/ (32)	possession	32 | *kim*$_i$: Con row + *r6* of i : Dep row 32 | *noun*$_i$: Con row + *r6* of i : Dep row	**Whose** idea is this? I borrowed **Rama's cycle**.
/recip_pro/ (33)	Reciprocal Pronouns	33 | *eka* + *dūsarā*$_i$: Con row	Rama and Sita like each other$_i$**.**

Table 5: APs under construction category

iv. Other

This category includes the APs which are unique for interrogative pronouns. For different types of interrogative pronouns different kinds of information from the USRs is utilized for the mapping.For instance, when the USR has *kim* in the Concept row, *krvn* relation in Dep row and *interrogative* in

Sent_Type then the AP /*unspec_manner/* will be postulated in the transfer module as shown in Table 6.

MRS quantifier	**Context**	**Rules**	**Example**
unspec_manner (34)	Interrogative Manner	34 | *kim*$_i$: Con row + *krvn* of **i** : Dep row + *interroga-tive* : Sent_type	**How**$_i$ **did you complete your work?**

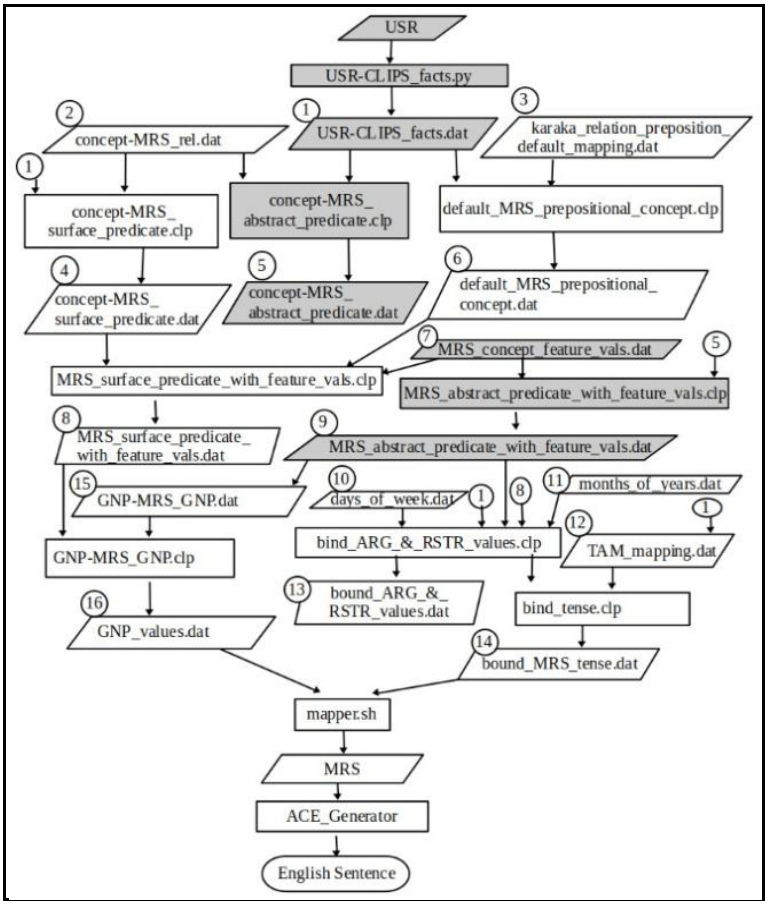
Table 6**:** Other APs

5. Implementation of Transfer Grammar for Abstract Predicates in CLIPS

The implementation is done at two levels: (a) Determining an AP (b) Specifying the feature structure description of the AP. For (a), information from USR has been utilized as discussed in the previous section (see Table 3 - Table 6). Once APs are identified for a given USR, we translate the sentence into English and run the ACE parser to find out the feature structure description of the targeted AP and add the AP along with its feature structure description in the dictionary if it is not already present there. This is our development stage for populating APs into the dictionary with appropriate feature structure description. Thus the lexicon for APs is created.

During English sentence generation from USR via MRS, this dictionary is consulted for framing the appropriate MRS for a given USR which in turn is used by ACE generator as an input and the English sentence is generated. The postulation of APs from USR is executed in CLIPS (Giarratano, J. C. 1993). The part with gray background of the flow chart describes the postulation of APs.

The USR_to_CLIPS_facts.py program converts a USR into CLIPS facts: CLIPS_ facts.dat (1). The concept-MRS_abstract_predicates.clp implements the rules (Table 3 - Table 6) and postulates AP types in concept-MRS_abstract_predicate.dat (5). Finally, the program MRS_abstract_predicate_with_-feature_vals.clp takes two files as input, one dictionary MRS_concept_ feature_vals.dat (7) and the output file (5), and returns MRS_abstract_ predicate_with_feature_vals.dat(9) that contains all APs with their feature structure description.



Flowchart 1: USR to MRS transfer module (APs in gray background)

We will explain Flowchart 1 with an example (3):

(3) īśā ji ne apane beṭe aura apanī beṭī ko somavāra ko
Isha respect k1 her son-k2 and her daughter k2 Monday k7t
kāśī ke sabase baḍe vidyālaya meṃ bharatī kiyā.
Kashi-place r6 most large school k7p admit do-past
'Ms. Isha admitted her son and her daughter, on Monday, in Kashi's largest school.'

#īśā ji ne apane beṭe aura apanī beṭī ko somavāra ko kāśī ke sabase baḍe vidyālaya meṃ bharatī kiyā.										
Concept	īśā	apanā/her	beṭā_1/son_1	apanā/her	beṭī_1/daughter_1	somavāra/Monday	kāśī	baḍā_2/big_1	vidyālaya_1/school_1	bharatī + kara_6-yā_1/admit_1-past

Index	1	2	3	4	5	6	7	8	9	10
Sem_ Cat	per fe ma le		anim male		anim female	dow	pl ac e			
Morp h_Sem								superl		
Dep	10: k1	3:r6	10:k2	5:r6	10:k2	10:k7t	9: r6	9:mod	10:k7p	0:main
Dis-course		1:co ref		1:co ref						
Speak er-view	re-spe ct									
Con-structi on	conjunction:[3,5]									
Sent_ Type	affirmative									

Table 7: USR for the sentence (3)

The rules for generating MRS APs from the information specified in USR are listed in Table 8 for the concepts given in the first column. For example, two APs occur for any named entity in MRS. For īśā, USR specifies per: yes (i.e., īśā is a person) which generates the AP /named/ which is bound by the quantifier /proper_q/. The 'yes' value for per (in Sem_Cat row), female (in Morpho Sem row) and respect (in speaker's view row generate two APs /compound/ and the quantifier /udef_q/ which in turn generates the lexical item "Ms." as a compound "Ms. Isha". Similarly we can examine the USR information utilized for generating APs in the context of other concepts in Table 8.

Concept label	**Eng. Eqiv.**	**Information specified in USR**	**Generating AP/s**	**Purpose**
īśā(1)	*īśā*	*per:yes*	/named/, /proper_q/,	*īśā*
		per:yes, respect:yes, feminine:yes	/compound/, /udef_q/,	Ms.
apanā(2)	*she*	*coreference to* 1, *feminine:yes*	/pronoun_q/,/pron/	her
		genitive to 3	/poss/, /def_explicit_q/	

		conj: [3,5]	/udef_q/	son and daughter
beṭā_1(3)	*son*			
apanā(4)	*she*	*coreference* to 1, *feminine:yes*	/pronoun_q/, /pron/	her
		genitive to 5	/poss/, /def_explicit_q/	
beṭī_1(5)	*daughter*			
somavāra(6)	*Monday*	*dow:yes*	/dofw/, /proper_q/	Monday
kāśī(7)	*kāśī*	*place:yes*	/proper_q/, /named/	*kāśī*
		genitive relation to noun (here *school*)	/def_explicit_q/, /poss/	*kāśī*'s
baḍā_2(8)	large	superl to 8:yes	/superl/	largest

Table 8**:** FlowChart 1 explained using example (3)

Surface predicates are handled separately using another CLIPS program. We assign feature values to surface predicates from the MRS_concept_feature_vals.dat(7) dictionary. After incorporating GNP values from the GNP dictionary, ARGument sharing will be done. Following this, binding of handle constraints LBL and RSTR values is done. Subsequently, mapper.sh will run for obtaining the complete MRS representation of a USR. The MRS representation then becomes input to the ACE generator for producing natural English sentences.

5.1 Statistical Observation on Transfer Rules

The implementation of rules for creating the APs include three types of mapping:

1. Direct Mapping: A relation or a lexical concept from USR is directly mapped to MRS AP;
2. Indirect Mapping: Information encoded at multiple layers in USR is used to postulate the AP;
3. Constraint based mapping: Where the rule includes constraints to prevent wrong or overgeneration of APs.

