

Examining Delayed Complements in Norwegian within an Incremental Left-Branching Grammar Framework

Petter Haugereid 

Western Norway University of Applied Sciences

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic


Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 104–117

Keywords: Norwegian, HPSG, Delayed Complements, Incremental Parsing

Haugereid, Petter. 2024. Examining Delayed Complements in Norwegian within an Incremental Left-Branching Grammar Framework. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 104–117. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.7. 

Abstract

In this paper, I will present an analysis of complement clauses in Norwegian that are licensed by the five adverbs *så/såpass* ‘so’ and *slik/sånn/sådan* ‘such’. It will be assumed that the licensed complement clause, although it is licensed by the adverb, is not a complement within the constituent with the adverb, but rather a complement of the clause. This opens for a uniform analysis of complement clauses licensed by adverbs, irrespective of their position with regard to the licensing adverb. The analysis will be conducted within the framework of an HPSG-inspired incremental typed feature structure grammar of Norwegian.

1 Introduction

A little studied, however not completely infrequent, phenomenon is that of *delayed complement clauses* in examples like (1) from Huddleston & Pullum (2002: 967).

- (1) So many people enrolled for the course that we had to move to a larger room.

So is here a degree adverb, modifying a degree determinative *many*, and it requires a complement clause *that we had to move to a larger room*. This complement clause is according to Huddleston & Pullum (2002) always at the end of the clause.

The construction is related to the more common construction where the complement clause appears adjacent to the phrase with *so*, referred to hereafter as the *so-phrase*. This is exemplified for Norwegian in (2a) where the *so-phrase* *så sen* ‘so late’ is directly followed by the complement clause *at jeg smiler* ‘that I smile’. This construction is semantically equal to the corresponding delayed complement construction demonstrated in (2b). In Norwegian it is always possible to front the *so-phrase*, and the complement clause then is left behind.

- (2) a. Han er så sen at jeg smiler.
he is so late that I smile
‘He is so late that I smile.’
b. Så sen er han at jeg smiler.
so late is he that I smile
‘He is so late that I smile.’

Probably the most famous sentence with a delayed complement clause is the first sentence of John 3:16 in the Bible, as shown in (3). In this sentence, the *so-phrase* *så høyt* ‘so highly’ is thematized, making the complement clause a delayed complement. The degree adverb here modifies an adverb *høyt* ‘highly’. Note that in the English translation, the *so-phrase* consists only of the adverb *so*.

[†]I would like to thank three anonymous reviewers and the audience at the HPSG 2024 conference in Olomouc, Czech Republic, for very useful comments and suggestions.

- (3) For så høyt har Gud elsket verden at han ga sin Sønn, den
 for so highly has God loved world-DEF that he gave REFL son, the
 enbårne [...]
 one and only [...]
 ‘For God so loved the world that he gave his one and only Son [...].’

There are also other (degree) adverbs that require complement clauses; *såpass* ‘so’, *slik* ‘such’, *sånn* ‘such’, and *sådan* ‘such’. While *så* and *såpass* function as degree adverbs modifying adjectives, adverbs, and prepositions, *slik* and *sånn* modify determiners (example (4a)) or function alone, as adverbs (example (4b)).

- (4) a. Med slik en kraft traff den at jeg falt.
 with such a force struck it that I fell
 ‘It struck with such a force that I fell.’
 b. For slik har vi elsket naturen at vi er nødt til å drepe
 for so have we loved nature-DEF that we are obliged to to kill
 den.
 it
 ‘For we have so loved the nature that we have to kill it.’

There are similar comparative constructions where a subordinate clause or an infinitival clause at the end of a sentence is dependent on a comparative element earlier in the sentence, as shown in (5).¹

- (5) a. More people enrolled for the course than we had expected.
 b. Too many people enrolled for the course to fit in the room.

