

An incremental approach to gapping and conjunction reduction

Petter Haugereid 

Western Norway University of Applied Sciences

Proceedings of the 24th International Conference on
Head-Driven Phrase Structure Grammar

University of Kentucky, Lexington


Stefan Müller (Editor)

2017

Stanford, CA: CSLI Publications

pages 179–198

Keywords: HPSG, Gapping, Conjunction Reduction, Incremental Parsing

Haugereid, Petter. 2017. An incremental approach to gapping and conjunction reduction. In Stefan Müller (ed.), *Proceedings of the 24th International Conference on Head-Driven Phrase Structure Grammar, University of Kentucky, Lexington*, 179–198. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2017.10. 

Abstract

In this paper I present an incremental approach to gapping and conjunction reduction where it is assumed that the first sentence in these constructions is fully parsed before the second sentence with the elided verb is parsed. I will show that the two phenomena can be given a uniform analysis by letting the construction type of the first conjunct be carried over to the second conjunct. This construction type imposes constraints on the arguments that the second conjunct can have. The difference between gapping and conjunction reduction is captured by the already existing constructions for sentence and VP coordination. The analysis is implemented in an HPSG grammar of Norwegian.

1 Introduction

Gapping and conjunction reduction are two out of more phenomena referred to as non-constituent coordination (NCC) in the literature. They pose a challenge to lexicalist approaches given the fact that the main verb of the second conjunct in these constructions is elided. The examples in (1)–(3) are taken from Sag et al. (1985). Example (1) shows the prototypical gapping construction with a transitive sentence in the first conjunct, and two arguments, but no verb, in the second conjunct. Example (2) demonstrates the fact that the gap may consist of a chain of control verbs. Example (3) demonstrates the conjunction reduction construction, where also the subject of the second conjunct is missing.

- (1) Kim likes Sandy, and Lee Leslie.
- (2) Pat wanted to try to go to Berne, and Chris $\left. \begin{array}{l} \text{to try to go to Rome.} \\ \text{to go to Rome.} \\ \text{to Rome.} \end{array} \right\}$
- (3) Kim gave a dollar to Bobbie and a dime to Jean.

1.1 Gapping in Norwegian

Gapping is possible with a range of constructions in Norwegian. In this section, I will present some of the constructions that have been considered in the implementation of the HPSG grammar Norsyg.

The constituents in a gapping construction may be a subject and an adverbial (see (4a)), and the adverbial may also come first, as shown in (4b).

- (4) a. Jeg **kom** i går og du i dag.
I arrived yesterday and you today
'I arrived yesterday and you today.'

[†]I would like to thank three anonymous reviewers and the audience at the HPSG 2017 conference in Lexington, Kentucky, for very useful comments and suggestions. In particular, I would like to thank Mark Steedman for his comments on an early version of this paper.

- b. I går **kom** jeg og i dag du.
 In yesterday arrived I and today you
 ‘Yesterday, I arrived, and today, you.’

In (5), two elements are gapped, the finite verb *tar* (‘takes’) and the particle *with* (‘med’). The particle cannot appear in the gapping construction.

- (5) Jeg **tar med** mat, og du (*med) drikke.
 I bring with food and you with drink
 ‘I will bring food, and you drinks.’

In (6), the reflexive verb *ønske seg* ‘wish for’ is gapped. The reflexive cannot appear in the gapping construction.

- (6) Jeg **ønsker meg** fisk, og du (*deg) steik.
 I wish REFL fish and you REFL roast
 ‘I want fish, and you roast.’

In (7), the reflexive particle verb *se seg ut* ‘pick out’ is gapped. Neither the reflexive nor the particle can appear in the gapping construction.

- (7) Jeg **ser meg ut** en fisk og du (*deg) (*ut) en steik.
 I see REFL out a fish and you REFL out a roast
 ‘I pick out a fish and you a roast.’

In transitive idiomatic expressions, all the idiomatic words are elided in the second conjunct (see (8)). It is not possible to elide just parts of the idiom.

- (8) Jeg **brakte på bane** isen, og du (*på) (*bane) sjokoladen.
 I brought on track ice-DEF and you on track chocolate-DEF
 ‘I brought up the ice cream, and you the chocolate.’

Verbs with selected prepositions, however, behave slightly different. If a verb has a selected preposition, the gapping construction is very odd if it does not have the preposition, as shown in (9a). However, when the gapping construction contains the selected preposition, as in (9b), it is much better.

- (9) a. ?? Jeg **hører på** Jon, og du Marit.
 I listen to Jon and you Marit
 ‘I listen to Jon, and you Marit.’
 b. Jeg **hører på** Jon, og du på Marit.
 I listen to Jon and you to Marit
 ‘I listen to Jon, and you (listen) to Marit.’

It is possible to have gapping with ditransitive verbs, as shown in (10a). We then get three constituents in the second conjunct. It is also possible to have two arguments and an adverb in a gapping construction, as shown in (10b).

- (10) a. Per **serverte** meg fisk, og Kari deg steik.
 Per served me fish and Kari you roast
 ‘Per served me fish, and Kari you roast.’
- b. Jeg **spiste** fisk i går og du steik i dag.
 I ate fish yesterday and you roast today
 ‘I ate fish yesterday and you (ate) roast today.’

