Announcements

• Homework 11 due Thursday 12/5 @ 11:59pm.
• No video of lecture on Friday 12/6.
  ▪ Come to class and take the final survey.
  ▪ There will be a screencast of live lecture (as always).
  ▪ Screencasts: http://www.youtube.com/view_play_list?p=XXv-cvA_iCIEwJhyDVdyLMiimv6Tup
• Homework 12 due Tuesday 12/10 @ 11:59pm.
  ▪ All you have to do is vote on your favorite recursive art.
• 29 review sessions next week! Come learn about the topics that interest you the most.
  ▪ See http://inst.eecs.berkeley.edu/~cs61a/fa13/exams/final.html for the schedule.
Natural Language Processing
Ambiguity in Natural Language

Unlike programming languages, natural languages are ambiguous.

**Syntactic ambiguity:**
- TEACHER STRIKES IDLE KIDS
- HOSPITALS ARE SUED BY 7 FOOT DOCTORS

**Semantic ambiguity:**
- IRAQI HEAD SEEKS ARMS
- STOLEN PAINTING FOUND BY TREE
Tasks in Natural Language Processing

Research in natural language processing (NLP) focuses on tasks that involve language:

**Question answering.** "Harriet Boyd Hawes was the first woman to discover and excavate a Minoan settlement on this island." Watson says, "What is Crete?"

**Machine Translation.** "Call a spade a spade!" Google Translate says, "Appeler un chat un chat."

**Semantic Parsing.** "When's my birthday?" Siri says, "Your birthday is May 1st."

Much attention is given to more focused language analysis problems:

**Coreference Resolution:** Do the phrases "Barack Obama" and "the president" co-refer?

**Syntactic Parsing:** In "I saw the man with the telescope," who has the telescope?

**Word Sense Disambiguation:** Does the "bank of the Seine" have an ATM?

**Named-Entity Recognition:** What names are in "Did van Gogh paint the Bank of the Seine?"
Machine Translation
Machine Translation

Target language corpus gives examples of well-formed sentences

- I will get to it later
- See you later
- He will do it

Parallel corpus gives translation examples

- I will do it gladly
- Yo lo haré de muy buen grado
- You will see later
- Después lo veras

Machine translation system:

- Source language: Yo lo haré después
- Model of translation
- Target language: I will do it later
Syntactic Agreement in Translation

I will do it gladly
Yo lo haré de muy buen grado

You will see later
Después lo veras

Machine translation system:
Syntactic Reordering in Translation
Context-Free Grammars
A context-free grammar models language generation.

A grammar contains rules that hierarchically generate word sequences using syntactic tags.

### Grammar Rules

- **S** → NP VP
- **NP** → PRP
- **VP** → VB NP
- **PRP** → you
- **VB** → know
- **VB** → help

### Lexicon

- **PRP** → I
- **PRP** → you
- **VB** → know
- **VB** → help
Probabilistic Context-Free Grammars

**Grammar Rules**

- **S**: S → NP VP
  - **NP**: NP → PRP
  - **VP**: VP → VB
    - **VP**: VP → VB NP
      - **PRP**: PRP → you
      - **MD**: MD → can

**Lexicon**

- **PRP**: PRP → I
- **PRP**: PRP → you
- **VB**: VB → know
- **VB**: VB → help
- **MD**: MD → can
Learning Probabilistic Context-Free Grammars

(Demo)
Parsing with Probabilistic Context-Free Grammars
 Parsing is Maximizing Likelihood

A probabilistic context-free grammar can be used to select a parse for a sentence.

Parse by finding the tree with the highest total probability that yields the sentence.

Algorithm: Try every rule over every span. Match the lexicon to each word.
Tree Transformations
Reordering Modal Arguments

Help you, I can! Yes! Mm!

When 900 years old you reach, look as good, you will not. Hm.

(Demo)