CS 6740/INFO 6300 Advanced Language Technologies

- Last class
  - General Intro to HLT
- This week: **word sense disambiguation**
  - Background from linguistics
    - Semantics
    - Lexical semantics
  - On-line resources
  - Computational approaches to WSD

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

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: polysemy

- Polysemy: the phenomenon of multiple related meanings within a single lexeme
  - 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.
  - While some ***banks*** furnish blood only to hospitals, others are much less restrictive.
  - New sense, e.g. bank[3]?
  - Polysemy requires that we associate a lexeme with a set of related senses.

Lexical semantic relations: 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. Smith'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.

CS 6740/INFO 6300 Advanced Language Technologies

- Last class
  - General Intro to HLT
- This week: **word sense disambiguation**
  - Background from linguistics
    - » Semantics
    - » Lexical semantics
  - On-line resources
  - Computational approaches to WSD

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

Sample entry

WordNet Search - 3.0 - WordNet home page - Glossary - Help

Word to search for: bass Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Noun

- S: (n) **bass** (the lowest part of the musical range)
- S: (n) **bass**, bass part (the lowest part in polyphonic music)
- S: (n) **bass**, basso (an adult male singer with the lowest voice)
- S: (n) sea bass, **bass** (the lean flesh of a saltwater fish of the family Serranidae)
- S: (n) freshwater bass, **bass** (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
- S: (n) **bass**, bass voice, basso (the lowest adult male singing voice)
- S: (n) **bass** (the member with the lowest range of a family of musical instruments)
- S: (n) **bass** (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Adjective

- S: (adj) **bass**, deep (having or denoting a low vocal or instrumental range) *"a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"*

Statistics

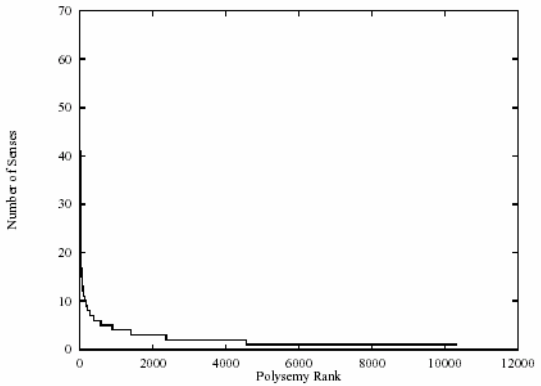
POS	Unique Strings	Synsets	Total Word-Sense Pairs
Noun	117798	82115	146312
Verb	11529	13767	25047
Adjective	21479	18156	30002
Adverb	4481	3621	5580
Totals	155287	117659	206941

More statistics

POS	Average Polysemy Including Monosemous Words	Average Polysemy Excluding Monosemous Words
Noun	1.24	2.79
Verb	2.17	3.57
Adjective	1.40	2.71
Adverb	1.25	2.50

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*

Possible uses for WordNet?

CS 6740/INFO 6300 Advanced Language Technologies

- Last class
  - General Intro to HLT
- This week: **word sense disambiguation**
  - Background from linguistics
    - » Semantics
    - » Lexical semantics
  - On-line resources
  - Computational approaches to WSD

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

Dictionary-based approaches

- Rely on machine readable dictionaries
- Initial implementation of this kind of approach is due to Michael Lesk (1986)
  - Given a word *W* to be disambiguated
    - » Retrieve all of the sense definitions, *S*, for *W* from the MRD
    - » Compare each *s* in *S* to the dictionary definitions of all the remaining words in the context
    - » Select the sense *s* with the most overlap with (the definitions of) these context words

Example

- Word: *cone*
- Context: *pine cone*
- Sense definitions

pine 1 kind of evergreen tree with needle-shaped leaves
2 waste away through sorrow or illness

cone 1 solid body which narrows to a point
2 something of this shape whether solid or hollow
3 fruit of certain evergreen trees

- Accuracy of 50-70% on short samples of text from *Pride and Prejudice* and an AP newswire article.

Next class

- Machine learning methods

Inductive ML framework