We can also have gapping when a verb is passivized, as shown in (11a). Then both the passive auxiliary and the main verb are elided in the second conjunct. If there is an expletive pronoun, this is also elided, as shown in (11b).

- (11) a. Jeg **ble servert** fisk, og du steik.
 I was served fish and you roast
 ‘I was served fish, and you roast.’
- b. I går **ble det servert** fisk og i dag steik.
 yesterday was it served fish and today roast
 ‘Yesterday, fish was served, and today roast (was served).’

1.2 Conjunction reduction in Norwegian

The examples we have looked at so far have been examples of gapping in sentence coordinations. (12a) and (12b) illustrate that it is possible to have gapping in cases where the topic is shared. In (12a) the two conjuncts share the subject. In the literature this is called conjunction reduction. I argue that (10a) and (12a) are examples of the same phenomenon, only that in (10a), we have sentence coordination and in (12b), we have coordination of sentences with a shared topic. As with other coordinations where the topic is shared, (12b) shows that it is also possible to let an adjunct be shared in gapping constructions (*i går* ‘yesterday’). I will show in Section 3.4 that no extra machinery is needed in order to account for gapping in coordinations where the topic is shared once the rules for vp coordination (or coordination with a shared topic) and gapping are in place.

- (12) a. Per **serverte** meg fisk og deg steik.
 Per served me fish and you roast
 ‘Per served me fish, and you roast.’
- b. I dag **ble jeg servert** fisk og du steik.
 Today was I served fish and you roast
 ‘Today I was served fish, and you roast.’

strained in order not to let the grammar create unwanted or unnecessary constituents.

2.2 HPSG

In the HPSG theory, Immediate Dominance schemata allows a grammar writer to specify constraints on a phrase and its immediate daughters without specifying the order of the daughters (Pollard & Sag, 1994). This makes it possible to account for free word order phenomena, but it is restricted to the immediate daughters of a phrase. In order to handle phenomena where the constituents involved are not immediate daughters of the same phrase, like discontinuous constituents and non-constituent coordination, the feature `DOM(ain)` has been introduced, where the linear order of the phonological items that a phrase consists of, is represented (Reape, 1994). The elements on the `DOM` list may be arranged in an order that is not reflected in the derivation tree. This separation of the order of phonological items from the constituent structure is referred to as linearization. Most approaches to non-constituent coordination makes use of the linearization approach (Kathol, 1995; Beavers & Sag, 2004; Chaves, 2005; Crysmann, 2008). The use of `DOM` to handle linearization phenomena is powerful, and although relational constraints may be added to the grammar in order to impose restrictions on the order of the phonological items, it may put a heavy burden on the parser if it is not properly constrained.

The distinction between phonological representation and constituent structure assumed in the linearization approach is not available in grammars written within the DELPH-IN network, like the `ERG` (Flickinger, 2000) and `JACY` (Siegel et al., 2016). These grammars use regular phrase structure rules where the phonology is simply concatenated, and constituents are reflected in the derivation tree. This is efficient, but it poses a challenge to phenomena like non-constituent coordination since the valence information of the verb in the first conjunct is not accessible at the point where the coordination happens (the valence requirements have been canceled off), and even if they were, there is no dummy verb in the second conjunct that can get these requirements.

3 An incremental approach

In this section, I will present an alternative, incremental approach, which makes use of regular phrase structure rules, like the DELPH-IN grammars just discussed, but which has in common with the linearization approach that the derivation tree is separated from the constituent structure (although in a different way). The constituent structure is reflected by the entering and popping of structure onto a `STACK` (Haugereid & Morey, 2012).

3.1 Incremental parsing and constituent structure

The approach assumes that parsing is done incrementally, that is, word by word. The parse tree of the transitive sentence in (13) is given in Figure 3.

- (13) Gutten spiser fisk.
boy-DEF eats fish
'The boy eats fish.'

The tree consists of unary and binary branching trees, and it is completely left-branching. In the bottom left corner is a start symbol, and all the words of the sentence attach to this symbol from the right, one by one. At the top of the tree is a unary *force* rule.

The grammar mainly has three types of rules:

1. Embedding structures – rules that initiate the processing a constituent
2. Adjunction structures – rules that continue the processing of an initiated constituent
3. Popping structures – rules that end the processing of a constituent

The tree in Figure 3 has two embedding structures, one for the subject *gutten* and one for the object *fisk*. The embedding structures put the parsing of the matrix constituent (the main clause) on hold while the embedded constituents (the NPs) are parsed. This is done by means of a feature *STACK*. An element with selected features of the matrix constituent (in Figure 3 represented by the *HEAD* feature) is added to the *STACK* whenever a new phrasal constituent is initiated. Since the NPs in Figure 3 only consist of one word, popping rules apply directly after the embedding rules, retrieving the features from the *STACK*. The rule that attaches the verb *spiser* ('eats') is an example of an *adjunction structure*. These rules attach words to the current constituent.

