

Lign 165: Computational Linguistics

Prof. Andrew Kehler
UCSD Department of Linguistics
kehler@ling.ucsd.edu
(858) 534-6239

Fall, 2008
MWF, 11:00-11:50, CSB 004
Office Hours: Wed 3-4, Fri 1-2, or by appt. (AP&M 4256)

TA: Henry Beecher, hbeecher@ling.ucsd.edu
Office Hours: TBA

Overview

This course provides an introduction to the fundamental concepts of computational linguistics, including computational approaches to language modelling, part-of-speech tagging, syntactic analysis, semantic interpretation, and applications.

Prerequisites

Ling 101 is strongly recommended. No programming experience assumed. Formal reasoning skills, an enthusiasm for natural language, and an ability to become computer savvy are all required.

Textbooks (One Optional, One Free)

Daniel Jurafsky and James H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, Prentice Hall, 2000, ISBN: 0130950696. [optional]

Fernando Pereira and Stuart Shieber, *Prolog and Natural-Language Analysis*, CSLI Publications, 1987, ISBN: 0937073180. An on-line version is available at <http://www.mtome.com/Publications/PNLA/pnla-digital.html> [free]

Administrivia

There will be five homework assignments, distributed approximately every four lectures, each worth 8% of your grade. They will include a mixture of paper-and-pencil and computer exercises. Assignments are due in class. Late assignments will be penalized at 5% up until 5pm on the due date, at 10% per day otherwise, and will not be accepted at all after the time at which graded assignments and answer keys are distributed.

There will be two exams: a midterm and a final, worth 25% and 35% of your grade respectively.

Students are permitted to consult with each other and/or work together in learning the concepts necessary for completing the homework, as long as each student: (1) writes up his or her own homework alone, using no notes resulting from the collaboration, and (2) lists the names of all other students involved in the collaboration prominently on the assignment. Collaborative efforts not meeting these restrictions are strictly forbidden.

Needless to say, please turn off your cell phones before entering the classroom.

Provisional Schedule

NOTE: There is no class on the first meeting day, September 26th.

- I. Introduction** (Monday, September 29)
 - II. Language Modeling, Part-of-Speech Tagging, Transformation-Based Learning, Empirical Evaluation** (Wednesday, October 1 – Wednesday, October 8)
Reading: J&M Chapter 8, Sections 5.1-5.5, Sections 6.1-6.3 (through pg. 210)
Monday, October 6: Assignment One Distributed (Due Wednesday, October 15)
 - III. Syntax and Parsing** (Friday, October 10 – Monday, October 20)
Reading: J&M Chapter 9, Sections 10.1-10.3
Wednesday, October 15: Assignment Two Distributed (Due Friday, October 24)
Assignment One Collected
 - IV. Introduction to Prolog** (Wednesday, October 22 – Wednesday October 29)
Reading: P&S Sections 2.1, 2.2, 2.5, and 2.7
Friday, October 24: Assignment Three Distributed (Due Monday, November 3)
Assignment Two Collected
 - V. Parsing with Prolog, Unification** (Friday, October 31 – Monday, November 17)
Reading: P&S Sections 3.7 and 4.2.1–4.2.5; excerpts from J&M Chapter 11
Monday, November 3: Assignment Three Collected
Wednesday, November 5: Midterm Review
Friday, November 7: Midterm Exam
Wednesday, November 12: Assignment Four Distributed (Due Monday, November 24)
 - VI. Semantic Interpretation** (Wednesday, November 19 – Wednesday, November 26)
Reading: P&S Section 4.1; excerpts from J&M Chapters 14 and 15
Monday, November 24: Assignment Five Distributed (Due Wednesday, December 3)
Assignment Four Collected
Friday, November 24: Thanksgiving Break
 - VII. Applications** (Monday, December 1 – Wednesday, December 3)
Wednesday, December 3: Assignment Five Collected
 - VIII. Summary and Review** (Friday, December 5)
- FINAL EXAM: Tuesday, December 9, 11:30-2:30**