UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Exam in INF5830 - Natural language processing
Day of exam: 14 juni 2012
Exam hours: 14.30-18.30 (4 hours)
This examination paper consists of 3 pages including this.
Appendices: Statistical table – 3 pages
Permitted materials: None

Make sure that your copy of this examination paper is complete before answering.

- **You may answer in English, Norwegian, Danish or Swedish.**
- **You should answer all questions. The weight of the various questions are indicated.**
- **You should read through the whole set to see whether anything is unclear so that you can ask your questions to the teachers when they arrive.**
- **If you think some assumptions are missing, make your own and explain them!**

1 Zipf's law (10%)

Explain in a few sentences Zipf's law for word frequencies.

2 Collocations (20%)

Explain what a collocation is. There have been proposed several different methods for finding collocations in texts. Explain two or three of them.

3 Classification (20%)

Explain how the k nearest neighbors classifier works. Explain how Rocchio classification works. Give an example where the two classifiers would yield different results.

4 Experiments (10%)

Kim has constructed two parsers A and B. Parser B is a refinement of parser A and Kim hopes it will perform significantly better than A. She is testing them on 2400 tree-banked sentences for sentence accuracy.

Parser A gets 2040 sentences correct and 360 sentences incorrect, i.e., it achieves 85%, while parser B gets 2064 sentences correct and 336 sentences incorrect, it achieves 86%. Can Kim from this conclude that parser B is better than parser A, and that this is statistically significant at the 0.05 level? State reasons for your answer.

5 Dependency parsing and semantic role labeling (40 %)

(a) Please provide a dependency graph for the following sentence:

1) The hungry children instantly consumed the cakes

You may use the following set of dependency labels:

subj – subject
obj – direct object
vmod – verbal modifier
nmod – nominal modifier
det – determiner

Explain the difference between *endocentric* and *exocentric* dependency constructions and provide examples of each type using the sentence above.

(b) Please explain the following terms in the context of syntactic parsing.

(i) robustness
(ii) disambiguation
(ii) accuracy

How may we characterize the differences between *grammar-driven* and *data-driven* approaches to parsing based on these three dimensions?

(c) You have been asked to annotate a small corpus of English sentences with semantic role information. Please choose an annotation scheme presented in the course and provide an annotation for the verbal arguments in the following two sentences:

2) The hungry children ate the cakes
3) The chocolate cake was eaten by a greedy boy

In what way does your assignment of semantic roles differ from a syntactic analysis of the same sentences?

(d) You want to create a semantic role labeling system using machine learning. An important step is to define a set of *features* to represent your data instances. Present at least three different features you would use in your system and show how these would be instantiated for the small corpus above (sentence 2-3).

END