A standard assumption in Scandinavian syntax since Diderichsen (1946) is that the constituent appearing before the finite verb in a main clause is topicalized. This also holds if the constituent is the subject. In the incremental approach, extraction is done by means of a unary extraction rule, which enters a feature structure on the *SLASH* list of its daughter, and a filler rule, which realizes the element on the *SLASH* list. The extraction rule applies in the "canonical position" of the constituent, and in a main clause, the canonical position of the subject is the position after the finite verb. This is illustrated in Figure 3, where the rule *extr-arg1-struct* enters a feature structure onto the *SLASH* list of the daughter. The filler rule *embedding-filler-struct* unifies the feature structure on the *SLASH* list with the second daughter *gutten*. Since it is an embedding rule, selected features of the filled-in constituent (here represented by the *HEAD* feature) are unified with those of the mother.

The constituent tree in Figure 4 is derived from the parse tree in Figure 3. Here we can see that there are two embedded structures (the two NPs) and that the verb is not embedded in any phrase.

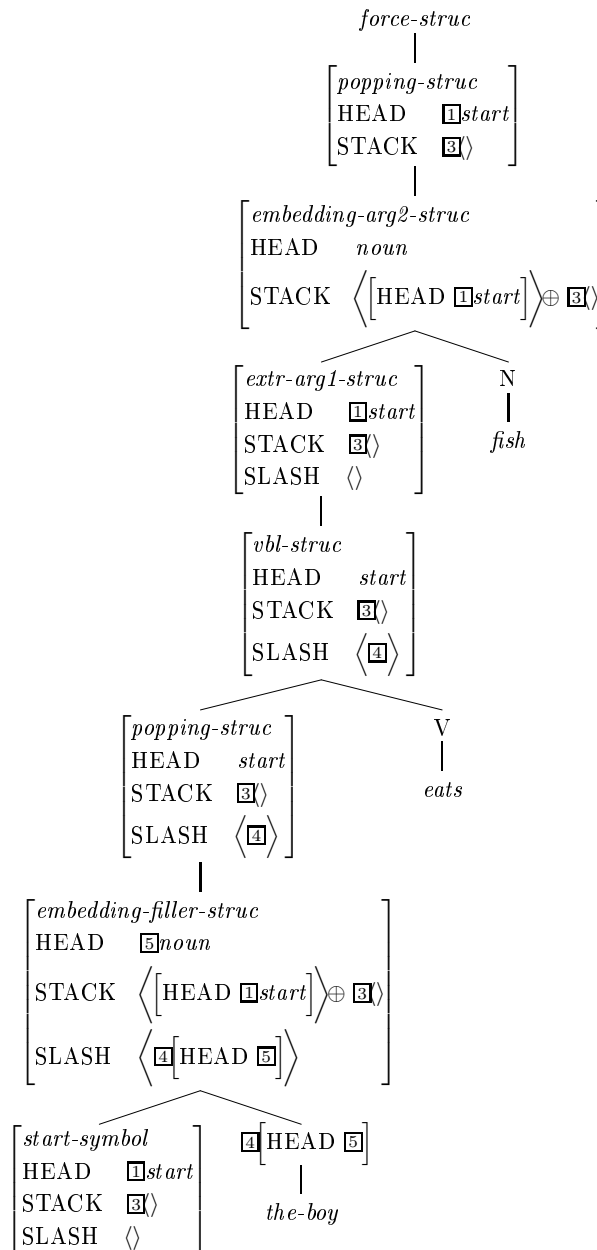


Figure 3: Parse tree of a transitive sentence

3.2 A constructionalist approach to argument structure

As mentioned, the derivation tree in the incremental approach is assumed to consist of binary and unary phrase structure rules where the binary rules have a word as their second daughter. A simplified representation of the transitive clause in Figure 3 is given in Figure 5.

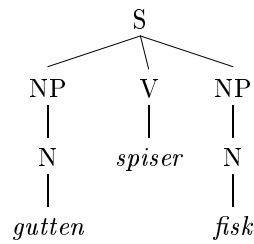


Figure 4: Constituent tree of a transitive sentence

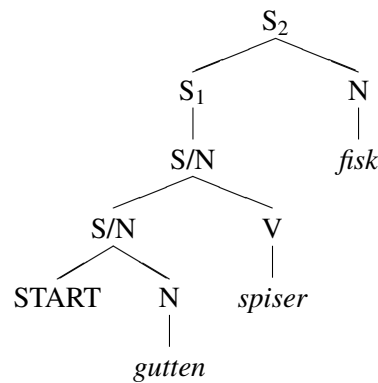


Figure 5: Simplified derivation tree of the transitive sentence *Gutten spiser fisk* ('The boy eats fish')

There are two things worth mentioning in connection to the tree in Figure 5. First, it is assumed that the topic of a main clause is extracted from its canonical position, as shown in Section 3.1.³

Second, valence requirements are handled by means of types (Haugereid, 2009, 2015). The verb *spiser* ('eats') is listed in the lexicon with an underspecified construction type *spise_prd*. This type is a part of a hierarchy of valence types that constrain which constellations of arguments the verb is allowed to appear with. A small part of this type hierarchy is illustrated in Figure 6. It shows that a verb with the type *spise_prd* is compatible with a transitive frame (*_spise_12_rel*) and an intransitive frame (*_spise_1_rel*).

