

Specification in the Scandinavian Grammar Matrix

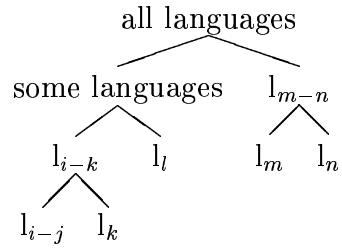
Anders Søgaard & Petter Haugereid
Linguistisches Graduiertenkolloquium, Universität Bremen

April 28-29 2005

- What motivates a grammar matrix?
- What principles should guide the construction of such a matrix?
- Machine translation as evaluation ... (demonstration)

Motivation

If languages have anything in common, and closely related languages have even more in common, this can be illustrated by an inheritance hierarchy:



... e.g. $l_k = \text{all languages} \cup \text{some languages} \cup l_{i-k} \cup l_k$.

Since Danish, Norwegian and Swedish are closely related languages, a matrix is motivated by (i) modularity, (ii) compatibility and (iii) multilingual applications.

Principles

The simple Scandinavian noun phrase, i.e. with one semantic index only, (roughly) has the word order: ?Q(uantifier), ?D(eterminer), ?Poss(essive), ?Num(eral), (*Adjectives), ?N(oun), as illustrated by:

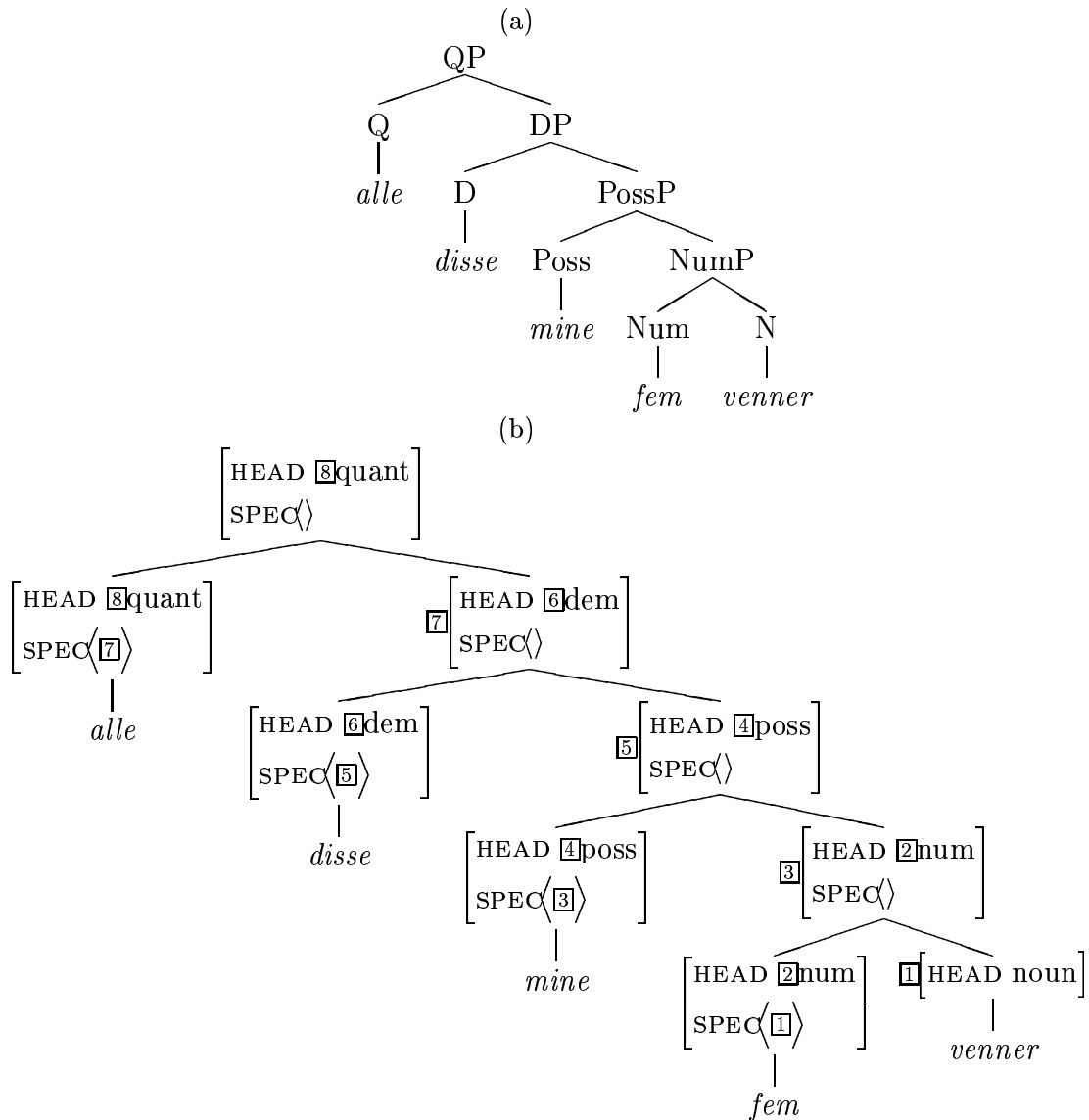
- (1) alle disse mine fem venner (all these my five friends)
- (2) alle mine fem venner (all these five friends)
- (3) alle disse mine venner (all these my friends)
- (4) alle disse fem venner (all these five friends)
- (5) alle disse mine fem (all these my five)
- (6) ...

