Syntactic Theory on Swedish

Mats Uddenfeldt Pernilla Näsfors

June 13, 2003

Report for Introductory course in NLP

Department of Linguistics
Uppsala University
Sweden

Abstract

Using the grammar presented in the second edition of "Syntactic Theory: A Formal Introduction" by Ivan A. Sag, Thomas Wasow and Emily M. Bender (2002) we studied a few Swedish grammatical structures and their semantic interpretation. This report contains examples and discussions regarding ambiguity problems for preposition phrases and noun phrases and their different semantical interpretations. It ends with a study of the Swedish gender agreement and discusses modifications to the grammar in regard to the required features.

Keywords: ambiguity, preposition phrases, coordination, noun phrases, gender agreement

Authors:

Mats Uddenfeldt (Mats.Uddenfeldt.9771@student.uu.se)

Pernilla Näsfors (Pernilla.Nasfors.5150@student.uu.se)

Supervisor:

Roussanka Loukanova (rloukano@stp.ling.uu.se)

Contents

1 Background

We are both studying computer science so when we first got in touch with Natural Language Processing (NLP) we had our eyes set on finding methods that would allow us to efficiently implement natural languages using computers. We had some previous experience of Context Free Grammars, since we had studied both automata theory and compiler techniques, but we realized this method was not computationally efficient for this purpose despite its logical appearance. CFG:s require a vast number of rules to implement a natural language, most of which would only be duplicates of others with minor changes (e.g. in person or number).

So after hearing about feature structures and unification grammars we were intrigued with the exciting possibilities it offered. Our main sources of information have been the second edition of "Syntactic Theory: A Formal Introduction" by Ivan A. Sag, Thomas Wasow and Emily M. Bender (2002) and on our lectures with Roussanka Loukanova from the Department of Linguistics at Uppsala University. At first we had planned to do a presentation of feature structures in the form of a review of the first chapters of the book, but as the course progressed we decided it would be more interesting to see if we could apply the grammar from the book on a few chosen Swedish grammatical structures. Seeing how the rules applied to our native tongue would give us more understanding of our own language and the principles behind feature structures, as well as finding out which rules would have to be modified to fit the Swedish grammar.

We have not done any statistic evaluation of our interpretations, but have used our friends and family to the extent it was possible. When we discuss how native readers and speakers react, this is our reference point. Validation of our results is another issue. This report contains our results in the form of discussion. Finally we provide a short list of other interesting Swedish sentence structures that we had to leave for the future.

2 Preposition phrases

Preposition phrases (PP:s) form an ambiguity problem since they can both be interpreted as modifiers and complements to a corresponding verb phrase. This raises the question of how we should analyze a particular PP. We can either introduce it as a complement licensed by a PP on the COMPS list of a verb or we can analyze it as a modifier introduced by another rule in the grammar. In some cases it is obvious which we should choose, since some verbs require a PP beginning with a certain word to function properly. This is true for the verb *lita* (Eng: rely), which requires the preposition *på* (Eng: on/upon).

(1) Hon litar på honom.
She relies on him.
*Hon litar.

In these mandatory cases we can safely choose to add the PP as a complement to the verb, but in other cases it is not as obvious as to how a PP is to be regarded. We wondered if it was possible to find a pattern for the semantic interpretation by testing a couple of sentences that should ensure ambiguity. We chose to study the Swedish preposition *på* (Eng: on/for) because it illustrates this ambiguity in a very good way; It is a very common preposition. It may be interpreted as the head of both modifier and complement PP:s. The verb *vänta* (Eng: wait) was chosen because it has both transitive and intransitive use. Although there exists a connection between the preposition *på* and the verb *vänta* we considered both interpretations to be equally frequent. In other words there is no easy way out as in (1) because semantically neither would have any advantage over the other.

(2) Hon väntar.
She waits.

If we translate the transitive use of *vänta* with the preposition phrase *på tåget* to English we end up with two possible translations. These were then used to create the ambiguous examples.

(3) på tåget
for the train
She waits for the train. (PP treated as a complement)

(4) på tåget
on the train
She waits on the train. (PP treated as a locative modifier)

2.1 Examples

The examples are first given in Swedish with corresponding English translations. We have given the translations a value between 1 and 5 to represent the likelihood of either semantical interpretation of the given examples.