The phrase at the top of the derivation of a clause is constrained to have negative valence types (*arg1-*, *arg2-*, *arg3-*, and *arg4-*), as shown in (14). As one goes down the tree, valence rules switch these types from negative in the mother to positive in the (first) daughter. This is spelled out in Figure 7. The types *arg1-* and *arg2-* are switched to *arg1+* and *arg2+* by the two valence rules (*embedding-arg2-struct* and *extr-arg1-struct*). At the bottom of the tree, the construction type of the verb, which is the PRED value of its KEYREL (*spise_prd*) is unified with the four

³Note that the use of '/' in the tree shows that there is an extracted element, and it must not be confused with the use of the slashes ('/' and '\') in CCG, even though the meaning is related.

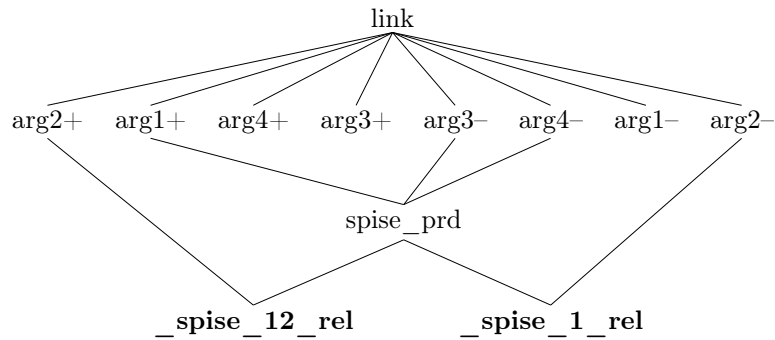


Figure 6: Type hierarchy of valence types

valence types $arg1+$, $arg2+$, $arg3-$, and $arg4-$, as shown in (15). This unification is allowed by the type hierarchy (the types $spise_prd$, $arg1+$, $arg2+$, $arg3-$, and $arg4-$ have a common subtype), and yields the construction type $_spise_12_rel$, which also serves as the predicate of the relation introduced by the verb.

$$(14) \left[\begin{array}{l} \text{VAL} \\ \text{CMP1|LINK } arg1- \\ \text{CMP2|LINK } arg2- \\ \text{CMP3|LINK } arg3- \\ \text{CMP4|LINK } arg4- \end{array} \right]$$

$$(15) \left[\begin{array}{l} \text{START} \\ \text{VAL} \\ \text{KEYREL|PRED } 0 \\ \text{SLASH } \langle \rangle \end{array} \left[\begin{array}{l} \text{CMP1|LINK } 0 \\ \text{CMP2|LINK } 0 \\ \text{CMP3|LINK } 0 \\ \text{CMP4|LINK } 0 \end{array} \right] \right]$$

3.3 Incremental parsing and coordination

An obvious challenge for the incremental approach is coordination. In HPSG, coordination of full constituents is straightforward, at least as long as the constituents are of the same category. It is the regular coordination rule that holds: $XP \Rightarrow XP \text{ Conj } XP$. Whether the coordinated constituent is a sentence, a VP or an NP, coordination is assumed to be captured by the same rule. In an incremental approach, however, one is forced to start building the second constituent on top of the first, as shown in Figure 8. This means that the rules involved in coordination of full constituents no longer is the combination of two equal constituents. Rather, they mark the end of one constituent and the beginning of a new constituent. This is illustrated in Figure 8 where the rule that adds the coordinator, also marks the be-