The point here is: *all the constituents are optional, but the order of specifiers is fixed.*

Specification can be analyzed in two ways. The first approach is called the Right Branching Analysis (RBA) here, while the second is called the Bundling Analysis.

Right Branching Analysis

The analysis (a) is the naïve phrase structure version of the Right Branching Analysis, but compare it to the full projection of the Icelandic DP in Vangsnes (2004): “[KP* [DP [CardP [Gen²P [Gen¹P [P²P [P²P [aP* [n²P [n¹P [DxP [WP [NP”]. The analysis (b) is found in Neville (2003), but see also van Eynde (2003).



which amounts to:

$$\text{R1 } QP \rightarrow Q DP$$

$$\text{R2 } DP \rightarrow D PossP$$

$$\text{R3 } PossP \rightarrow Poss NumP$$

$$\text{R4 } NumP \rightarrow Num N$$

$$\text{R5 } QP \rightarrow Q PossP$$

$$\text{R6 } QP \rightarrow Q NumP$$

$$\text{R7 } QP \rightarrow Q N$$

$$\text{R8 } DP \rightarrow D NumP$$

$$\text{R9 } DP \rightarrow D N$$

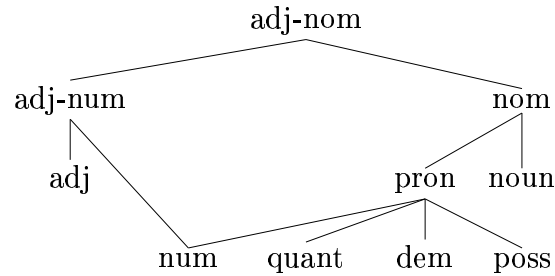
$$\text{R10 } PossP \rightarrow Poss N$$

or:

$$\text{H1 } X''[\text{SPR } \langle \rangle] \rightarrow \square Y''[\text{SPEC } \square] + \square X'[\text{SPR } \langle \square \rangle]$$

Right Branching Analysis (HPSG)

While the Head Specifier Schema allows her to minimize the set of grammar rules radically, Neville (2003) is instead forced to introduce multiple lexical entries. This is a consequence of the way she defines the syntactic categories (the type hierarchy has been simplified somewhat):



The lexical entries will be:

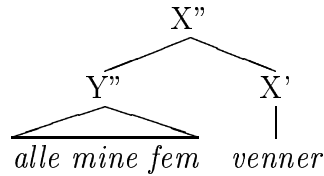
alle ₁	[SPEC dem]
alle ₂	[SPEC poss]
alle ₃	...

It is not possible to underspecify the lexical entry directly, because the immediate supertype is “pron” which also subsumes “quant”. (In addition, if higher types are used for specification, word order is left unconstrained.) The total number of lexical entries for our fragment is then 11.

