Natural Language Processing
Assignment 3: Syntax

---

1 Introduction

This assignment involves material from classes 8, 9 and 10. You should have watched the relevant videos, read the relevant chapters in the textbook and made a serious attempt at completing the relevant labs before you attempt this assignment. If you feel you have done that and still find the instructions unclear, you are welcome to email the course teachers and/or go to office hours to ask for help. We indicate after each subtask how much we expect you to write. If you are interested in receiving a VG grade, you may complete the *optional* exercise available to you, which is related to syntactic formalisms. Note that the exercise must be completed in full in order to receive VG credit. Please do not submit more than 5 pages overall. All answers should be self-contained.

2 Assignment

The assignment is based on the sentences in the **en10** set, which you annotated in Lab 8, parsed in Lab 9 and wrote a grammar for in Lab 10:

1. He worked for the BBC for a decade.
2. She spoke to CNN Style about the experience.
3. Global warming has caused a change in the pattern of the rainy seasons.
4. I also wonder whether the Davis Cup played a part.
5. The scheme makes money through sponsorship and advertising.
6. If a Turkish employee quits, then the Turkish work councils come.
7. A witness told police that the victim had attacked the suspect in April.
8. Mr Osborne signed up with a US speakers agency after being sacked in July.
9. The RHS collected comments sent in by schoolchildren and teachers involved in the experiment.
10. National reaction to the events in Kansas demonstrated how deeply divided the country had become.

For this assignment, you should submit answers to the following questions:

1. Discuss three cases, involving (at least) three different dependency relations, where your annotation in Lab 8 differed from that of another group or from the gold standard. Cite definitions or examples from the UD guidelines to support your conclusion about which annotation is correct in each case. Are there any mismatches between your linguistic intuition and the UD guidelines? Are some sentences ambiguous and do you think they could receive different interpretations?

2. Discuss two different errors involving (at least) two different dependency relations in the output of the parser trained in Lab 9 for the **en10** sentences. Discuss possible reasons the parser made these mistakes.

3. The parser trained in Lab 9 uses the arc-eager transition system for dependency parsing described in Kübler et al. Briefly describe how transition-based parsing works and write down the transition sequence used to construct the parse tree for sentence 1 to illustrate your description. List the content of the stack (S), buffer (B) and arc set (A) after each transition.

4. List the grammar you designed in Lab 10 after verifying that it does in fact generate sentences 1–5 with reasonable analyses. (By reasonable, we mean that it would be an acceptable tree according to some linguistic theory, it does not matter which one.) Draw all the parse trees that the grammar assigns to sentence 3 and describe the essential differences between the trees. Discuss what parts of the tree you think are correct and why your grammar also produces incorrect trees. If your grammar only produces one tree, discuss why this is the case and possible advantages and drawbacks of your grammar.

5. Using the grammar from lab 10, draw the completed chart for a parse of sentence 1 using the CKY algorithm described in Jurafsky and Martin and explain briefly how it works, including mentioning in which order the cells are filled. (Hint: have a serious look at the textbook for this question! It is easy to miss a step!)

We expect between half a page and a page in response to each question. Please do not submit more than 5 pages overall. Note that figures are excluded towards the page count. Your answers to each question should be self-contained.

3   VG: Constituency vs. Dependency

Describe the concept of dependency and constituency within linguistics —both formally and in your own words. What do you perceive to be the advantages and disadvantages of each formalism — e.g. for languages with variable word order? Is one formalism a better representation for NLP tasks than another? Why? For which tasks?

4   Grading Criteria

Basic Criteria

- Answers are given in understandable English.
- Answers are stated clearly and coherently.
- Answers are essentially correct.

Additional Criteria

- Answers are well motivated.
- Answers are well illustrated.
- Answers reveal extensive knowledge of the textbook chapter(s).

To pass the assignment, you must meet all the basic criteria on all subparts of the assignment. To get VG, you must in addition meet some of the additional criteria for most of subparts.

5   Submission

Submit your assignment as a pdf file named firstname_lastname_assignment_3.pdf. It should follow the style and margins given in the example submission even if not created with LaTeX. See deadline on studentportalen.