UNIVERSITY OF MALTA

Institute of Linguistics

LIN2160 Introduction to Computational Linguistics Test

Wednesday, 28th May 2008 10:30 – 11:30

Name: _____Course: _____Year: _____

Answer one question from section A and one from section B.

Section A

1. What is the relevance of Chomsky Hierarchy of grammars for natural language? discuss. [50 marks]

2.

a) Discuss the phenomenon of grammatical agreement in natural language with examples, and outline one way in which one could account for agreement in the present tense in English. [26 marks]
b) Define the concept of a context-free grammar (CFG). What characteristic makes it context-free? [8 marks]
c) Give an example indicating why agreement is difficult to handle using a purely context-free grammar [8 marks]
d) Argue for or against the proposition that mechanisms available in PATR2 over and above pure CFG are adequate for handling agreement? [8 marks]

Section B

1. The grammar/lexicon shown in figure 1 generates sentences such as “John sleeps”.

Grammar	Lexicon	
Rule s -> np vp Rule np -> n Rule np -> d n Rule vp -> v	\w John \c n \f <lex> = John \w Mary \c n \f <lex> = Mary \w the \c d	\w pig \c n \f <lex> = pig \w horse \c n \f <lex> = horse \w sleeps \c v \f <lex> = sleep

Figure 1

a) How many sentences do the grammar and lexicon generate altogether?
b) How many of these are ungrammatical and why?
c) *By adding new constraints*, make changes that eliminate the ungrammatical sentences.
d) Extend the grammar/lexicon to accept “John smokes cigars” and to reject “John sleeps cigars”.
e) Extend the grammar to allow for zero or any number of adjectives inside noun phrases e.g. “the large black horse” (ignore punctuation issues). Show the corresponding parse tree for this example.
f) Add further constraints to noun phrases to create a representation of the semantics, so that parsing the noun phrase in (e) would yield the representation

```
[cat    np
 semantics [type   horse
            size   large
            colour black] ]
```

g) What further changes are necessary to create a representation for the sentence that includes slots for the semantic representations of subject and object e.g. for the sentence “the large black horse smokes a cuban cigar”

```
[cat    s
 subj semantics [type   horse
                 size   large
                 colour black]
 obj  semantics [type   cigar
                 origin Cuba]]
```

2.

a) Explain what you understand by (i) regular expressions and (ii) finite state machines. What is the relationship between the two?
b) Assuming an alphabet consisting of all alphanumeric characters, explain in your own words the languages denoted by the following regular expressions. In each case draw the corresponding finite state automaton.
   i. [0-9]*
   ii. [A-Z]+[0-9A-Za-z]*
   iii. ^[^aeiou] [^aeiou] ^[^aeiou] [^aeiou] ^[^aeiou]
c) Describe three underlying problems of Computational Morphology. Give examples of each.
d) What meant by two-level morphology? Why is it necessary?
e) Englex consists of a rules file, a grammar file, and various sublexicons. Explain the function of each, giving examples where possible.
f) A typical entry (in this case for the initial sublexicon) looks like this:

```
\lf  fox
\lx  N
\alt Suffix
\gl  fox
```

Explain the meaning of the different fields.

g) Assuming the lexicon consists of sublexicons for optional prefixes, roots and optional suffixes, the initial lexicon below

```
\lf 0
\lx INITIAL
\alt Prefix
\gl
```

and a final lexicon called “#”, write other entries so that the lexicon can handle the words “large”, “enlarge”, “enlargement”, “mice”