In (5a) the comparative governor *more* licenses the comparative complement *than we had expected* at the end of the clause. In (5b) the comparative governor *too* licenses the comparative complement *to fit in the room*. The difference from the delayed complement clauses like (1), apart from the licensing element, is that while the complement clause in delayed complement clauses is a regular that-clause, the complement in (5a) is a subordinate clause introduced by *than*, and the complement in (5b) is an infinitival clause. It shows that the licensing element has a syntactic requirement for the comparative complement.

However, whereas the complement clause required by *so* must appear at the end of the clause, as illustrated by (6a), the comparative complements licensed by *more* and *too* may appear within the sentence, as shown in (6b) and (6c).

- (6) a. * So many people that we had to move to a larger room, enrolled for
 the course.
 b. More people than we had expected enrolled for the course.
 c. Too many people to fit in the room enrolled for the course.

¹Thanks to Dan Flickinger for pointing this out to me.

In this paper, the focus will be on the delayed complement construction illustrated in (1)–(3). However, it will also be outlined how the analysis can be extended to comparative constructions, as illustrated in (5).

2 Corpus Searches

A search for the words *så/såpass* ‘so’ and *slik/sånn/sådan* ‘such’ followed by the complementizer *at* ‘that’ within a window of the following 10 words in the 100 million word Leksikografisk bokmåskorpus (Fjeld et al. 2020) yielded the number of matches shown in Table 1.² A manual inspection of the first 50 matches in each search revealed that a significant number of the complement clauses were licensed by the adverb.³ The total number of complement clauses licensed by the five adverbs is estimated to be about 40,000. The total number of complement clauses with the complementizer *at* in the corpus is 1,025,355. This implies that about 4% of the *at* complement clauses are licensed by an adverb. Among these, about 1,000 (1 of 40) is a delayed complement construction.

	Matches	Manual inspection	Estimate
<i>så</i> ... <i>at</i>	59,671	29/50	34,609
<i>såpass</i> ... <i>at</i>	1,346	46/50	1,238
<i>slik</i> ... <i>at</i>	9,723	19/50	3,694
<i>sånn</i> ... <i>at</i>	1,260	22/50	554
<i>sådan</i> ... <i>at</i>	65	10/65	10
Total			40,105

Table 1: Estimated number of complement clauses licensed by adverbs in Leksikografisk bokmåskorpus

3 HPSG Analysis

The ERG (Flickinger 2000) provides an analysis for sentences like (2a) where the complement clause is adjacent to the *so*-phrase. The complement clause is then treated as a complement of *so*, as shown in Figure 1. The MRS (Copestake et al. 2005) of the sentence is given in Figure 2. It shows how the predicate of *so* (*_so_x_comp*) takes *late* as its first argument (11) and the subordinate clause as its second argument (12).

However, in cases where the *so*-phrase is nonadjacent to the licensed CP, as in (1), the ERG lacks an analysis where the CP is an argument of *so*. The delayed complement construction poses a challenge for regular HPSG grammars given that the element that selects for the complement clause, *so*, occurs at the top of the tree,

²The corpus is not syntactically annotated, so a syntactic search is not possible.

³All the 65 matches with *sådan* were manually inspected.

