INTRODUCTION:
COMMUNICATION AND LANGUAGE

Theme
Language is not simply the words that you use, syntactically structured. Language is a method of communicating information so as to achieve goals to arrive at desired states and interpretations of the world. It is a highly complex procedure involving various types of knowledge and processes.

Summary of Contents
1. Introduction: What is language?
Compare language of humans to Natural Language Processing (NLP). Do this in a sentence-by-sentence readthrough and discussion of a text, with stepwise illustration and gradual building up of pertinent processes, memory structures, goals, affects, inferences, etc., in order to illustrate diversity of knowledge and processes required to understand a question, formulate an answer, and generate it. (Have three volunteers to blackboard; one records words, one records the concepts ‘behind’ words, and one records interrelationships among concepts to get overall meanings.)

Note all relevant information on blackboard as it is encountered, into one of several columns:
• **World knowledge** of various kinds; (info)
• **Episodic knowledge and remindings**
• **Situational effects**
• **Interpersonal effects**
and then the process:
• **Expectations**: activated, and realized or flushed.

2. Stages and levels of processing
Draw out the NLP pipeline: from words through parsing etc. to internal representation, and then out again to words. Draw some applications in the middle.

Show the substages: POS tagging, discourse analysis, macro- and micro-planning.

Describe lexis, syntax, discourse, semantics, pragmatics.
• **nonsense** — *green deeply ideas the the sleep colorless*
• **syntax**: — *colorless green ideas sleep furiously* (Chomsky)
• **semantics** — *happy tired children sleep deeply*
• **pragmatics** — *terrorist =? freedom fighter =? guerrilla* or *get off my foot! =? you’re heavy*

Types/levels of representation of language
Many types, for different purposes, including:
Lemmatizing demorphed, individuated word tokens
Tagging Part of Speech (POS) tags
Syntactic parsing Parse trees (many theories of grammar)
Semantic analysis Semantic readings (many theories of semantics)
Discourse analysis Discourse structures (trees?)
Pragmatic analysis Speech Acts; Locution/Illocution/Perlocution
Cohesiveness graphs (lexical chains)
Information structure (theme/rheme, given/new, topic/focus)

3. The world of Natural Language Processing

Short history of NLP. NLP ⇐ CS, Linguistics, Philosophy, Cognitive Psychology
Major themes of NLP: parsing, generation, discourse
Current status and capabilities
Major institutions and activities, conferences and journals
NLP at ISI

4. Course outline

Hand out the course schedule and go through the topics.

Assignments

• Homework 1: parsing (30%)
• Homework 2: generation (30%)
• Homework 3: statistical processing (30%)
• Final: written assignment (optional: coded) (10%)

Our goals

• Teach the basics of Natural Language Processing (NLP)
• Build NLP software, including a small parser and generator
• Search for students to help with our research (and their Ph.D.s)

5. Textbook

  Benjamin/Cummings Publishing.
READINGS
Chapter 1 of Jurafsky and Martin.

OPTIONAL ASSIGNMENT
Create a two- to three-page textual or pictorial design of a system to perform the kinds of processing required to handle the example text and one other text (chosen by the student). Design must contain functionally motivated modules and show information flow and dependencies. A description of the processing of the chosen text must accompany the assignment.
While window shopping in Sunset Plaza, you may recognize something missing as you pass the sidewalk cafes.

Letter grades once posted prominently in restaurant windows have all but disappeared into the hands of health inspectors from the Los Angeles County Health Department.

But that’s expected to change in three to four months as West Hollywood comes into compliance with the county's revised restaurant grading system.

Since July, the more than 120 food establishments in the 1.9-square-mile city were restricted from displaying letter grades A, B, or C in restaurant windows city-wide.

The county decided that West Hollywood, along with a number of other cities, must adopt the Los Angeles County health code in its entirety, with no amendments, to display the restaurant placards. Before July, displaying the grades was voluntary for restaurants.

“The county basically said all or nothing”, said Hillary Selvin, executive director of the West Hollywood Chamber of Commerce.

In 1988, West Hollywood Mayor Jeff Prang said the county's health code reinspection policies were penalizing restaurants rather than helping ensure high and safety standards because if a restaurant was given a grade lower than an A, it had to wait as much as a year to be reevaluated.

The West Hollywood City Council voted to amend the county's health code in 1998 with a proviso allowing restaurants that received a grade lower than A to request and pay for a reinspection in 15 days.

(LA Times, Spring 2001)

Questions to ponder:

• Sentence 1: given what “recognize” means, how can you recognize something that is missing? What do “while” and “may” mean?

• Sentence 2: how do you connect “letter grades” to “something missing”? Why would health inspectors take the grades—shouldn’t they give them instead? There should be a big unexplained question hanging in your mind. What does ‘having a question in your mind’ mean? How do you program a question?

• Sentences 3–5: the question gets answered. How is the answer assembled? What other questions remain?