Topics for today

- General introduction to NLP
  - Why study NLP?
  - Why is NLP such a challenging task?
- Class description and syllabus

Natural language and NLP

- “natural” language
  - Languages that people use to communicate with one another
- Ultimate goal
  - To build computer systems that perform as well at using natural language as humans do
- Immediate goal
  - To build computer systems that can process text and speech more intelligently

**Dialogue systems**

- Require both understanding and generation
  - Dave: Open the pod bay doors, HAL.
  - HAL: I'm sorry Dave, I'm afraid I can't do that.
  - Dave: What's the problem?
  - HAL: I think you know what the problem is just as well as I do.

Why study NLP?

- Useful applications...
  - E.g. information retrieval

  Topic: Advantages and disadvantages of using potassium hydroxide in any aspect of organic farming, especially...
Why study NLP?

- Useful applications…
  - E.g. question answering systems
    » How many calories are there in a Big Mac?
    » Who is the voice of Miss Piggy?
    » Who was the first American in space?
  - Retrieve not just relevant documents, but return the answer

- Interdisciplinary…
  - Linguistics
    » models for language
  - Psychology and psycholinguistics
    » models of cognitive processes/language
  - Mathematics
    » studies properties of formal models, methods of inference from these models
  - vs. NLP
    » Computational study of language use
    » Definite engineering aspect in addition to a scientific one
      ◆ Engineering: to enable effective human-machine communication
      ◆ Scientific: to explore the nature of linguistic communication
    » Emphasis on computational, not cognitive plausibility
    » Models of language: optional

Bill Gates, 1997 “…now we’re betting the company on these natural interface technologies”
Why study NLP?

- Challenging…
  - AI-complete
    » To solve NLP, you’d need to solve all of the problems in AI
  - Turing test
    » Posits that engaging effectively in linguistic behavior is a sufficient condition for having achieved intelligence.

- …But little kids can “do” NLP…
  - Why is NLP hard?

Why is NLP hard?

Ambiguity!!!! …at all levels of analysis 😞

- Phonetics and phonology
  - Concerns how words are related to the sounds that realize them
  - Important for speech-based systems.
    » "I scream" vs. "ice cream"
    » "nominal egg"
  - Moral is:
    » It’s very hard to recognize speech.
    » It’s very hard to wreck a nice beach.

- Morphology
  - Concerns how words are constructed from sub-word units
  - Unionized
    » un-ionized in chemistry?

- Syntax
  - Concerns sentence structure
  - Different syntactic structure implies different interpretation
    » Squad helps dog bite victim.
    ◆ [np squad] [vp helps [np dog] [inf-clause bite victim]]
    ◆ [vp squad] [vp helps [np dog] [inf-clause bite victim]]
  - Visiting relatives can be trying.

Why is NLP hard?

- Semantics
  - Concerns what words mean and how these meanings combine to form sentence meanings.
    » Jack invited Mary to the Halloween ball.
    ◆ dance vs. some big sphere with with Halloween decorations?
    » Visiting relatives can be trying.
    » Visiting museums can be trying.
    ◆ Same set of possible syntactic structures for this sentence
    ◆ But the meaning of museums makes only one of them plausible
Why is NLP hard?

Ambiguity!!!! …at all levels of analysis 😊

- Discourse
  - Concerns how the immediately preceding sentences affect the interpretation of the next sentence
    - Merck & Co. formed a joint venture with Ache Group, of Brazil. It will be called Prodome Ltd.
    - Merck & Co. formed a joint venture with Ache Group, of Brazil. It will own 50% of the new company to be called Prodome Ltd.
    - Merck & Co. formed a joint venture with Ache Group, of Brazil. It had previously teamed up with Merck in two unsuccessful pharmaceutical ventures.

- Pragmatics
  - Concerns how sentences are used in different situations and how use affects the interpretation of the sentence.
    - "I just came from New York."
      - Would you like to go to New York today?
      - Would you like to go to Boston today?
      - Why do you seem so out of it?
      - Boy, you look tired.

Syllabus (tentative)

- History and state-of-the-art
- Lexical semantics and word-sense disambiguation
- Part-of-speech tagging and HMMs
- Morphology
- Noisy channel model
- Language modeling
- Parsing
- Discourse processing
- Generation
- Inference and world knowledge
- Semantic analysis
- Information retrieval models
- Text categorization
- Question answering systems
- Summarization systems
- Dialogue systems
- Information extraction
- Machine Translation

Additional Course Info

- Time: Tuesdays and Thursdays, 1:25-2:40
  Place: 110 Olin Hall
Instructor: Claire Cardie, 5161 Upson Hall
Office hours: see the top of my home page

- Lecture Notes, Readings, Assignments

- Course Management System (CMS): We'll be using the CS department course management system for submission of assignments, grading, etc. You can get to CMS via the above link. You'll need your Cornell netid and password.

- Resources:
  - Lillian Lee's list of general NLP resources
  - NLP resources available locally are listed under the local resources link of the Cornell NLP home page.
Reference Material

- Required text book:

- Other useful references:
  - Others listed on course web page...

Prereqs and Grading

- Prerequisites
  - Elementary computer science background.

- Grading
  - 15%: critiques of selected readings and research papers
  - 40%: programming assignments
  - 10%: midterm
  - 25%: final examination
  - 10%: participation
  
  You'll be expected to participate in class discussion and class exercises or otherwise demonstrate an interest in the material studied in the course.

- Final exam
  - Weds, Dec 6, 7-9:30pm

Readings and Critiques