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:

  The book won't be 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.](http://nlp.stanford.edu/fsnlp/) 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](http://nlp.stanford.edu/fsnlp/)

Of course, I'm also fond of:

  Buy it at the Stanford Bookstore (recommended class text) or Amazon ($64 new).
  Please see [http://nlp.stanford.edu/fsnlp/](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:


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

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Date</th>
<th>Topic</th>
<th>Assigned Reading</th>
<th>Optional Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4/2/08</td>
<td>Introduction Overview of NLP, Statistical machine translation. Language models and their role in speech processing. Course introduction and administration.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4/2/08</td>
<td><strong>Good background reading:</strong> M&amp;S 1.0-1.3, 4.1-4.2, Collaboration Policy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Optional reading on Unix text manipulation (useful skill!):</strong> Ken Church's tutorial Unix for Poets</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(If your knowledge of probability theory is limited, also read M&amp;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.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Distributed today:</strong> Programming Assignment 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4/7/08</td>
<td><strong>N-gram Language Models and Information Theory</strong></td>
<td>J&amp;M ch. 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4/7/08</td>
<td><strong>Assigned reading:</strong> J&amp;M ch. 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Alternative reading:</strong> M&amp;S 1.4, 2.2, ch. 6.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Optional reading:</strong> Joshua Goodman (2001), A Bit of Progress in Language Modeling, Extended Version</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Optional reading:</strong> Stanley Chen and Joshua Goodman (1998), An empirical study of smoothing techniques for language modeling</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Optional reading:</strong> Teh, Yee Whye. 2006. A Hierarchical Bayesian Language Model based on Pitman-Yor Processes. EMNLP 2006.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4/14/08</td>
<td>Statistical Alignment Models and Expectation Maximization (EM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4/14/08</td>
<td><strong>Assigned reading:</strong> Kevin Knight. A Statistical MT Tutorial Workbook MS., August 1999. Sections 15-37 (get the free beer!).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(read also the relevant Knight Workbook FAQ)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Further reading:</strong> M&amp;S 13.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4/16/08</td>
<td>Putting together a complete statistical MT system</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4/16/08</td>
<td><strong>Required reading:</strong> J&amp;M, secs 25.7-10, 25.12.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Seminal background reading:</strong> 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....]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>K. Yamada and K. Knight. 2002. A Decoder for Syntax-Based Statistical MT. ACL.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Due today:</strong> Programming Assignment 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Distributed today:</strong> Programming Assignment 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section 2</td>
<td>Fri</td>
<td>The EM algorithm</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lecture 6
4/21/08
Mon
Information Extraction (IE) and Named Entity Recognition (NER).
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 Naïve 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:
Douglas E. Appelt. 1999. Introduction to Information Extraction Technology

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 Naïve 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:
Douglas E. Appelt. 1999. Introduction to Information Extraction Technology

Lecture 7
4/23/08
Wed
Maximum Entropy Classifiers
Assigned Reading:
J&M secs. 6.6-7 (maximum entropy models)
Additional references:
M&S section 16.2

Lecture 8
4/28/08
Mon
Maximum Entropy Sequence Classifiers
Assigned Reading:
J&M secs. 6.0-6.4 and 6.8-6.9 (HMMs in detail and then MEMMs).
Adam Berger, A Brief Maxent Tutorial
Distributed today: Final project guide

Lecture 9
4/30/08
Wed
IE and text mining
Assigned reading: J&M secs. 22.2, 22.4.
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 3
Fri
4/25/08
Corpora and other resources

Lecture 10
5/5/08
Mon
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.
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
unlexicalized PCFG parsing.

Assigned reading: J&M sec. 13.4


Due today: final project proposals

Section S  Parsing, PCFGs

Fri 5/9/08

Lecture [Moved forward from 5/19/08] Semantic Role Labeling

Assigns reading: J&M secs. 19.4, 20.9

Further reading:


Lecture Lexicalized Probabilistic Context-Free Grammars (LPCFGs)

Assigns reading: J&M ch. 14 (you can stop at the end of sec. 14.7, if you'd like!)

Optional readings:

• Eugene Charniak (1997), Statistical techniques for natural language parsing, AI Magazine.


Lecture Modern Statistical Parsers

Assigns reading: J&M ch. 14 (you can stop at the end of sec. 14.7, if you'd like!)

Optional readings:


### Lecture 15
**Wed 5/21/08**  
**Computational Semantics**  
Semantic representations, lambda calculus, compositionality, syntax/semantics interfaces, logical reasoning.  
**Assigned reading:**  
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 16
**Wed 5/28/08**  
**Compositional Semantics II**  
Semantic representations, lambda calculus, compositionality, syntax/semantics interfaces, logical reasoning.  
**Assigned reading:**  
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**  
Reading: (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. Further reading: 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  

### Lecture 18
**Wed 6/4/08**  
**Question Answering (QA)** TREC-style robust QA, textual inference  
**Assigned reading:** J&M secs 23.0, 23.2  

### 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.