Introduction to lexical semantics

- Lexical semantics
  - 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

Lexical semantic relations: homonymy

- Homonyms: words that have the same orthographic 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.

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?
    » 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**

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

CS6740/INFO6300

- Short intro to word sense disambiguation
  - Lexical semantics
  - Lexical semantic resources: WordNet
    - Word sense disambiguation
      » Supervised machine learning methods
      » WSD evaluation

WordNet

- Handcrafted database of lexical relations
- Three separate databases: nouns; verbs; adjectives and adverbs
- Each database is a set of lexical entries (unique orthographic forms)
  - Entries described and indexed in terms of synsets, i.e., sets of synonyms (lexemes with the “same meaning”)
Sample WordNet 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)

Some 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 (among synsets)

- Nouns
- Verbs
- Adjectives/adverbs

<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 subordinate events</td>
<td>walk → stroll</td>
</tr>
<tr>
<td>Entails</td>
<td>From events to their subtypes</td>
<td>snore → sleep</td>
</tr>
<tr>
<td>Antonym</td>
<td>From events to the events they entail</td>
<td>increase → decrease</td>
</tr>
<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>
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.

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.

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 in context $C$
    » Retrieve all of the sense definitions, $S$, for $W$ from the MRD
    » Compare each $s$ in $S$ to the dictionary definitions $D$ of all the remaining words $c$ in the context $C$
    » Select the sense $s$ with the most overlap with $D$ (the definitions of the context words $C$)

Machine learning approaches

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

- Examples of task
  - (features + class)
  - description of context
  - correct word sense

ML Algorithm

Novel example
- (features)
- learn one such classifier for each lexeme to be disambiguated

Classifier
- (program)
- class

Running example

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.

1 Fish sense
2 Musical sense
3 …

Feature vector representation

- **target**: the word to be disambiguated
- **context**: portion of the surrounding text
  - Select a “window” size
  - Tagged with part-of-speech information
  - Stemming or morphological processing
  - Possibly some partial parsing
- Convert the context (and target) into a set of features
  - Attribute-value pairs
    - Numeric, boolean, categorical, …

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.

| pre2-word | pre2-pos | pre1-word | pre1-pos | fol1-word | fol1-pos | fol2-word | fol2-pos | guitar | NN1 | and | CJC | player | NN1 | stand | VVB |
Co-occurrence features

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

0 0 0 1 0 0 0 0 1 0

Inductive ML framework

Examples of task
(features + class) correct word sense

description of context
ML Algorithm

Novel example (features) → Classifier (program) → class

learn one such classifier for each lexeme to be disambiguated

Lots of options!!!!

CS6740/INFO6300

- Short intro to word sense disambiguation
  - Lexical semantics
  - Lexical semantic resources: WordNet
  - Word sense disambiguation
    » Supervised machine learning methods
    » WSD evaluation

SENSEVAL

- Three tasks (originally)
  - Lexical sample
  - All-words
  - Translation
- Multiple (12+) languages
- Lexicon
  - SENSEVAL-1: from HECTOR corpus
  - SENSEVAL-2: from WordNet 1.7
- Lots of community participation
  - SENSEVAL-1 (1998): 93 systems from 34 teams
Lexical sample task

- Select a sample of words from the lexicon
- Systems must then tag instances of the sample words in short extracts of text
- SENSEVAL-1: 35 words
  - 700001 John Dos Passos wrote a poem that talked of "the bitter beat look, the scorn on the lip."
  - 700002 The beans almost double in size during roasting. Black beans are over roasted and will have a bitter flavour and insufficiently roasted beans are pale and give a colourless, tasteless drink.

### Lexical sample task: SENSEVAL-1

<table>
<thead>
<tr>
<th>Nouns</th>
<th>Verbs</th>
<th>Adjectives</th>
<th>Indeterminates</th>
</tr>
</thead>
<tbody>
<tr>
<td>-n</td>
<td>-v</td>
<td>-a</td>
<td>-p</td>
</tr>
<tr>
<td>accident</td>
<td>amaze</td>
<td>brilliant</td>
<td>band</td>
</tr>
<tr>
<td>behaviour</td>
<td>bet</td>
<td>deaf</td>
<td>bitter</td>
</tr>
<tr>
<td>bet</td>
<td>bother</td>
<td>floating</td>
<td>hurdle</td>
</tr>
<tr>
<td>disability</td>
<td>bury</td>
<td>generous</td>
<td>sanction</td>
</tr>
<tr>
<td>excess</td>
<td>calculate</td>
<td>giant</td>
<td>shake</td>
</tr>
<tr>
<td>float</td>
<td>consume</td>
<td>modest</td>
<td></td>
</tr>
<tr>
<td>giant</td>
<td>derive</td>
<td>slight</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>TOTAL</td>
<td>TOTAL</td>
<td>TOTAL</td>
<td>TOTAL</td>
</tr>
<tr>
<td>2756</td>
<td>2501</td>
<td>1406</td>
<td>1785</td>
</tr>
</tbody>
</table>

All-words task

- Systems must tag almost all of the content words in a sample of running text
  - sense-tag all predicates, nouns that are heads of noun-phrase arguments to those predicates, and adjectives modifying those nouns
  - ~5,000 running words of text
  - ~2,000 sense-tagged words

Translation task

- SENSEVAL-2 task
- Only for Japanese
- word sense is defined according to translation distinction
  - if the head word is translated differently in the given expressional context, then it is treated as constituting a different sense
- word sense disambiguation involves selecting the appropriate English word/phrase/sentence equivalent for a Japanese word
### SENSEVAL-2 results (2001)

