

# Natural Language Processing Exam

Gosse Bouma

29-6-2001

1. (a) What is the difference between a deterministic and a non-deterministic finite state automaton?  
(b) What is a stemmer?
2. Consider input string consisting only of *a*'s and *b*'s. Give regular expressions (in van Noord's FSA notation or some (unix/perl) notation) for recognizing the following three cases:
  - (a) The total number of characters in the input is even.
  - (b) The total number of *a*'s in the input is even.
  - (c) The total number of *a*'s is even and the total number of *b*'s is even.
3. Give the finite state automaton for the regular expression you gave in question (2b).
4. Consider the following quote from Groucho Marx:
  - (a) One morning I shot an elephant in my pajamas.
  - (b) (How he got into my pajamas I don't know. )

Give a context-free grammar for (a fragment of) English which produces the first sentence in (at least) two ways, and explain the joke in terms of this grammar.

5. (a) Give an example which illustrates why a simple backtracking parser encounters efficiency problems caused by repeated parsing of subtrees.  
(b) Explain why a *chart (or tabular or dynamic programming) parser* does not encounter this problem.
6. (a) When does a feature structure A subsume a feature structure B?  
(b) Give a definition of (feature) unification in terms of subsumption.