CS474 Natural Language Processing

- Last class
  - History
  - Tiny intro to semantic analysis
- Next lectures
  - Word sense disambiguation
    » Background from linguistics
      ◆ Lexical semantics
    » On-line resources
    » Computational approaches [next class]

Semantic analysis

- Assigning meanings to linguistic utterances
- **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 **bank**[1] can hold the investments in a custodial account in the client's name.
  - But as agriculture burgeons on the east **bank**[2], 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. bank[3]?
  - 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.

Polysemous lexemes

- For any given single lexeme we would like to be able to answer the following questions:
  – What distinct senses does it have?
    » generally rely on lexicographers
  – How are these senses related?
    » relatively little work in this area
  – How can they be reliably distinguished?
    » this is the task of **word sense disambiguation**

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?

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 are these senses related?

- Hasn't received much attention from lexicographers
- Important as systems begin to handle a wider variety of input texts…and encounter novel uses of words
  - Metaphor
  - Metonymy

Metaphor

- Situations where we refer to, and reason about, concepts using words and phrases whose meanings are appropriate to *other completely different kinds of concepts.*
  - Love is a rose. Time is money.
- Conventional metaphors
  - That doesn't **scare** Digital, which has grown to be the world's second-largest computer maker by poaching customers of IBM's mid-range machines.
  - COMPANY AS PERSON metaphor
  - Fuqua Industries Inc. said Triton Group Ltd., a company it helped **resuscitate**, has begun acquiring Fuqua shares.
  - And Ford was **hemorrhaging**; its losses would hit $1.54 billion in 1980.

Metonymy

- Situations where we denote a concept by naming some other concept *closely related to it.*
  - He likes Shakespeare.
    - » AUTHOR FOR AUTHOR'S WORKS
  - The White House had no comment.
    - » PLACE FOR INSTITUTION
  - Give the coke to the ham sandwich.
    - » ???

Computational approaches

- Convention-based approaches
  - Rely on formal representations of conventional metaphors and metonymies
  - Assumes that a small set of these will suffice
  - Semantic analysis applies them to figurative language
- Reasoning-based approaches
  - View metaphor and metonymy interpretation as general analogical reasoning tasks rather than as problems specific to language processing
  - Assume that metaphors depend on inherent structural similarities between the meaning representations derived compositionally from the input and the correct representations that capture the intended meaning of the input.
- No large-scale solutions to either problem to date.

Word sense disambiguation

- Given a *fixed* set of senses associated with a lexical item, determine which of them applies to a particular instance of the lexical item
- Two fundamental approaches
  - WSD occurs during semantic analysis as a side-effect of the elimination of ill-formed semantic representations
  - Stand-alone approach
    » WSD is performed independent of, and prior to, compositional semantic analysis
    » Makes minimal assumptions about what information will be available from other NLP processes
    » Applicable in large-scale practical applications

CS474 Natural Language Processing

- Last class
  - History
  - Tiny intro to semantic analysis
- Next lectures
  - Word sense disambiguation
    » Background from linguistics
      ◆ Lexical semantics
    On-line resources
    » Computational approaches [next class]

WordNet

- Handcrafted database of lexical relations
- Three separate databases: nouns; verbs; adjectives and adverbs
- Each database is a set of lexical entries (according to unique orthographic forms)
  - Set of senses associated with each entry

| Category	Unique Forms	Number of Senses
Noun	94474	116317
Verb	10319	22066
Adjective	20170	29881
Adverb	4546	5677

Synonymy

- Lexemes with the same meaning
- Invoke the notion of **substitutability**
  - Two lexemes will be considered synonyms if they can be substituted for one another in a sentence without changing the meaning or acceptability of the sentence
    » How *big* is that plane?
    » Would I be flying on a *large* or small plane?
    » Miss Nelson, for instance, became a kind of *big* sister to Mrs. Van Tassel's son, Benjamin.
    » We frustrate 'em and frustrate 'em, and pretty soon they make a *big* mistake.
    » Also issues of **register**
      ◆ Social factors that surround the use of possible synonyms, e.g. politeness, group status.

Hyponymy

- Pairings where one lexeme denotes a subclass of another

vehicle (hypernym)

↑

car (hyponym)
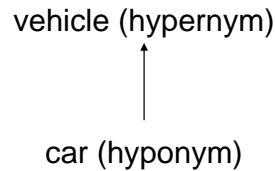
Sample entry

The noun "bass" has 8 senses in WordNet.
1. bass - (the lowest part of the musical range)
2. bass, bass part - (the lowest part in polyphonic music)
3. bass, basso - (an adult male singer with the lowest voice)
4. sea bass, bass - (flesh of lean-fleshed saltwater fish of the family Serranidae)
5. freshwater bass, bass - (any of various North American lean-fleshed freshwater fishes especially of the genus Micropterus)
6. bass, bass voice, basso - (the lowest adult male singing voice)
7. bass - (the member with the lowest range of a family of musical instruments)
8. bass - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Distribution of senses

- Zipf distribution of senses
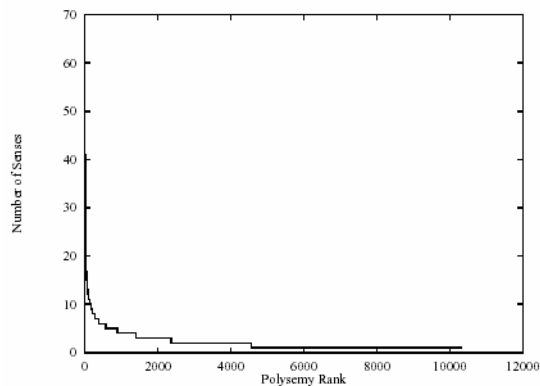
WordNet relations

- Nouns

Relation	Definition	Example
Hypernym	From concepts to superordinates	breakfast → meal
Hyponym	From concepts to subtypes	meal → lunch
Has-Member	From groups to their members	faculty → professor
Member-Of	From members to their groups	copilot → crew
Has-Part	From wholes to parts	table → leg
Part-Of	From parts to wholes	course → meal
Antonym	Opposites	leader → follower

- Verbs

Relation	Definition	Example
Hypernym	From events to superordinate events	fly → travel
Troponym	From events to their subtypes	walk → stroll
Entails	From events to the events they entail	snore → sleep
Antonym	Opposites	increase ⟺ decrease

- Adjectives/adverbs

Relation	Definition	Example
Antonym	Opposite	heavy ⟺ light
Adverb	Opposite	quickly ⟺ slowly