



CS 224N / Ling 280 – Natural Language Processing

Course Description

This course is designed to introduce students to the fundamental concepts and ideas in natural language processing (NLP), and to get them up to speed with current research in the area. It develops an in-depth understanding of both the algorithms available for the processing of linguistic information and the underlying computational properties of natural languages. Word-level, syntactic, and semantic processing from both a linguistic and an algorithmic perspective are considered. The focus is on modern quantitative techniques in NLP: using large corpora, statistical models for acquisition, disambiguation, and parsing. Also, it examines and constructs representative systems.

Prerequisites

- Adequate experience with programming and formal structures (e.g., CS106B/X and CS103B/X).
- Programming projects will be written in Java 1.5, so knowledge of Java (or a willingness to learn on your own) is required.
- Knowledge of standard concepts in artificial intelligence and/or computational linguistics (e.g., CS121/221 or Ling 180).
- Basic familiarity with logic, vector spaces, and probability.

Intended Audience

Graduate students and advanced undergraduates specializing in computer science, linguistics, or symbolic systems.

Textbook and Readings

This year, the required text will be:

- Daniel Jurafsky and James H. Martin. 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Second Edition. Prentice Hall. The book won't be able in time for the class. (June 2008 update: it's now available for purchase!) We will use a reader containing parts of the second edition. The reader is available for ordering at [University Readers](#). You order it online and they ship it to you. The cost is \$40.58. [[Detailed purchasing instructions](#).] Once you've ordered it, you can have access to the first couple of chapters that we'll use online for free. If you have any difficulties, please e-mail orders@universityreaders.com or call 800.200.3908, and email the class email list. It's referred to as J&M in the syllabus. [[Book website](#)]

Of course, I'm also fond of:

- Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press. Buy it at the Stanford Bookstore (recommended class text) or [Amazon](#) (\$64 new). Please see <http://nlp.stanford.edu/fsnlp/> for supplementary information about the text, including errata, and pointers to online resources.

Other useful reference texts for NLP are:

- James Allen. 1995. *Natural Language Understanding*. Benjamin/Cummings, 2ed.
- Gerald Gazdar and Chris Mellish. 1989. *Natural Language Processing in X*. Addison-Wesley. [Where X = Prolog, Lisp, or, I think, Snobol.

- Frederick Jelinek. 1998. *Statistical Methods for Speech Recognition*. MIT Press.

Other papers with relevant material will occasionally be posted or distributed for appropriate class lectures.

Copies of in-class hand-outs, such as readings and programming assignments, will be posted on the syllabus, and hard copies will also be available outside Gates 158 (in front of Prof. Manning's office) while supplies last.

Assignments and Grading

There will be three substantial programming assignments, each exploring a core NLP task. They are a chance to see real, close to state-of-the-art tools and techniques in action, and where students learn a lot of the material of the class.

There will be a final programming project on a topic of your own choosing.

Finally, there will be simple weekly online quizzes, which will aim to check that you are thinking about what you hear/read.

Course grades will be based 60% on programming assignments (20% each), 8% on the quizzes, and 32% on the final project.

Be sure to read the policies on late days and collaboration.

