Course Overview
Lecture #1

Computational Linguistics
CMPSCI 591N, Spring 2006
University of Massachusetts Amherst

Andrew McCallum

http://www.cs.umass.edu/~mccallum/courses/cl2006

Where to find syllabus, announcements, slides, homeworks
Today’s Main Points

• Why is natural language interesting and difficult, complex and ambiguous.
  – Why? How to we resolve this ambiguity?

• Six “layers” of natural language

• Computational Linguistics history, an overview, current successes

• Get to know each other, and our motivations for being here

• Course mechanics; what you can expect
1967

Stanley Kubrick, filmmaker
1928 - 1999

Arthur C. Clarke, author, futurist
1917 -
HAL
HAL’s Capabilities

- Display graphics
- Play chess
- *Natural language production and understanding*

- Vision
- Planning
- Learning
- …
Graphics

HAL

Now
Chess

HAL

Now

Andrew McCallum, UMass Amherst,
including material from Chris Manning and Jason Eisner
Natural Language Understanding

**HAL**

*David Bowman:*
Open the pod bay doors, Hal.

**HAL:**
I’m sorry, Dave, I’m afraid I can’t do that.

*David Bowman:*
What are you talking about, Hal?

...**HAL:**
I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.

Many useful tools, but none that come even close to HAL’s ability to communicate in natural language.
I propose to consider the question "Can machines think?"…
We can only see a short distance ahead, but we can see plenty there that needs to be done.
Layers of Computational Linguistics

1. Phonetics & Phonology
2. Morphology
3. Syntax
4. Semantics
5. Pragmatics
6. Discourse
1. Phonetics & Phonology

The study of: language sounds, how they are physically formed;
systems of discrete sounds, e.g. languages’ syllable structure.

\[ \text{dis-k\&-'nekt} \]
\[ \text{disconnect} \]

“It is easy to recognize speech.”
“It is easy to wreck a nice beach.”
2. Morphology
The study of the sub-word units of meaning.

disconnect

“not”    “to attach”

Even more necessary in some other languages, e.g. Turkish:

uygarlastiramadiklarimizdanmissinizcasina

uygar las tir ama dik lar imiz dan mis siniz casina

(behaving) as if you are among those whom we could not civilize
3. Syntax
The study of the structural relationships between words.

I know that you and Frank were planning to disconnect me.

Not same structure:
You know me--Frank and I were planning to disconnect that.
4. Semantics
The study of the literal meaning.

I know that you and Frank were planning to disconnect me.

ACTION = disconnect
ACTOR = you and Frank
OBJECT = me
5. Pragmatics
The study of how language is used to accomplish goals.

What should you conclude from the fact I said something? How should you react?

I’m sorry Dave, I’m afraid I can’t do that.

Includes notions of polite and indirect styles.
6. Discourse
The study of linguistic units larger than a single utterance.

The structure of conversations:
  turn taking, thread of meaning.

David Bowman:
  Open the pod bay doors, Hal.
HAL:
  I’m sorry, Dave, I’m afraid I can’t do that.
David Bowman:
  What are you talking about, Hal?
...HAL:
  I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.
Linguistic Rules

E.g. Morphology

To make a word plural, add “s”

- dog → dogs
- baby → babies
- dish → dishes
- goose → geese
- child → children
- fish → fish (!)
Inherent Ambiguity in Syntax

Fed raises interest rates 0.5% in effort to control inflation

NY Times headline 17 May 2000
Where are the ambiguities?

Part-of-speech ambiguities

Fed raises interest rates 0.5 % in effort to control inflation

Syntactic attachment ambiguities

Word sense ambiguities: Fed → "federal agent"
interest → a feeling of wanting to know or learn more

Semantic interpretation ambiguities above the word level.
Effects of V/N Ambiguity (1)

S

NP  VP

NNP   V   NP

Fed raises NN NN

interest rates
Effects of V/N Ambiguity (2)

```
S
  NP  VP
    N   V  NP
Fed raises interest N
  rates
```
Effects of V/N Ambiguity (3)

S
  /  
NP    VP
 /    /
N    N    V    NP
Fed raises interest rates CD N
   /    /    |   |
  0.5  %    |   |
          |   |
Ambiguous Headlines

- Iraqi Head Seeks Arms
- Juvenile Court to Try Shooting Defendant
- Teacher Strikes Idle Kids
- Stolen Painting Found by Tree
- Kids Make Nutritious Snacks
- Local HS Dropouts Cut in Half
- British Left Waffles on Falkland Islands
- Red Tape Holds Up New Bridges
- Clinton Wins on Budget, but More Lies Ahead
- Ban on Nude Dancing on Governor’s Desk
What is grammatical and what isn’t?

