Introduction to Computational Linguistics SS2016 - Homework Assignment 3
Due: Before the beginning of the lecture on 12 May, 2016

- Must not be handed in individually but **by workgroup**. Each submission must contain the names of all workgroup participants.
- Identical solutions from different workgroups will not receive credit (so don't copy the solutions from other workgroups and don't allow others to copy your solution).
- Pls. hand in in the form of a PDF document sent to your tutor by email.

1. Grammar Writing
A. Write a simple CFG that generates at least the following sentences, but (ideally) no ungrammatical sentences. If you can't avoid generating ungrammatical sentences, give examples of such ungrammatical sentences that your grammar generates and comment briefly on why it is hard to avoid them.

Bert admires Mary.
Charles eats hot chips with a fork.
John pets the small cat.

The terminal symbols of your grammar must be words, not phrases.

Provide a **formal specification of the grammar** in Chomsky normal form, giving
- the set of non-terminal symbols,
- the set of terminal symbols,
- the start symbol,
- and the set of production rules (same format as used in Exercise 2 below).

Use the following non-terminal symbols: S, NP, VP, PP, V, N' (a nominal expression that contains a noun; e.g., an adjective plus a noun; pronounced: "N-bar"), N, P (preposition), Adj (adjective), D (determiner), PN (proper name)

There are many possible solutions. Pls. provide only one. (No extra credit for additional solutions!)

B. Draw the trees that your grammar generates for the three above sentences.

2. CYK Parsing
Consider the context-free grammar G = <N, Σ, S, P> in Chomsky normal form, defined as follows:

N={S,G,O,L,E}
Σ = {p,e,m,o}
$S \in N$
P = { $S \rightarrow O\ O$,
$G \rightarrow L\ G \mid m$,
$O \rightarrow E\ G \mid o$,
$L \rightarrow E\ E \mid p$,
$E \rightarrow L\ L \mid e$ }

Does the grammar G generate both the strings ("sentences") peepmo and peppmo, (consisting of the "words" p, e, m, and o)?

The task is focussed on making you familiar with the CYK algorithm, without letting any grammatical intuitions interfere. This is the reason for choosing a somewhat abstract example that has nothing to do with natural language.

Use the CYK algorithm to answer this question and document your solution step by step for each of the two strings.

A. set up the two charts and fill in the words (lexical chart fill, as on lecture slides)

B. perform the syntactic chart fill (as on lecture slides)
Make sure that you include in the CYK charts all possible constituents that the grammar provides for these strings, not only the ones that form part of a complete analysis.