Course Schedule

- Lecture 1**
Wed
4/2/08
Introduction Overview of NLP. Statistical machine translation. Language models and their role in speech processing. Course introduction and administration.
No required reading.
Good background reading: M&S 1.0-1.3, 4.1-4.2, [Collaboration Policy](#)
Optional reading on Unix text manipulation (useful skill!): Ken Church's tutorial [Unix for Poets](#)
(If your knowledge of probability theory is limited, also read M&S 2.0-2.1.7. If that's too condensed, read the probability chapter of an intro statistics textbook, e.g. Rice, *Mathematical Statistics and Data Analysis*, ch. 1.)
Distributed today: [Programming Assignment 1](#)
- Lecture 2**
Mon
4/7/08
N-gram Language Models and Information Theory
n-gram models. Entropy, relative entropy, cross entropy, mutual information, perplexity. Statistical estimation and smoothing for language models.
Assigned reading: J&M ch. 4
Alternative reading: M&S 1.4, 2.2, ch. 6.
Tutorial reading: Kevin Knight. A Statistical MT Tutorial Workbook MS., August 1999. Sections 1-14.
Optional reading: Joshua Goodman (2001), [A Bit of Progress in Language Modeling, Extended Version](#)
Optional reading: Stanley Chen and Joshua Goodman (1998), [An empirical study of smoothing techniques for language modeling](#)
Optional reading: Teh, Yee Whye. 2006. A Hierarchical Bayesian Language Model based on Pitman-Yor Processes. EMNLP 2006.
- Lecture 3**
Wed
4/9/08
Statistical Machine Translation (MT), Alignment Models
Assigned reading: J&M ch. 25, sections 25.0-25.5, 25.11.
- Section 1**
Fri
4/11/08
Smoothing
Smoothing: absolute discounting, proving you have a proper probability distribution, Good-Turing implementation. Information theory examples and intuitions. Java implementation issues.
- Lecture 4**
Mon
4/14/08
Statistical Alignment Models and Expectation Maximization (EM)
EM and its use in statistical MT alignment models.
Assigned reading: Kevin Knight. A Statistical MT Tutorial Workbook. MS., August 1999. Sections 15-37 (get the free beer!).
(read also the relevant [Knight Workbook FAQ](#))
Reference reading: Geoffrey J. McLachlan and Thiriyambakam Krishnan. 1997. *The EM Algorithm and Extensions*. Wiley
Further reading: M&S 13.
Moore, Robert C. 2005. [Association-Based Bilingual Word Alignment](#). In Proceedings, Workshop on Building and Using Parallel Texts: Data-Driven Machine Translation and Beyond, Ann Arbor, Michigan, pp. 1-8.
Moore, Robert C. 2004. [Improving IBM Word Alignment Model 1](#). In Proceedings, 42nd Annual Meeting of the Association for Computational Linguistics, Barcelona, Spain, pp. 519-526.
- Lecture 5**
Wed
4/16/08
Putting together a complete statistical MT system
Decoding and A* Search. Recent work in statistical MT: statistical phrase based systems and syntax in MT.
Required reading: J&M, secs 25.7-10, 25.12.
"Seminal" background reading: Brown, Della Pietra, Della Pietra, and Mercer, [The Mathematics of Statistical Machine Translation: Parameter Estimation](#). *Computational Linguistics*.
[After their work in speech and language technology, the team [turned to finance...](#)]
Ulrich Germann, Michael Jahr, Kevin Knight, Daniel Marcu, and Kenji Yamada. 2001. [Fast Decoding and Optimal Decoding for Machine Translation](#). ACL.
K. Yamada and K. Knight. 2002. [A Decoder for Syntax-Based Statistical MT](#). ACL.
David Chiang. 2005. [A hierarchical phrase-based model for statistical machine translation](#). ACL 2005, pages 263-270.
Due today: [Programming Assignment 1](#)
Distributed today: [Programming Assignment 2](#)
- Section 2**
Fri
The EM algorithm

4/18/08

Lecture 6 **Information Extraction (IE) and Named Entity Recognition (NER).**
Information sources, rule-based methods, evaluation (recall, precision). Introduction to supervised machine learning methods. Naive Bayes (NB) classifiers for entity classification.
Mon 4/21/08
Assigned reading:
J&M secs 22.0-22.1 (intro to IE and NER).
J&M secs. 5.5 and 5.7 (introduce HMMs, Viterbi algorithm, and experimental technique). If you're not familiar with supervised classification and Naive Bayes, read J&M sec 20.2 before the parts of ch. 5.
Alternative reading: M&S 8.1 (evaluation), 7.1 (experimental methodology), 7.2.1 (Naive Bayes), 10.2-10.3 (HMMs and Viterbi)
Background older IE reading:
Peter Jackson and Isabelle Moulinier. 2007. *Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization*. John Benjamins. 2nd edition. Ch. 3.
Ion Muslea (1999). [Extraction Patterns for Information Extraction Tasks: A Survey](#), *AAAI-99 Workshop on Machine Learning for Information Extraction*.
Douglas E. Appelt. 1999. [Introduction to Information Extraction Technology](#)

Lecture 7 **Maximum Entropy Classifiers**
Assigned Reading:
Wed 4/23/08 class slides.
J&M secs 6.6-7 (maximum entropy models)
Additional references:
M&S section 16.2
Adwait Ratnaparkhi. [A Simple Introduction to Maximum Entropy Models for Natural Language Processing](#). Technical Report 97-08, Institute for Research in Cognitive Science, University of Pennsylvania.

