1 684.02: Statistical Natural Language Processing, Winter 2002

Instructor:	Chris Brew, Oxley Hall, Linguistics.
Course location:	12:30-2:18 TR, 218 Cockins Hall
Office hours	Wednesday 1-2:30 or (preferably) by appointment
Email:	`cbrew@acm.org`
Course alias	`snlp@ling.ohio-state.edu`

2 Aims and General Description

This class has three main aims: familiarity with tools and techniques for handling text corpora, knowledge of the characteristics of some of the available corpora, and a secure grasp of the fundamentals of statistical natural language processing. Specific topics include:

1. probability and information theory as they have been applied to computational linguistics.
2. knowledge of fundamental techniques of probabilistic language modeling.
3. experience of working with corpus data.
4. Knowledge of statistical approaches to representative language engineering tasks.

3 Prerequisites

Basic background in computational linguistics will be assumed. Programming experience, mathematical skills and familiarity with text corpora are all desirable.

4 Syllabus

The course provides both awareness of the theoretical issues in statistical language modeling and practical experience of applying the methods of the subject. We also introduce a sample of its main application areas.

We always emphasize the connection between the needs of the task in hand and the underlying probabilistic or information theoretic model. After completing the course, students will be in a position to understand and critique research papers in statistical computational linguistics. They will also be aware of the trade-offs involved in selecting statistical techniques for practical language engineering tasks.

5 Assessment

There will be regular weekly homework assignments, varying from corpus search exercises through probability problems, practice in statistical modeling and other small activities.

Formal assessment and grading will be via assignments and/or an optional final project. The project will be worth the same as three assignments.

6 Study Materials

The course book is Manning and Schu̇tze *Foundations of Statistical Natural Language Processing*. It's a big, dense textbook, with lots of stuff in it. If you read most of it, understand most of that, then dip back into it as needed we will have succeeded.

here are several good introductions to probability and information theory. I'll be happy to recommend something appropriate to your background on an individual basis.

7 References

- C. Manning and H.Schu̇tze (1999) *Foundations of Statistical Natural Language Processing*, MIT Press