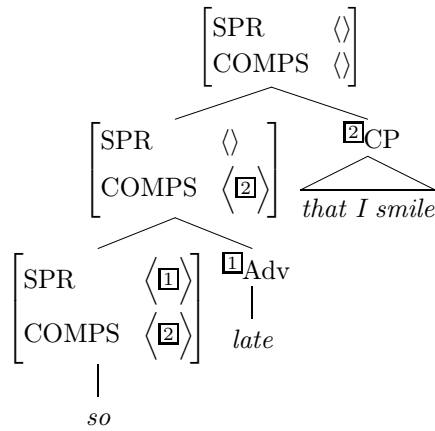


Figure 1: Analysis of *so late that I smile* by the ERG

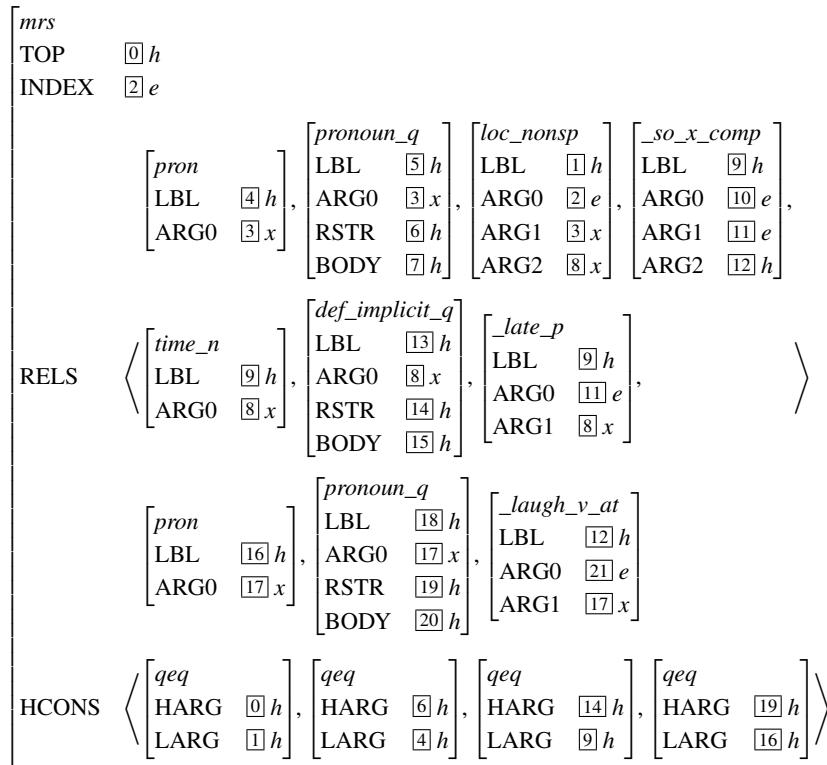


Figure 2: MRS of the sentence *He was so late that I laughed.*

as part of a subject or a filler, and even though one would allow for a complement clause to be realized after the the subject or filler, once the phrase containing *so* has been realized, there is no way to access the valence requirements of *so*.

If we assume that the canonical position of the complement clause in a delayed complement construction is adjacent to the *so*-phrase, as implied in the analysis in Figure 1, the delayed complement construction could be considered a case of extraposition. However, since the complement clause consistently appears at the end of the matrix clause, there is no evidence supporting such an analysis. A version of the delayed complement construction where the complement clause is not at the end, would be ungrammatical, as illustrated in (6a).

4 Analysis

In this section, the delayed complement construction and the more regular construction, with the complement clause adjoined to the *so*-phrase, will be given a uniform analysis.

4.1 Incremental analysis

The analysis is conducted within the framework of an HPSG-inspired incremental typed feature structure grammar for Norwegian (Haugereid 2009), implemented using the LKB system (Copestake 2002) as part of the Delph-In effort.⁴ This approach assumes a distinction between a parse tree and a constituent tree (Haugereid & Morey 2012), where utterances are parsed incrementally in a bottom-up fashion from left to right, resulting in a completely left-branching tree structure. The grammar functions similarly to a shift-reduce parser, utilizing a *STACK* feature to monitor matrix constituents during the parsing of embedded constituents. Upon completing the parse, the constituent structure of the clause can be inferred from examining the *STACK* feature of each node in the parse tree. As will be shown, this incremental approach is well-suited for the analysis of delayed complement constructions.

Figure 3 provides the parse tree for example (2b), repeated below as (7). In this figure, the stacked constituent (V) is put on a list while the subordinate clause is parsed. At the top of the tree, the V constituent is popped from the list. A more detailed analysis of the same sentence is provided later, in Section 4.6.

- (7) Så sen er han at jeg smiler.
so late is he that I smile
'He is so late that I smile.'