Section 3 **Corpora and other resources**
Fri 4/25/08

Lecture 8 **Maximum Entropy Sequence Classifiers**
Assigned Reading:
Mon 4/28/08 class slides.
J&M secs. 6.0-6.4 and 6.8-6.9 (HMMs in detail and then MEMMs).
Other references: Adwait Ratnaparkhi. [A Simple Introduction to Maximum Entropy Models for Natural Language Processing](#). Technical Report 97-08, Institute for Research in Cognitive Science, University of Pennsylvania.
Adam Berger, [A Brief Maxent Tutorial](#)
Distributed today: [Final project guide](#)

Lecture 9 **IE and text mining**
Assigned reading: J&M secs. 22.2, 22.4.
Wed 4/30/08 *HMMs for IE reading:* Dayne Freitag and Andrew McCallum (2000), [Information Extraction with HMM Structures Learned by Stochastic Optimization](#), AAAI-2000
Maxent NER reading: Jenny Finkel et al., 2005. [Exploring the Boundaries: Gene and Protein Identification in Biomedical Text](#)
Due today: [Programming Assignment 2](#)
Distributed today: [Programming Assignment 3](#)

Section 4 **Maximum entropy sequence models**
Fri 5/2/08

Lecture 10 **Syntax and Parsing for Context-Free Grammars (CFGs)** Parsing, treebanks, attachment ambiguities. Context-free grammars. Top-down and bottom-up parsing, empty constituents, left recursion, and repeated work.
Mon 5/5/08 Probabilistic CFGs.
Assigned reading: J&M ch. 13, secs. 13.0-13.3.
Background reading: J&M ch. 9 (or M&S ch. 3). This is especially if you haven't done any linguistics courses, but even if you have, there's useful information on treebanks and part-of-speech tag sets used in NLP.

Lecture **Dynamic Programming for Parsing** Dynamic programming for parsing. The CKY algorithm. Accurate

11 unlexicalized PCFG parsing.
Wed *Assigned reading:* J&M sec. 13.4
5/7/08 *Additional information:* Dan Klein and Christopher D. Manning. 2003. [Accurate Unlexicalized Parsing](#). ACL 2003, pp. 423-430.
Due today: final project proposals

Section 5 Parsing, PCFGs
Fri
5/9/08

Lecture *[Moved forward from 5/19/08]* **Semantic Role Labeling**
12 *Assigned reading:* J&M secs. 19.4, 20.9
Mon *Further reading:*
5/12/08 Daniel Gildea and Daniel Jurafsky. 2002. [Automatic Labeling of Semantic Roles](#). Computational Linguistics 28:3, 245-288.
Kristina Toutanova, Aria Haghighi, and Christopher D. Manning, 2005. [Joint Learning Improves Semantic Role Labeling](#). Proceedings of 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005), pp. 589-596.
Pradhan, S., Ward, W., Hacioglu, K., Martin, J., Jurafsky, D., "[Semantic Role Labeling Using Different Syntactic Views](#)", in Proceedings of the Association for Computational Linguistics 43rd annual meeting (ACL 2005), Ann Arbor, MI, June 25-30, 2005.
V. Punyakanok, D. Roth, and W. Yih, [The Necessity of Syntactic Parsing for Semantic Role Labeling](#). Proc. of the International Joint Conference on Artificial Intelligence (IJCAI) (2005) pp. 1117--1123.

Lecture **Lexicalized Probabilistic Context-Free Grammars (LPCFGs)**
13 Lexicalization and lexicalized parsing. The Charniak, Collins/Bikel, and Petrov & Klein parsers.
Wed *Assigned reading:* J&M ch. 14 (you can stop at the end of sec. 14.7, if you'd like!)
5/14/08 *Alternative reading:* M&S Ch. 11
Optional readings:

- Eugene Charniak (1997), [Statistical techniques for natural language parsing](#), *AI Magazine*.
- Eugene Charniak (1997), [Statistical parsing with a context-free grammar and word statistics](#), *Proceedings of the Fourteenth National Conference on Artificial Intelligence*. AAAI Press/MIT Press, Menlo Park (1997).
- Eugene Charniak (2000), [A Maximum-Entropy-Inspired Parser](#), *Proceedings of NAACL-2000*.