Bundling Analysis

Another possibility subsumed by H1 is to bundle the specifiers in a Specifier Phrase:

$$H1 \quad X''[\text{SPR } \langle \rangle] \rightarrow \boxed{1} Y''[\text{SPEC } \boxed{2}] + \boxed{2} X'[\text{SPR } \langle \boxed{1} \rangle]$$



The trouble then is to decide what the head of the Specifier Phrase is. Intuitively, what you would like to say is that any of the specifiers may function as the (single) head, and all subceding ones function as adjuncts. Unfortunately, the adjuncts are then unsaturated. Eventually this analysis thus requires a duplication of the lexical entries. The specifier constituents must be entered as both specifiers and adjuncts. The total number of lexical entries under this analysis is 9.

Computational Grammars

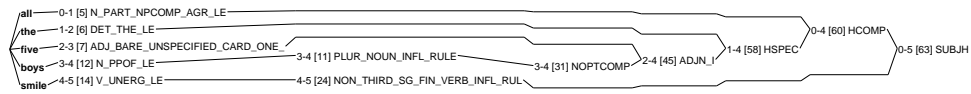
The English Resource Grammar (Flickinger, 2002) sticks to the naïve structure, but implements some bundling:

E1 [all [the [five boys]_{adj-n1}]_{h-spec}]_{h-comp}

E2 [all [my [five boys]_{adj-n1}]_{h-spec}]_{h-comp}

E3 [[all my]_{hspechc} [five boys]_{adj-n1}]_{h-spec}

e.g.,

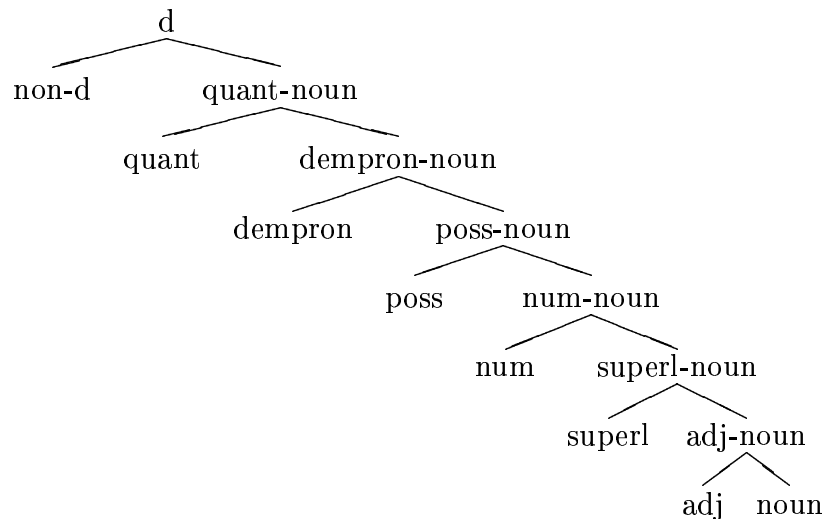


Our approach

Our approach is an improvement of the HPSG-version of the Right Branching Analysis. Only one rule is introduced to handle the fragment, but without multiplication of lexical entries. This is done by altering the type hierarchy of Neville (2003). In addition, it is shown how this rule - call it the Extended Head Specifier Rule - can be even further generalized to cover adjuncts in noun phrases (a similar claim, but in different terms, is presented by van Eynde, 2003), e.g.:

(7) alle disse mine fem gode venner (all these my five good friends)

The type hierarchy of relevance here is (“d” is not the value of HEAD, but the head factor D, e.g. $\begin{bmatrix} D & d \\ V & \dots \end{bmatrix}$; this is only motivated by facts outside of our fragment):



References

- van Eynde, Frank. 2003. Prenominals in Dutch. Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar. Seoul, Korea.
- Flickinger, Dan. 2002. On building a more efficient grammar by exploiting types. In Stephan Oepen et al. (eds.), Collaborative Language Engineering. Stanford: CSLI.
- Neville, Anne. 2003. Danish pre-nominals as specifier heads. Proceedings of Scandinavian Conference of Linguistics 19. Tromsø, Norway. *Also in:* Tromsø University Working Papers on Language & Linguistics 31 (2).
- Vangsnes, Øystein. 2004. Rolling up the Scandinavian noun phrase. Proceedings of the 27th Generative Linguistics in the Old World. Thessaloniki, Greece.