VP fronting in Norwegian

Petter Haugereid
Western Norway University of Applied Sciences
petter.haugereid@hvl.no

Abstract

This paper presents an account of fronted VPs in Norwegian and their implementation in an incremental typed feature structure grammar. The analysis covers sentences such as *I've promised to help them [and help them I* WILL*]* (Huddleston & Pullum, 2002, 1376), where the VP *help them* is fronted while the auxiliary WILL remains in its canonical position. I show how a constructional approach captures filler–gap dependencies, the role of the dummy verb *gjøre* 'do', and the identity of semantic representations for fronted and non-fronted variants, thereby avoiding the need for multiple lexical entries or non-inflecting lexical rules.

1 Introduction

As shown in Faarlund et al. (1997, 876–877), verb phrases in Norwegian can be topicalized (see (1a)). The complements and adjuncts of the verb phrase are typically topicalized together with the main verb, as illustrated in (1b), where the direct object *pengene* 'the money' and the adjunct *tilbake* 'back' are topicalized along with the main verb *sende* 'send'.[1]

(1) a. Syngje kan eg ikkje
sing kan I not
Sing, I can not

b. Sende pengene tilbake vil eg ikkje
send money-DEF back will I not
Send the money back, I will not

Müller & Ørsnes (2013) provide an HPSG analysis of VP fronting in Danish by means of an empty VP category that gets its valence features from the topicalized VP, where the elements on the COMPS list are realized, but the SPR list is non-empty. The empty VP category combines with the element(s) on the SPR list in the same way as other verb projections. Their analysis also accounts for partial VP fronting. In this paper I will give an alternative analysis without empty categories, but with a VP extraction rule that dominates the topicalized VP.

1.1 Analysis in LFG

An analysis of VP fronting is implemented in the Norwegian LFG grammar NorGram (Dyvik, 2000). In NorGram, a finite verb in the canonical position (a modal auxiliary or the dummy verb *gjøre* 'do') is analyzed as the matrix predicate, and the fronted VP as a predicate complement (XCOMP) of of this verb. The SUBJ of the finite verb in the canonical position is identical to the SUBJ of the fronted VP, as in a raising construction. This is illustrated for the minimal

[1] The examples in (1) are taken from Faarlund et al. (1997, 876), and the examples in (3) – (5) as well as example (i) in Footnote 6 are from the NorGramBank corpus.

example (2) in Figure 1.

(2) Spise gjør jeg.
eat-INF do-PRES I
Eat, I do.

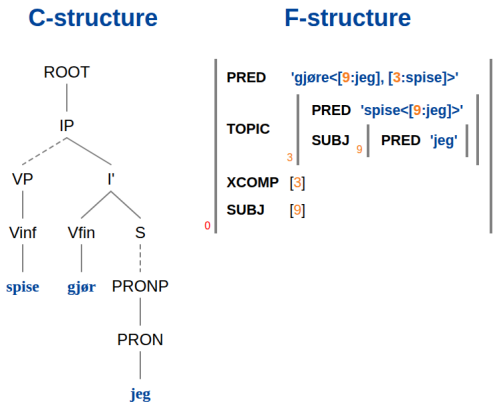

Figure 1: LFG analysis of *Spise gjør jeg* 'Eat, I do'

1.2 Search results from NorGramBank

The NorGram analysis facilitates a search for the construction in the NorGramBank corpus (Rosén et al., 2012; Dyvik et al., 2016). A search for VP topics yields 1,604 hits, and a manual inspection of the first 100 shows that about half are true examples of fronted VPs (46 out of 100). Assuming that this proportion holds for the full set of hits gives roughly 800 examples of fronted VPs in a corpus of 13,235,938 sentences, corresponding to about one fronted VP per 17,000 sentences.

In the examples shown in (1), the finite verb is an auxiliary. The auxiliary is never topicalized with the main verb. When the main verb is the finite verb, the dummy verb *gjøre* 'do' is used, as illustrated in (3).

(3) a. Hører dårlig gjorde han også.
hears bad did he also
He also heard bad.

b. Men tenke gjorde jeg.
but think did I
But think I did.

A search in NorGramBank for fronted VPs where the dummy verb 'gjøre' acts as the verb of the matrix clause, gives 204 matching sentences. A manual inspection reveals that 116 of the sentences are actual examples of fronted VPs. 50 of these sentences have a finite verb in the topic, as shown in (3a) where the verb *hører* has present tense. 66 of the sentences have an infinitive verb form in the topic, illustrated by (3b) where the verb *tenke* 'think' is in the infinitive.