Examples for each type of mapping are given in Table 9 to Table 11.

USR information utilized	**Context**	**Example**	**MRS AP**
Morpho-semantic row	Superlative degree	*superl*	/superl/
Dependency relation row	Sequence of events	*rpk*	/subord/

Table 9: Examples of Direct Mapping

USR info utilized	**Context**	**Ex.**	**MRS AP**
Concept & de-	Interrogative pro-	*where*	/which_q/, /loc_nonsp/, /place_n/

pendency	noun		
Concept & dependency	Implicit quantifier	*whose*	/def_implicit_q/, /poss/, /person/, /which_q/

Table 10: Examples of Indirect Mapping

In Constraint-based mapping, we take into account the mapping rules that apply some constraints for generating an AP. For example, the generation of /person/ and /thing/ use the same information, *kim* in Con row and *k2* relation in Dep row. The distinguishing factor that works is the *animacy* feature in Sem_Cat row. The presence of an *animacy* feature triggers the postulation of /person/, that along with /which_q/ generates '*who'* in English sentences. On the other hand, the absence of an *animacy* feature in Sem_Cat row postulates /thing/ which along with /which_q/ generate the English word '*What'.*

USR information utilized	**Context**	**Example**	**MRS AP**
Concept row	*kim*	*vahām̐*	/place_n/, /def_implicit_q/, /loc_nonsp/
Semantic Category row	*per/place/org/ne*	*Sanju*	/proper_q/, /named/

Table 11: Examples of Constraint-based Mapping

Rule	**Number of Rules**	**Percentage**
Direct mapping	10	27.78%
Indirect mapping	15	41.67%
Constraint based mapping	11	30.55%

Table 12: Statistical observation on transfer rules in CLIPS

We observe that rules written for Indirect mapping are the highest in number. Thus, we conclude that information used for postulating MRS APs is distributed at different layers of the USRs.

6. Experimental Setup, Result and Error Analysis

Preparing a test suite for APs is challenging. Although experienced linguists have been given the task, they do not have an idea of APs which is a framework internal feature of MRS. To address this issue, we have prepared a short guideline for the linguists who created the test suite. For each AP, we created 5 sample sentences with the word under consideration underlined and asked the linguist to create 10 more sentences in which the underlined words must be present. For example, for /card/ one of the 5 sample sentences was 'the boy ate five mangoes'. For some cases where instruction statements can easily be prepared, we have given the instruction statement along with 5 sample sentences. For example, for person names, we have given one sample sentence, 'Hari came home', and also the following instruction: "The sentence should

have a person name". Following the guidelines, our linguist team has created 262 test suite sentences. USR annotators were then asked to create USRs for these sentences. After obtaining the USRs, we proceeded to execute them using the transfer module and ACE Generator. Following this, we meticulously examined whether the anticipated APs were accurately generated. The Results and Error analysis has been given in Table 13.

Total USRs	**Total expected APs**	**APs generated**	**Error Analysis**	
			Concept missing	**Typographical error**
262	491	469	16	6
		95.5%	3.25%	1.22%

Table 13: Result and Error Analysis for APs

The result shows quite a promising conversion rate from USR to MRS as far as APs are concerned. The errors occur mainly due to wrong USR input as indicated in Table 13.

Table 14 shows results for each class of APs.

	Quantifiers	**Abstraction**	**Construction**	**Others**	**Total APs**
	110	281	70	30	**491**
Total errors	10	8	2	2	**22**
Accuracy	90.9%	97.15%	97.14%	93.33%	**95.5%**

Table 14 Results for each class of APs

7. Conclusion

This paper presents an architecture and implementation of converting the semantic representation USR to another semantic representation MRS to generate Natural language English using the open-source ACE generator. The focus of the paper has been on postulating APs which is a theory internal construct of MRS. USR is based on Indian Grammatical Tradition and Panini, while MRS is rooted in HPSG. It is interesting to note that USR does capture almost complete information that APs tend to represent. That is why we get 95% accuracy in postulating APs from USR. The only exception is the Quantifier APs of MRS. USR does not work with the assumption that every noun or noun phrase will have to be bound by a quantifier. Nevertheless, we were successful in generating all *_q predicates. The result of the work is surely motivating enough to develop a full-fledged transfer grammar module from USR to MRS for English and other languages as well for which MRS-based grammar exists.

Acknowledgement
We are thankful and acknowledge that this work was carried out as a part of the research project 'Language Communicator Tool for End Users' under the Project titled 'National Language Translation Mission (NLTM), BHASHINI' funded by the Ministry of Electronics and Information Technology (MeitY), Government of India.

References

Abend, O., & Rappoport, A. (2017, July). The state of the art in semantic representation. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 77-89).

Bharati, A., & Kulkarni, A. (2008). Information Coding in a language: Some insights from Paninian Grammar. Dhimahi, Journal of Chinmaya International Foundation Shodha Sansthan I (1), 77–91

Boas, H. C. (2005). Semantic frames as interlingual representations for multilingual lexical databases. International Journal of Lexicography, 18(4), 445-478.

Boguslavsky, I. Semantic analysis supported by inference in a functional model of language [Semanticheskij analiz s oporoj na umozakljuchenija v funktsional'noj modeli jazyka]. Problems of linguistics [Voprosy jazykoznanija], (1-2021), 29-56.

Cabezudo, M. A. S., & Pardo, T. A. S. (2019). Natural language generation: recently learned lessons, directions for semantic representation-based approaches, and the case of Brazilian Portuguese language. In Proceedings.

Copestake, A., Flickinger, D., Malouf, R., Riehemann, S., & Sag, I. (1995). Translation using minimal recursion semantics. In Proceedings of the Sixth Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages.

Copestake, A., Flickinger, D., Pollard, C., & Sag, I. A. (2005). Minimal recursion semantics: An introduction. Research on language and computation, 3, 281-332.

Flickinger, D., Bender, E. M., & Oepen, S. (2014, May). Towards an Encyclopedia of Compositional Semantics: Documenting the Interface of the English Resource Grammar. In LREC (pp. 875-881).

Garg, K., Paul, S., Sukhada, Kumari, R., and Bahawir, F. 2023. Evaluation of Universal Semantic Representation (USR). Proceedings of the 4th International Workshop on Designing Meaning Representations. 2023.

Giarratano, J. C. (1993). CLIPS user's guide. NASA Technical Report, Lyndon B Johnson Center.

Hajic, J. (2002, May). Tectogrammatical representation: towards a minimal transfer in machine translation. In Proceedings of the Sixth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+ 6) (pp. 216-226).

Kim, J. B. (2006). Minimal recursion semantics: An application into Korean. Journal of the Linguistic Association of Korea, 14(2), 59-85.

Kiparsky, P. (1995). Paninian Linguistics. In Concise History of the Language Sciences (pp. 59-65). Pergamon.

Kulkarni, A. (2008). Panini's Ashtadhyayi A Computer Scientist's viewpoint, 4th APCAP NIAS Bangalore

Malarkodi C S and Sobha Lalitha Devi. 2020.A Deeper Study on Features for Named Entity Recognition. In Proceedings of the WILDRE5– 5th Workshop on Indian Language Data: Resources and Evaluation, pages 66–72, Marseille, France. European Language Resources Association (ELRA).

Ranta, A., Angelov, K., Gruzitis, N., & Kolachina, P. (2020). Abstract syntax as interlingua: Scaling up the grammatical framework from controlled languages to robust pipelines. Computational Linguistics, 46(2), 425-486.

Song, L., Zhang, Y., Wang, Z., & Gildea, D. (2018). A graph-to-sequence model for AMR-to-text generation. arXiv preprint arXiv:1805.02473.

Sukhada, Paul, S., and Bhattacharjee, B. 2023. Theoretical Foundation of Universal Semantic Representation. IIIT/TR/2023/11142.

Further evidence for an HPSG-based theory of the semantics of *different* and *the same*

Shûichi Yatabe
University of Tokyo

Proceedings of the 30th International Conference on
Head-Driven Phrase Structure Grammar

University of Massachusetts Amherst

Stefan Müller, Elodie Winckel (Editors)

2023

Frankfurt/Main: University Library

pages 142–162

Keywords: HPSG, Minimal Recursion Semantics, reciprocals, cumulative readings, right-node raising

Yatabe, Shûichi. 2023. Further evidence for an HPSG-based theory of the semantics of *different* and *the same*. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 142–162. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2023.8.