4.2 Licensed Complement

The central assumption of the analysis presented in this paper, drawn from Huddleston & Pullum (2002: 967), is that the complement clause consistently appears at the end of the clause and that it is a complement of clause structure, rather than the licensing adverbs (in Norwegian, *så/såpass* 'so' and *slik/sånn/sådan* 'such').

⁴<https://github.com/delph-in/docs/wiki>

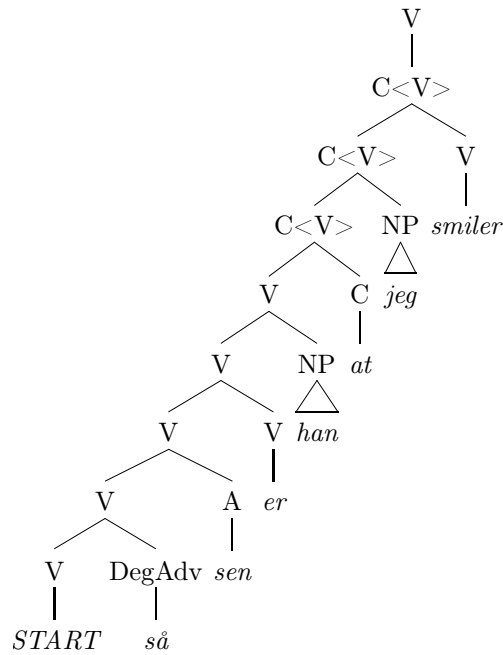


Figure 3: Incremental parse tree

This is achieved by allowing the feature licensing the complement clause ascend the tree from the point where the licensing adverb is realized until it triggers a rule, initiating the parsing of a complement clause. The analysis encompasses lexical entries for the licensing degree adverbs, a rule for the licensing adverbs, a feature LC (Licensed Complement), and a rule for the licensed complement.

4.3 Degree Adverbs

The lexical entry for the degree adverb *så* ‘so’ is given in (8). It modifies an adjective, adverb or preposition.

$$(8) \left[\begin{array}{l} \textit{degadv-word} \\ \text{STEM} \quad \langle \textit{“s\aa”} \rangle \\ \text{HEAD} \quad \left[\begin{array}{l} \textit{degadv} \\ \text{MOD} \langle \left[\text{HEAD } \textit{adj-adv-prep} \right] \rangle \end{array} \right] \\ \text{KEYREL} \quad \left[\text{PRED } \textit{s\aa_deg} \right] \end{array} \right]$$

The predicate of *så*, *så_deg*, is an underspecified type with two possible subtypes, *så_deg_rel* and *så_deg_cp_rel*, as illustrated in Figure 4. The regular degree adverb type *så_deg_rel* inherits from the type *comp-*, which means that it is not compatible with a complement argument, while the type *så_deg_cp_rel* inherits from the type

comp+, which means that it requires a CP complement. This underspecification is unique to the five CP-licensing adverbs.

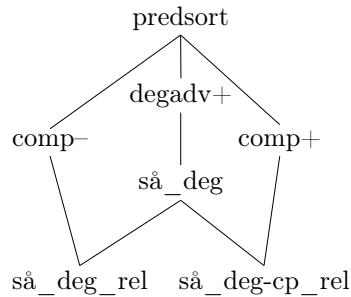


Figure 4: Type hierarchy of predicate types for the degree adverb *så* ‘so’

4.4 Degree Adverbs Requiring CP Complements

The rule responsible for attaching degree adverbs that require a CP complement is presented in Figure 5. Given the left branching structures in this approach, coupled with the leftward attachment of degree adverbs, these adverbs are parsed before the modified word. To accommodate this, the element on the MOD list of the degree adverbs is unified with the feature *PREMOD* in the mother node. The subsequent rule that attaches the modified word unifies the *PREMOD* feature of its first daughter with the *LOCAL* value of its second daughter. Consequently, the degree adverb has the *LOCAL* features of the word it modifies on its MOD list. This is demonstrated later, in Figure 7.

