

Syllabus for CG136/CS146

“Introduction to Computational Linguistics”

Instructor: Eugene Charniak and Mark Johnson

Spring semester, 2009

Tuesday and Thursday, 2:30–4pm, Watson Center (CIT) 506

The goal of this class is to develop computational models that do interesting and possibly even useful things with language, with an emphasis on statistical methods. Topics include basic probability and statistics, language modeling, machine translation, hidden Markov models, syntactic parsing, and pronoun reference.

We’re writing the textbook for this class as we go, and chapters will be made available in the course directory on the CS machines:

`/course/cs146`

You will need to have a CS account to take this course; see me if you don’t have one.

Grading is based on computer projects (40%), problem sets (25%), midterm (15%) and final (20%) (subject to adjustment during the semester). No collaboration of any kind is allowed on problem sets. Collaboration up to, but not including, sharing of code, is allowed on computer projects, unless explicitly told otherwise in class.

Class Schedule, Spring 2009

This schedule is tentative, and may be adjusted (e.g., topics inserted, deleted or reordered) as the class proceeds.

Week 1: Why this will be a great class.

Week 2: Probability, Statistics and Language Modeling

Week 3: Language Modeling

Week 4–5: Machine Translation

Week 6–8: Hidden Markov Models

Week 9–11: Syntactic Parsing

Week 12–14: Pronominal Coreference

The mid-term will be some time around week 7 or 8, and the in-class final will probably be during reading period.

Programming Assignments

All programming is to be done in C++. This will allow students to use C++ classes supplied by the course. There will be approximately 7 programming assignments for this course, i.e., one every two weeks. In general each assignment will have several parts, which you'll be expected to complete in stages.