• John I believe Sally said Bill believed Sue saw.
• What did Sally whisper that she had secretly read?
• John wants very much for himself to win.
• Who did Jo think said John saw him?
• The boys read Mary’s stories about each other.
• Mary, while John had had had had had had had had had was the correct answer.
What is grammatical and what isn’t?

• John I believe Sally said Bill believed Sue saw.
• What did Sally whisper that she had secretly read?
• John wants very much for himself to win.
• Who did Jo think said John saw him?
• The boys read Mary’s stories about each other.
• Mary, while John had had “had” had had “had had;” “had had” was the correct answer.
Language Evolves

• Morphology
  – We learn new words all the time: bioterrorism, cyberstalker, infotainment, thumb candy, energy bar

• Part-of-speech
  – Historically: “kind” and “sort” were always *nouns*: “I knowe that sorte of men ryght well.” [1560]
  – Now also used as *degree modifiers*: “I’m sort of hungry.” [Present]
    “It sort o’ stirs one up to hear about old times.” [1833]
Natural Language Computing is hard because

• Natural language is:
  – highly ambiguous at all levels
  – complex and subtle
  – fuzzy, probabilistic
  – interpretation involves *combining evidence*
  – involves reasoning about the world
  – embedded a social system of people interacting
    • persuading, insulting and amusing them
    • changing over time
To handle this **ambiguity** and to **integrate evidence** from multiple levels we turn to:

The tools of probability:

- Bayesian Classifiers (not rules)
- Hidden Markov Models (not DFAs)
- *Probabilistic* Context Free Grammars

- …other tools of Machine Learning, AI, Statistics
Another Area where Probabilistic Combination of Evidence Won
Natural Language Processing

• Natural Language Processing (NLP) studies how to get computers to do useful things with natural languages:
  – Most commonly Natural Language Understanding
  – The complementary task is Natural Language Generation

• NLP draws on research in Linguistics, Theoretical Computer Science, Artificial Intelligence, Mathematics and Statistics, Psychology, Cognitive Science, etc.
What & Where is NLP

• Goals can be very far-reaching
  – True text understanding
  – Reasoning and decision-making from text
  – Real-time spoken dialog
• Or very down-to-earth
  – Searching the Web
  – Context-sensitive spelling correction
  – Analyzing reading-level or authorship statistically
  – Extracting company names and locations from news articles.
• These days, the later predominate (as NLP becomes increasingly practical, focused on performing measurably useful tasks now).
• Although language is complex, and ambiguity is pervasive, NLP can also be surprisingly easy sometimes:
  – rough text features often do half the job
Linguistics

• Linguistics is the study of natural languages:
  – Understanding this naturally-occurring phenomenon.
  – Structure, meaning, how acquired, differences and commonalities across languages.

• Linguistics draws on research in Natural Language Processing, Theoretical Computer Science, Artificial Intelligence, Mathematics and Statistics, Psychology, Cognitive Science, etc.
Some brief history: 1950s

- Early CL on machines less powerful than pocket calculators.
- Foundational work on automata, formal languages, probabilities and information theory.
- First speech systems (Davis et al, Bell Labs).
- MT heavily funded by military, but basically just word substitution programs.
- Little understanding of natural language syntax, semantics, pragmatics.
Some brief history: 1960s

- Alvey report (1966) ends funding for MT in America - the lack of real results realized
- ELIZA (MIT): Fraudulent NLP in a simple pattern matcher psycholtherapist
  - It’s true, I am unhappy.
  - *Do you think coming here will make you not to be unhappy?*
  - I need some help; that much is certain.
  - *What would it mean to you if you got some help?*
  - Perhaps I could earn to get along with my mother.
  - *Tell me more about your family.*
- Early corpora: Brown Corpus (Kudera and Francis)
Some brief history: 1970s

- Could interpret questions, statements commands.
  - Which cube is sitting on the table?
  - The large green one which supports the red pyramid.
  - Is there a large block behind the pyramid?
  - Yes, three of them. A large red one, a large green cube, and the blue one.
  - Put a small one onto the green cube with supports a pyramid.
  - OK.
Some brief history: 1980s

- Procedural --> Declarative (including logic programming)
- Separation of processing (parser) from description of linguistic knowledge.
- Representations of meaning: procedural semantics (SHRDLU), semantic nets (Schank), logic (perceived as answer; finally applicable to real languages (Montague))
- Perceived need for KR (Lenat and Cyc)
- Working MT in limited domains (METEO)
Some brief history: 1990s