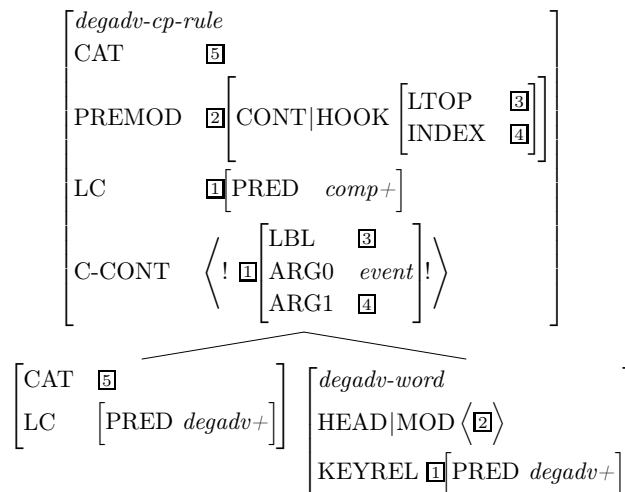


Figure 5: Rule for attaching degree adverb that requires a complement clause

The degree adverb rule unifies the KEYREL of the degree modifier with the feature LC (Licensed Complement) in the mother. This relation is specified to have the PRED value *comp+*. The corresponding value in the first daughter is underspecified, allowing for more than one CP-licensing adverb.

4.5 Delayed Complement Rule

The final part of the analysis involves a rule that initiates the parsing of a CP required by an adverb, given in Figure 6.

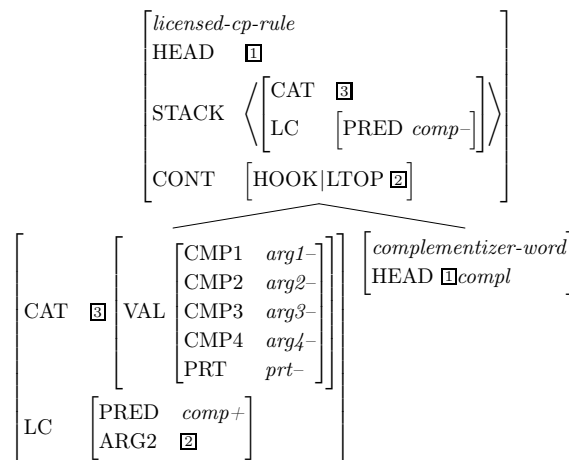


Figure 6: Rule for attaching complementizer initiating CP licensed by degree modifier

The first daughter of the rule is a clause where all the arguments and particles are realized,⁵ and which licenses a complement clause (the LC|PRED value is *comp+*). The second daughter is a complementizer. In the mother node, the CAT features of the initial daughter are placed on a STACK.⁶ Additionally, the ARG2 of the LC relation in the first daughter is unified with the LTOP of the complement clause (the mother).

4.6 Analysis of Sentence with Delayed Complement

The analysis of sentence (2b)/(7) with a delayed complement is illustrated in Figure 7. It demonstrates the incremental parsing of the sentence, detailing how the degree adverb *så* ‘so’ licenses the delayed complement clause through the feature LC. The figure also depicts how the relation of the degree adverb is linked to the modified

⁵The negative values of CMP1, CMP2, CMP3, CMP4, and PRT indicate that all the dependents of the main verb are realized.

⁶The STACK feature allows for parsing of embedded structures, see Haugereid & Morey (2012).

adjective and the complement clause, and that the relation is added to RELS via C-CONT|RELS.

