Semantic analysis

- Assigning meanings to linguistic utterances
  - **Compositional semantics**: we can derive the meaning of the whole sentence from the meanings of the parts.
    - Marseille 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

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

Caveats

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

Last class
- Introduction to the field of NLP

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

Semantic analysis
- Assigning meanings to linguistic utterances
- **Compositional semantics**: we can derive the meaning of the whole sentence from the meanings of the parts.
- Marseille 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

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

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

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

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?
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.

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

Next lectures

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

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

<table>
<thead>
<tr>
<th>Category</th>
<th>Unique Forms</th>
<th>Number of Senses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>94474</td>
<td>116317</td>
</tr>
<tr>
<td>Verb</td>
<td>10319</td>
<td>22066</td>
</tr>
<tr>
<td>Adjective</td>
<td>20170</td>
<td>29881</td>
</tr>
<tr>
<td>Adverb</td>
<td>4546</td>
<td>5677</td>
</tr>
</tbody>
</table>

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)

Statistics

<table>
<thead>
<tr>
<th>POS</th>
<th>Unique Strings</th>
<th>Synsets</th>
<th>Total word+sense pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>117798</td>
<td>82115</td>
<td>146312</td>
</tr>
<tr>
<td>Verb</td>
<td>11529</td>
<td>13767</td>
<td>25047</td>
</tr>
<tr>
<td>Adj</td>
<td>21479</td>
<td>18156</td>
<td>30002</td>
</tr>
<tr>
<td>Adv</td>
<td>4546</td>
<td>3621</td>
<td>5580</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>155287</strong></td>
<td><strong>11765</strong></td>
<td><strong>206941</strong></td>
</tr>
</tbody>
</table>
More WordNet Statistics

<table>
<thead>
<tr>
<th>Part-of-speech</th>
<th>Avg Polysemy</th>
<th>Avg Polysemy w/o monosemous words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>1.24</td>
<td>2.79</td>
</tr>
<tr>
<td>Verb</td>
<td>2.17</td>
<td>3.57</td>
</tr>
<tr>
<td>Adjective</td>
<td>1.40</td>
<td>2.71</td>
</tr>
<tr>
<td>Adverb</td>
<td>1.25</td>
<td>2.50</td>
</tr>
</tbody>
</table>

Distribution of senses

- Zipf distribution of senses

WordNet relations

- Nouns
  - Hyponym
  - Hypernym
  - Has-Member
  - Member-Of
  - Has-Part
  - Part-Of
  - Antonym

<table>
<thead>
<tr>
<th>Relation</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyponym</td>
<td>From concepts to superordinates</td>
<td>breakfast → meal, meal → lunch</td>
</tr>
<tr>
<td>Hypernym</td>
<td>From concepts to subtypes</td>
<td>faculty → professor</td>
</tr>
<tr>
<td>Has-Member</td>
<td>From groups to their members</td>
<td>captain → crew</td>
</tr>
<tr>
<td>Member-Of</td>
<td>From members to their groups</td>
<td>table → leg</td>
</tr>
<tr>
<td>Has-Part</td>
<td>From wholes to parts</td>
<td>course → meal</td>
</tr>
<tr>
<td>Part-Of</td>
<td>From parts to wholes</td>
<td>leader → follower</td>
</tr>
<tr>
<td>Antonym</td>
<td>Opposites</td>
<td></td>
</tr>
</tbody>
</table>

- Verbs
  - Hypernym
  - Hyponym
  - Entails
  - Antonym

<table>
<thead>
<tr>
<th>Relation</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypernym</td>
<td>From events to superordinate events</td>
<td>fly → travel</td>
</tr>
<tr>
<td>Hyponym</td>
<td>From events to their subtypes</td>
<td>walk → stroll</td>
</tr>
<tr>
<td>Entails</td>
<td>From events to the events they entail</td>
<td>snore → sleep</td>
</tr>
<tr>
<td>Antonym</td>
<td>Opposites</td>
<td>increase ⇔ decrease</td>
</tr>
</tbody>
</table>

- Adjectives/adverbs
  - Antonym
  - Adverb

<table>
<thead>
<tr>
<th>Relation</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antonym</td>
<td>Opposite</td>
<td>heavy ⇔ light</td>
</tr>
<tr>
<td>Adverb</td>
<td>Opposite</td>
<td>quickly ⇔ slowly</td>
</tr>
</tbody>
</table>

Next week

- Computational methods for WSD!!