Abstract

Yatabe (2021) presents a theory according to which the meaning of a word like *different* in a sentence like *Anna and Bill like different films* contains the meaning of a reciprocal pronoun. Since the postulated reciprocal meaning inside the meaning of a word like *different* requires the presence of a semantic antecedent, the theory entails that the apparent internal reading of a sentence like *John saw and reviewed different films*, which does not contain a plural DP that could serve as the semantic antecedent of the postulated reciprocal meaning, must be licensed in a way that is entirely different from the way in which the internal reading of a sentence like *Anna and Bill like different films* is licensed. In the present paper, I adduce additional pieces of evidence for this theory. In order to enhance the plausibility of the proposed theory, I also show how the collective interpretation of reciprocals and the interaction of reciprocals and cumulative interpretation can be accounted for within the theory.

1 Introduction

A sentence like (1) has two distinct readings, which are called the external reading and the internal reading respectively in the literature.

(1) Anna and Bill like different films.

In the external reading, the sentence means "Anna and Bill like films that are different from the contextually salient film or films", and in the internal reading, the sentence means "The film or films that Anna likes and the film or films that Bill likes are different from each other". There is a sense in which the internal reading is licensed by the presence of the plural DP *Anna and Bill* here; the internal reading becomes unavailable when the DP is replaced with a singular DP like *Anna*. The DP whose presence licenses an internal reading of a sentence containing a word like *different* in this sense will be referred to as the *antecedent* of that word in what follows.

In Brasoveanu (2011), it is argued, convincingly in my view, that we need to distinguish two types of internal readings. The first type is exemplified by a sentence like (1), in which the antecedent of *different* is a plural noun phrase. The second type of internal reading is exemplified by the sentence, *Every student read a different book*, in which the antecedent of *different* is a singular, distributive quantifier. The same word, *different*, is used in both types of sentences in English, but as argued in Beck (2000) and Brasoveanu (2011), there are languages that use distinct words in these two types of sentences. This paper is primarily about the semantics of words that give rise to the first type of internal reading, that is, the type of internal reading that is licensed by the presence of a plural antecedent.

[†]I thank Steve Wechsler for invaluable discussion and David Beaver and Kyle Johnson for bringing to my attention some shortcomings of an earlier version of the theory presented here.

Yatabe (2021) presents a theory according to which one of the meanings of a word like *different* contains the meaning of a reciprocal pronoun inside it. In this theory, a sentence like (1) is associated with a semantic representation like (2) by mechanisms provided by HPSG and Minimal Recursion Semantics.

(2) a = Anna $\wedge$ b = Bill
$\wedge$ some(X, $X = a + b$,
every(y, member_of(y, X),
some(z, and(film_or_films(z),
every(w, other(w, X, y),
the(v, film_or_films(v) $\wedge$ like(w, v),
different(z, v)))),
like(y, z))))

Lines 5–7 more or less correspond to the contribution that the adjective *different* makes to the meaning of the sentence, and line 5 more or less corresponds to what I claim to be the reciprocal meaning contained in the meaning of *different*. As is standard in MRS representations, a quantifier meaning is expressed by a three-place predicate whose three arguments are the variable it binds, its restrictor, and its nuclear scope respectively. The predicate *member_of* is assumed to hold of its two arguments if and only if the denotation of the first argument is a member of the group denoted by the second argument. The predicate *other* is assumed to hold of its three arguments if and only if the denotation of the first argument is a member of the group denoted by the second argument other than the denotation of the third argument. (The symbol *and* represents conjunction that is expressed by an elementary predication whose RELN value is *and*, while the symbol "$\wedge$" represents conjunction that is expressed by a shared handle.) In this paper, I will refer to the analysis illustrated in (2) as the hidden-reciprocal analysis of internal readings.

According to the hidden-reciprocal analysis of internal readings, the antecedent of a word like *different* is the semantic antecedent of a reciprocal meaning, and therefore has to be a DP. The analysis is thus not applicable to the apparent internal readings of sentences like (3) and (4), which do not contain a DP that could serve as the antecedent of *different*. Accordingly, it is claimed in Yatabe (2021) that the apparent internal readings of sentences like (3) and (4) are licensed in a way that is entirely different from the way in which the internal readings of sentences like (1) are licensed. I will refer to this claim as the non-uniformity claim about apparent internal readings.

(3) Different people discovered America and invented bifocals.

(4) John saw and reviewed different films.

In this paper, I will present some new evidence for the hidden-reciprocal analysis of internal readings and for the non-uniformity claim about apparent internal readings. In addition, in order to enhance the plausibility of the overall theory, I will also show how the collective interpretation of reciprocals and the interaction of

reciprocals and cumulative interpretation can be accounted for within the theory. In exemplifying phenomena involving internal readings, I will mostly rely on Japanese examples because Japanese is one of the languages that make a morphological distinction between the two types of internal readings mentioned above.

2 Additional evidence for the hidden-reciprocal analysis

2.1 Distributed internal readings

First, consider the range of interpretations that a sentence like (5) can have.

(5) [Gakusei-tachi ga] [betsu-betsu no hon o] yonda.
[students NOM] [mutually different book ACC] read-PAST
'The students read different books.'

The expression *betsu-betsu no*, which I have glossed as "mutually different", can only give rise to an internal reading, unlike the expression *betsu no*, which can give rise to an external reading as well as an internal reading, as shown in (6) and (7).

(6) Kaoru ga betsu no hon o yonda.
Kaoru NOM different book ACC read-PAST
'Kaoru read a book different from the contextually salient book.'

(7) Gakusei-tachi ga hitori-hitori betsu no hon o yonda.
students NOM each different book ACC read-PAST
'The students each read a book different from the books that the other students read.'

When sentence (5) is presented without any context, the most salient reading is probably one in which it means "No two of the students read the same book or books". This, however, is not the only reading the sentence has. Suppose that a group consisting of 15 students had been divided into groups of three and that each student had been told not to read the same book or books as the other two students in the same group. In such a context, (5) has a reading in which it is true if and only if each student obeyed the instruction and read a book or books different from each of the books read by the other two students in the same group. This is a reading in which the core meaning of the expression *betsu-betsu no* 'mutually different' is required to hold, in a distributed way, in each of the subgroups that together constitute the group that the sentence is talking about. I will refer to a reading like this as a distributed internal reading.

The existence of distributed internal readings is predicted by the hidden-reciprocal analysis of internal readings. In order to see how, we first need to take a look at a certain type of interpretation that can be assigned to reciprocal pronouns. Consider the sentence in (8).

(8) [Gakusei-tachi ga] [otagai no kaado ni] sain shita.
[students NOM] [each other GEN card DAT] sign do-PAST
'The students signed each other's cards.'

Suppose that a group consisting of 15 students had been divided into groups of three and that each student had been told to sign the cards of the two other students in the same group. In such a context, sentence (8) has a reading in which it is true if and only if each student signed the cards of the two other students in the same group. This is a reading embodying what Dalrymple et al. (1998) call Distributed Strong Reciprocity.

The theory presented in Yatabe (2021) does not take the existence of Distributed Strong Reciprocity into account, but we can easily rectify that shortcoming by modifying slightly the meaning assigned to the predicate *other*, which is used to express the meaning of reciprocals. Consider (9), which is the semantic representation assigned to sentence (8) in the theory under discussion.

(9) the(X, students(X),
every(y, member_of(y, X),
every(w, other(w, X, y),
the(z, card_or_cards_of(z, w),
signed(y, z)))))

I now propose interpreting the predicate symbol *other* as a predicate that holds of its three arguments if and only if (i) the denotation of the first argument is a member of the group formed by those members of the group denoted by the second argument that are "closely related", in the contextually relevant sense, to the denotation of the third argument and (ii) the denotation of the first argument does not overlap with the denotation of the third argument. In the case at hand, each student could be viewed as "closely related" to his or her two groupmates, and the elementary predication "other(w, X, y)" could thus function as a formula that is true if and only if the denotation of w is one of the two groupmates of the denotation of y. When that interpretation is given to this elementary predication, the semantic representation expresses the Distributed Strong Reciprocity reading of the sentence, whereas the same semantic representation expresses what is called Strong Reciprocity in the literature when each student is viewed as "closely related" to all the students, rather than just to his or her two groupmates.

What I called a distributed internal reading above is the reading that results when the reciprocal meaning inside the meaning of a word like *different* is that of Distributed Strong Reciprocity. Consider, for example, the semantic representation shown in (10), which is associated with sentence (5).