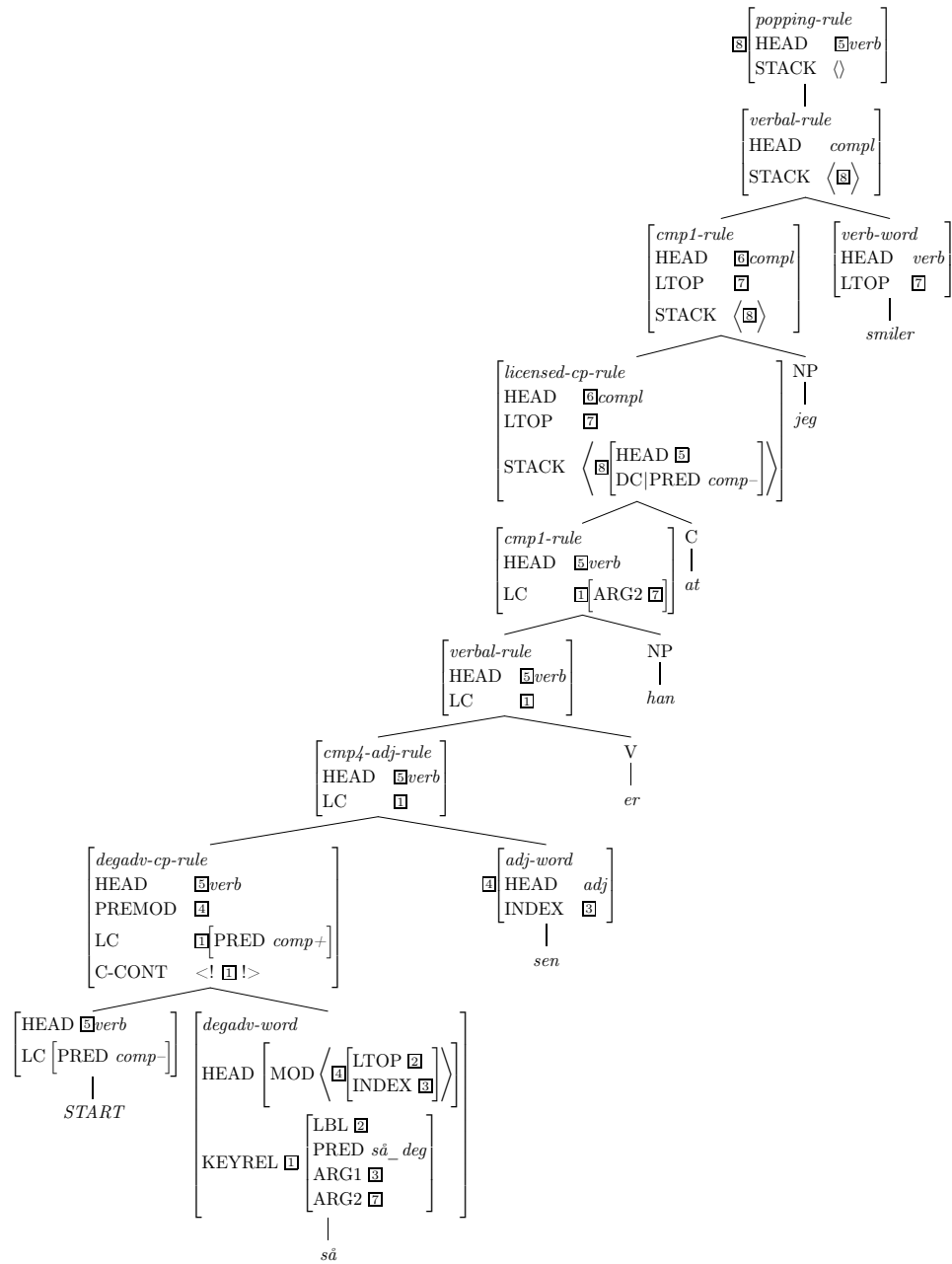


Figure 7: Incremental analysis of sentence with delayed complement

The MRS resulting from the analysis in Figure 7 is given in Figure 8. It illustrates how the relation of the degree adverb *sã_deg-cp_rel* has two arguments.

The first (5) is the index of of the modified adjective *_sen_a_rel*, and the second (6) is the index of the the delayed complement clause.

