If yot think some assumptions are missing, make your own and explain them!

1 Context-free grammars and parsing (45%)

(a) **(Counts 10%)** In English compund nouns are most often written as two words, as in "mountain bike". As the first noun may be modified, as in "natural language processing" it is most natural to consider this a syntactic phenomenon. The following small example grammar, call it G1, is meant to capture this phenomenon.

i) $S \rightarrow NP\ VP$
ii) $NP \rightarrow DET\ N$
iii) $VP \rightarrow TV\ NP$
iv) $N \rightarrow N\ N$
v) $N \rightarrow A\ N$
vi) $NP \rightarrow Mary|John$
vii) $DET \rightarrow a|the|her|his$
viii) $TV \rightarrow bought|loves|misses$
ix) $N \rightarrow bike|jersey|mountain|sleeve|brake|$
x) $A \rightarrow long|hydraulic|knitted|expensive|steep$

Draw all the trees the grammar ascribes to the sentnence

(1) *Mary bought a mountain bike hydraulic brake.*

(b) **(Counts 5%)** Some parsers will have problems with this grammar. Which parsers? Explain shortly why it is a problem.

(c) **(Counts 10%)** Some constructions are very ambiguous with this grammar. For example sentence (2) gets 7 different analyses. You are not supposed to show this.

(2) *John bought a mountain bike long sleeve jersey.*

This ambiguity counts in favor of using a table parser when we are to decide whether sentences are accepted by grammar G1. Give a complete CKY-parse of (2). You may do this by drawing a table and filling in the cells. Please number the symbols as they are added to the table to make it possible to see how you proceed.

(d) **(Counts 10%)** Show how a chart parser for this grammar accepts sentenece (3).

(3) *John misses his expensive bike.*

(e) **(Counts 10%)** Let G2 be the grammar one gets by removing rule (iv) $N \rightarrow N\ N$. Is the language described by G2 regular? Is the language described by the whole grammar G1 regular? State the reasons for yor answers shortly.

2 Type Hierarchies and Multiple Inheritance (20%)

(a) **(Counts 5%)** Following is the description of a type hierarchy (taken from one of our grammars) in the Type Description Language (TDL). Draw the graph corresponding to this hierarchy, i.e. a diagram of super- and sub-type relationships, extending downwards from the `*top*` type. For the purpose of this exercise, we ignore all declarations of appropriate features on types.

```
pos := *top*.
agr-pos := pos.
det := agr-pos.
modable := agr-pos.
noun := modable.
verb := modable.
premodifier := pos.
postmodifier := pos.
prep := postmodifier.
adj := premodifier.
adv := premodifier & postmodifier.
```

(b) **(Counts 5%)** In this hierarchy, are there any pairs of types that cannot be unified, i.e. for which no greatest lower bound exists? If so, please name an example pair or two.

(c) **(Counts 10%)** In two or three sentences, what is the interpretation of the vertical dimension of the type hierarchy, i.e. what does it mean for a type to be a sub-type of another type? In this light, using one sentence each, explain any two out of the following four terms *specificity*, *inheritance*, *monotonicity*, or *subsumption*.

3 Unification-Based Grammar (15%)

(a) **(Counts 10%)** Recall that all unification-based grammars of English that we developed throughout the semester (using the TDL formalism) contained rules called *specifier – head rule* and *head – complement rule*. Show the (approximate) feature structure representations of these rules. Explain the three most important features involved, relevant value constraints, and percolation of information between the daughters of each rule and the rule mother. For each rule, give examples of two phrases built using this rule.

(b) **(Counts 5%)** Name at least two phenomena in natural languages that are (at best) unpleasant to account for in plain context-free grammars (CFGs). Show a few examples, indicating the relevant range of variation in each case, and summarize in a few sentences why these are challenging for a pure CFG analysis. Thinking about these phenomena, what is the main gain in moving to unification-based grammars (in no more than two sentences)?

4 Formal Morphology and Syntax (20%)

(a) **(Counts 10%)** We made a sharp distinction in our type hierarchies between objects of type `lexeme` vs. ones of type `word`. We further subdivided morphological rules into derivational vs. inflectional ones. In a few sentences, characterize key distinctions between derivation and inflection, give one example of each type of morphological rules (in English), and explain briefly what types of input (i.e. argument) each rule accepts, and what types of output it produces.

(b) **(Counts 5%)** In terms of its morphological structure, is there ambiguity in analyzing the English form *barking*? Using informal language (i.e. suitable abbreviations of syntactic and morphological categories, as you see fit), sketch all possible morphological analyses of this form.

(c) **(Counts 5%)** Sketch constituent trees for each of the possible readings of the following sentence:

(4) *The fierce cat chased the dog near the aardvark.*

To label the nodes of the tree, use abbreviatory labels like 'Det', 'N', 'NP', 'VP', et al. Furthermore, please annotate each branch as to whether the constituent dominated by it acts as a *head*, *specifier*, *complement*, or *modifier*.

END