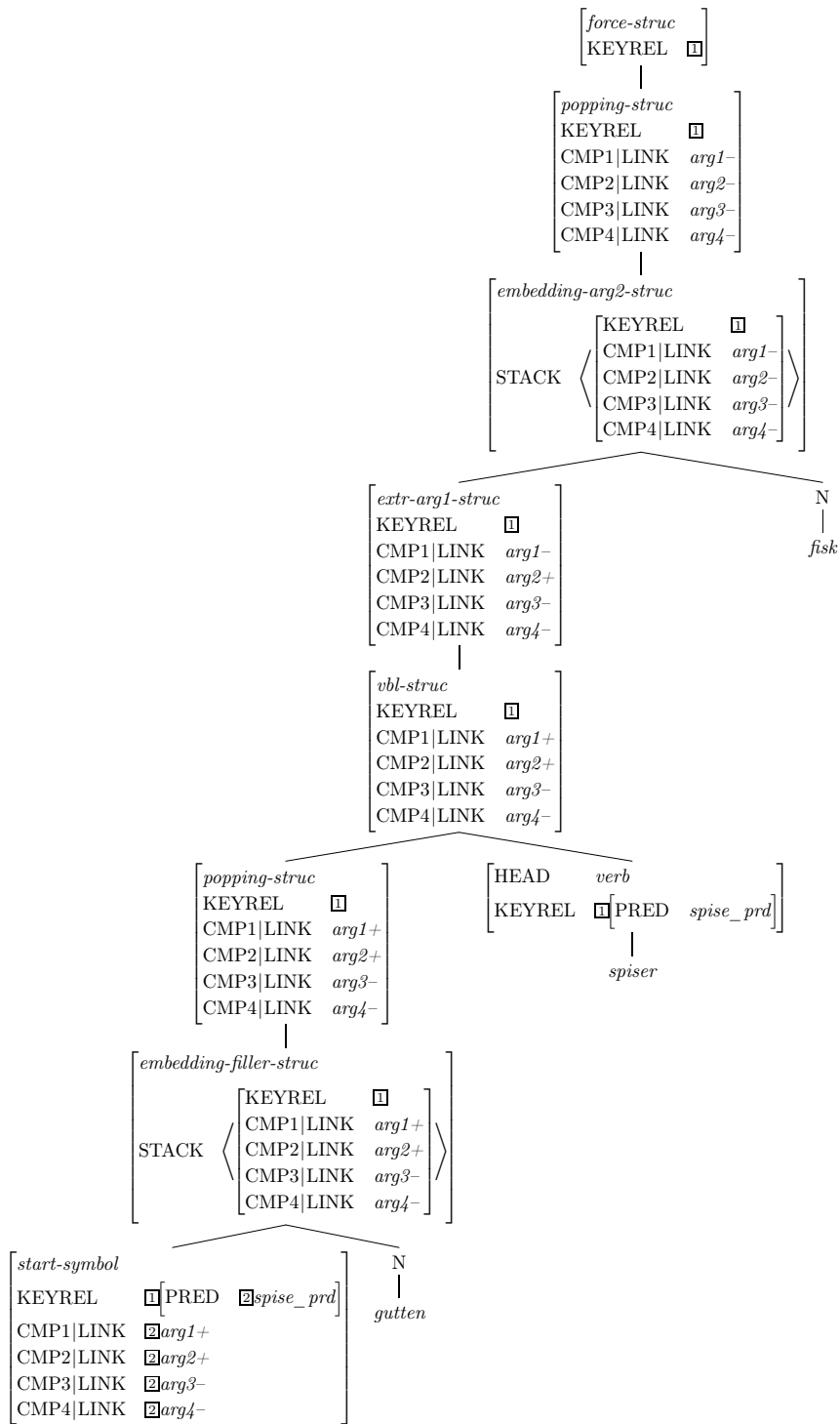


Figure 7: Valence types of a transitive sentence

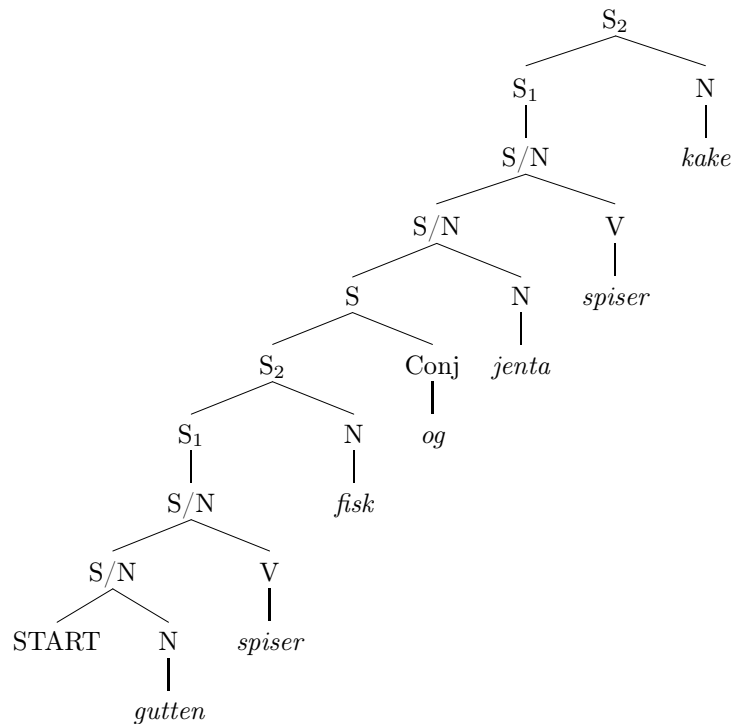


Figure 8: Derivation tree of two coordinated transitive sentences

ginning of a new clause. In practice, the mother of the coordinator has the same constraints as *START* (see (15)).

In order to account for coordination of main clauses and VPs, the hierarchy of phrase types in Figure 9 is created. Most of the constraints of the two types of coordination rules (*conj-s-struct* and *conj-top-struct*) are captured in a supertype *conj-struct*. *conj-struct* takes as its first daughter a structure that has realized all its arguments, that is, the valence types are all negative. The second daughter is a conjunction item, which can be either a conjunction or a comma, in case there are more than two conjuncts. The mother unifies the valence types of the sentence that is to be built next. The type links the two conjuncts with a conjunction relation that is entered onto the C(onstructional)-CONT RELS list. *conj-struct* is underspecified with regard to whether there is an element on the SLASH list or not.

