CS 674: Natural Language Processing
Spring 2000

Thackara said, "Look – I'm a postmodernist. Is all this grammar really necessary?" "James," di Giovanni said, "we're talking on this phone *in grammar*." –John Walsh, "A Legend in His Own Mind", *The New Yorker*, Dec. 22 1997

Time and Place Hollister 314, MW 10:10–11am.

Instructor Lillian Lee. Upson 4152, `llee@cs.cornell.edu`. Office hours: MW 11-12, or by appt.

Reference Materials

- Allen, *Natural Language Understanding*, 2^{nd} ed., 1995. [C]
- Manning and Schütze, *Foundations of Statistical Natural Language Processing*, 1999.
- Grosz, Sparck Jones, and Webber, *Readings in Natural Language Processing*, 1986. [C]
- Jelinek, *Statistical Methods for Speech Recognition*, 1997.
- Jurafsky and Martin, *Speech and Language Processing*, 2000. (I have not seen a copy of this book)
- Conference proceedings: ACL proceedings dating back to 1982; ANLP proceedings from 1983 on; and COLING proceedings from 1988 on. [C]
- Journal: *Computational Linguistics*

Items marked with a [C] are on reserve at Carpenter Engineering Library.

Prerequisites Elementary computer science background (ability to analyze the space and time complexity of an algorithm, for example), elementary knowledge of probability (e.g., what conditional probability is). You may wish to skim Chapter 2 of Allen or Chapter 3 of Manning and Schütze for background linguistic knowledge.

Grading Policy

- Reaction essays: 30%. Every Monday for the first half of the semester (starting with the third week), a reaction essay of one to two pages, based on a paper from the on-line reading list (`http://courses.cs.cornell.edu/cs674/2000sp/readings.html`; hardcopies will be on reserve at Carpenter), is due. The intent is to help you find a project topic, and to acquaint you with some recent papers in the field.
- Final project: 60%. An independent project on a topic of your choice is due on May 16. Both programming and non-programming projects, including (thoughtful) surveys of papers, are welcome; all must include a careful writeup. Projects that attempt novel algorithms, implementations, proofs, analyses, and so on are especially encouraged – NLP is a fairly easy field to start doing research in!
- Participation: 10%.

Course homepage `http://courses.cs.cornell.edu/cs674/2000sp/`. This page contains links to handouts, the reaction essay reading list, references mentioned in class, online resources, and the course newsgroup, `cornell.class.cs674`, where announcements will be posted. Feel free to post relevant news (talk announcements, questions about papers, and so on).

Academic Integrity You are encouraged to discuss issues, papers, projects, etc. with others. However, the work you submit, if not otherwise attributed, must be your own. *Be sure to credit your sources correctly.* Violating the Code of Academic Integrity may result in failing the course. When in doubt about an issue, ask beforehand!

Syllabus

References in round brackets are to chapters in Allen; square brackets refer to Manning and Schütze. Due dates for reaction essays are underlined.

Jan 24	Introduction. (Ch 1)
Jan 26	The Chomsky hierarchy. (Ch 3.1; box 3.1, p. 46)
Jan 31-Feb 2	Context-free parsing. (Ch 3.3-3.4, 3.6)
Feb 7	Probabilistic CFG's. (Ch 7.1, 7.5)
Feb 9	Feature-based CFG's. (Ch 4.1-4.2,4.7-4.8)
Feb 14-16	Tree adjoining grammars. (Box 5.5, pg. 153)
Feb 21-23	Finite-state methods (Ch 3.7, 7.3) [Ch 10, 6.1]
Feb 28-Mar 1	Semantics. (Ch 8.1, 8.3; box 8.1 (pg 239); Ch 12.1-12.2, 12.4, 14.6)
Mar 6-8	Global discourse phenomena. (Ch 14.2-14.4, 16.1-16.4, 16.6-16.7)
Mar 13	Word sense disambiguation: knowledge-based methods. (Ch 10.1-10.3) [ch 7.3]
Mar 15	WSD: statistical and machine-learning methods. (Ch 10.4) [Ch. 7.1-7.2, 7.5]
	Project proposals due (may be turned in up to Friday at noon).
Mar 27	Smoothing. (Ch 7.2; box 7.1 (pg 205)) [Ch. 6.2-6.3]
Mar 29	Clustering. [Ch. 14]
Apr 3	Machine translation: classic approaches.
Apr 5	Machine translation: statistical approaches. [Ch. 13.3]
Apr 10	Machine learning of language: theoretical results.
	Project preliminary literature survey due.
Apr 12	Grammar induction [Ch. 6.2.3-6.2.4]
Apr 17	Classification: nearest neighbors [Ch. 16.4]
Apr 19	Statistics vs. linguistics: a case study in POS tagging
Apr 24	The Federalist Papers.
Apr 26	The Turing Test.
May 4-9	**Project presentations.** (tentative)
May 16	**Projects due.**