

Norwegian Word Order in HPSG

Liv Ellingsen
Universitetet i Oslo
liv.ellingsen@ilf.uio.no

31st May 2004

Abstract

This paper presents an HPSG analysis for main word order phenomena in Norwegian. The analysis has been built on the language independent basis grammar the Matrix (Bender, Flickinger, & Open, 2002), and it has been implemented in the LKB system (Copestake, 2001). The main phenomena treated in the analysis are the V2 constraint in Norwegian main clauses, the distribution of certain types of sentence adverbials and the contrast between main and subordinate clause structures.

1 Data

Norwegian is a V2 language, i.e. the finite verb always occupies the second position in a Norwegian declarative main clause, as shown in (1). Combined with the possibility of topicalization, this implies that the subject has more than one possible position in this type of clause. In a clause where another element has been topicalized, the canonical subject position is after the finite verb, before any of the verb's complements, as seen in (1b).

- (1) a. Gyrd ga boka til Inge.
Gyrd gave the-book to Inge
b. Til Inge ga Gyrd boka.

Normally, subordinate clauses differ structurally from main clauses. Topicalization is not allowed in a subordinate clause structure, and so it has the strict SVO order shown in (2).

- (2) a. [da] Gyrd ga boka til Inge.
[when] Gyrd gave the-book to Inge
b. *[da] til Inge ga Gyrd boka.

Another interesting structural difference between the two clause types is the distributional possibilities of sentence adverbials. In main clauses, a sentence adverbial like *sikkert* (*surely*) can be placed directly after the finite verb as seen in 3, or directly after an inverted subject. In subordinate clauses, the sentence adverbial can be placed directly before the finite verb, as the V2 constraint does not apply to this type of structure.

- (3) a. Gyrd ga sikkert boka til Inge.
b. *Gyrd sikkert ga boka til Inge.
c. [da] Gyrd sikkert ga boka til Inge.
d. *[da] Gyrd ga sikkert boka til Inge.

Sentence adverbials can mainly be placed in three different positions in a Norwegian clause: (i) in the sentence-initial position (topicalized) (ii) in the middle field after the finite verb in main clauses or directly before the finite verb in subordinate clauses and (iii) at sentence end. By comparing the three sentence adverbials *sikkert*, *heldigvis* (*luckily*) and *allikevel* (*still*), the sentences in (4) show how sentence adverbials differ regarding their distributional possibilities.

- (4) a. Til Inge ga Gyrd heldigvis/sikkert/allikevel boka.
b. Gyrd ga boka til Inge *heldigvis/*sikkert/allikevel .
c. Heldigvis/*Sikkert/Allikevel ga Gyrd boka til Inge.

A very small group of sentence adverbials (like *allikevel*) can be placed in all the three possible positions in a clause. The large remaining group of sentence adverbials can not be placed at sentence end. This group is then divided into ‘heavy’ sentence adverbials that can be stressed, and so also topicalized (like *heldigvis*), and ‘light’ sentence adverbials that can only stand in the middle field of the sentence.

Sentence adverbials that can stand at sentence end in a main clause can also stand in this position in a subordinate clause structure. All sentence adverbials can stand directly to the left of the finite verb in a subordinate clause, while attempts at placing them to the left of the subject result in ungrammatical or highly doubtful constructions for most adverbials¹, as shown in (5).

- (5) [da] *heldigvis/*sikkert/*allikevel Gyrd har gitt boka til Inge.

2 Analysis and implementation

The analysis of these phenomena and its implementation are built on the LinGO Grammar Matrix (Bender et al., 2002), a language-independent ‘starter kit’ for grammar developers. This core grammar provides basic types for the necessary constructional types, such as head-subject, head-complement and head-adjunct phrases.

The LinGO Grammar Matrix is mainly based on the English LinGO ERG (Copestake & Flickinger, 2000). This makes it interesting to see whether the basic types can be directly transferred to a Norwegian grammar, or if they need to be adjusted in any way before they are put to use.

Concerning the possibility of subject inversion in Norwegian, this construction is also found in English and it should therefore be possible to analyse and implement it based on the Matrix’ construction types. One possible way to handle inversion in English is to use a lexical rule that moves the subject to the front of the verb’s complement list (Sag & Wasow, 1999). As the order of the inverted subject and the complements are fixed in Norwegian, this would be a possibility. When the sentence adverbials are introduced, this solution becomes problematic, though: Because the sentence adverbials can be placed on each side of an inverted subject, but not between two complements, it is necessary to explicitly keep track of the subject.