The value of the SLASH list is specified on the two subtypes, *conj-s-struct* and *conj-top-struct*. The type *conj-s-struct* has an empty SLASH list. This means that it has the same status as *START* (see 15), and it initiates the building of a new sentence.

The second subtype, *conj-top-struct*, has an element on the SLASH list which is the topic. This gives it the status of a structure where the topic is realized, but where it is yet to be extracted. The topic of the first daughter is also the topic of the mother, which means that the two sentences will share topic. This accounts for

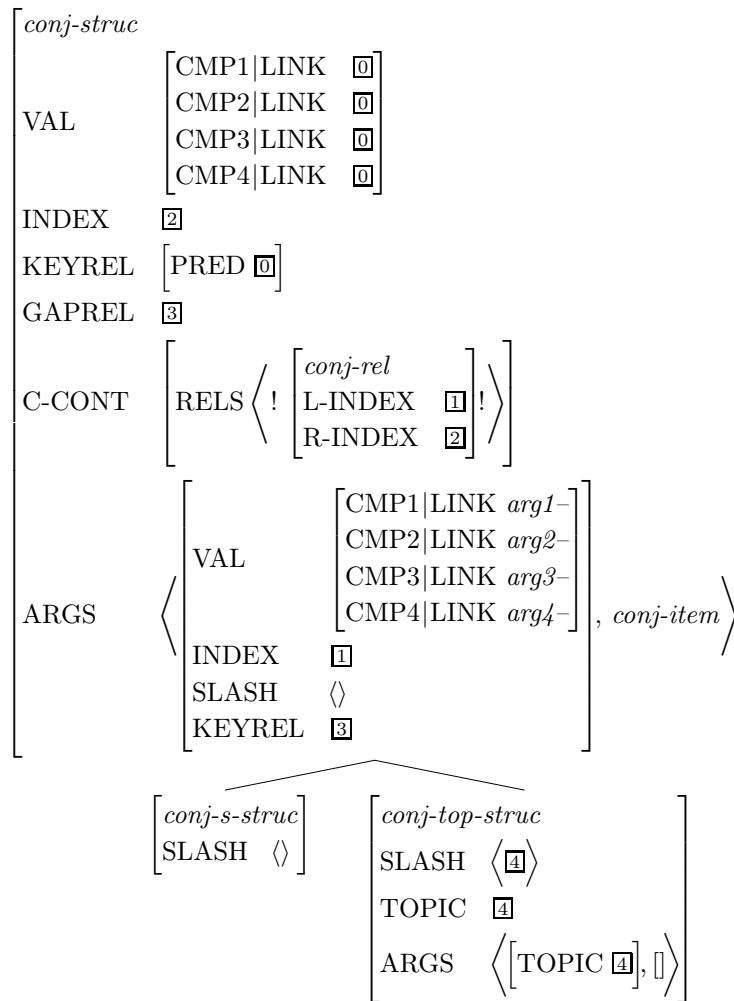


Figure 9: Hierarchy of coordination rules

VP coordination, where the shared topic is the subject, but also similar kinds of coordination where the shared topic is an object or an adjunct.

3.4 Analysis of gapping

In order to account for the gapping phenomena presented in Section 1, I introduce a set of unary rules corresponding to the rules that attach verbs, particles, reflexives and idiomatic words. The rule for eliding verbs is given in Figure 10. It takes as its only daughter a structure that requires a verb (the VBL value is *synsem*), and gives a new structure where there is no longer a verb requirement (the VBL value is *anti-synsem*). In addition, the value of GAPREL of the daughter is unified with the KEYREL. As shown in Figure 9, the GAPREL has as value the KEYREL relation of the first conjunct in a conjunction. This relation is the relation contributed by the

main verb. This ensures that the gapping construction has the same relation (and syntactic construction) as the first conjunct.

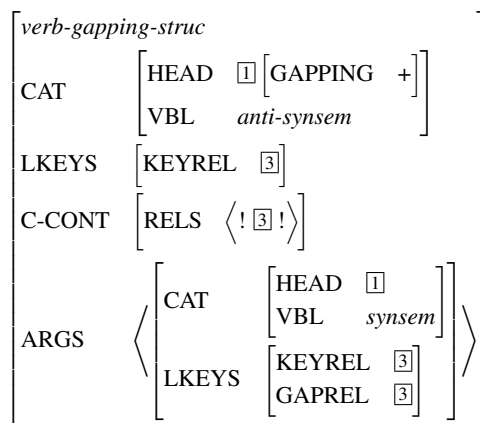


Figure 10: Type for elided verbs

The incremental design where verbs are treated as a kind of obligatory adjuncts, makes an account of gapping constructions relatively straightforward. Since the contribution of a (main) verb in a regular main clause is to contribute a type which constrains what kinds of constructions it can appear in, the only addition needed is to make the construction type available in the gapping construction. As shown in Figure 10, this type comes from the GAPREL feature. In this way the gapping rule substitutes the verb. The construction type carried over from the first conjunct guarantees that the valence rules that apply in the first conjunct also apply in the second conjunct.

