

1 Phrase structure trees and latex issues

I started this assignment by looking at the structure of the grammar in English.dbnf to see the type of format that was expected for the output. Then, I chose three sentences to analyse, in both English and Swedish, in the way I would expect/hope they would be analysed by that grammar. I was hoping to be able to compare them to the output of the parse2pdf command later in the assignment. However, I was working on eduserv, which does not have access to my computer screen to show me the pdf output. I tried to solve this by using the parse2latex command instead and then paste the result in overleaf, but the command took a very long time to run, and when I compiled the resulting file the output was in UD tree format and not phrase structure trees as I was expecting (or maybe I just misunderstood how to use the commands). Finally, I just gave up and drew the trees by hand based on the bracket notation in the automatically parsed file. This had the added benefit of making them easier to compare to my hand-drawn gold standard trees. The resulting trees can be seen in §4.

2 Testing the grammar

I extracted the words and their parts of speech from the annotated corpora using

```
cat comp-syntax-corpus-english-std.conllu | gfud extract-pos-words  
> extracted-pos
```

and

```
cat comp-syntax-corpus-swedish-std.conllu | gfud extract-pos-words  
> extracted-pos-swe
```

Then I used the grammar English.dbnf to parse the sentences:

```
cat extracted-pos | gfud dbnf English.dbnf Utt > new-english.conllu  
and
```

```
cat extracted-pos-swe | gfud dbnf English.dbnf Utt > new-swedish.conllu
```

These files can be found in the GitHub repository. Finally, I evaluated the result against the original conllu files:

```
gfud eval macro LAS comp-syntax-corpus-english-std.conllu new-english.conllu  
and
```

```
gfud eval macro LAS comp-syntax-corpus-swedish-std.conllu new-swedish.conllu
```

The result for English was

```
udScore = 0.7013706905011253, udMatching = 24, udTotalLength =  
339, udSamesLength = 237, udPerfectMatch = 3
```

and for Swedish

```
udScore = 0.5619417537083945, udMatching = 24, udTotalLength =  
304, udSamesLength = 169, udPerfectMatch = 3
```

3 Modifying the grammar

I did some small modifications to the grammar (mainly to the lexicon), used it on the pos-tagged words and evaluated the resulting parses using `gfud eval` again. Unfortunately, the result was even worse than for the English grammar:

```
udScore = 0.5411084203750612, udMatching = 24, udTotalLength =  
304, udSamesLength = 165, udPerfectMatch = 2
```

One possible reason for this (and also a reason why I did not change much in the grammar) is that English and Swedish are already quite similar grammatically, and changing the grammar to reflect the ways in which I know that they differ unavoidably caused problems somewhere else in the grammar, without making much of a positive difference. For example, I tried making some changes to accommodate the Swedish V2 rule, but I could not figure out how to make that work without messing up some other part of the grammar. While I think that it is definitely possible to modify the grammar in a way that makes it better, or even write a completely new one, it is quite a project and frankly I don't think I could do it within the given time frame.

For comparison, I also used the Swedish grammar from `gf-ud` and evaluated the result in the same way:

```
udScore = 0.5701961729003138, udMatching = 24, udTotalLength =  
304, udSamesLength = 169, udPerfectMatch = 3
```

This worked slightly better than the English grammar (and much better than my modified Swedish grammar), but the improvement is small enough that I think that I would have had to work very hard and spend a lot of time and effort to be able to achieve a similar improvement by modifying the English grammar.

4 Phrase structure trees

Below are the gold standard phrase structure trees I drew based on my understanding of the sentences, and the corresponding automatic parses (based on the original English grammar).

The most obvious error it makes is in distinguishing between questions and declarative sentences, which shows in how it e.g. misidentifies interrogative pronouns as expletives (2), and other pronouns as interrogatives (8).

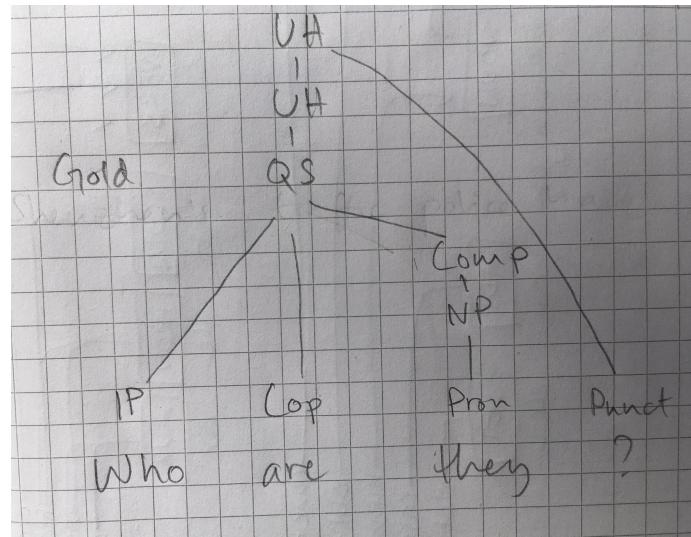


Figure 1: Gold standard parse for "Who are they?"