<table>
<thead>
<tr>
<th>Language</th>
<th>Task</th>
<th>No. of submissions</th>
<th>No. of teams</th>
<th>IAA</th>
<th>Baseline</th>
<th>Best system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Czech</td>
<td>AW</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>94</td>
</tr>
<tr>
<td>Basque</td>
<td>LS</td>
<td>3</td>
<td>2</td>
<td>.75</td>
<td>.65</td>
<td>76</td>
</tr>
<tr>
<td>Estonian</td>
<td>AW</td>
<td>2</td>
<td>2</td>
<td>.72</td>
<td>.85</td>
<td>67</td>
</tr>
<tr>
<td>Italian</td>
<td>LS</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>39</td>
</tr>
<tr>
<td>Korean</td>
<td>LS</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>.71</td>
<td>74</td>
</tr>
<tr>
<td>Spanish</td>
<td>LS</td>
<td>12</td>
<td>5</td>
<td>.64</td>
<td>.48</td>
<td>65</td>
</tr>
<tr>
<td>Swedish</td>
<td>LS</td>
<td>8</td>
<td>5</td>
<td>.95</td>
<td>-</td>
<td>70</td>
</tr>
<tr>
<td>Japanese</td>
<td>LS</td>
<td>7</td>
<td>3</td>
<td>.86</td>
<td>.72</td>
<td>78</td>
</tr>
<tr>
<td>Japanese</td>
<td>TL</td>
<td>9</td>
<td>8</td>
<td>.81</td>
<td>.37</td>
<td>79</td>
</tr>
<tr>
<td>English</td>
<td>AW</td>
<td>21</td>
<td>12</td>
<td>.75</td>
<td>.57</td>
<td>69</td>
</tr>
<tr>
<td>English</td>
<td>LS</td>
<td>26</td>
<td>15</td>
<td>.86</td>
<td>.51/16</td>
<td>64/40</td>
</tr>
</tbody>
</table>

### SENSEVAL-2 de-briefing

- **Where next?**
  - Supervised ML approaches worked best
    - Looking at the role of feature selection algorithms
  - Need a well-motivated sense inventory
    - Inter-annotator agreement went down vs. SENSEVAL-1 (moved to WordNet senses)
  - Need to tie WSD to real applications
    - The translation task was a good initial attempt

### SENSEVAL-3 2004

- 14 core WSD tasks including
  - All words (Eng, Italian): 5000 word sample
  - Lexical sample (7 languages)
- Tasks for identifying semantic roles, for multilingual annotations, logical form, subcategorization frame acquisition
- Evaluations now called SEMEVAL

### Results

- 27 teams, 47 systems
- Most frequent sense baseline
  - 55.2% (fine-grained)
  - 64.5% (coarse)
- Most systems significantly above baseline
  - Including some unsupervised systems
- Best system
  - 72.9% (fine-grained)
  - 79.3% (coarse)
**SENSEVAL-3 lexical sample results**

<table>
<thead>
<tr>
<th>System/Team</th>
<th>Description</th>
<th>F1</th>
<th>R</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCCS (3)</td>
<td>A M fois system, with correction of the a priori disagreement, by dividing the output confidence of the senses by frequency (c = 0.2)</td>
<td>72.9</td>
<td>72.0</td>
<td>79.3</td>
</tr>
<tr>
<td>BCCS (4)</td>
<td>A second system, with different correction fraction of a priori disagreement, and reweighting word confidence (LSA)</td>
<td>72.6</td>
<td>72.8</td>
<td>79.5</td>
</tr>
<tr>
<td>BCCS</td>
<td>A combination of two neural networks, using a neural network to predict the most likely sense, and a second neural network to predict the confidence of each sense</td>
<td>73.4</td>
<td>73.4</td>
<td>78.8</td>
</tr>
<tr>
<td>BCCS</td>
<td>Similar to BCCS, but with different number of features</td>
<td>72.3</td>
<td>72.3</td>
<td>78.8</td>
</tr>
<tr>
<td>NIST (4)</td>
<td>A neural network classifier (LSTM), using local and topical features, with a term weighting scheme</td>
<td>73.2</td>
<td>72.2</td>
<td>78.4</td>
</tr>
<tr>
<td>NIST</td>
<td>Similar to NIST, but with different number of features</td>
<td>72.2</td>
<td>72.1</td>
<td>78.6</td>
</tr>
</tbody>
</table>

**SENSEVAL-3 results (unsupervised)**

<table>
<thead>
<tr>
<th>System/Team</th>
<th>Description</th>
<th>F1</th>
<th>R</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>unison</td>
<td>An unsupervised system using a Link-like similarity between context of ambiguous word and dictionary definitions. Experiments performed for various window sizes, various similarity measures</td>
<td>73.0</td>
<td>73.0</td>
<td>78.1</td>
</tr>
<tr>
<td>ITT Bombay</td>
<td>A Maximum entropy model for unsupervised clustering, using neighboring words and syntactic structures as features. A few selected instances are used to map context classes to WordNet/Semcor senses.</td>
<td>56.3</td>
<td>56.3</td>
<td>60.4</td>
</tr>
<tr>
<td>L2S LNA</td>
<td>A combination of two unsupervised modules, using base part of speech and frequency information</td>
<td>54.7</td>
<td>54.7</td>
<td>60.6</td>
</tr>
<tr>
<td>L2S LNA</td>
<td>An unsupervised system relying on definition properties (synonym, antonym, subordination patterns, etc.) to distinguish senses. Performance is generally a function of how well senses are distinguished</td>
<td>45.6</td>
<td>45.6</td>
<td>55.5</td>
</tr>
<tr>
<td>L2S LNA</td>
<td>An unsupervised system that combines the conceptual senses with the frequency of words to disambiguate, information about domain also taken into account</td>
<td>50.2</td>
<td>41.7</td>
<td>59.1</td>
</tr>
</tbody>
</table>