• Resurgence of finite-state methods for NLP: in practice they are incredibly effective.
• Speech recognition becomes widely usable.
• Large amounts of digital text become widely available and reorient the field. The Web.
• Resurgence of probabilistic / statistical methods, led by a few centers, especially IBM (speech, parsing, Candide MT system), often replacing logic for reasoning.
• Recognition of *ambiguity* as key problem.
• Emphasis on machine learning methods.
Some brief history: 2000s

• A bit early to tell! But maybe:
  – Continued surge in probability, Bayesian methods of evidence combination, and joint inference.
  – Emphasis on meaning and knowledge representation.
  – Emphasis on discourse and dialog.
  – Strong integration of techniques, and levels: bringing together statistical NLP and sophisticated linguistic representations.
  – Increased emphasis on unsupervised learning.
Example Applications of NLP

Natural Language Processing
Natural Language Processing should make it possible for people to use computers in much the same way that they would use a human assistant to get their work done. Research.microsoft.com/nlp/- 28k - Cached - Similar pages

ISI's Natural Language Group
Overview of Research Environment Natural Language Processing at USC/ISI. USC offers a wide range of courses in areas related to natural language processing. Description: The Natural Language Processing group at the Information Sciences Institute of the University of Southern Cal. Category: Computers > Artificial Intelligence > ... > Research Groups
www.isi.edu/natural-language/nlp-at-isi.html - 15k - Cached - Similar pages

Foundations of Statistical Natural Language Processing
Foundations of Statistical Natural Language Processing. Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press. nlp.stanford.edu/fsnlp/ - 7k - Cached - Similar pages

Yahoo! Directory Artificial Intelligence > Natural Language...
Artificial Intelligence > Natural Language Processing Directory > Science > Computer Science > Artificial Intelligence > Natural Language Processing, ... dir.yahoo.com/Science/Computer/Science/Artificial_Intelligence/Natural_Language_Processing/
Example Applications of NLP: MSWord spelling correction, grammar checking
Example Applications of NLP:

News categorization and summarization
Example Applications of NLP
Information Extraction: Find experts, employees

Dr. Andrew McCallum
Action Editor
Journal of Machine Learning Research
Last Mentioned on 10/12/2003

<table>
<thead>
<tr>
<th>Other Titles Held:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Member, Editorial Board</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Additional Current Employment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carnegie Mellon University</td>
</tr>
<tr>
<td>Post-Doctoral Fellow</td>
</tr>
<tr>
<td>Adjunct Faculty Member</td>
</tr>
<tr>
<td>Adjunct Faculty Position</td>
</tr>
<tr>
<td>University of Massachusetts Amherst, CO</td>
</tr>
<tr>
<td>Research Associate Professor</td>
</tr>
<tr>
<td>Adjunct Faculty</td>
</tr>
<tr>
<td>Research Scientist</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Board Memberships and Affiliations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inteliseek Inc</td>
</tr>
<tr>
<td>Member of Advisory Board</td>
</tr>
<tr>
<td>IJCAI</td>
</tr>
<tr>
<td>Member, Program Committees (past)</td>
</tr>
<tr>
<td>AAAI</td>
</tr>
<tr>
<td>Member, Program Committees (past)</td>
</tr>
<tr>
<td>ICMML</td>
</tr>
<tr>
<td>Member, Program Committees (past)</td>
</tr>
<tr>
<td>NIPS</td>
</tr>
<tr>
<td>Member, Program Committees (past)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Past Employment History</th>
</tr>
</thead>
<tbody>
<tr>
<td>WhizBang Labs Inc</td>
</tr>
<tr>
<td>Vice President of Research and Development</td>
</tr>
<tr>
<td>Just Research</td>
</tr>
<tr>
<td>Research Scientist</td>
</tr>
<tr>
<td>Biomedical Information Communication Center of Oregon Health Sciences University</td>
</tr>
<tr>
<td>Machine Learning Researcher</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Rochester</td>
</tr>
<tr>
<td>Dartmouth College</td>
</tr>
</tbody>
</table>

Information about Andrew McCallum was compiled from 6 sources:

http://www.eliyon.com
Example Applications of NLP: Information Extraction: Job Openings

foodscience.com-Job2

JobTitle: Ice Cream Guru
Employer: foodscience.com
JobCategory: Travel/Hospitality
JobFunction: Food Services
JobLocation: Upper Midwest
Contact Phone: 800-488-2611
DateExtracted: January 8, 2001
Source: www.foodscience.com/jobs_midwest.html

