Word sense disambiguation

- » Background from linguistics
  - ◆ Lexical semantics
- » On-line resources
- » Computational approaches

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

- **Homonyms:** lexemes that have the same orthographic and phonological 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?

Spelling correction

A. **Homonyms:** lexemes that have the same orthographic and phonological 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.

B. **Homophones:** distinct lexemes with a shared pronunciation
   - E.g. *would* and *wood*, *see* and *sea.*

C. **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.

Speech recognition

A. **Homonyms:** lexemes that have the same orthographic and phonological 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.

B. **Homophones:** distinct lexemes with a shared pronunciation
- E.g. *would* and *wood*, *see* and *sea*.

C. **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.

Text-to-speech

A. **Homonyms:** lexemes that have the same orthographic and phonological 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.

B. **Homophones:** distinct lexemes with a shared pronunciation
  - E.g. *would* and *wood*, *see* and *sea.*

C. **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.

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

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 sentences 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.

An aside: Demetri Martin

- **Lexical humor**
  - Mind your own *business*. Mind your own *beeswax*.
    - » But not exchangeable
    - » Never hear "After working a year or two, I plan to go to beeswax school."
  - You know what's weird? Food trucks. They don't deliver.
  - "Nope" == "no", but with positive overtones.
    - » E.g. Are you still serving pancakes? Nope!
    - » Never hear "unfortunately, nope."

Word sense disambiguation

- Given a *fixed* set of senses associated with a lexical item (usually an orthographic form rather than a lexeme),

- determine which sense applies to a particular instance of the lexical item in context (i.e., in running text)

Marseille played along the **bank** of the Seine
where he found and ate a smelly **bass**.

- **Word sense disambiguation**
  - Background from linguistics
    - ◆ Lexical semantics
  - → On-line resources
  - Computational approaches

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	# of Senses
Noun	117798	82115
Verb	11529	13767
Adjective	21479	18156
Adverb	4481	3621

WordNet Browser

Word to search for: `bass` [Search WordNet]

Display Options: [(Select option to change) ▲▼] [Change]

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)
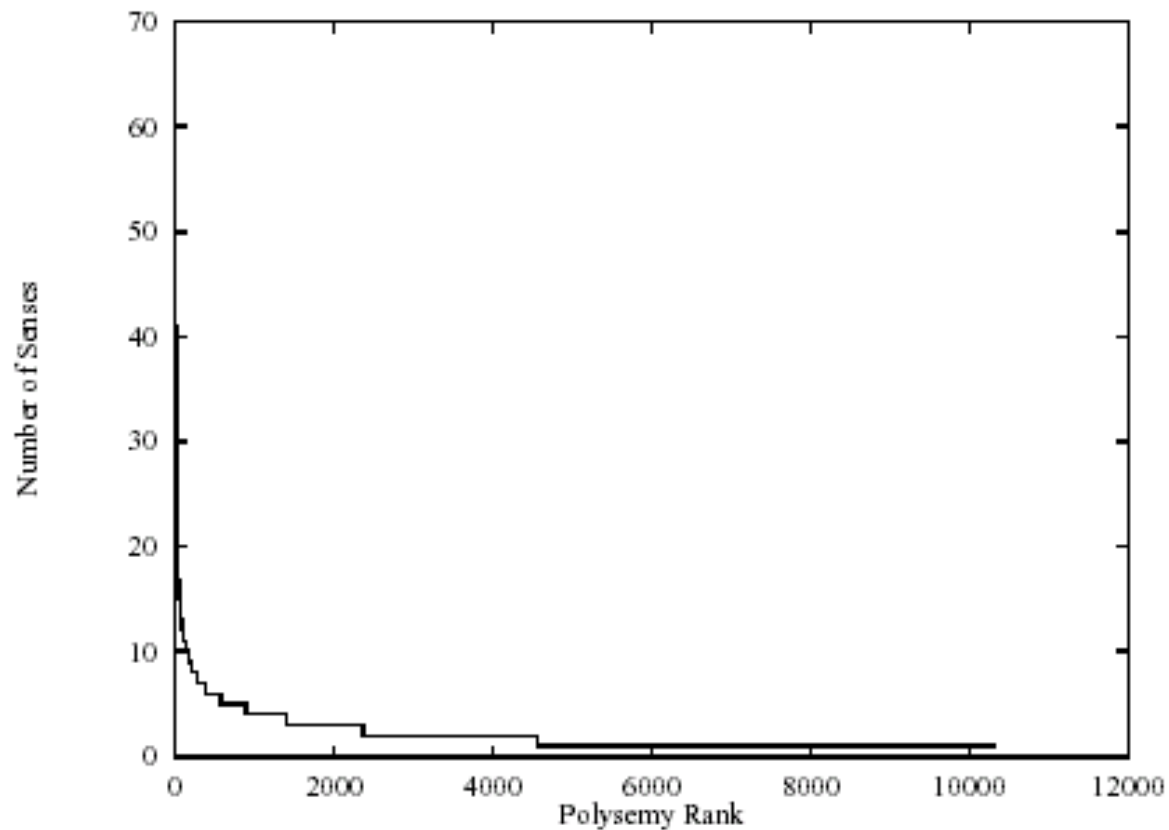
Adjective

- 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"*

Some WordNet Statistics

Part-of-speech	Avg Polysemy	Avg Polysemy w/o 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 verb senses

WordNet relations (between **synsets**)

- 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*

- Word sense disambiguation
  - Background from linguistics
    - Lexical semantics
  - On-line resources
  - Computational approaches