(10) the(X, students(X),
every(y, member_of(y, X),
some(z, and(book_or_books(z),
every(w, other(w, X, y),
the(v, book_or_books(v) $\wedge$ read(w, v),

different(z, v))))),
read(y, z))))

Here again, the elementary predication "other(w, X, y)" could function as a formula that is true if and only if the denotation of w is one of y's two groupmates. And when that elementary predication is given that interpretation, the semantic representation expresses the distributed internal reading of the sentence.

The proposed theory makes correct predictions about the interpretation of sentences in which the antecedent of a word like *different* is a quantificational noun phrase as well. Consider (11) and (12), for example.

(11) [Hotondo no gakusei ga] [otagai no kaado ni] sain shita.
[most student NOM] [each other GEN card DAT] sign do-PAST
'Most students signed each other's cards.'

(12) [Hotondo no gakusei ga] [betsu-betsu no hon o] yonda.
[most student NOM] [mutually different book ACC] read-PAST
'Most students read different books.'

Since in this theory the quantifier meaning inside the reciprocal meaning is assumed to be copied from the semantic antecedent of the reciprocal, these sentences are associated with the semantic representations shown in (13) and (14) respectively.

(13) the(X, students(X),
most(y, member_of(y, X),
most(w, other(w, X, y),
the(z, card_or_cards_of(z, w),
signed(y, z)))))

(14) the(X, students(X),
most(y, member_of(y, X),
some(z, and(book_or_books(z),
most(w, other(w, X, y),
the(v, book_or_books(v) ∧ read(w, v),
different(z, v)))),
read(y, z))))

The predicate *most* in line 3 of (13) is copied there from line 2, and the same predicate in line 4 of (14) is copied there from line 2.[1] Suppose that the sentences

[1] In the theory proposed, the quantifier meaning that gets copied into the meaning of a reciprocal must be that of a distributive quantifier. It is assumed that quantifiers that are composed of numeral determiners such as *at most eight* are not distributive and that their meanings are therefore never copied into the meaning of a reciprocal. I speculate that a sentence like *At most eight books were written* is associated with a semantic representation that means "It is not the case that a book was written, with the exception of at most eight books", in which the meaning of the quantifier *at most eight books* is decomposed into negation, an existential quantifier, and a statement of exceptions. Given such an analysis, a reciprocal sentence like *At most eight students knew each other* is predicted to be associated with a semantic representation that means "It was not the case that a student knew any other student, with the exception of at most eight students".

are both talking about a group consisting of 12 students that had been divided into pairs, so that each student had one partner who was "closely related" to him or her. In such a situation, the representation in (13) is true if and only if most students signed their partner's card, assuming that an elementary predication of the form "most(x, R, S)" is true if and only if the number of possible values of x that make both R and S true is larger than the number of possible values of x that make R true but S false. This corresponds to one possible reading of sentence (11). Likewise, (14), which is true in the situation described above if and only if most students read a book or books different from the book or books their partner read, expresses a reading that sentence (12) can have in such a situation.

Thus, when combined with the natural hypothesis that the reciprocal meaning contained in the meaning of a word like *different* can express Distributed Strong Reciprocity as well as Strong Reciprocity, the hidden-reciprocal analysis of internal readings makes correct predictions about what I have called distributed internal readings.

2.2 Collective interpretation of reciprocals

In the remainder of this section, namely in subsections 2.2 and 2.3, I will discuss some inadequacies of the theory of reciprocals that I have presented so far, and will show how they can be taken care of by adding some new lexical entries to the grammar and modifying some peripheral if not minor aspects of the theory.

In the present subsection, I will discuss problems posed by the types of interpretations of reciprocals that I have been ignoring up to this point. In subsection 2.1, I only considered two kinds of interpretations that can be given to reciprocals, viz. Distributed Strong Reciprocity and Strong Reciprocity. The analysis that I have proposed deals with these two interpretations of reciprocals in a unified way. It has been noted in the previous literature, however, that reciprocals can be given some other types of interpretations as well. Some of these other types of interpretations turn out to be problematic for the proposed theory.

I will begin by examining what has been called Weak Reciprocity in the literature. Weak Reciprocity is a term that has been used in relation to the fact that reciprocals appear to be able to express an existential quantifier even when its semantic antecedent has the quantificational force of a universal quantifier. Consider the sentences in (15) and (16).

(15) They scratched one another's backs. (from Langendoen (1978))

(16) The children give each other a present. (from Beck (2001))

Example (15) is judged to be true if there were four people forming something like a circle and the first person scratched the second person's back, the second person scratched the third person's back, the third person scratched the fourth person's back, and the fourth person scratched the first person's back. In other words, the sentence can mean "*Every* person scratched *some* other person's back, and

every person's back has been scratched by some person", rather than "*Every* person scratched *every* other person's back". The reciprocal appears able to express an existential quantifier in sentence (16) as well. Beck (2001) says the following about this example.

> One way to do the presents is that everybody brings a present and drops it into a big bag. Later you get to close your eyes and choose a present from the bag. Alternatively, you write your name on a piece of paper, the papers go into a bag, and you have to bring a present for the person whose name you draw. In either case, [(16)] can truthfully describe the procedure because every child gives and receives a present.

In other words, this sentence can mean "*Every* child gives a present to *some* other child, and every child receives a present from some child", rather than "*Every* child gives a present to *every* other child". Both in (15) and in (16), the idea that the quantificational force of the semantic antecedent is copied into the meaning of a reciprocal does not seem to work at first blush, although it seemed to work in earlier cases, which exemplified either Strong Reciprocity or Distributed Strong Reciprocity.

Not all cases of Weak Reciprocity are problematic for the theory. For instance, the interpretation of sentence (15) is in fact correctly accounted for in the proposed theory as long as it is assumed that there are situations where only the person directly in front of you is to be regarded as "closely related" to you. Given such an assumption, the semantic representation assigned to sentence (15) in accordance with the theory presented so far can mean "Each person scratched the back of the person directly in front of him or her", and the sentence is therefore expected to be usable in the kind of situation described above. The same analysis applies to an example like (17).

(17) Walking down Mass. Ave. from Arlington to Boston the sociologist found out: The residents on the eastern side of Mass. Ave. know each other. (from Sauerland (1998))

According to Sauerland (1998), the second sentence in (17) can be true even if every resident on the eastern side of Mass. Ave. only knows his or her neighbors. This is expected in my account because a resident's neighbors are the people who are easiest to regard as "closely related" to that resident.

The theory as it has been presented, however, cannot account for all cases of Weak Reciprocity. For example, sentence (16) does pose a problem for the account. Since the child who receives each present is not known in advance, there is no sense in which the child who ends up receiving a present is in any way "closely related" to the child who has brought that present. This means that the solution that is available in the case of sentence (15) is not available in this case and that some modification must be made to the theory.

I propose adding the following hypotheses to the theory. Reciprocal pronouns allow two distinct interpretations: non-collective interpretation, which is the one I have been discussing, and collective interpretation. Weak Reciprocity is one possible result of a reciprocal receiving collective interpretation. The two interpretations of reciprocal pronouns are each associated with a separate lexical entry, although the quantificational force of the semantic antecedent is copied into the meaning of a reciprocal irrespective of which interpretation the reciprocal receives. The meaning of a word like *different* is assumed to contain the non-collective meaning of a reciprocal; the theory of internal readings thus remains unaffected. I will flesh out these hypotheses in the remainder of this subsection.

It has been noted in the literature that reciprocal pronouns can receive collective interpretation. The sentences in (18) and (19) are some of the relevant examples.

(18) The satellite, called Windsock, would be launched from under the wing of a B-52 bomber and fly to a 'liberation point' where the gravitational fields of the Earth, the Sun and the Moon cancel each other out. (from Dalrymple et al. (1998))

(19) The children painted a picture of each other. (from Sauerland (1998))

In (18), the gravitational field of the Earth is supposed to cancel out the combined gravitational fields of the Sun and the Moon, and so on. Likewise, (19) has a reading in which it means that each child painted one picture showing all the other children.

It turns out that some instances of Weak Reciprocity can be regarded as cases of collective interpretation. For instance, the example in (16) above can be taken to be saying "Each child gave a present to the group consisting of all the others". This is arguably a natural account, given the kinds of situations in which the sentence can truthfully be used.

What is shown in (20) is the semantic representation that I propose to assign to sentence (19); it illustrates the way that I propose to analyze the collective interpretation of a reciprocal in general.