A search in the corpus for fronted VPs topicalized from a subordinate clause or an infinitival clause yields the sentences in (4). In (4a) the topic *overnattet* 'slept over' is extracted from a subordinate clause, and in (4b) the topic *forlate ham* 'leave him' is extracted from an embedded infinitival clause.

(4) a. Overnattet hendte det også at han
slept-over happened it also that he
gjorde
did
It also happened that he slept over

b. Forlate ham kom hun ikke til å gjøre.
Leave him came she not to to do
She would not leave him.

While the complements of the main verb are topicalized together with the main verb, adverbials, on the other hand, often remain in the canonical VP position, as illustrated in (5), where the adjunct *dag og natt* 'day and night' appears adjacent to *gjøre* 'do'.

(5) Kaste terninger kan en gjøre dag og natt
throw dices can one do day and night
One can roll dice day and night

2 An incremental TFSG grammar

The VP fronting construction is implemented in an HPSG-inspired, incremental, typed feature structure grammar for Norwegian (Haugereid, 2009), using the LKB system (Copestake, 2002) as part of the Delph-In effort.[2] There are two fundamental aspects of the grammar that make possible the analysis presented in Section 3.

Firstly, the grammar assumes a distinction between a parse tree and a constituent tree (Haugereid & Morey, 2012). In the parse tree, utterances are parsed incrementally in a bottom-up, left-to-right fashion. During parsing, the constituent structure is tracked by means of a feature STACK. In this way, the grammar resembles a shift-reduce parser. This is illustrated later in Figure 3.

Secondly, the grammar assumes a hierarchy of (sub)construction types (Haugereid, 2007). Verbs, prepositions, particles, and sometimes nouns and adjectives contribute subconstruction types that, together with subconstruction types contributed by the syntactic rules, make up the construction type of a sentence. This construction type is also the semantic predicate of the clause. This is illustrated by the type hierarchy in Figure 2, where the upper types *arg1+*, *arg1–*, *arg2+*, *arg2–*, and so on indicate whether an argument-realizing syntactic rule has applied,[3] and the lower types *_spise_1_rel* and *_spise_12_rel* are construction types (unifications of syntactic linking types and the underspecified predicate of the verb) that also function as the predicate of the clause. In this way, a verb can be underspecified with regard to which syntactic frame it appears in, and there is no need for multiple lexical entries or valence changing rules in order to account for valence alternations.

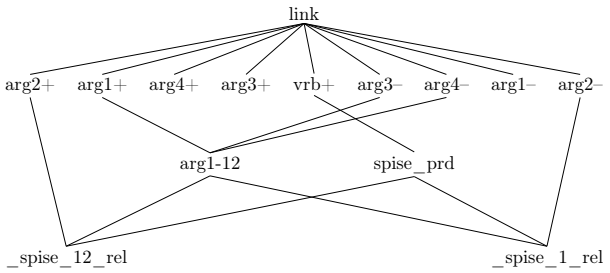

Figure 2: Hierarchy of subconstruction types

In addition to valence alternations, the constructional approach to argument structure is motivated by a range of phenomena, such as scrambling, gapping, multi-word expressions, and head-final languages. It implies that the lexical items, including verbs, do not introduce a relation; rather, it is the syntactic construction that introduces the relations. The lexical entry for the verb *spise* 'eat', for example, is given in (6). As can be seen, the RELS list is empty, and the KEYREL|PRED value is the underspecified *spise_prd* type.[4]

(6)
$$\begin{bmatrix} \textit{verb-lxm} & \\ \text{STEM} & \langle \text{“spise”} \rangle \\ \text{HEAD} & \textit{verb} \\ \text{KEYREL} & \left[\text{PRED} \quad \textit{spise\_prd}\right] \\ \text{RELS} & \langle !! \rangle \end{bmatrix}$$

[2] https://github.com/delph-in/docs/wiki

[3] The grammar has valence rules for four different kinds of arguments; CMP1 arguments (agents/experiencers), CMP2 arguments (themes/patients), CMP3 arguments (beneficiaries) and CMP4 arguments (obliques).

[4] The RELS list is empty in order to allow for verbs, that otherwise would fuction as main verbs, to function as dummy verbs or light verbs, where their semantic contribution is different from regular main verbs. If the verb is realized as a main verb, the relation will be entered by the construction that attaches the verb (see the rule that attaches *spise* in Figure 3).