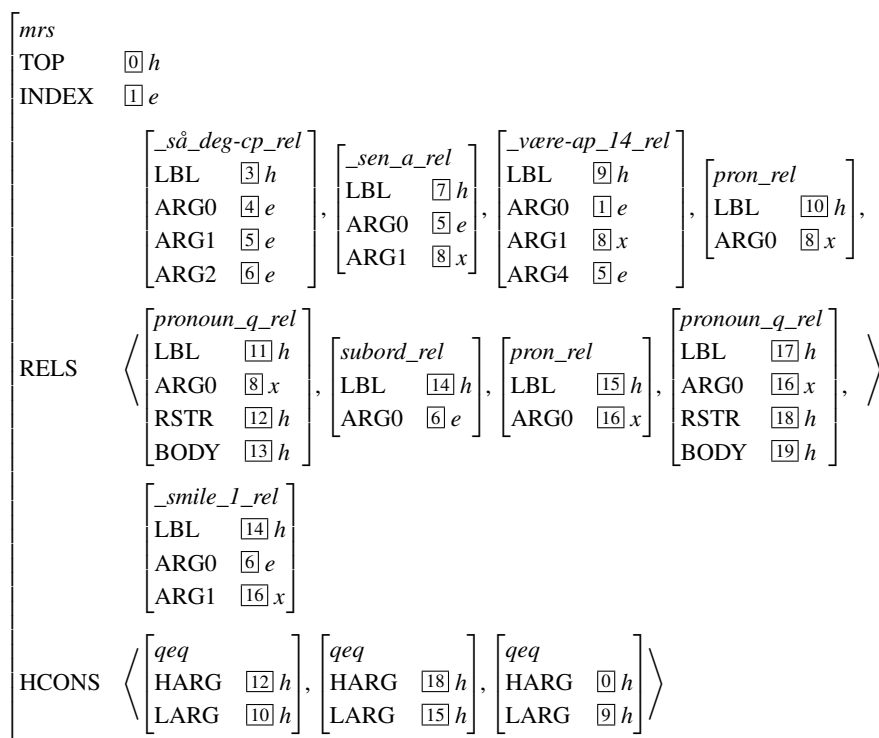


Figure 8: MRS of the sentence *Så sen er han at jeg smiler* ‘He is so late that I laugh.’

4.7 Comparative Structures

When it comes to the comparative structures exemplified in (5), repeated below as (9), they can be analyzed in a similar fashion to the delayed complement constructions.

- (9) a. More people enrolled for the course than we had expected.
 b. Too many people enrolled for the course to fit in the room.

Just like *so*, the comparative governors *too* and *more* would have a relation with an underspecified PRED value, as illustrated in the simplified type hierarchy in Figure 9.

There would be separate rules for attaching *too* and *more* as comparative governors requiring a complement clause. Additionally, there would be separate rules for initiating an infinitival clause if the LC feature has a *vp+* requirement, and a *than-clause* if the LC feature has a *than-cp+* requirement.

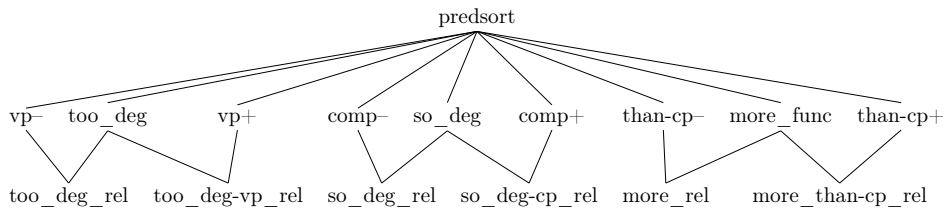


Figure 9: Simplified type hierarchy of predicate types for *too*, *so*, and *more*

The analysis of the comparative structures differs from the analysis of delayed complements in that the licensed clause does not need to be sentence-final. The rules that trigger the parsing of the licensed clause can also apply within the sentence, accommodating examples like (6b) and (6c).

5 A Challenge for Regular HPSG Grammars

The analysis presented in Section 4 can be adapted to regular HPSG grammars. One would then have to assume that the feature *LC* ascends to the sentence level and there combines the sentence with the complement clause.