(20) $\text{the}(X, \text{children}(X),$
$\quad \text{every}(y, \text{member\_of}(y, X),$
$\quad\quad \text{some}(Z, \text{every}(w, \text{other}(w, X, y), \text{member\_of}(w, Z))$
$\quad\quad\quad \wedge \text{every}(w', \text{member\_of}(w', Z), \text{other}(w', X, y)),$
$\quad\quad\quad \text{some}(u, \text{picture\_of}(u, Z),$
$\quad\quad\quad\quad \text{painted}(y, u)))))$

In a situation where each child can be viewed as "closely related" to all the children, this semantic representation means that each of the children painted a picture of the group containing all the other children and no one else. Shown in Fig. 1 is the lexical entry for *each other* that produces semantic representations like this.

Notice that the semantic representation shown in (20) still involves copying of the quantificational force of the semantic antecedent into the reciprocal meaning;

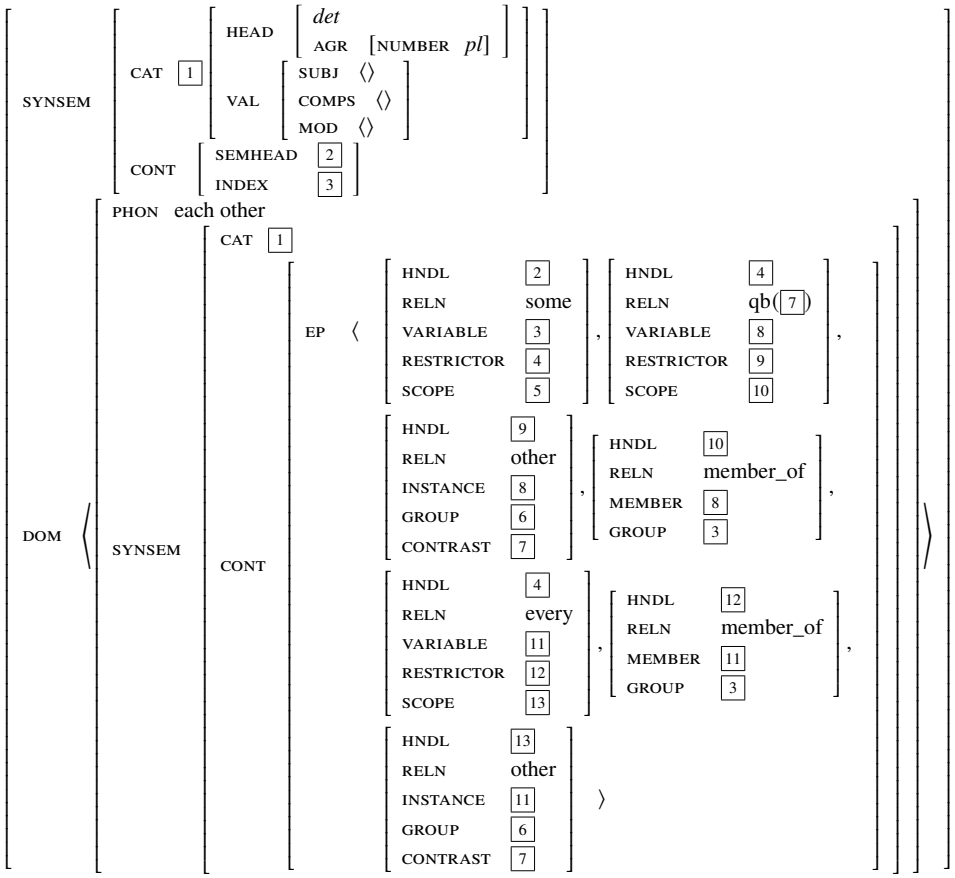


Figure 1: The lexical entry for *each other* that gives rise to collective interpretation

the predicate *every* in line 3 was copied there from line 2 on account of the *qb* function in the lexical entry in Fig. 1. (See Yatabe (2021) for the definition of the *qb* function.) This way of representing the collective interpretation receives support from the example in (21), assuming that the sentence has a reading in which it is true if and only if most children painted one picture showing most of the other children.

(21) Most children painted a picture of each other.

The proposed account assigns to this example the semantic representation shown in (22), in which the predicate *most* in line 3 has been copied there from line 2.

(22) the(X, children(X),
most(y, member_of(y, X),
some(Z, most(w, other(w, X, y), member_of(w, Z))
$\wedge$ every(w', member_of(w', Z), other(w', X, y)),
some(u, picture_of(u, Z),
painted(y, u)))))

It is my contention that the theory that I am proposing is capable of accounting for the entire range of reciprocal interpretations except those interpretations mentioned in footnote 2 below that arguably involve idiomatization, although here I can only discuss two additional examples, namely (23) and (24).

(23) The telephone poles are spaced five hundred feet from each other. (from Dalrymple et al. (1998))

(24) The members of this family have inherited the shop from each other for generations. (from Beck (2001))

The sentence in (23) illustrates the type of interpretation that Dalrymple et al. (1998) call Intermediate Reciprocity, while the sentence in (24) is supposed to illustrate the interpretation that Dalrymple et al. (1998) named Inclusive Alternative Ordering.[2]

Sentence (23) can be analyzed the same way that sentence (17) is; it can be interpreted as saying that each of the telephone poles is spaced five hundred feet from its neighbor(s). Such an interpretation arises when for each telephone pole x, only the telephone poles that are closest to x are taken to be "closely related" to x.

Sentence (24) can be viewed as involving collective interpretation of a reciprocal. Specifically, the sentence can be interpreted as saying that each member of the family has inherited the shop from the collective formed by all the other family members for generations. This way of looking at (24) helps make sense of the contrast between (24) and (25), noted in Beck (2001).

(25) ??These three people inherited the shop from each other.

When only three people are involved as in (25), it is difficult to view the people as consisting of one individual and a large, stable collective from which one can inherit a shop. That difficulty can be the source of the low acceptability of (25).

To summarize the discussion in this subsection, some instances of the so-called Weak Reciprocity readings and other readings of reciprocal pronouns pose a problem for the theory of reciprocals proposed in Yatabe (2021) and subsection 2.1, but the problem can be taken care of by adding to the grammar a new lexical entry for reciprocals that gives them collective interpretation.

2.3 Interaction with cumulative interpretation

Another inadequacy of the theory of reciprocals presented in Yatabe (2021) and subsection 2.1 concerns sentences like the following, discussed in Sternefeld (1998) and Sauerland (1998).

(26) John read the letters they wrote to each other.

[2] Beck (2001) identifies two distinct types of Inclusive Alternative Ordering readings. Sentence (24) is an example of one of those two types, and the other type is exemplified by sentences like *The two books are lying on top of each other*, discussed by Langendoen (1978) and others. Both Langendoen (1978) and Beck (2001) adduce evidence that suggests, to my mind, that the second type is best understood as involving idiomatization of sorts.

(27) They wrote these six letters to each other.

Both these sentences allow the antecedent of a reciprocal and another DP to jointly receive cumulative interpretation. In (27), for example, the antecedent of the reciprocal and the DP *these six letters* can receive cumulative interpretation, giving the sentence a reading in which it is true only if there are six letters such that each of those six letters was sent by one of the people to one of the others. This observation is incompatible with the theory proposed in Yatabe (2021), because in that theory the semantic antecedent of a reciprocal is taken to have the quantificational force of a distributive universal quantifier, which never gives rise to cumulative interpretation when associated with a grammatical subject position. Sentence (28), for instance, does not have a cumulative reading, i.e. a reading in which it is true if and only if every student wrote a letter or letters and there were six letters in total which were each written by a student.

(28) Every student wrote six letters.

In order to account for the availability of cumulative interpretation in (26) and (27), below I will propose modifications to some peripheral if not minor aspects of my theory of reciprocals and delineate a novel account of cumulative interpretation that takes advantage of those modifications.

The first thing that I need to do is to modify the meaning assigned to the silent distributive operators that are assumed to be responsible for the distributive readings of sentences like *The residents wrote letters*. In Yatabe (2021), the silent distributive operators, both the one that is assumed to be adjoined to a subject-seeking expression like a VP and the one that is assumed to be adjoined to a DP, are taken to have the quantificational force of a distributive universal quantifier, causing the problem just described. The new lexical entries that I now propose to associate with the two distributive operators are shown in Fig. 2 and Fig. 3, respectively.

These lexical entries give rise to syntactic structures like (29) and (30), where the symbols "dist1" and "dist2" are used respectively to stand for the silent distributive operators given in Fig. 2 and Fig. 3.

(29) [The residents [$_{\mathrm{VP}}$ dist1 [$_{\mathrm{VP}}$ wrote letters]]]

(30) [[$_{\mathrm{DP}}$ [$_{\mathrm{DP}}$ The residents] dist2] [wrote letters]]

Both syntactic structures are capable of producing the semantic representation shown in (31).