(5) a. Hon väntar på tåget.

b. She waits for the train. (3/5)

c. She waits on the train. (2/5)

(6) a. Hon väntar på honom.

b. She waits for him. (4/5)

c. She waits on top of him. (1/5)

(7) a. Hon väntar på honom på tåget.

b. She waits for him on the train. (5/5)

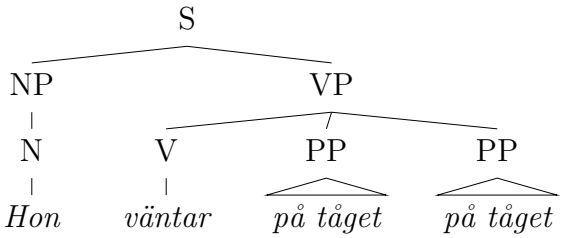
c. She waits on top of him for the train. (0/5)

(8) a. Hon väntar på tåget på tåget.

b. She waits for the train on the train. (5/5)

c. She waits on the train for the train. (0/5)

(9)



2.2 Discussion

As we can see in the examples above there seems to be an inclination to interpret a PP directly following the verb as a complement. We are used to modifiers being optional, so our intuitive interpretation takes this into account.

However in (6a) the connection is especially strong. This is because the modifier interpretation (6c) where she "waits on/on top of him" has a very weak inference with the reader. Of course she could be waiting for the train while sitting on someone, but it is very unlikely for a reader to have experienced this. But why are we more likely to interpret the single PP in (5a) as a complement? The inference in this case is equally strong for both interpretations. We will see if we can answer this by looking at the other examples.

When we combine the first two sentences into (7a) we get a more interesting case. A Swedish reader will always interpret the meaning of this sentence as shown in (7b). This follows the same logic as when we discussed the first two sentences apart, but put together the opposite interpretation is not even considered possible. When we have two PP:s it becomes all the more clear as to which is complement and which is modifier due to extended context given. On the other hand we have the special case of (8a) which is also illustrated in the tree structure (9). This example is especially interesting since the two PP:s are in fact identical. Even though the "sum" of the two interpretations is the same in both interpretations, we found that a native reader interprets the sentence in a given order. The reader will seek to attach a complement to the verb before looking for a modifier. This implies that phrase structure is a key to understanding how PP:s are interpreted semantically. We can try this theory by combining (5a), which is ambiguous, and (6a), which is very likely to be interpreted as a complement, but placing (6a) in the "modifier" position:

(10) a. *Hon väntar på tåget på honom.

b. She waits for the train on top of him.

c. *She waits on the train for him.

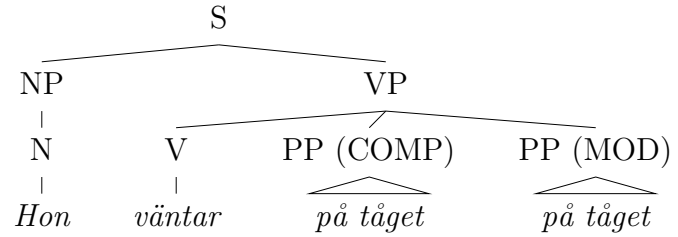
A native reader will react strongly to this sentence. It sounds terribly wrong even though it is in fact grammatically correct. The first PP is immediately interpreted as a complement when read, which leaves the final PP to be interpreted as a modifier. Since the inference is so weak for this interpretation, it becomes unsemantical. In reality the reader would perhaps ignore this error and interpret the sentence as in (10c).

If we attach two PP:s to the verb, with one acting as a complement and the other as a modifier, it surely seems as if there is no real semantic ambiguity. In this case the first PP is always interpreted as the complement. This connection is most likely the cause of the increased likelihood for a single

PP as in (5a) to be interpreted semantically as a complement, even though it is equally "likely" that it was meant to be a locative modifier. For spoken sentences we have a slightly different situation since intonation can be used to emphasize the complement even though it is placed some distance from the verb.