Figure 11 shows how a gapped conjunct is analyzed. The rule *conj-s-struct* unifies the KEYREL value of its first daughter with its GAPREL value. Further up the tree, the unary rule *verb-gapping-struct* unifies the GAPREL value with the KEYREL. In this way, the construction type of the first conjunct also becomes the construction type of the second conjunct.⁴

The implemented grammar produces the MRS given in Figure 13 for a sentence with a gapping construction.^{5,6} Note that the predicate *_spise_12* appears twice. This is a result of the unification of the construction type of the first conjunct with the construction type of the second conjunct.

3.5 Analysis of conjunction reduction

The examples of conjunction reduction (see (12a) and (12b)) are accounted for by the combination of the *conj-top-struct* rule and the *verb-gapping-struct* rule. The

⁴The embedding and popping rules as well as the features CMP3 and CMP4 are in Figure 11 omitted for expository reasons.

⁵The MRS display is made by Michael Goodman: <https://github.com/goodmami/demophin>.

⁶The grammar currently labels the last event as the top relation. This should rather be the conjunction relation *_og_c*, and will be fixed in a future version of the grammar.

produces the MRS given in Figure 14 for a sentence with a conjunction reduction construction.

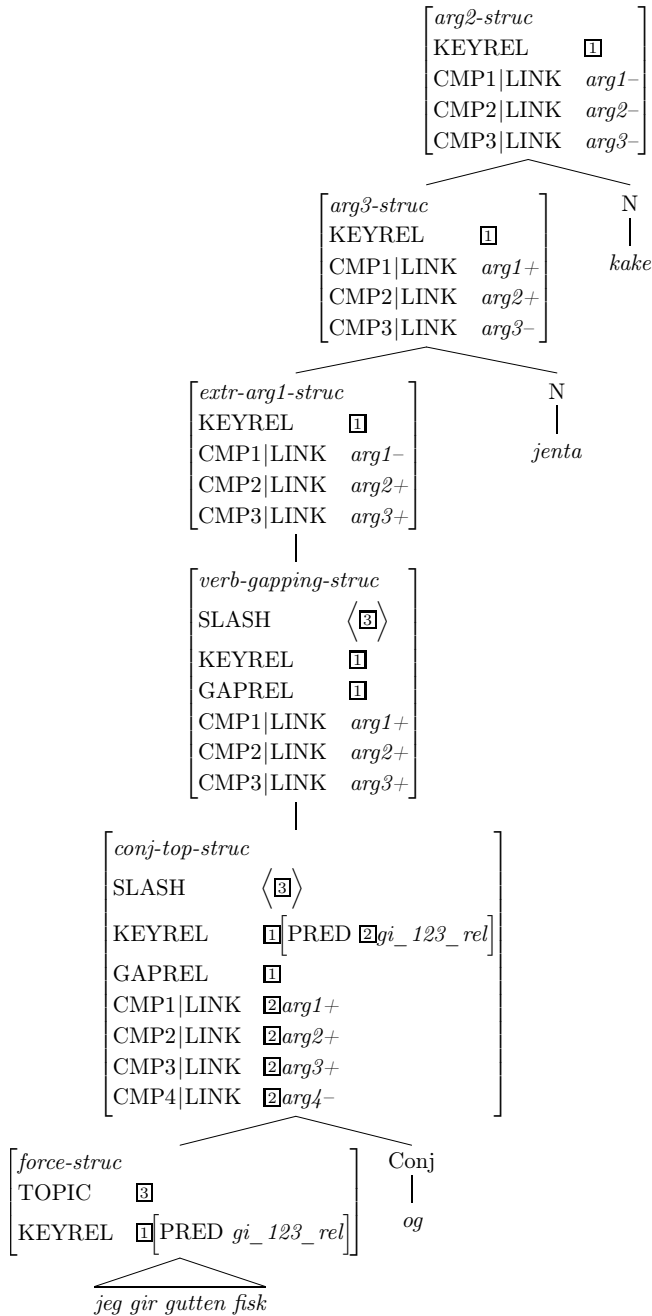


Figure 12: Analysis of conjunction reduction in *Jeg gir gutten fisk, og jenta kake* 'I give the boy fish, and the girl cake.'

4 Discussion

4.1 Forward gapping and backward gapping

This paper has presented an analysis of gapping and conjunction reduction for Norwegian, which is an SVO language, like English. According to Ross (1970), gapping operates forward in SVO languages (see (16)). However, in Japanese, which is a VSO language, the verb appears in the last conjunct in gapping constructions, and not the first (Ross, 1970). This phenomenon is referred to as backward gapping (see (17)).

- (16) a. SVO + SVO + SVO + ... + SVO \Rightarrow
b. SVO + SO + SO + ... + SO
- (17) a. SOV + SOV + SOV + ... + SOV \Rightarrow
b. SO + SO + SO + ... + SOV

The grammar presented in this paper is designed in such a way that a clause in principle can be parsed without a verb. The argument structure is assumed to originate from the syntactic rules, and the verb is treated as a kind of obligatory modifier. If there is no verb, the parse will result in an underspecified type which only reflects the argument structure of the clause, but not the predicate of the main verb. (Not having a verb will of course increase the search space, but it will be manageable. Still, it should probably be combined with some kind of statistical "guesser".)