(31) $\mathrm{the}(x, \mathrm{residents}(x),$
$\quad \mathrm{some}(c, \mathrm{cover\_list}(c, x),$
$\quad\quad \mathrm{every}(x', \mathrm{ith\_element}(x', c),$
$\quad\quad\quad \mathrm{some}(y, \mathrm{letters}(y),$
$\quad\quad\quad\quad \mathrm{wrote}(x', y)))))$

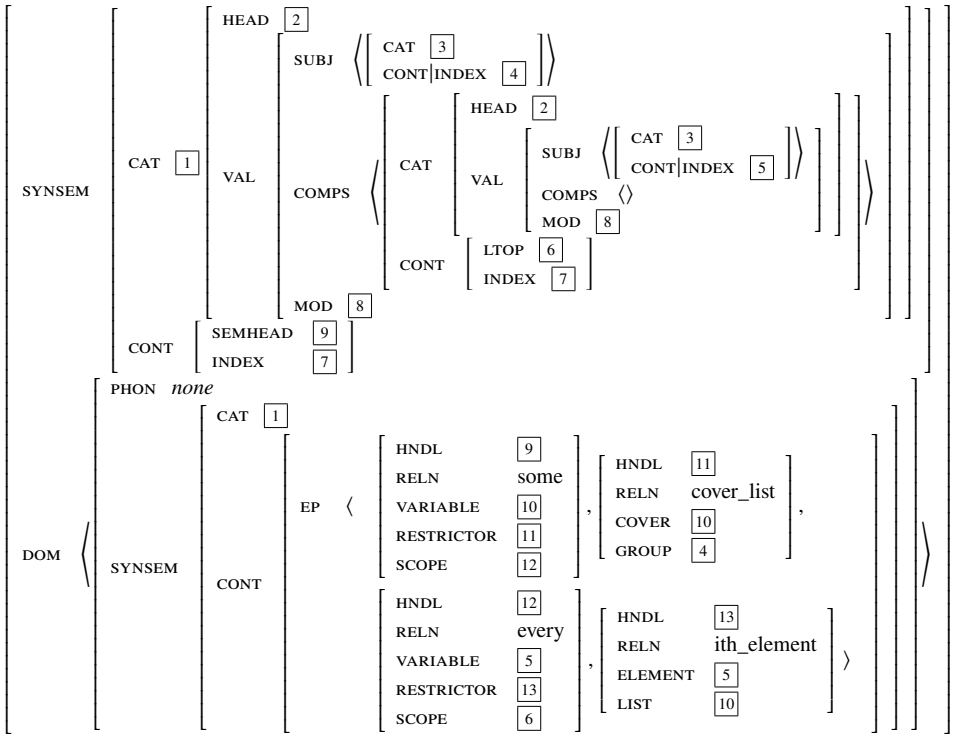


Figure 2: The lexical entry for a silent distributive operator that combines with subject-seeking expressions

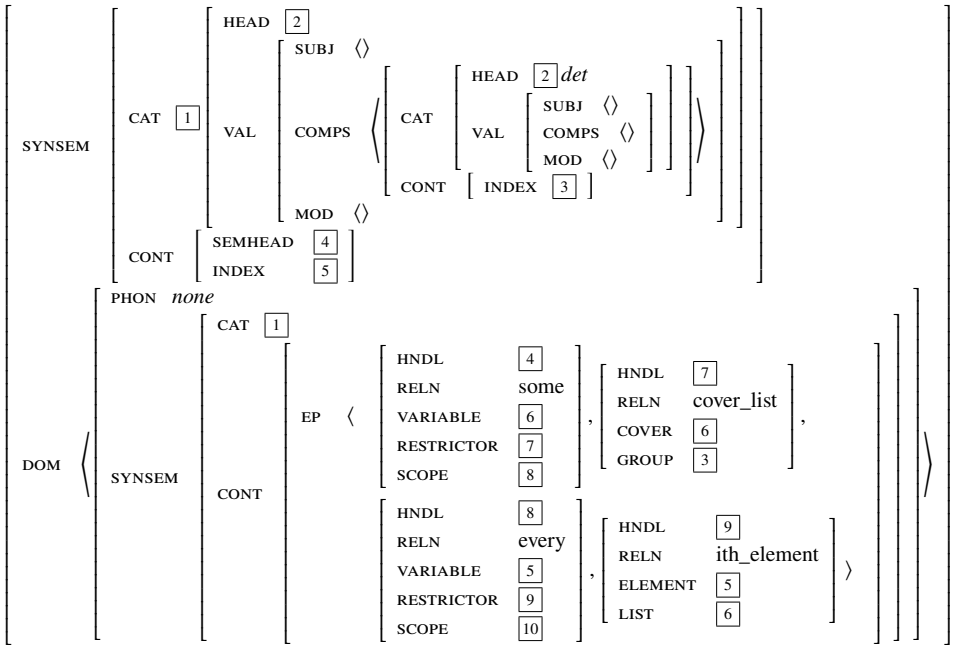


Figure 3: The lexical entry for a silent distributive operator that combines with DPs

The predicate symbol *cover_list*, used in (31), is given the interpretation defined in (32).

(32) A formula of the form "cover_list(x, y)" is true if and only if the denotation of x is a cover list of the denotation of y. We say that a list C is a *cover list* of an entity E if and only if the sum of all the elements of C amounts to E.

A cover list is so called because it is similar to what is called a cover in the literature (see Schwarzschild (1996, Chapter 5) and the references cited there) but is a list rather than a set. One possible cover list of a group consisting of three books, say Book1, Book2, and Book3, is a list whose first element is Book1 and whose second element is Book2 + Book3, i.e. the group consisting of Book2 and Book3.

The predicate symbol *ith_element*, also used in (31), is to be interpreted according to the definition in (33).

(33) Definition of the interpretation of the predicate symbol *ith_element*:
Suppose that one or more elementary predications of the form "ith_element(_, _)" are labeled by the same handle and that no other elementary predication is labeled by that handle. Then the denotation of those elementary predications, i.e. "ith_element(a_1, b_1) $\wedge \cdots \wedge$ ith_element(a_n, b_n)" (where $n \geq 1$), is defined only if there is an integer m larger than 1 such that $b_1, \ldots, b_n$ are each a list of length m. Provided it is defined, the denotation is 1 if there is an integer i ($1 \leq i \leq m$) such that, for each j ($1 \leq j \leq n$), a_j is the ith element of b_j, and is 0 otherwise.

Let me illustrate the consequences of these definitions using the sentence *The residents wrote letters* as an example. Suppose the DP *the residents* refer to three people, Amy, Bill, and Chris. One possible cover list for the denotation of this DP is, then, the list <Amy + Bill, Chris>, where "Amy + Bill" is meant to be the sum of Amy and Bill. When the variable c denotes that cover list, the formula "ith_element(x', c)" denotes 1 if and only if x' denotes either Amy + Bill or Chris. Thus, the semantic representation shown in (31) above can be true if Amy and Bill jointly wrote some letters and Chris wrote some letters too.

I am now in a position to present my account of cumulative readings. In the account that I propose, cumulative interpretation is generated when quantificational elementary predications coming from two or more silent distributive operators are merged with each other by the rule given in (34), which is one of the rules that Yatabe and Tam (2021) propose in order to assign appropriate truth conditions to sentences such as *Every woman is smiling and every man is frowning who came in together*, discussed in Fox and Johnson (2016). (What is given in (34) is the version of the rule presented in Yatabe (2021), which is different from the original version only in minor respects.)

(34) MRS Adjustment Rule 2:
Suppose that a given MRS representation contains n elementary predications of the following form, that these n elementary predications appear in

this order in the MRS representation, and that there is no other elementary predication in the representation that shares the same RESTRICTOR value and the same SCOPE value with these n elementary predications.

$$\begin{bmatrix} \text{HNDL} & \boxed{h_1} \\ \text{RELN} & \boxed{1} \\ \text{VAR} & \boxed{v_1} \\ \text{RESTRICTOR} & \boxed{2} \\ \text{SCOPE} & \boxed{3} \end{bmatrix}, \cdots, \begin{bmatrix} \text{HNDL} & \boxed{h_n} \\ \text{RELN} & \boxed{1} \\ \text{VAR} & \boxed{v_n} \\ \text{RESTRICTOR} & \boxed{2} \\ \text{SCOPE} & \boxed{3} \end{bmatrix}$$

Then these n elementary predications can be replaced by a single elementary predication of the following form, if $\boxed{h_1}$, ..., $\boxed{h_n}$ are known to be identical with each other and $\boxed{v_1}$, ..., $\boxed{v_n}$ are distinct from each other.

$$\begin{bmatrix} \text{HNDL} & \boxed{h_1} \\ \text{RELN} & \boxed{1} \\ \text{VAR} & \boxed{v_1 + \cdots + v_n} \\ \text{RESTRICTOR} & \boxed{2} \\ \text{SCOPE} & \boxed{3} \end{bmatrix}$$

The newly created elementary predication is to be placed at the position where the leftmost of the deleted elementary predications was located, and is to be interpreted in the expected way. For instance, when the denotation of $\boxed{1}$ is "every", the resultant elementary predication is interpreted as saying "Every n-tuple that makes the restrictor true makes the nuclear scope true as well".

