

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

Petter Haugereid

Western Norway University of Applied Sciences

petter.haugereid@hvl.no

## 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 and 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 and 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.*

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 or function alone, as adverbs.

## 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. 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.<sup>1</sup> 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 ... at</i>	59,671	29/50	34,609
<i>såpass ... at</i>	1,346	46/50	1,238
<i>slik ... at</i>	9,723	19/50	3,694
<i>sånn ... at</i>	1,260	22/50	554
<i>sådan ... 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.<sup>2</sup>

<sup>1</sup>All the 65 matches with *sådan* were manually inspected.

<sup>2</sup>If we assume that the canonical position of the complement clause in a delayed complement construction is adjacent to the

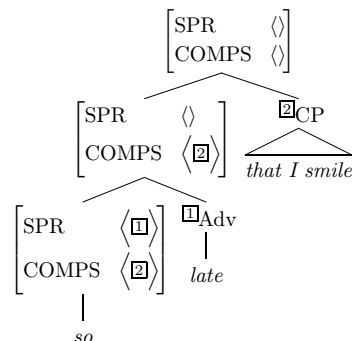


Figure 1: Analysis of *so late that I smile* by the ERG

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, as part of a subject or a filler.

## 4 Incremental 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. 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 a part of the Delph-

*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, like *\*So many people that we had to move to a larger room, enrolled for the course*, would be ungrammatical.

In effort.<sup>3</sup> The approach assumes a division between a parse tree and a constituent tree (Haugereid and Morey, 2012), where utterances are parsed in a bottom-up fashion, incrementally, from left to right, resulting in a completely left-branching tree structure.<sup>4</sup>

The central assumption of the analysis, drawn from Huddleston and 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’). 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.

The lexical entry for the degree adverb *så* ‘so’ is given in (3). It modifies an adjective, adverb or preposition.

$$(3) \left[ \begin{array}{l} \text{degadv-word} \\ \text{STEM} \quad \langle \text{“så”} \rangle \\ \text{HEAD} \quad \left[ \begin{array}{l} \text{degadv} \\ \text{MOD} \langle \text{[HEAD adj-adv-prep]} \rangle \end{array} \right] \\ \text{KEYREL} \quad \left[ \text{PRED } s\grave{a}\_deg \right] \end{array} \right]$$

The predicate of *så*, *så\_deg*, is an underspeci-

<sup>3</sup><https://github.com/delph-in/docs/wiki>

<sup>4</sup>The grammar resembles a shift-reduce parser, utilizing a STACK feature to monitor matrix constituents during parsing of embedded constituents. Upon completing the parse, the constituent structure of the clause can be inferred from the resulting AVM by examining the STACK feature of each node of the parse tree. This incremental approach is well-suited for the analysis of delayed complement constructions (although the analysis can be extended to other approaches as well).

fied type with two possible subtypes, *så\_deg\_rel* and *så\_deg\_cp\_rel*, as illustrated in Figure 2. 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.

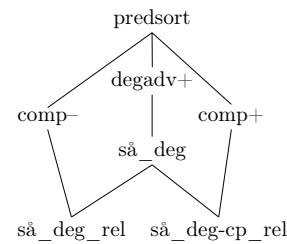


Figure 2: Type hierarchy of predicate types for the degree adverb *så* ‘so’

The rule responsible for attaching degree adverbs that require a CP complement is presented in Figure 3. 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 5.

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.

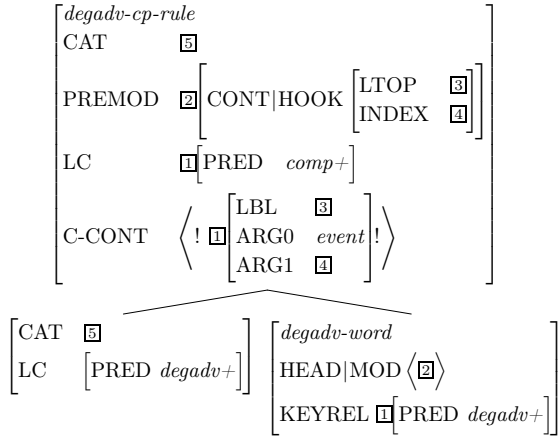


Figure 3: Rule for attaching degree adverb that requires a complement clause

Finally, the rule that initiates the parsing of a CP required by an adverb, is given in Figure 4.

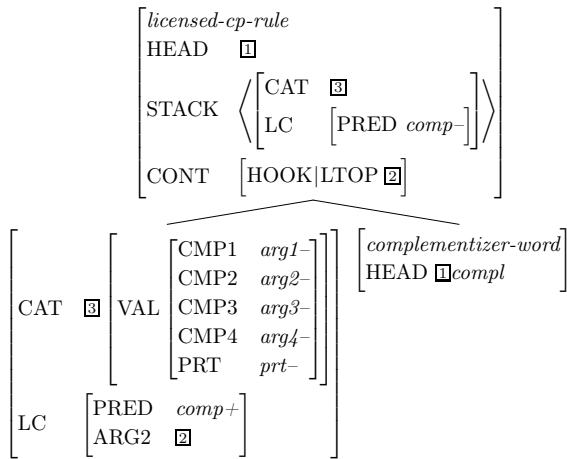


Figure 4: Rule for attaching complementizer initiating CP licensed by degree modifier

The first daughter of the rule is a clause where all

the arguments and particles are realized,<sup>5</sup> 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.<sup>6</sup> Additionally, the ARG2 of the LC relation in the first daughter is unified with the LTOP of the complement clause (the mother).