However, the position of the licensing element in a *so*-phrase is not always the same. For example, in (10a) the licensing element *so* comes after the determiner (*a*), and in (10b) the licensing element *such* comes before the determiner.

- (10) a. a so good sound
 b. such a good sound

This means that the NP would get the licensing feature *LC* from the right daughter if the licensing element follows the determiner, as exemplified in Figure 10, while it would get the *LC* feature from the left daughter if the licensing element precedes the determiner (see Figure 11). This would require two different NP rules.

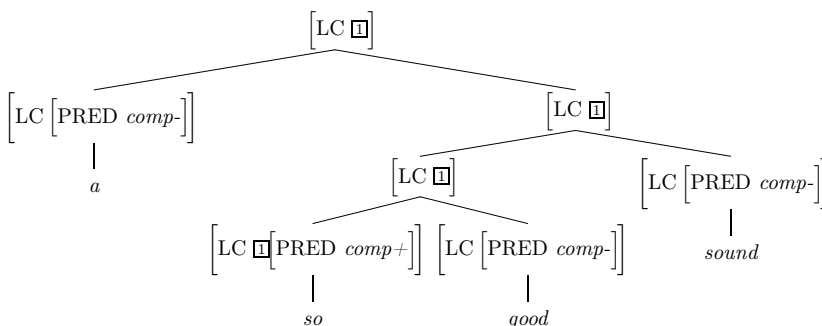


Figure 10: Getting the *LC* feature from the right daughter of an NP

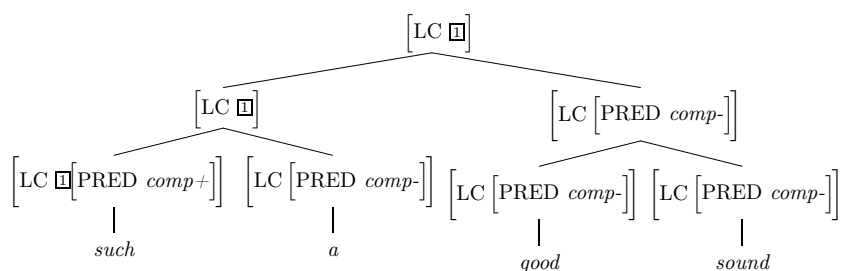


Figure 11: Getting the LC feature from the left daughter of an NP

With the incremental approach, however, the LC feature is just passed up from first daughter to mother once the degree adverb has been realized, and there is no need for separate NP rules.

6 Conclusion

An analysis of delayed complement constructions is provided, wherein the complement clause appearing at the end of the sentence is assumed to be a complement of the clause licensed by an adverb like *så* 'so', rather than by the adverb itself.

By adopting this assumption, a consistent analysis can be applied regardless of the proximity between the phrase with *så* 'so' and the complement clause. This approach allows for flexibility, accommodating scenarios where the phrase with the licensing adverb is either adjacent to or distant from the complement clause, while maintaining a uniform analysis throughout.

The left-branching structures assumed in this approach facilitate an analysis where the licensing adverb's requirement is registered and passed up from the daughter node to the mother node, regardless of whether the adverb appears at the beginning of the sentence or not. This method can be adapted to a standard HPSG grammar. However, this adaptation poses challenges, as it necessitates considering that the licensing condition (LC) feature could originate from either the first or the second daughter of an NP.

References

- Copestake, Ann. 2002. *Implementing typed feature structure grammars* (CSLI Lecture Notes 110). Stanford, CA: 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.
- Fjeld, Ruth Vatvedt, Anders Nøklestad & Kristin Hagen. 2020. Leksikografisk bokmålskorpus (LBK) – bakgrunn og bruk. *Oslo Studies in Language* 11(1). 47–59.

- 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 & Technology. (Doctoral 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 (eds.). 2002. *The Cambridge grammar of the English language*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/9781316423530.