So what can we conclude from this? After examining the patterns of how PP:s act in regard to verbs in Swedish we think that it can be regarded as safe to interpret the first of several PP:s in written text as a complement, as long as the verb is compatible of course. Thus the tree structure in (9) would have the following interpretation:

(11)



3 Coordination of noun phrases

Coordination of noun phrases (NP:s) is an interesting construction in natural language which we thought would be good to explore further. Most often the scope of attributes to an NP is clear to a human reader, but it is not always as easy to analyze computationally. If you have a strictly syntactical point of view you will always end up with ambiguity. The grammar in "Syntactic Theory: A Formal Introduction" has features for semantics so we can cancel some ambiguities without actual knowledge of the context. Qualified guesses can almost always be made. We will look at the following types of coordinations: distributive coordinations of non-full NP:s and distributive coordinations of full NP:s.

3.1 Examples

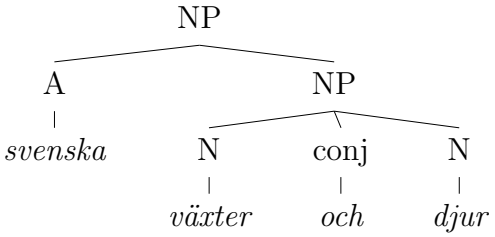
The examples are first given in Swedish with corresponding English translations. Coordinations can theoretically contain any number of items using commas as conjunctions, but we have limited our examples to conjunctions of two entities. They cover the following types of coordination: distributive of non-full NP:s (12) and (13) and distributive of full NP:s (16), (14), (15) and (17).

(12) min plånbok och mobiltelefon
my wallet and cellular phone

(13) mitt körkort eller pass
my driver's license or passport

(14) svenska växter och djur
Swedish plants and animals

(15)



(16) Pernillas lärare och vänner
Pernilla's teachers and friends

(17) Visa ansökan och betyg för Pernilla och Mats.
Show application and grades for Pernilla and Mats.

3.2 Discussion

The most simple example of coordination of NP:s are when you attach two nouns using a conjunction. Taking this one step further we add a distributed determiner to the conjunction as in sentence (12) and (13). For the grammar to accept this unification the word has to agree in both form and case. All determiners in Swedish are different for neuter (Sv: neutrum) and non-neuter (Sv: utrum) words. This means we also have to require agreement in gender. We will return to this question in the next chapter. In example (12) we have *telefon* and *plånbok*, both non-neuter nouns. In example (13) we have *körkort* and *pass*, both neuter nouns. Consider the following examples with the possessive determiners *min* (non-neuter) and *mitt* (neuter):

(18) a. Min mobiltelefon och plånbok.

b. *Mitt mobiltelefon och plånbok.

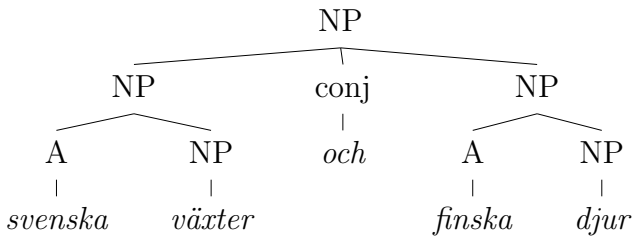
c. *Mitt körkort eller plånbok.

d. Mitt körkort eller pass.

Distributive coordination of full NP:s pose a scope ambiguity problem. Since each part of the coordination can function as an NP on its own, how do we know the scope of a given attribute (14) or possessive (16). We found that if you insert a new attribute or possessive to the left of another noun in the sentence you break the distribution.

(19) svenska växter och finska djur
Swedish plants and Finnish animals

(20)



To most people there is no doubt of what is meant in sentence (17). It is clear that the person uttered this statement wants to retrieve information about four things: *ansökan för Pernilla, betyg för Pernilla, ansökan för Mats* and *betyg för Mats*. In other words the PP attribute *för Pernilla och Mats*, which contains a coordination itself, scopes over the whole NP coordination

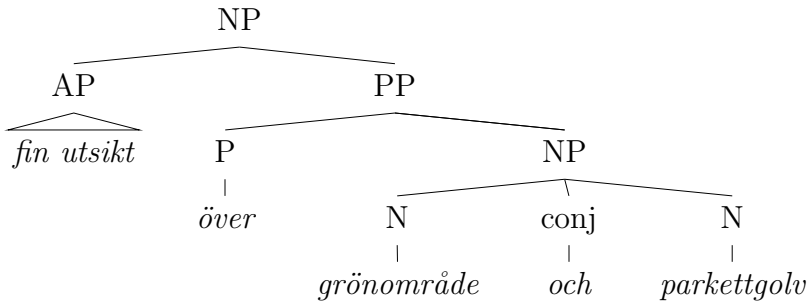
ansökan och betyg, and not only over the preceding noun *Pernilla*. This decision is based on our knowledge that applications and grades are something that is connected to people. From a computational point of view we have no knowledge of this so we cannot eliminate the ambiguity. The other interpretation asks for three things: *ansökan för Pernilla, betyg för Pernilla* and *Mats*. If you look at example (21) this alternative interpretation is more understandable.

