CS674 Natural Language Processing

- Last class
  - Introduction to lexical semantics
    - » Compositional semantics
    - » Homonymy
    - » Polysemy
- Today
  - Metaphor
  - Synonymy, hyponymy
  - Lexical semantic resources
  - Word sense disambiguation

Polysemous lexemes

- For any given single lexeme we would like to be able to answer the following questions:
  - What distinct senses does it have? [last class]
  - 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.

Topics for today

- Metaphor
- Synonymy, hyponymy
- Lexical semantic resources
- Word sense disambiguation

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)

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

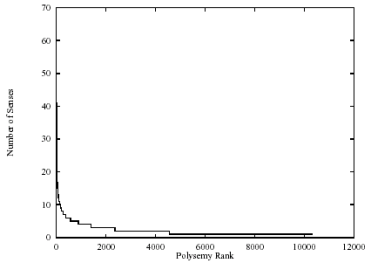
Category	Unique Forms	Number of Senses
Noun	94474	116317
Verb	10319	22066
Adjective	20170	29881
Adverb	4546	5677

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*

Topics for today

- Metaphor
- Synonymy, hyponymy
- Lexical semantic resources
- Word sense disambiguation

Word sense disambiguation

- Given a *fixed* set of senses is 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

Machine learning approaches

- Inductive machine learning methods
  - Supervised
  - Bootstrapping
  - Unsupervised
- Emphasis is on acquiring the knowledge needed for the task from data, rather than from human analysts.

Feature vector input

- **target:** the word to be disambiguated
- **context** : portion of the surrounding text
  - Tagged with part-of-speech information
  - Select a "window" size
  - Stemming or morphological processing
  - Possibly some partial parsing
- Convert the context into a set of features
  - Attribute-value pairs
    - » Numeric or nominal values

Collocational features

- Encode information about the lexical inhabitants of *specific* positions located to the left or right of the target word.
  - E.g. the word, its root form, its part-of-speech
  - *An electric guitar and **bass** player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.*
  - [guitar, NN1, and, CJC, player, NN1, stand, VVB]

Co-occurrence features

- Encodes information about neighboring words, ignoring exact positions.
  - **Features**: the words themselves (or their roots)
  - **Values**: number of times the word occurs in a region surrounding the target word
  - Select a small number of frequently used content words are selected for use as features
    - » 12 most frequent content words from a collection of *bass* sentences drawn from the WSJ: *fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band*
    - » Co-occurrence vector (window of size 10) for the previous example:

      [0,0,0,1,0,0,0,0,0,0,1,0]