Lecture **Modern Statistical Parsers**
14 Search methods in parsing: Agenda-based chart, A*, and "best-first" parsing. Dependency parsing. Discriminative parsing. *Assigned reading:* J&M ch. 14 (you can stop at the end of sec. 14.7, if you'd like!)
Mon *Alternative, less up-to-date reading:* M&S 8.3, 12
5/19/08 *Optional readings:*

- Dan Klein and Christopher D. Manning. 2002. [A Generative Constituent-Context Model for Improved Grammar Induction](#). Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 128-135.
- Dan Klein and Christopher D. Manning. 2002. [Natural Language Grammar Induction using a Constituent-Context Model](#). In Thomas G. Dietterich, Suzanna Becker, and Zoubin Ghahramani (eds), Advances in Neural Information Processing Systems 14 (NIPS 2001). Cambridge, MA: MIT Press, vol. 1, pp. 35-42.
- Dan Klein and Christopher D. Manning. 2003. [Factored A* Search for Models over Sequences and Trees](#). IJCAI 2003.
- Dan Klein and Christopher D. Manning. 2003. [A* Parsing: Fast Exact Viterbi Parse Selection](#). HLT-NAACL 2003.
- Kristina Toutanova, Christopher D. Manning, Stuart M. Shieber, Dan Flickinger, and Stephan Oepen. 2002. [Parse Disambiguation for a Rich HPSG Grammar](#). First Workshop on Treebanks and Linguistic Theories (TLT2002), pp. 253-263. Sozopol, Bulgaria.
- Kristina Toutanova, Christopher D. Manning, Dan Flickinger, and Stephan Oepen. 2005. [Stochastic HPSG Parse Disambiguation using the Redwoods Corpus](#). Research in Language and Computation 2005.

- B. Taskar, D. Klein, M. Collins, D. Koller and C. Manning. [Max-Margin Parsing](#). Empirical Methods in Natural Language Processing (EMNLP04), Barcelona, Spain, July 2004. Received best paper award.
- Eugene Charniak and Mark Johnson (2005). [Coarse-to-Fine n-Best Parsing and MaxEnt Discriminative Reranking](#), Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)
- Ryan McDonald, Koby Crammer and Fernando Pereira (2005). [Online Large-Margin Training of Dependency Parsers](#). 43rd Annual Meeting of the Association for Computational Linguistics, ACL 2005.

Lecture 15 Wed 5/21/08	Computational Semantics Semantic representations, lambda calculus, compositionality, syntax/semantics interfaces, logical reasoning. <i>Assigned reading:</i> An Informal but Respectable Approach to Computational Semantics J&M ch. 18 (you can skip secs. 18.4 and 18.6-end, if you wish).
Mon 5/26/08	Memorial Day no class
Lecture 16 Wed 5/28/08	Compositional Semantics II Semantic representations, lambda calculus, compositionality, syntax/semantics interfaces, logical reasoning. <i>Assigned reading:</i> An Informal but Respectable Approach to Computational Semantics J&M ch. 18 (you can skip secs. 18.4 and 18.6-end, if you wish).
Lecture 17 Mon 6/2/08	Lexical Semantics <i>Reading:</i> (Okay, I'm not so naive as to think that you'll actually read this in week 9 of the quarter....) J&M secs. 19.0-9.3. <i>Further reading:</i> J&M secs 20.0-20.8 (not included in reader, I'm afraid). <!--[slides: see last time] I. Androutsopoulos et al., Language Interfaces to Databases Luke S. Zettlemoyer and Michael Collins. Learning to Map Sentences to Logical Form: Structured Classification with Probabilistic Categorical Grammars . In Proceedings of the Twenty First Conference on Uncertainty in Artificial Intelligence (UAI-05), 2005. -->
Lecture 18 Wed 6/4/08	Question Answering (QA) TREC-style robust QA, textual inference <i>Assigned reading:</i> J&M secs 23.0, 23.2 <i>Further reading:</i> Marius Pasca, Sanda M. Harabagiu. High Performance Question/Answering . SIGIR 2001: 366-374. <i>Due today:</i> Final project reports
Monday 6/9/08	Final Project Presentations Students will give short (~5 min) presentations on their final projects during the time slot allocated for a final exam.