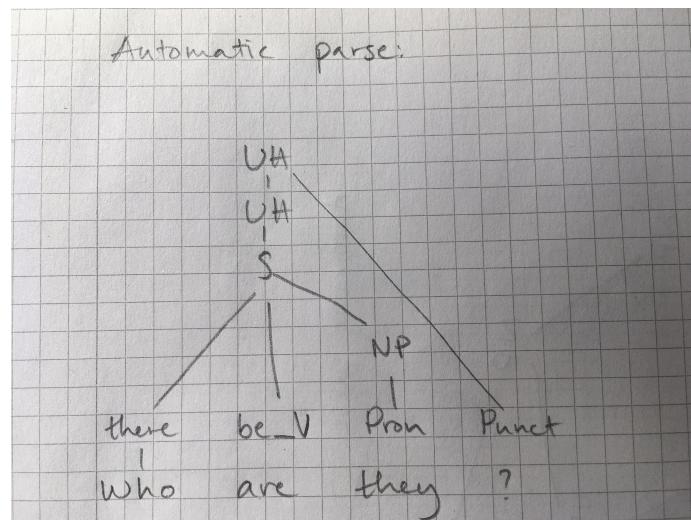


Figure 2: Automatic parse for 1.

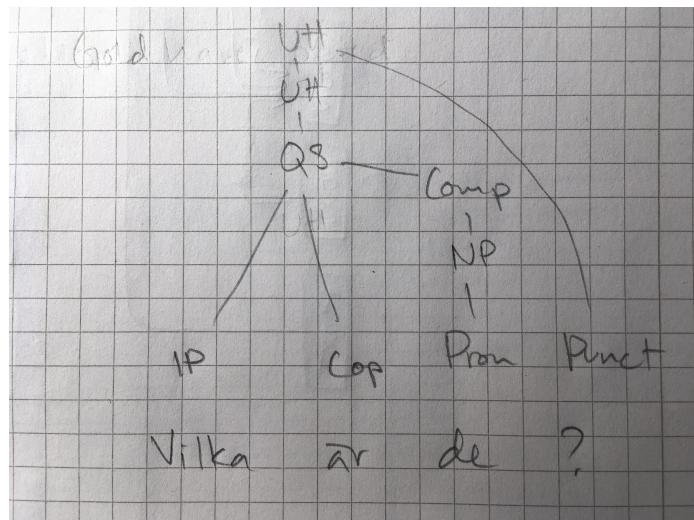


Figure 3: Gold standard parse for the Swedish version of 1.

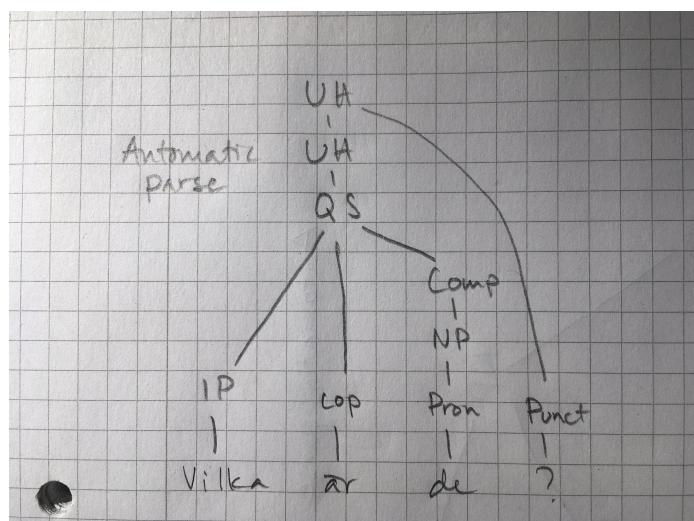


Figure 4: Automatic parse for the Swedish version of 1.

