Methods in Computational Linguistics
Master of Science *Computational Linguistics*
Exercise: Parsing

apl. Prof. Dr. Sabine Schulte im Walde
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart

January 27+28, 2020

In this exercise, you manually apply various parsing algorithms relying on constituency and dependency grammars, and you run a dependency parser.

1 Parsing – Do it Yourself

(1) Write down the steps that a *top-down parser* would perform to parse the given sentence, if it always chooses the top-most rule first.

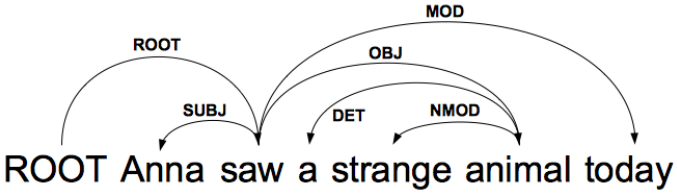
Input sentence: "the man saw a strange animal on the street"

Grammar:

1. S → NP VP
2. NP → DET N
3. NP → DET ADJ N
4. NP → DET ADJ N PP
5. VP → V NP
6. VP → VP PP
7. PP → P NP
8. DET → the
9. DET → a
10. ADJ → strange
11. N → animal
12. N → man
13. N → street
14. V → saw
15. P → on

(2) Write down the steps that a *bottom-up parser* would perform to parse the same input sentence with the same grammar, if it always chooses the top-most rule first.

(3) Write down the transition sequence of the *arc-standard parsing strategy* for the following tree:



2 Parsing: Application

Your task is to apply the MATE dependency parser to either German or English input sentences (your choice).

Here's a starting point to access the online interfaces and some documentation:
`https://www.ims.uni-stuttgart.de/en/research/resources/tools/matetools/`.

The interfaces themselves are available here: `http://de.sempar.ims.uni-stuttgart.de/` for German and here: `http://en.sempar.ims.uni-stuttgart.de/` for English.

(1) Generate sentences for the language of your choice that include the following linguistic constructions:

(a) PP-attachment
(b) Passivisation
(c) Verb complex including a modal verb
(d) Relative clause

(2) Parse the sentences. The output is provided in `tab-separated CoNNL format`, containing

- the position of the word (column 1),
- the word form (column 2),
- the lemma (column 4),
- the part-of-speech (column 6),
- the morphological analysis (column 8),
- the head of the word (column 10),
- the dependency relation (column 12), and
- duplicated/other information in the remaining columns that you can ignore.

(3) Choose at least two of your sentences and manually draw the dependency trees for the parses.

(4) Analyse the dependency structures in the parses:

- Describe how the parser behaves with regard to the four constructions specified in (1) above.
  (a) Where does the PP attach to?
  (b) Do surface subjects and objects attach to the main (passivised) verb?
  (c) Which is the head within a modal verb construction?
  (d) Which is the head of a relative clause, and where does the clause attach to?
- Does the parser parse your sentences correctly, or are there cases where you expected different parses?