

# For Monday

- Read chapter 26
- No homework

# Program 5

- Any questions?

# Semantics

- Most work probably hand-constructed systems
- Some more interested in developing the semantics than the mappings
- Basic question: what constitutes a semantic representation?
- Answer may depend on application???

# Possible Semantic Representations

- Logical representation
- Database query
- Case grammar

# Distinguishing Word Senses

- Use context to determine which sense of a word is meant
- Probabilistic approaches
- Rules
- Issues
  - Obtaining sense-tagged corpora
  - What senses do we want to distinguish?

# Semantic Demos

- <http://www.cs.utexas.edu/users/ml/geo.html>
- <http://www.cs.utexas.edu/users/ml/rest.html>
- <http://www.ling.gu.se/~lager/Mutbl/demo.html>

# Information Retrieval

- Take a query and a set of documents.
- Select the subset of documents (or parts of documents) that match the query
- Statistical approaches
  - Look at things like word frequency
- More knowledge based approaches interesting, but maybe not helpful

# Information Extraction

- From a set of documents, extract “interesting” pieces of data
- Hand-built systems
- Learning pieces of the system
- Learning the entire task (for certain versions of the task)
- Wrapper Induction

# Question Answering

- Given a question and a set of documents (possibly the web), find a small portion of text that answers the question.
- Some work on putting answers together from multiple sources.

# QA Demos

- [http://l2r.cs.uiuc.edu/~cogcomp/qa\\_news\\_demo.php](http://l2r.cs.uiuc.edu/~cogcomp/qa_news_demo.php)
- <http://demos.inf.ed.ac.uk:8080/qualim/>

# Text Mining

- Outgrowth of data mining.
- Trying to find “interesting” new facts from texts.
- One approach is to mine databases created using information extraction.

# Pragmatics

- Distinctions between pragmatics and semantics get blurred in practical systems
- To be a practically useful system, some aspects of pragmatics must be dealt with, but we don't often see people making a strong distinction between semantics and pragmatics these days.
- Instead, we often distinguish between **sentence** processing and **discourse** processing

# What Kinds of Discourse Processing Are There?

- Anaphora Resolution
  - Pronouns
  - Definite noun phrases
- Handling ellipsis
- Topic
- Discourse segmentation
- Discourse tagging (understanding what conversational “moves” are made by each utterance)

# Approaches to Discourse

- Hand-built systems that work with semantic representations
- Hand-built systems that work with text (or recognized speech) or parsed text
- Learning systems that work with text (or recognized speech) or parsed text

# Issues

- Agreement on representation
- Annotating corpora
- How much do we use the modular model of processing?

# Pronoun Resolution Demo

- <http://www.clg.wlv.ac.uk/demos/MARS/index.php>

# Summarization

- Short summaries of a single text or summaries of multiple texts.
- Approaches:
  - Select sentences
  - Create new sentences (much harder)
  - Learning has been used some but not extensively

# Machine Translation

- Best systems must use all levels of NLP
- Semantics must deal with the overlapping senses of different languages
- Both understanding and generation
- Advantage in learning: bilingual corpora exist--but we often want some tagging of intermediate relationships
- Additional issue: alignment of corpora

# Approaches to MT

- Lots of hand-built systems
- Some learning used
- Probably most use a fair bit of syntactic and semantic analysis
- Some operate fairly directly between texts

# Generation

- Producing a syntactically “good” sentence
- Interesting issues are largely in choices
  - What vocabulary to use
  - What level of detail is appropriate
  - Determining how much information to include