3 Analysis of VP fronting

3.1 Syntax

There are three additions made to the grammar in order to account for fronted VPs: a VP filler rule (entering the fronted VP), a VP popping rule (exiting the fronted VP), and a VP extraction rule (see Figure 3).

Given that the grammar parses a sentence incrementally, bottom-up, the filler–gap dependency is the opposite of the mechanism assumed in HPSG (Borsley & Crysmann, 2024), where the extraction site is dominated by the filler rule. In the present approach, the filler rule applies first, at the bottom left of the parse tree, and then the SLASH feature percolates up the tree with the relevant syntactic and semantic information until it reaches the extraction site, where the constituent is linked by an extraction rule.

The use of the features STACK and SLASH in the analysis in Figure 3 can be tracked by following the value of the VAL feature of the *start-symbol* (2) up the tree until the *arg1-rule* alters its COMP1|LINK value in the mother node. The altered VAL (5) is entered on the SLASH by the extraction rule, thereby ending up as the VAL value of the filler rule.

At the bottom of the tree, the subconstruction types *arg1+*, *arg2–*, and *spise_prd* (as well as *arg3–* and *arg4–*) are unified.[5] The type hierarchy in Figure 2 shows that when these subconstruction types are unified, they yield the construction type *_spise_1_rel*, which becomes the semantic predicate of the clause.

3.2 Semantics

The analysis in Figure 3 gives the MRS (Copestake et al., 2005) in (7). When the grammar generates from this MRS, all the sentences in (8) are produced, since their semantic representations are identical.

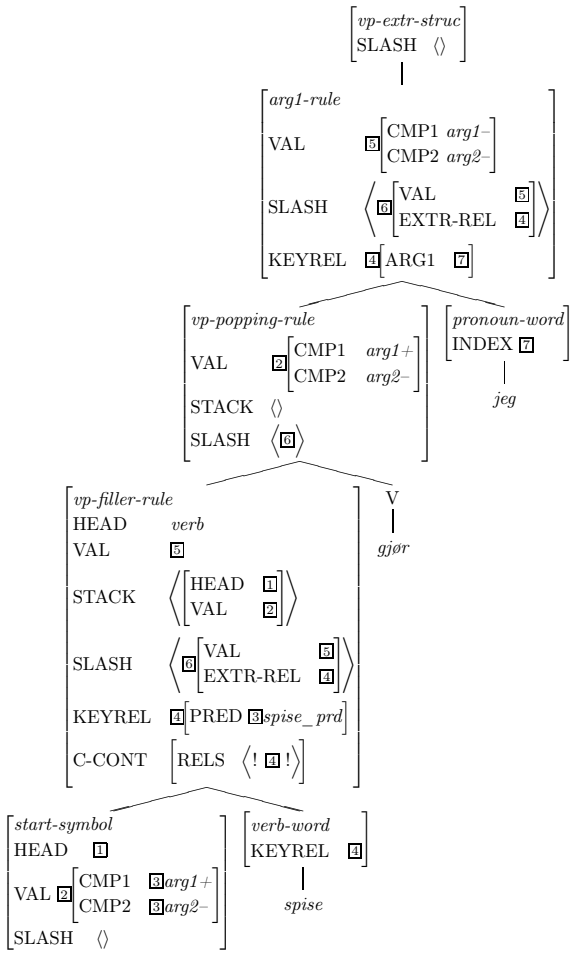

Figure 3: Analysis of *Spise gjør jeg* 'Eat, I do'

(7) $$\left\langle \text{h1,e2,} \left\{ \begin{array}{l} \text{h3:\_spise\_1(e2:PRES,x4),} \\ \text{h5:pers-pron(x4),} \\ \text{h6:pronoun\_q(x4, h8, h7)} \end{array} \right\}, \left\{ \begin{array}{l} \text{h8 qeq h5,} \\ \text{h1 qeq h3} \end{array} \right\} \right\rangle$$

(8) a. Spise gjør jeg.
eat-INF do-PRES I
Eat, I do.

[5] The linking types of beneficiary arguments and oblique arguments are not displayed in the analysis.

b. Spiser gjør jeg.
eat-PRES do-PRES I
Eat, I do.

c. Jeg spiser.
I eat-PRES
I eat.

