1. (a) Draw a tree for the following sentence, using whatever features are necessary to make subcategorization and the long-distance dependency work out:

(1) Kim, Dana believes Chris knows Sandy trusts ... i

(b) Describe how subcategorization is handled here.

(c) Describe how the trace is linked to Kim.

2. Here's a set of CFG rules that don’t use feature structures:

NP_{1sg} → Det N_{1sg}  
NP_{2sg} → Det N_{2sg}  
NP_{3sg} → Det N_{3sg}  
NP_{1pl} → Det N_{1pl}  
NP_{2pl} → Det N_{2pl}  
NP_{3pl} → Det N_{3pl}

And here's the same set using a feature structure notation:

NP → Det N
<NP PERSON> = <N PERSON>
<NP NUMBER> = <N NUMBER>

Assume we’re using the Earley parser, and we’ve already processed the input from position 0 to position 1, using the rule Det → the.

(a) Describe the current state of the Earley parser for the CFG rules without feature structures.

(b) Describe the current state of the Earley parser for the CFG rules with feature structures.

3. Is the language a^n b^2 a^n context-free? (Jurafsky and Martin, question 16.1)

4. Assuming that the parser finds the correct dependency analysis for the German sentence in (2), walk through the steps of how Nivre’s parser produces a parse.

(2) Die Wirtschaftspolitik läßt auf sich warten.

'Economic policy is a long time coming.'

5. Obtain a copy of MaltParser (http://maltparser.org/) and see if you can run the commands under Start using MaltParser at http://maltparser.org/userguide.html

Using MaltParser and following the user guide, do the following:

(a) Pick a language from the Universal Dependencies treebank, either version 1.0 or 2.0 (https://github.com/ryanmcd/uni-dep-tb)

(b) Train MaltParser on the available training data. Depending upon the exact data size, this will likely take several hours. If you face significant problems, go ahead and reduce the size of the training data for this experiment. Be clear in your write-up as to what happened.

(c) Parse the available testing data using MaltParser and the model you trained with.

(d) Evaluate the results. You can use eval.pl (http://ilk.uvt.nl/conll/software.html#eval) or MaltEval (http://www.maltparser.org/malteval.html), and/or you can qualitatively evaluate a set number of sentences.

You may work in groups for this question, if you prefer.