(21) Visa X och Y för Z och W.
Show X and Y for Z and W.

How do we reach this conclusion using computational techniques? One way would be to add semantic information to each word in the dictionary and add corresponding requirements when analyzing coordination of NP:s. But would this be enough? Consider the next example found in a Swedish newspaper:

(22) Lägenheten har fin utsikt över grönområde och parkettgolv.
The apartment has a nice view over a park and a parquet floor.

(23)



Canceling ambiguities of the type shown in (23) is very difficult. A human reader can understand that the notice simply wants to inform us of the existence of the parquet floor. It would be interesting to devote more time into finding a computational solution. However, it is beyond the scope of this report.

4 Gender agreement

The most obvious difference between English and Swedish is that all common nouns in Swedish are assigned one of four gender categories: neuter, non-neuter, feminine or masculine. Gender distinctions in Swedish are shown on determiners, adjectives and pronouns. This leads to difficulties in gender agreement.

4.1 Examples

The examples are first given in Swedish with corresponding English translations. Words in the feminine and masculine gender categories often behave like non-neuter words, so when we don't mention feminine and masculine we refer to them as non-neuter. Since there are many distinctions here that are not obvious in the English translation we have decided to comment the examples.

(24) ett bord
a table

(25) en stol
a chair

For determiners you have two different forms for gender. The determiner *ett* in (24) is used for nouns in neuter, while the determiner *en* in (25) is used in non-neuter.

(26) bordet
the table

(27) stolen
the chair

For definite nouns, it is a bit more complicated. In Swedish you do not always need to use a determiner like *the* to put nouns in definite form. What you do to put a noun in definite form, as you can see in example (26) and (27), is to append a suffix to the original noun. The pattern for this differs from word to word, but basically you put *-et* or *-t* after nouns in neuter and *-en* or *-n* after nouns in non-neuter. If you build a sentence in definite form with only a determiner and a definite noun, the determiner takes on a more concrete demonstrative meaning:

(28) det bordet
that table

(29) den stolen
that chair

For adjectives you have several different forms for gender, but in indefinite form you only have the two possibilities with neuter like in (30) and non-neuter like in (31). This gender distinction is on the other hand only applicable to the singular forms of the words. In plural, (32) and (33), you have only one option for the adjective. The same plural form *gamla* is also used for all definite adjectives in plural form.

(30) ett gammalt bord
an old table

(31) en gammal stol
an old chair

(32) gamla bord
old tables

(33) gamla stolar
old chairs

Finally we have the definite adjectives in the singular form and this is the first place where you can see a difference between neuter, the common non-neuter, the feminine and the masculine. As you can see in these examples we have the same form *gamla* in (34) for neuter, in (35) for non-neuter and in (36) for feminine, but in (37) for masculine you have the form *gamle*.

(34) det gamla bordet
the old table

(35) den gamla stolen
the old chair

(36) den gamla kvinnan
the old woman

(37) den gamle mannen
the old man

Among the pronouns you can also find examples of differences in gender. Here you can see that you have the pronoun *detta* in (38) for neuter, *denna* in (39) for non-neuter and in (40) for feminine and in (41) for masculine you have the pronoun *denne*.

(38) detta bordet
this table

(39) denna stolen
this chair

(40) denna kvinnan
this woman

(41) denne mannen
this man

4.2 Discussion

Due to the gender distinctions for determiners, nouns, adjectives and pronouns it could be quite complicated to create gender agreement for Swedish grammar. The grammar in chapter 4 of "Syntactic Theory: A Formal Introduction" would have to be expanded in several ways. In the type hierarchy, we would have to add a GEND(er) feature to *noun*, *det* and *adj*. GEND would have to be expanded so it would contain *neut, non-neut, fem, masc*. The gender agreement would help us to distinguish which constructions that are correct:

