CS674 Natural Language Processing

- Semester so far…
  - Introduction and history
  - Computational morphology
  - Noisy channel model
    - » Spelling correction
    - » Pronunciation variation
  - Language modeling (n-gram models)
- Today
  - Lexical semantics

Semantic analysis

- **Compositional semantics**: we can derive the meaning of the whole sentence from the meanings of the parts.
  - Max ate a green apple.
- Relies on knowing:
  - the meaning of individual words
  - how the meanings of individual words combine to form the meaning of groups of words
  - how it all fits in with syntactic analysis

Caveats

- Problems with a compositional approach
  - a former congressman
  - a toy elephant
  - kicked the bucket

Introduction to lexical semantics

- Lexical semantics is the study of
  - the systematic meaning-related connections among words and
  - the internal meaning-related structure of each word
- Lexeme
  - an individual entry in the lexicon
  - a pairing of a particular orthographic and phonological form with some form of symbolic meaning representation
- Sense: the lexeme's meaning component
- Lexicon: a finite list of lexemes

Dictionary entries

- right *adj.* located nearer the right hand esp. being on the right when facing the same direction as the observer.
- left *adj.* located nearer to this side of the body than the right.
- red *n.* the color of blood or a ruby.
- blood *n.* the red liquid that circulates in the heart, arteries and veins of animals.

Lexical semantic relations: homonymy

- Homonyms: *words that have the same form and unrelated meanings*
  - Instead, a **bank¹** can hold the investments in a custodial account in the client's name.
  - But as agriculture burgeons on the east **bank²**, the river will shrink even more.
- Homophones: distinct lexemes with a shared pronunciation
  - E.g. *would* and *wood*, *see* and *sea.*
- Homographs: identical orthographic forms, different pronunciations, and unrelated meanings
  - The expert angler from Dora, Mo., was fly-casting for **bass** rather than the traditional trout.
  - The curtain rises to the sound of angry dogs baying and ominous **bass** chords sounding.

Why do these distinctions matter?

- One type or another is more likely to affect specific NLP applications.
  - Spelling correction?

  - Speech recognition?

  - Text-to-speech?

  - Information retrieval?

Lexical semantic relations: polysemy

- Polysemy: the phenomenon of multiple related meanings within a single lexeme
  - Example: While some **banks** furnish blood only to hospitals, others are much less restrictive.
  - New sense, e.g. bank³?
  - Polysemy allows us to associate a lexeme with a set of related senses.
- Distinguishing homonymy from polysemy is not always easy. Decision is based on:
  - Etymology: history of the lexemes in question
  - Intuition of native speakers

Polysemous lexemes

- For any given single lexeme we would like to be able to answer the following questions:
  - What distinct senses does it have?
  - How are these senses related?
  - How can they be reliably distinguished?
- Answers dictate how well semantic analyzers, search engines, NL generators, and MT systems perform their tasks.

How many word senses per polysemous lexeme?

- Use as many senses as necessary to account for all the fine distinctions in meaning observed in some very large corpus of examples.
- Too many senses
- Example: *serve*
  - They rarely *serve* red meat, preferring to prepare seafood, poultry or game birds.
  - He *served* as U.S. ambassador to Norway in 1976 and 1977l.
  - He might have *served* his time, come out and led an upstanding life.
- Zeugma: combine two separate uses of a lexeme into a single example using a conjunction
  - » Which of those flights *serve* breakfast?
  - » Does Midwest Express *serve* Philadelphia?
  - » ?Does Midwest Express *serve* breakfast or Philadelphia?