The analysis of sentence (2b) with a delayed complement is illustrated in Figure 5. 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 adjective and the complement clause, and that the relation is added to RELS via C-CONT|RELS.

## 5 Conclusion

By assuming that CPs licensed by adverbs like *så* ‘so’ are complements of the clause rather than complements of the adverb, a consistent analysis can be applied regardless of the proximity between the so-phrase and the CP. 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.

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

<sup>6</sup>The STACK feature allows for parsing of embedded structures, see Haugereid and Morey (2012).

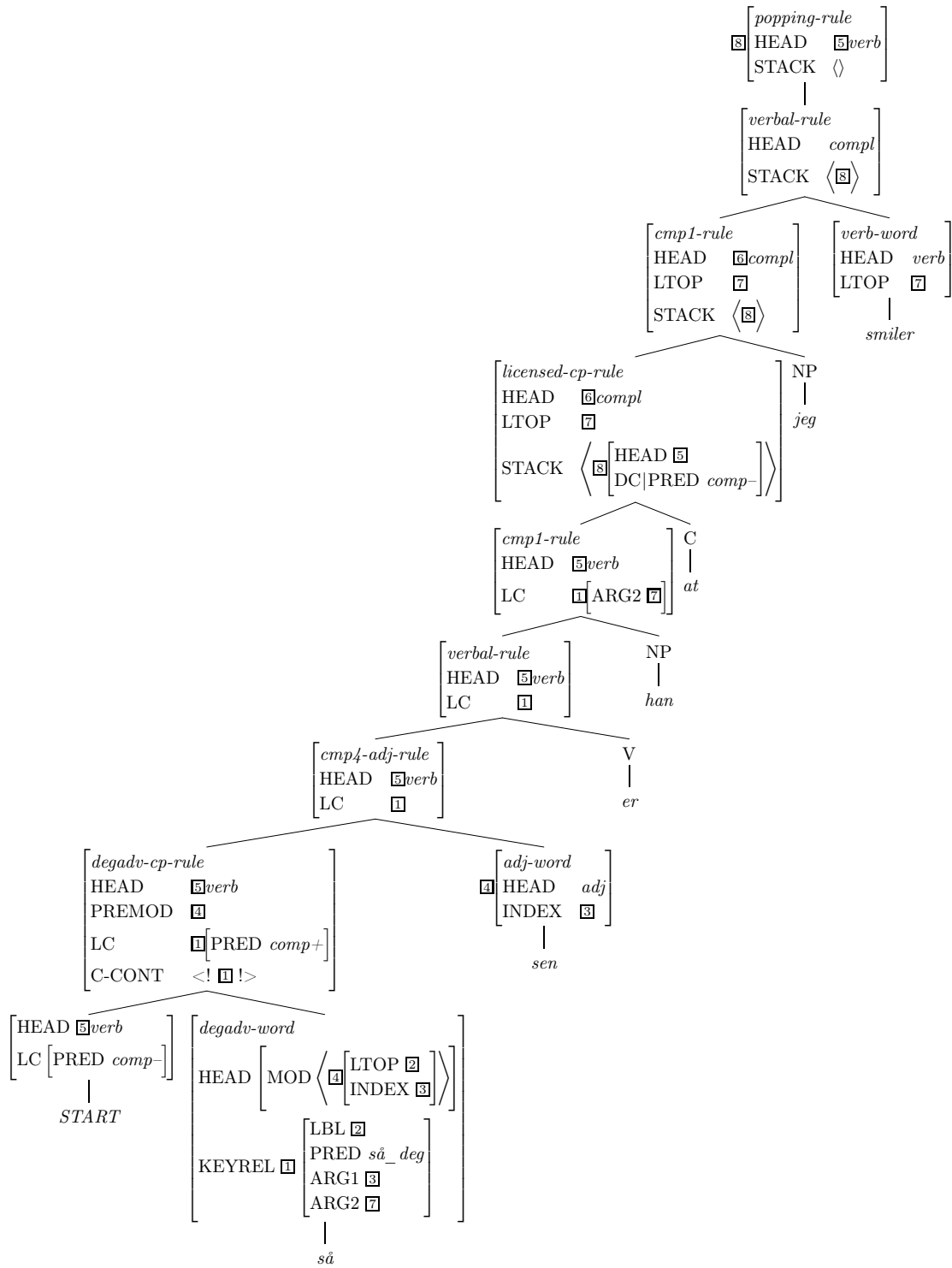


Figure 5: Incremental analysis of sentence with delayed complement

## References

- Copestake, A. (2002). *Implementing Typed Feature Structure Grammars*. CSLI publications.
- Fjeld, R. V., Nøklestad, A., and Hagen, K. (2020). Leksikografisk bokmålskorpus (LBK) – bakgrunn og bruk. *Oslo Studies in Language*, **11**(1), 47–59.
- Flickinger, D. P. (2000). On building a more efficient grammar by exploiting types. *Natural Language Engineering*, **6**(1), 15–28.
- Haugereid, P. (2009). *Phrasal subconstructions: A constructionalist grammar design, exemplified with Norwegian and English*. Ph.D. thesis, Norwegian University of Science and Technology.
- Haugereid, P. and Morey, M. (2012). A left-branching grammar design for incremental parsing. In S. Müller, editor, *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar, Chungnam National University Daejeon*, pages 181–194.
- Huddleston, R. and Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press.