The generation of all the items in (8) from the same MRS is possible because the dummy verb *gjøre* does not contribute a relation to the semantics in the VP fronting construction, and because the tense of the event in fronted VP constructions is determined by the dummy verb, not the main verb. Even though *gjøre* plays a special role in the fronted VP construction, this does not mean that there is a need for two lexical entries for the verb. If *gjøre* is analyzed as a main verb (which by far is its most frequent use), its PRED value will be unified with the KEYREL|PRED value of the clause, as for any other verb. Thus, given the constructional approach to semantic composition, there is need for only one lexical entry for *gjøre* 'do'. This analysis of *gjøre* differs from the LFG analysis in Figure 1, where *gjøre* contributes a predicate that takes the fronted VP as an argument.

4 Evaluation

With the new rules presented in this paper, the grammar now accounts for all regular cases of VP fronting in Norwegian.[6] A comparison with a version of the grammar without the VP fronting rules on a set of test sentences shows that the added analysis comes with a cost: an 8.3% increase in tasks and a 4.0% increase in space. An inspection of the results shows that items that start with a verb, such as yes-no questions and certain imperatives, require more processing with the added rules.

[6]There exist a few examples of partial VP fronting in Norwegian, such as (i), where the object of the fronted verb, a subordinate clause, is not fronted. The approach is suited for an analysis of these as well.

(i) Glemmes skal det heller ikke at Beethoven satte i
forget-PASS shall it also not that Beethoven set in
gang en pengeinnsamling til Bachs yngste datter
walk a fundraiser to Bachs youngest daughter
It should also not be forgotten that Beethoven initiated a fundraiser for Bach's youngest daughter

5 Summary

This paper has shown that Norwegian VP fronting can be captured straightforwardly in an incremental, HPSG-inspired typed feature structure grammar. The analysis derives VP filler–gap dependencies and the behavior of dummy *gjøre* without extra lexical stipulation. Fronted and non-fronted variants share the same MRS, supporting the view that VP fronting is a constructional variant of clause structure rather than a separate lexical pattern.

References

Borsley, Robert D. & Berthold Crysmann. 2024. Unbounded dependencies. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-driven phrase structure grammar* (Empirically Oriented Theoretical Morphology and Syntax 9), 571–634. Berlin: Language Science Press. doi:10.5281/zenodo.13644930. https://doi.org/10.5281/zenodo.13644930.

Copestake, Ann. 2002. *Implementing typed feature structure grammars*. CSLI publications.

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan A. Sag. 2005. Minimal Recursion Semantics: An introduction. *Research on Language and Computation* 3(2–3). 281–332. doi: 10.1007/s11168-006-6327-9.

Dyvik, Helge. 2000. Nødvendige noder i norsk: Grunntrekk i en leksikalsk-funksjonell beskrivelse av norsk syntaks. In Øivin Andersen, Kjersti Fløttum & Torodd Kinn (eds.), *Menneske, språk og felleskap*, Novus forlag.

Dyvik, Helge, Paul Meurer, Victoria Rosén, Koenraad De Smedt, Petter Haugereid, Gyri Smørdal Losnegaard, Gunn Inger Lyse & Martha Thunes. 2016. Norgrambank: A 'deep' treebank for norwegian. In *Proceedings of the tenth international conference on language resources and evaluation (lrec'16)*, 3555–3562. Portorož, Slovenia: ELRA.

Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.

Haugereid, Petter. 2007. Decomposed phrasal constructions. In Stefan Müller (ed.), *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar, Stanford Department of Linguistics and CSLI's LinGO Lab*, 120–129. Stanford, CA: CSLI Publications. doi:10.21248/hpsg.2007.7.

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 & 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. Stanford, CA: CSLI Publications. doi:10.21248/hpsg.2012.11.

Huddleston, Rodney & Geoffrey K Pullum. 2002. *The cambridge grammar of the english language*. Cambridge University Press.

Müller, Stefan & Bjarne Ørsnes. 2013. Towards an HPSG analysis of object shift in Danish. In Glyn Morrill & Mark-Jan Nederhof (eds.), *Formal Grammar: 17th and 18th International Conferences, FG 2012, Opole, Poland, August 2012, revised selected papers, FG 2013, Düsseldorf, Germany, August 2013: Proceedings* (Lecture Notes in Computer Sciecne 8036), 69–89. Berlin: Springer. doi:10.1007/978-3-642-39998-5_5.

Rosén, Victoria, Koenraad De Smedt, Paul Meurer & Helge Dyvik. 2012. An open infrastructure for advanced treebanking. In Jan Hajič, Koenraad De Smedt, Marko Tadić, & António Branco (eds.), *Meta-research workshop on advanced treebanking at lrec2012*, 22–29. Istanbul, Turkey.