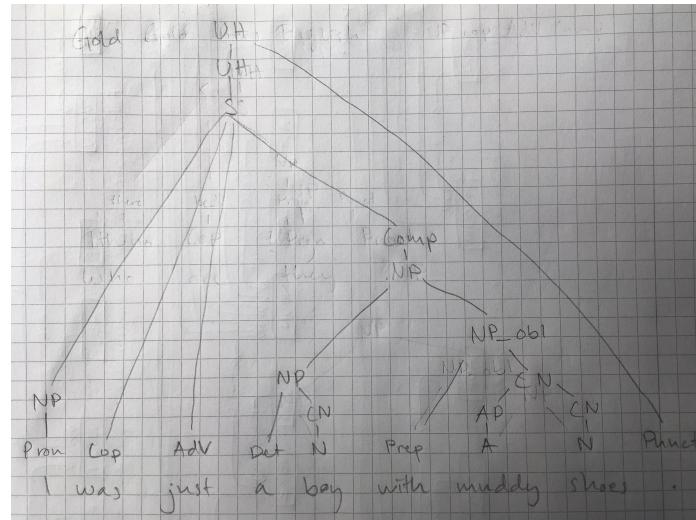


Figure 5: Gold standard parse for "I was just a boy with muddy shoes"

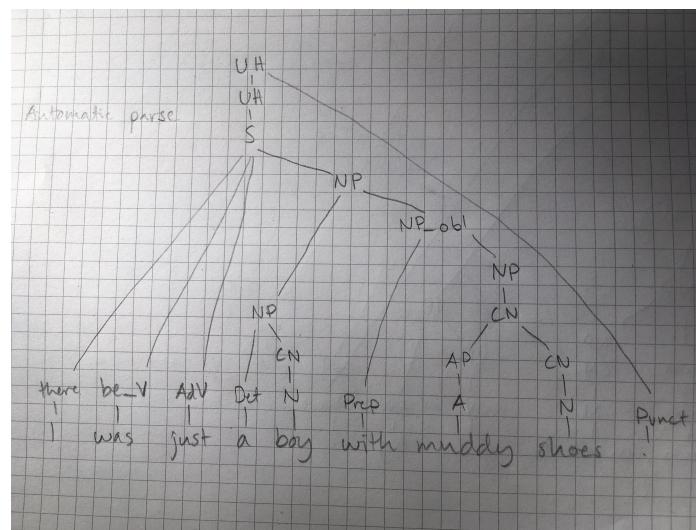


Figure 6: Automatic parse for 5

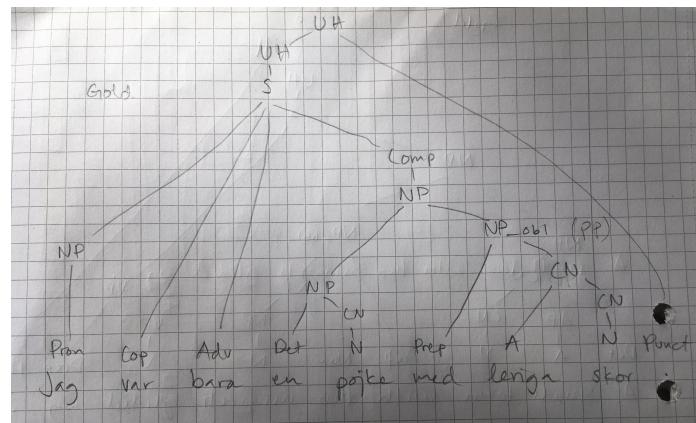


Figure 7: Gold standard parse for Swedish version of 5.

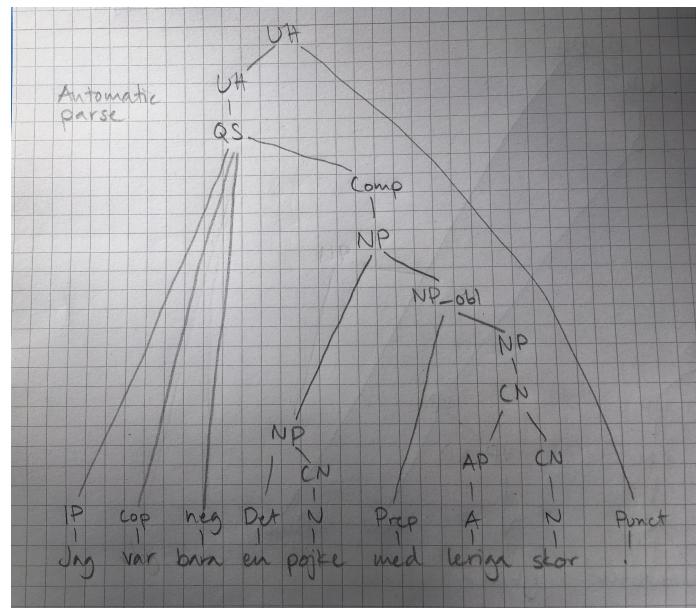


Figure 8: Automatic parse for Swedish version of 5.