I will use sentence (35) to illustrate the workings of the proposed account.

(35) The residents wrote four letters.

(36) [[The residents] [dist1 [wrote [[four letters] dist2]]]]

The sentence can have the syntactic structure shown in (36), and the two quantificational elementary predications that come from *dist1* and the two quantificational elementary predications that come from *dist2* can be merged pairwise with each other by the MRS Adjustment Rule 2. The resulting semantic representation will look like (37).

(37) the(x, residents(x),
  some(y, four(y) ∧ letters(y),
    some(c+d, cover_list(c, x) ∧ cover_list(d, y),
      every(x'+y', ith_element(x', c) ∧ ith_element(y', d),
        wrote(x', y')))))

We can associate with the subject DP the cover list <Amy + Bill, Chris>, and with the object DP the cover list <letter1, letter2 + letter3 + letter4>. When the variables

c and d are assigned these two cover lists as their values respectively, the formula "ith_element(x', c) $\wedge$ ith_element(y', d)" denotes 1 if and only if either x' denotes Amy + Bill and y' denotes letter1 or x' denotes Chris and y' denotes letter2 + letter3 + letter4. Therefore the sentence is predicted to be true if Amy and Bill jointly wrote one letter and Chris wrote three letters. The proposed theory thus accounts for the availability of the cumulative reading of sentence (35).

In order for the semantic representation in (37) to express adequate truth conditions, it needs to be assumed that the denotation of a formula of the form "every(x, P, Q)" is undefined when the denotation of the second argument, P, is undefined. In conjunction with the stipulation (stated in (33)) that the denotation of a formula of the form "ith_element(a_1, b_1) $\wedge \cdots \wedge$ ith_element(a_n, b_n)" is undefined unless $b_1, \ldots, b_n$ are lists of the same length, this assumption ensures that (37) is true if and only if the cumulative reading of the sentence is true.

The proposed account of cumulative interpretation is arguably capable of dealing with sentences in which DPs like *at most eight books* contribute to cumulative interpretation, as long as it is assumed (i) that, as proposed in footnote 1, the meaning of such a DP is decomposed into negation, an existential quantifier, and a statement of exceptions and (ii) that negations that come from two or more such DPs can be merged into one. It would then be possible to interpret a sentence like *Exactly two residents wrote exactly four letters* as meaning "It is not the case that a resident wrote a letter, with the exception of exactly two residents and exactly four letters".

Let us now see how we can deal with sentences like (26) and (27). I will focus on sentence (27). This sentence can have a syntactic structure like (38), where the subject and the reciprocal each combine with *dist2* once and the DP *these six letters* combines with *dist2* twice.

(38) [[They dist2] [wrote [[[these six letters] dist2] dist2] to [[each other] dist2]]]

Given the account of cumulative interpretation proposed here and given the account of the collective interpretation of reciprocals proposed in Sect. 2.2, this syntactic structure can be associated with the semantic representation shown in Fig. 4. The formula "$x \rightarrow j+b+t$" on line 1 means that $j+b+t$ (which could denote John, Bill, and Tom) has been selected as the antecedent of *they* (see Yatabe (2022)). Lines 3 and 4 express the cumulative interpretation of *they* and *these six letters*. Lines 5 and 6 express the collective interpretation of *each other*. And lines 7 and 8 express the cumulative interpretation of *these six letters* and *each other*.

This semantic representation is true in a situation where each of the three people referred to by the subject DP sent the other two people a letter each. Let us refer to the three people being talked about as John, Bill, and Tom, and to the six letters being talked about as Letter1, ..., Letter6. The values of the variables c and d, which are meant to denote cover lists, can be <John, Bill, Tom> and <Letter1 + Letter2, Letter3 + Letter4, Letter5 + Letter6>, respectively. When the values of the variables x' and y' are John and Letter1 + Letter2 respectively, the value of z will

$$
\begin{aligned}
&\text{some}(x, x \rightarrow j + b + t,\\
&\quad \text{some}(y, \text{these}(y) \wedge \text{six}(y) \wedge \text{letters}(y),\\
&\quad\quad \text{some}(c + d, \text{cover\_list}(c, x) \wedge \text{cover\_list}(d, y),\\
&\quad\quad\quad \text{every}(x' + y', \text{ith\_element}(x', c) \wedge \text{ith\_element}(y', d),\\
&\quad\quad\quad\quad \text{some}(z, \text{every}(w, \text{other}(w, x, x'), \text{member\_of}(w, z))\\
&\quad\quad\quad\quad\quad \wedge \text{every}(w', \text{member\_of}(w', z), \text{other}(w', x, x')),\\
&\quad\quad\quad\quad\quad \text{some}(e + f, \text{cover\_list}(e, y') \wedge \text{cover\_list}(f, z),\\
&\quad\quad\quad\quad\quad\quad \text{every}(y'' + z', \text{ith\_element}(y'', e) \wedge \text{ith\_element}(z', f),\\
&\quad\quad\quad\quad\quad\quad\quad \text{wrote}(x', y'', z')))))))
\end{aligned}
$$

Figure 4: A semantic representation that can be assigned to (38)

be Bill + Tom, the values of e and f can be <Letter1, Letter2> and <Bill, Tom> respectively, and the last two lines of the representation can therefore mean "John wrote Letter1 to Bill and Letter2 to Tom". What happens when the value of x' is either Bill or Tom is analogous.

Incidentally, the modifications that have been proposed in this subsection allow the theory to assign adequate truth conditions to sentences like (39).

(39) They released one another.

As noted in Langendoen (1978), sentence (39) can be true when there were three prisoners A, B, and C, A and B jointly released C, and then C released A and B. In the theory proposed, this reading can be obtained by associating with the subject DP a cover list of the form <A + B, C>.

There is one problem that I have been ignoring so far. The problem is that the semantic representation in Fig. 4 in fact violates the condition on the relationship between the second and the third argument of the predicate *other* formulated in Yatabe (2021) (stated in (44) of that paper). I thus propose that that constraint be replaced with the constraint given in (41).[3] The term *source variable*, used in (41), is defined in (40).[4]

(40) Definition of *source variable*:
A variable x is a *source variable* of a variable y if and only if
(a) x and y are the same variable, or
(b) y is bound by a quantifier whose restrictor is of the form "member_of(y, x)", or
(c) there is a variable c such that (i) y is bound by a quantifier in whose restrictor a formula of the form "ith_element(y, c)" is conjoined with zero or more other formulas and (ii) c is bound by a quantifier in

[3] Unlike the condition stated in (44) of Yatabe (2021), the condition stated in (41) here does not take into account examples like *Tom shouted and Mary cried each other's names*, discussed in Chaves (2014).

[4] Clause (d) is necessary in dealing with cases involving *respectively* interpretation. See Yatabe (2022) for discussion.

whose restrictor a formula of the form "cover_list(c, x)" is conjoined with zero or more other formulas, or

(d) y is bound by a quantifier whose restrictor is of the form "$y = x$", or

(e) x is a source variable of a variable that is a source variable of y.

(41) In each elementary predication whose RELN value is *other*, the GROUP value (i.e. the second argument in the linear notation) must be a source variable of, but must not be the same variable as, the CONTRAST value (i.e. the third argument in the linear notation).

3 Evidence for the non-uniformity claim

The hidden-reciprocal analysis of internal readings is not applicable to sentences like (3) and (4), which do not contain a DP that could serve as the antecedent of *different*. The analysis thus entails that the grammatical mechanism that gives rise to the apparent internal readings of these sentences is different from the mechanism that gives rise to the internal readings of sentences like (1). In Yatabe (2021), the apparent internal readings of (3) and (4) are accounted for by hypothesizing (i) that (3) involves left-node raising (LNR) of *different people* out of two clauses, (ii) that (4) involves right-node raising (RNR) of *different films* out of two VPs, and (iii) that the word *different* in these sentences denotes a one-place predicate that is satisfied if and only if (a) its sole argument has a referential equivalent of the form $x_1 + \cdots + x_n$, where $n > 1$, and (b) for each i and j such that $1 \leq i \leq n$, $1 \leq j \leq n$, and $i \neq j$, the denotation of x_i and the denotation of x_j are different from each other.[5] The term *referential equivalent*, which I have just used, is defined as follows (see Yatabe (2022)).