I would assume for a head final language like Japanese, that the mechanism I describe in Section 3.2 would be "turned around", so that the unification of the valence types and the predicate of the main verb would happen at the top of the tree, rather than at the bottom. And the gapping rule would get its constraints from the opposite direction, from "above" rather than "below". This would account for backward gapping.

It would be possible to design the grammar in such a way that it was just like English, with SVO word order, but with backward gapping. However, if I assume that the mechanism involving valence types and the passing of the predicate type is reversed, it would exclude backward gapping for a language like English.

4.2 Coverage

The analysis of gapping and conjunction reduction presented in this paper is far from exhaustive. However, it accounts for a number of challenging phenomena that proves hard to account for within standard lexicalist approaches without resorting to powerful mechanisms that may lead to drastic decreases in parser efficiency.

All the sentences in Sections 1.1 and 1.2 are accounted for, and most of the analyses have been implemented in the Norwegian HPSG grammar Norsyng. This

includes analyses of sentences with transitive and ditransitive verbs (see (10)), particle verbs (see (5)), verbs with selected prepositions (see (9)), sentences with non-subject topics and passive sentences (see (11)). Also analysis of gapping with multiple conjuncts, like in *John ate fish, Mary beef, and Sandy chicken* is implemented. There are ongoing experiments to also include analyses of sentences with reflexive verbs (see (6) and (7)) and VP idioms (see (8)). Some preliminary tests have been done to check the impact that the inclusion of the analysis has on parser efficiency. A test on 333 test sentences shows an increase of processing effort of 46%. The increase was mainly due to one sentence.

References

- Beavers, John & Ivan A. Sag. 2004. Coordinate ellipsis and apparent non-constituent coordination. In Stefan Müller (ed.), *Proceedings of the hpsg-2004 conference, center for computational linguistics, katholieke universiteit leuven*, 48–69. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/HPSG/5/>.
- Chaves, Rui Pedro. 2005. A linearization-based approach to gapping. In G. Jäger, P. Monachesi, G. Penn & S. Wintner (eds.), *Fg-mol 2005: The 10th conference on formal grammar and the 9th meeting on mathematics of language*, 14. CSLI. <http://errepe.no.sapo.pt/glin.pdf>.
- Crysmann, Berthold. 2008. An asymmetric theory of peripheral sharing in HPSG: Conjunction reduction and coordination of unlikes. In Gerald Penn (ed.), *Proceedings of fgvienna : The 8th conference on formal grammar, aug 16–17 2003, vienna*, 45–64. Stanford: CSLI publications. <http://csli-publications.stanford.edu/FG/2003/crysmann.pdf>.
- Diderichsen, Paul. 1946. *Elementær dansk grammatik*. København: Gyldendal.
- Flickinger, Daniel P. 2000. On building a more efficient grammar by exploiting types. *Natural Language Engineering* 6(1). 15–28.
- Haugereid, Petter. 2009. *Phrasal subconstructions: A constructionalist grammar design, exemplified with norwegian and english*: Norwegian University of Science and Technology dissertation. <http://urn.kb.se/resolve?urn=urn:nbn:no:ntnu:diva-5755>.
- Haugereid, Petter. 2015. Unique lexical entries in a subconstructional grammar. In Stefan Müller (ed.), *Proceedings of the 22nd international conference on head-driven phrase structure grammar, nanyang technological university (ntu), singapore*, 110–121. Stanford, CA: CSLI Publications. <http://csli-publications.stanford.edu/HPSG/2015/haugereid.pdf>.

- Haugereid, Petter & Mathieu Morey. 2012. A left-branching grammar design for incremental parsing. In Stefan Müller (ed.), *Proceedings of the 19th international conference on head-driven phrase structure grammar, chungnam national university daejeon*, 181–194. <http://csli-publications.stanford.edu/HPSG/2012/haugereid-morey.pdf>.
- Kathol, Andreas. 1995. *Linearization-based German syntax*: Ohio State University dissertation.
- Pollard, Carl J. & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.
- Reape, Mike. 1994. Domain union and word order variation in German. In John Nerbonne, Klaus Netter & Carl J. Pollard (eds.), *German in Head-Driven Phrase Structure Grammar* (CSLI Lecture Notes 46), 151–197. Stanford University: CSLI Publications.
- Ross, John Robert. 1970. Gapping and the order of constituents. In M. Bierwisch & K. Heidolph (eds.), *Progress in linguistics*, 249–259. The Hague: Mouton.
- Sag, Ivan A, Gerald Gazdar, Thomas Wasow & Steven Weisler. 1985. Coordination and how to distinguish categories. *Natural Language & Linguistic Theory* 3(2), 117–171.
- Siegel, Melanie, Emily M. Bender & Francis Bond. 2016. *Jacy: An implemented grammar of Japanese* CSLI Studies in Computational Linguistics. Stanford University: CSLI Publications.
- Steedman, Mark. 2000. *The syntactic process*. Cambridge, London: MIT Press.
- Steedman, Mark. 2017. Gapping and information structure. Manuscript.