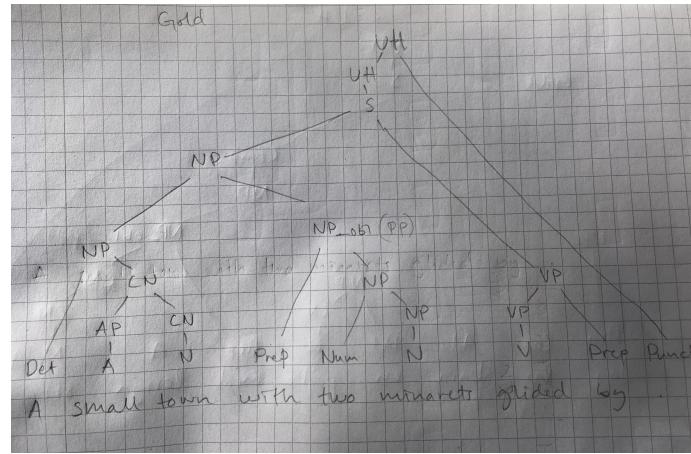


Figure 9: Gold standard parse for "A small town with two minarets glided by."

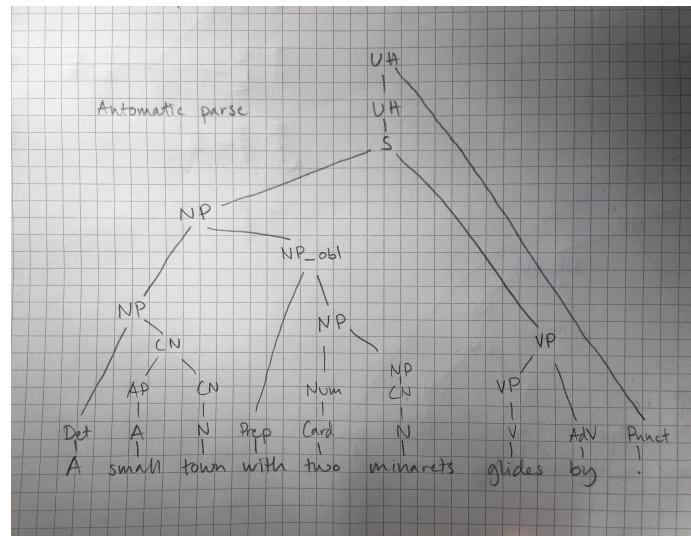


Figure 10: Automatic parse for 9.

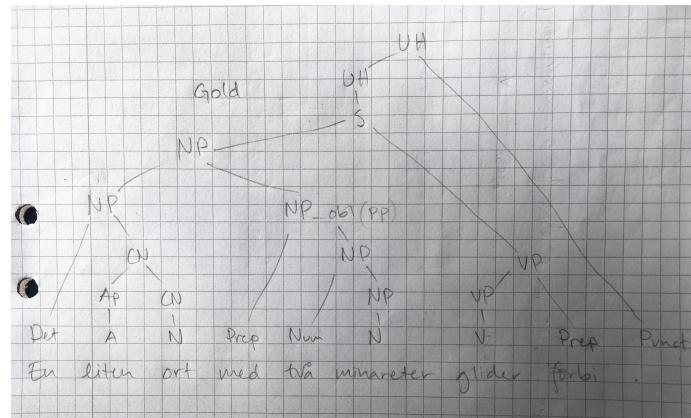


Figure 11: Gold standard parse for Swedish version of 9.

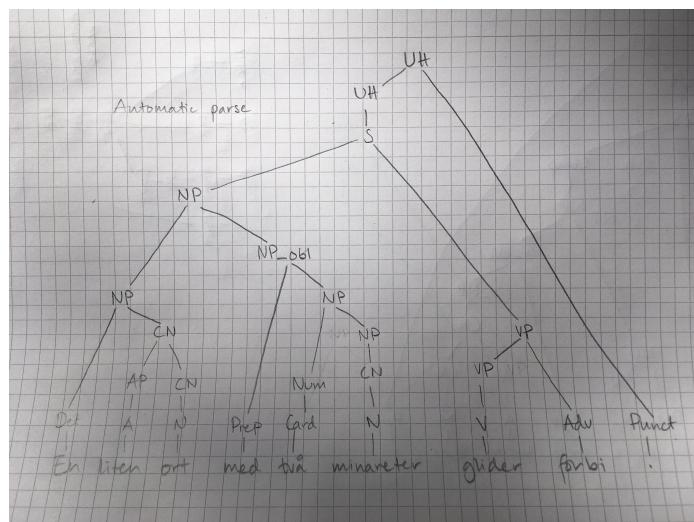


Figure 12: Automatic parse for Swedish version of 9.