(42) Definition of *referential equivalent*:
A variable x is a *referential equivalent* of a variable y if and only if (i) x and y are the same variable, or (ii) x is bound by a quantifier whose restrictor is of the form "$x \rightarrow y$" or of the form "$x = y$", or (iii) y is bound by a quantifier whose restrictor is of the form "$y \rightarrow x$" or of the form "$y = x$", or (iv) x is a referential equivalent of some variable that is a referential equivalent of y.

The details of this account are as follows. In the theory of RNR and LNR defended in Yatabe and Tam (2021), a right- or left-node-raised expression may be given a composite index (i.e. an index of the form $x_1 + \cdots + x_n$) whose components (i.e. $x_1, \ldots, x_n$) are the indices that the expression is given before application of right- or left-node raising. As a result, sentence (3) can be associated with a semantic representation that means "for some $x + y$ such that $x + y$ are different people, x discovered America and y invented bifocals", if *different people* is taken to have been left-node-raised. Likewise, sentence (4) can be associated with a

[5] Hypothesis (iii) is hinted at but not articulated in Yatabe (2021).

semantic representation that means "for some $x + y$ such that $x + y$ are different films, John saw x and reviewed y", if *different films* is taken to have been right-node-raised. The grammatical mechanism that licenses this interpretation is the same mechanism that licenses the most salient interpretation of the example in (43), discussed in Abbott (1976).

(43) I borrowed, and my sister stole, a total of $3000 from the bank.

The example in (44) below, discussed in Kubota and Levine (2016) and Kubota and Levine (2020), could be taken to be circumstantial evidence for such an account, because the acceptability of a sentence like this is expected in a theory based on the view that the apparent internal readings of the sentences under discussion are resulting from LNR and RNR, but unexpected in other theories, such as Kubota and Levine's, according to which apparent internal readings are always licensed by the presence of some sort of conjunction or plurality in the same sentence.

(44) John defeated, whereas/although Mary lost to, the exact same opponent.

Example (44) involves RNR out of non-coordinate structure, and the account described above is the only currently available account of the apparent internal reading that this sentence has. The theory advocated in Kubota and Levine (2016) and Kubota and Levine (2020) does not contain a mechanism that licenses non-coordinate RNR, and is therefore incapable of handling the apparent internal reading of a sentence like this.

Kubota and Levine, however, say the following about examples like (44).

> We think that the relevant generalization is whether the construction in question has the meaning of conjunction. *Whereas* and *although* are truth-conditionally equivalent to conjunction, with an extra pragmatic function of indicating a particular discourse relation (some kind of contrast) between the two clauses. Since the analysis we present below is predicated of the conjunctive meaning of *and* rather than its syntactic coordinatehood, the examples in [(44)], rather than undermining our analysis, in fact provide further corroboration for it. (Kubota and Levine (2020, p. 123))

I take them to be saying here something like the following: if a sentence like (44) is possible only when the syntactic structure involved has conjunctive meaning, then the semantic part of their theory can be kept intact, and it might even be the case that a non-coordinate syntactic structure that has conjunctive meaning can be treated, on some temporary basis, as a type of coordinate structure, allowing the syntactic part of their theory to be kept intact as well. In other words, I interpret the authors as saying, in effect, that non-coordinate RNR and LNR are coordinate RNR and LNR in disguise.

There is a reason to believe that non-coordinate RNR and LNR cannot be explained away as coordinate RNR and LNR in disguise. Consider sentence (45).

(45) [Onaji apaato no betsu-betsu no heya ni], Tanaka san
[same apartment building GEN mutually different room DAT] Tanaka san
ga hikkoshite kita no to onaji hi ni Yamada san mo
NOM move-GER come-PAST NML with same day DAT Yamada san also
hikkoshite kita no desu.
move-GER come-PAST NML be.POL.PRES
'Yamada san also moved into, on the same day that Tanaka san moved into, different units in the same apartment building.'

This sentence shows that an example like (44) is possible even when the syntactic structure involved does not have conjunctive meaning (at least in Japanese). Sentence (45) involves non-coordinate LNR of the dative noun phrase *onaji apaato no betsu-betsu no heya ni* out of an adjunct clause and out of the clause modified by that adjunct clause. The meaning expressed by the adjunct clause is non-conjunctive, unlike that expressed by the adjunct clause in (44). The adjunct clause in (45) means that the event denoted by the main clause took place on a certain day, and that temporal meaning is clearly part of the truth conditions of the sentence, not any kind of implicature.

A similar example can be constructed in English as well, as shown in (46), an example that I owe to Steve Wechsler (personal communication).

(46) Chris moved into, on the same day that Pat moved out of, the same apartment building, but different units.

These observations are consistent with the theory proposed in Yatabe (2021), which incorporates the non-uniformity claim about apparent internal readings, and are problematic for theories like those proposed in Carlson (1987), Barker (2007), and Kubota and Levine (2020), in which the presence of coordination or plurality is taken to be the source of all apparent internal readings.

4 Concluding remarks

In summary, I have shown the following two things. First, the hypothesis that the reciprocal meaning contained inside the meaning of a word like *different* can be that of Distributed Strong Reciprocity as well as that of Strong Reciprocity serves to expand the empirical coverage of the hidden-reciprocal analysis of internal readings, arguably boosting the plausibility of that analysis. And second, cases of apparent internal readings involving non-coordinate RNR or LNR indicate that not all apparent internal readings involve coordination or plurality, lending support to the non-uniformity claim about apparent internal readings. In order to enhance the credibility of the overall theory, I have also shown how the collective interpretation of reciprocals and the interaction of reciprocals and cumulative interpretation can be accounted for within the theory.

References

Abbott, Barbara (1976). Right node raising as a test for constituenthood. *Linguistic Inquiry*, 7:639–642.

Barker, Chris (2007). Parasitic scope. *Linguistics and Philosophy*, 30:407–444.

Beck, Sigrid (2000). The semantics of *different*: Comparison operator and relational adjective. *Linguistics and Philosophy*, 23:101–139.

Beck, Sigrid (2001). Reciprocals are definites. *Natural Language Semantics*, 9:69–138.

Brasoveanu, Adrian (2011). Sentence-internal *different* as quantifier-internal anaphora. *Linguistics and Philosophy*, 34:93–168.

Carlson, Greg N. (1987). Same and different: Some consequences for syntax and semantics. *Linguistics and Philosophy*, 10:531–565.

Chaves, Rui P. (2014). On the disunity of right-node raising phenomena: Extraposition, ellipsis, and deletion. *Language*, 90(4):834–886.

Dalrymple, Mary, Makoto Kanazawa, Yookyung Kim, Sam Mchombo, and Stanley Peters (1998). Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy*, 21:159–210.

Fox, Danny and Kyle Johnson (2016). QR is restrictor sharing. In *Proceedings of the 33rd West Coast Conference on Formal Linguistics*, pages 1–16, Cascadilla Proceedings Project, Somerville, MA.

Kubota, Yusuke and Robert Levine (2016). The syntax-semantics interface of 'respective' predication: A unified analysis in Hybrid Type-Logical Categorial Grammar. *Natural Language & Linguistic Theory*, 34:911–973.

Kubota, Yusuke and Robert D. Levine (2020). *Type-Logical Syntax*. MIT Press, Cambridge, Massachusetts.

Langendoen, D. Terence (1978). The logic of reciprocity. *Linguistic Inquiry*, 9(2):177–197.

Sauerland, Uli (1998). Plurals, derived predicates, and reciprocals. In *The Interpretive Tract (MIT Working Papers in Linguistics 25)*, pages 177–204.

Schwarzschild, Roger (1996). *Pluralities*. Kluwer, Dordrecht.

Sternefeld, Wolfgang (1998). Reciprocity and cumulative predication. *Natural Language Semantics*, 6:303–337.

Yatabe, Shûichi (2021). A representationalist theory of the semantics of *different*, *the same*, and *each other*. https://ling.auf.net/lingbuzz/006007.

Yatabe, Shuichi (2022). Respectively interpretation and binding conditions A and B. In *Proceedings of the 29th International Conference on Head-Driven Phrase Structure Grammar*, pages 135–155. https://proceedings.hpsg.xyz/article/view/905.

Yatabe, Shûichi and Wai Lok Tam (2021). In defense of an HPSG-based theory of non-constituent coordination: A reply to Kubota and Levine. *Linguistics and Philosophy*, 44:1–77.