Another possible solution for English is to use flat rules that bind the subject and the complements at the same time, as sisters of the verb (Ginzburg & Sag, 2001). This is also a possibility for Norwegian, but again the sentence adverbials complicate the picture. Because the adjuncts would have to be taken into account in a flat rule, a large number of rules would be needed to account for all possible combinations of subject, adjuncts and complements².

A traditional and well-used way to describe Scandinavian clause structure originates in Diderichsen’s clause schematas (Diderichsen, 1962). Figure (1) shows Diderichsen’s clause schematas for main and subordinate clause structures, slightly adapted for Norwegian (Faarlund, Lie, & Vannebo, 1997). It is possible to use these schematas as a basis for an alternative binary HPSG analysis of the phenomena they are describing.

¹Some specific exceptions to this pattern exist, including adverbials like *ikke* (*not*).

²This approach would grow even more complicated if the analysis was to be extended to cover the phenomenon of object shift as well.

| Main clause schemata | | | | | | | |
|----------------------|-------------|----|------|---------|---------------|------|---|
| Fund.field | Nexus field | | | | Content field | | |
| F | v | a1 | n | a2 | V | N | A |
| Til Inge | ga | - | Gyrd | sikkert | - | boka | - |

| Subordinate clause schemata | | | | | | | |
|-----------------------------|-------------|------|---------|----|---------------|---------------|---|
| Comp.field | Nexus field | | | | Content field | | |
| f | a1 | n | a2 | v | V | N | A |
| At | - | Gyrd | sikkert | ga | - | boka til Inge | - |

Figure 1: Diderichsen's clause schematas, adapted for Norwegian

It is necessary to define the subject position in the nexus field (after the finite verb) in a main clause structure as a canonical subject position for this type of clause. In a subordinate clause structures, the only possible subject position is the position before the finite verb in the nexus field.

It is thus necessary at least to define two head-subject rules, one head-final, combining the finite verb with its subject to the left, and one head-initial, combining the finite verb with its inverted subject to the right and marking the structure as a main clause structure. This marking can for instance be done via the feature MC (main clause).

It is also possible to follow Diderichsen even more closely and say that a subject in the sentence-initial position is extracted from its canonical inverted position and topicalized in the same way as complements and adjuncts. In that case, the head-final version of the head-subject rule marks the structure as a subordinate clause structure. This analysis has been implemented in the grammar on which this paper is based and it results in the two structures shown in figure (2) for the (correctly syntactically ambiguous) sentence *Inge beundrer Gyrd* (*Inge admires Gyrd*).

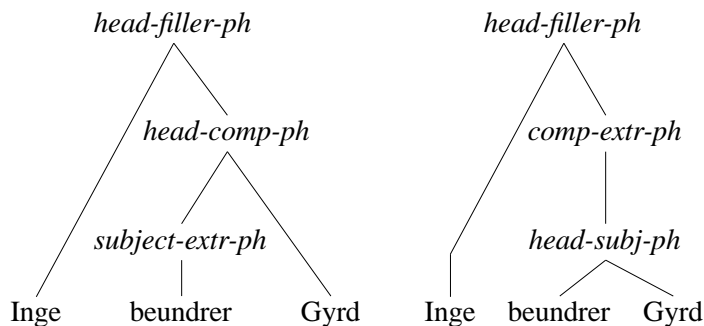


Figure 2: Main clause structure with subject and object extraction

To accommodate the head-initial head-subject rule, a few changes have to be made to the basic types in the Matrix, for instance when it comes to the treatment of complements. The Matrix does not expect the subject to be combined with its verb before all complements are gone and a VP has been built. When the verb is combined with its subject before any of the complements, destroying the traditional VP, the complements must also be passed up from the head daughter to the mother in a head-subject phrase.

This approach lets us keep track of the subject in a clause, but it is not quite enough to ensure correct placement of sentence adverbials in the nexus field. The nexus field of a main clause structure spans from the finite verb to the content field, which starts with the first complement of the finite verb. But as the Matrix uses binary head-complement phrases, cancelling the complements of the COMPS list one by one, it is not possible to see whether a structure still has a complete COMPS list, or whether one or more complements have already been cancelled of the list.

A solution to this problem is to define a feature NUC^3 (nucleus) in the verb's feature structure, whose value tells us whether we still are in the nexus field or not. All finite verbs then start out with the NUC value $+$. The inverted head-subject rule leaves the NUC value of the phrase unchanged, while any version of a head-complement rule changes the value to $-$. Modification by a sentence adverbial in the nexus field also leaves the NUC value unchanged.

Head-modifier rules handling adverbials in the nexus field marks the structure as a main or subordinate clause structure in the same way as the different types of head-subject rules, as the head-final/head-initial distinction plays the same role in these cases.