(42) a. den gamle mannen

b. *det gamle mannen

c. *den gamla mannen

d. *den gamle mannet

We would also have to add a feature that would tell if nouns, determiners and adjectives are in definite form or not. We could call this feature DEFIN and create it in a similar way as the feature COUNT described in section 4.6.3 in "Syntactic Theory: A Formal Introduction". Definite nouns and adjectives and the determiners *den* and *det* would be specified as [DEFIN +] and indefinite nouns and adjectives and the determiners *en* and *ett* would be specified as [DEFIN -]. This feature would force nouns in definite form to co-occur with a definite determiner and would stop constructions like these:

(43) a. *ett bordet

b. *det bord

c. *ett gamla bord

d. *det gammalt bordet.

We finish this section with two examples of lexical entries that include these new features. The non-definite determiner *en* is only for singular nouns in non-neuter, feminine or masculine form. The noun *stol* would be compatible with this determiner, as we can see from the examples:

(44) a.

$$\left\langle \text{en}, \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \textit{det} & \\ \text{AGR} & \begin{bmatrix} \text{NUM} & \textit{sl} \\ \text{GEND} & \textit{non-neut} \mid \textit{fem} \mid \textit{masc} \end{bmatrix} \\ \text{COUNT} & + \\ \text{DEFIN} & - \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

b.

$$\left\langle \text{stol}, \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \boxed{1}\begin{bmatrix} \text{PER} & \textit{3rd} \\ \text{NUM} & \textit{sl} \\ \text{GEND} & \textit{non-neut} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & \left\langle \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{det} & \\ \text{AGR} & \boxed{1} \\ \text{COUNT} & + \\ \text{DEFIN} & - \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

5 Summary

5.1 Preposition phrases

We looked at the ambiguity problem when a preposition phrase can either be interpreted as a complement or as a modifier to a corresponding verb phrase. Studying example sentences with the Swedish preposition *på* (Eng: on/for) and the verb *vänta* (Eng: wait) we arrived at the conclusion that the inclination to interpret a single PP directly following the VP as a complement can be expanded into a rule whenever there are two PP:s present and the verb in question has a compatible COMPS list. The only exception to this rule occurs in spoken Swedish. If we place intonation on either preposition it becomes emphasized and takes on the role of a complement despite being placed at some distance of the verb. We leave that for future discussion.

5.2 Coordination of noun phrases

Coordination is a very interesting construction in natural language and it poses a number of ambiguity problems. We constructed examples of the following types of coordination: distributive coordination of non-full NP:s and distributive coordination of full NP:s. By investigating the scope ambiguity problem we found that you could break the scope by inserting an attribute to the left of a previously coordinated noun. Finally we looked at some tricky semantic ambiguities that can arise when coordinating NP:s to the right of preposition phrases.

5.3 Gender agreement

Gender agreement is a very important part of the Swedish language and we tried to study different examples of how gender distinction affects determiners, adjectives and pronouns. Swedish have four gender categories which would have to be included in the grammar: neuter, non-neuter, feminine or masculine. We also found the need of feature to distinguish definite forms of words from others: [DEFIN +] or [DEFIN -]. Finally we constructed two lexical entries to show the applications of these changes.

5.4 Further work

Since this was only our first glance at the discipline of Natural Language Processing we found many interesting things that could have made it into this report, but had to be left out. We figured it would be a good idea to

publish the sentence structures we had in mind so that we or others may use it as reference for future studies. To make a long list short here it is:

1. Number agreement on adjectives, adverbials, determiners, verbs.
2. Investigate lexical entries that are modifiers (adverbials, adjectives, ...)
3. Investigate phrasal modifiers for different phrasal categories
4. Adjectives after noun (Mats röde, Lasse liten, flicka lilla)
5. Possessives after noun (huset ditt, flickan min)
6. Adverbials before the verb they modify (subclauses: som inte sprang, ofta cyklade omkull)
7. Number distinctions on determiners (individual: en öl, collective: ett öl)

5.5 Final words

We would like to thank Roussanka Loukanova for assisting us through the course and steering our curious minds in the right direction. This introductory course has certainly given us a taste for NLP and hopefully we will receive another dose the next semester. Now it is time to break up for summer and see what ambiguities real life will present us with.

Uppsala, June 13 2003

Mats Uddenfeldt and Pernilla Näsfors