OtherCompanyJobs: foodscience.com-Job1

If you dream of cold, creamy chocolate or ooey-gooey cookie dough, this may be the job for you. As the ice cream guru, you will work closely with the marketing and sales teams to develop new flavors and improve existing ones. This is a unique opportunity for someone with a passion for ice cream and a desire to contribute to a successful company. The position is based in the Upper Midwest for about a year and requires a BS in food science or a related field. Contact us at 1-800-488-2611 for more information.
Example Applications of NLP: Information Extraction: Job Openings
Example Applications of NLP: Automatically Solving Crossword Puzzles

Having trouble getting the last word in that puzzle? Having trouble getting the first? See if our search engine can help! Unlike pure pattern dictionary searches, we actually analyze the clue as well.

How to Search: Enter a clue and either the length of the answer or an answer pattern. For unknown letters in the word pattern, you can use a question mark.

Clue: Trout Basket Clue: Cut
Pattern: 5 Pattern: ???n
Clue: Scheme Clue: ?a?T??a??ke
Pattern: F...
Example Applications of NLP: Question Answering

who is married to bill gates?

Type in your question in English, French, Spanish, German, Italian or Portuguese.

Question:

who is married to bill gates

Possible answers:  XML  TXT

- Bill was married to Melinda French Gates in 1994 in Hawaii.
- Mary Gates, Bill's mother, biggest fan, and strongest prodger, finally laid down an ultimatum in 1993. She was dying of cancer, and wanted to see her only son married.
- Bill Gates married Melinda French in Hawaii on January 1, 1994, and his mother died a few months later.
- 1994 Bill Gates and Melinda French married in Hawaii on New Years Day.

Try your question on other engines:

Alta Vista  |  CNN News Engine  |  Ask Jeeves  |  Excite  |  Google  |  HotBot  |  Lycos  |  Start  |  Yahoo
Example Applications of NLP: Machine Translation
Example Applications of NLP:
Automatically generate Harlequin Romance novels?
Goals of the Course

• Introduce you to Computational Linguistics problems and solutions.

• Ultimate focus on handling ambiguity by probabilistic integration of evidence.

• Using these tools to empower Linguistics’ goals. (More on this next class.)

• Give you some hands-on practice with data and a handful of methods.
This Class

• Assumes you come with some skills…
  – Some basic statistics, decent programming skills (in a language of your choice--although solutions will be in Java)
  – Some ability to learn missing knowledge

• Teaches key theory and methods for language modeling, tagging, parsing, etc.

• But it’s something like an “AI Systems” class:
  – Hands on with data
  – Often practical issues dominate over theoretical niceties
Course Logistics

- **Professor:** Andrew McCallum
- **TAs:** Pallika Kanani
  Greg Druck
  Assistant: Gideon Mann

- **Time:** Tue/Thu 2:30-3:45pm

- **Mailing list:** complying-class@cs.umass.edu

- **More information on Web site:**
  http://www.cs.umass.edu/~mccallum/courses/cl2006
Grading

• 6 short written homework / programming assignments.
  – no way to really internalize without doing it
  – some hands-on experience
  – should be fun!
  – should take about 1-2 hours each.

• Random, informal in-class “collaborative quizzes”
  – help you set expectations for the mid-term and final

• Final project: with a small team, mixed CS & Ling
  – chance to explore a special interest at end of term

• Midterm & Final, and classroom participation
Programming? Yipes!

- Yes, but with *extensive* support for those w/out experience.

- Historically popular language for CL courses:
  - Prolog (clean, hard to learn, counter-intuitive)
  - Perl (quick, but obfuscated syntax, messy to read)
  - Interpreted, rapid prototyping

- Why **Python** is better-suited:
  - easy to learn, clean syntax, powerful features
  - becoming increasingly popular in CompLinguistics!
  - Extensive tutorials, CompLing support, toolkits, data, etc.

- Many CS students don’t know it either: put you on more equal footing.
Syllabus Outline

• Two halves:
  – First: hands-on course, introductory, methods, HW
  – Second: more like a seminar + project

• First half:
  – Language, structures, and computation
  – Foundation of probability and information theory
  – Use those foundations to work with language

• Example topics:
  – Language models, language prediction, spam filtering.
  – Collocations, word clustering, word sense disambiguation.
  – Finite state machines, Markov models, Part-of-speech tagging.
  – Modern parsing techniques.
  – Information extraction, semantics, question answering, discourse.
This Week

• Visit course Web site, browse around.

• Read Chapters 1 and 2 in Jurafsky & Martin textbook

• Install Python on your computer
  – Get extensive help from the TAs if you like!
Thank you!