The different types of adverbials themselves can be defined as shown in the type hierarchy in figure (3). The hierarchy defines the types of modifiers according to their distributional possibilities only, so cross-classification with types that contribute information from other dimensions (for instance part-of-speech) is necessary.

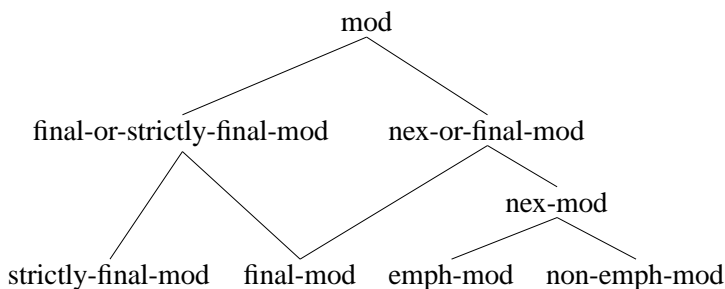


Figure 3: Typehierarki under *mod*

The type *nex-or-final-mod* has no constraints on its placement, and sentence adverbials that can be placed in all possible positions of a clause are defined as instances of this type. They are then forced down to the appropriate subtype by rule application, to receive constraints necessary for obtaining correct parses.

The type *nex-mod*, sentence adverbials that can not stand at sentence end, are constrained to modify elements with NUC value $+$. Its two subtypes, *emph-mod* and *non-emph-mod*, have no constraints on their own, but the extraction rules makes use of this distinction to make sure that only heavy sentence adverbials are topicalized. Finally, the type *final-or-strictly-final-mod* are constrained to only modify elements with an empty $COMPS$ list⁴.

Figure (4) shows two parse trees taken from the LKB where the type of clause is decided on by the placement of the sentence adverbials. The parse tree for the left sentence, *den boka likte Gyrd neppe* (*that book liked Gyrd probably-not*), unambiguously shows a main clause structure because of the *nex-mod* sentence adverbial placed after the subject. The structure to the right can only be a subordinate clause structure due to the left-modification of the verb.

If we compare the main clause structure in figure (4) with the two possible structures for the simple sentence in figure (2), we see that the placement of the *nex-mod* adverbial *neppe* also disambiguates the structure, deciding that *Gyrd* has to be the subject of the clause.

The analysis presented here covers the main phenomena of Norwegian word order, and although this is not yet implemented, it is also very likely that it can be extended to cover more special, but important phenomena like object-shift, where light, unstressed subjects and complements like pronouns demand to be placed before any sentence adverbials in the nexus field.

³The feature name NEX was already taken.

⁴The distinction between *strictly-final-mod* and *final-mod* is made for purely technical reasons, to avoid spurious parses in cases where an adverbial can be extracted from more than one position in the sentence.

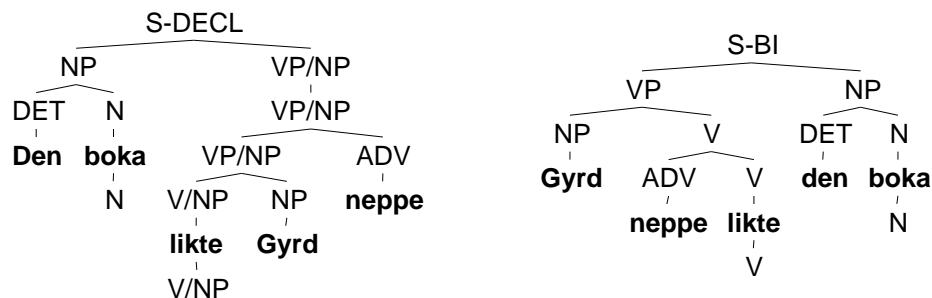


Figure 4: Main and subordinate clause structure, type of structure decided by adverbial placement

References

- Bender, E. M., Flickinger, D., & Oepen, S. (2002). The grammar Matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammar. In *Proceedings of the 19th international conference on computational linguistics*. Taipei, Taiwan.
- Copestake, A. (2001). *Implementing typed feature structure grammars*. Stanford, California: Center for the Study of Language and Information.
- Copestake, A., & Flickinger, D. (2000). An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the 2nd linguistic resources and evaluation conference* (pp. 591–600). Athens, Greece: ELDA.
- Diderichsen, P. (1962). *Elementær dansk grammatik* (3 ed.). København: Gyldendal.
- Faarlund, J. T., Lie, S., & Vannebo, K. I. (1997). *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Ginzburg, J., & Sag, I. A. (2001). *Interrogative investigations: the form, meaning, and use of English interrogatives*. Stanford, California: CSLI Publications.
- Sag, I. A., & Wasow, T. (1999). *Syntactic theory: A formal introduction*. Stanford: Center for the Study of Language and Information.