Outline of today’s lecture

Lecture 6: Lexical semantics
  - Words and concepts
  - Semantic relations
  - Polysemy
  - Word sense disambiguation
Semantics

**Compositional** semantics:
- studies how meanings of phrases are constructed out of the meaning of individual words
- principle of compositionality: meaning of each whole phrase derivable from meaning of its parts
- sentence structure conveys some meaning: obtained by syntactic representation

**Lexical** semantics:
- studies how the meanings of individual words can be represented and induced
What is lexical meaning?

- recent results in psychology and cognitive neuroscience give us some clues
- but we don’t have the whole picture yet
- different representations proposed, e.g.
  - formal semantic representations based on logic,
  - *or* taxonomies relating words to each other,
  - *or* distributional representations in statistical NLP
- but none of the representations gives us a complete account of lexical meaning
How to approach lexical meaning?

- **Formal semantics**: set-theoretic approach
e.g., cat’: the set of all cats; bird’: the set of all birds.
- meaning postulates, e.g.

\[ \forall x [\text{bachelor}'(x) \rightarrow \text{man}'(x) \land \text{unmarried}'(x)] \]

- Limitations, e.g. *is the current Pope a bachelor?*
- Defining concepts through enumeration of all of their features in practice is highly problematic
- How would you define e.g. *chair, tomato, thought, democracy?* – impossible for most concepts
- **Prototype theory** offers an alternative to set-theoretic approaches
How to approach lexical meaning?

- **Formal semantics**: set-theoretic approach
e.g., cat′: the set of all cats; bird′: the set of all birds.
- meaning postulates, e.g.

  \[
  \forall x [\text{bachelor}'(x) \rightarrow \text{man}'(x) \land \text{unmarried}'(x)]
  \]
- Limitations, e.g. *is the current Pope a bachelor?*
- Defining concepts through enumeration of all of their features in practice is highly problematic
- How would you define e.g. *chair, tomato, thought, democracy?* – impossible for most concepts
- **Prototype theory** offers an alternative to set-theoretic approaches
Prototype theory

- introduced the notion of graded semantic categories
- no clear boundaries
- no requirement that a property or set of properties be shared by all members
- certain members of a category are prototypical – or instantiate the prototype

Prototype theory (continued)
Prototype theory (continued)
Prototype theory (continued)
Prototype theory (continued)

- Categories form around prototypes; new members added on basis of resemblance to prototype
- Features/attributes generally graded
- Category membership a matter of degree
- Categories do not have clear boundaries
Hyponymy: IS-A

- *dog* is a hyponym of *animal*
- *animal* is a hypernym of *dog*
- hyponymy relationships form a taxonomy
- works best for concrete nouns
Some issues concerning hyponymy

- not useful for all words: *thought, push, sticky*?
- individuation differences: is *table* a hyponym of *furniture*?
- multiple inheritance: e.g., is *coin* a hyponym of both *metal* and *money*?
- what does the top of the hierarchy look like?
Other semantic relations

Classical relations:

**Meronomy**: PART-OF  e.g., *arm* is a **meronym** of *body*, *steering wheel* is a meronym of *car* (piece vs part)

**Synonymy**  e.g., *aubergine/eggplant*.

**Antonymy**  e.g., *big/little*

Also:

**Near-synonymy/similarity**  e.g., *exciting/thrilling*

  e.g., *slim/slender/thin/skinny*
WordNet

- large scale, open source resource for English
- hand-constructed
- wordnets being built for other languages
- organized into synsets: synonym sets (near-synonyms)

S: (v) interpret, construe, see (make sense of; assign a meaning to) "What message do you see in this letter?"; "How do you interpret his behavior?"

S: (v) understand, read, interpret, translate (make sense of a language) "She understands French"; "Can you read Greek?"
WordNet tree for verbs

```
understand_1
  └── interpret_1, see_2, construe_1, comprehend_1, grasp_2, ...
  |  └── re-interpret_2, read_1, allegorise_1
      |  └── re-read_1, spell out_2, decipher_2, skim_7, trace_8
          |  └── skim over_2
```
Polysemy and word senses

The children ran to the store
If you see this man, run!
Service runs all the way to Cranbury
She is running a relief operation in Sudan
the story or argument runs as follows
Does this old car still run well?
Interest rates run from 5 to 10 percent
Who’s running for treasurer this year?
They ran the tapes over and over again
These dresses run small
Polysemy

- **homonymy**: unrelated word senses. *bank* (raised land) vs *bank* (financial institution)

- *bank* (financial institution) vs *bank* (in a casino): related but distinct senses.

- **regular polysemy** and sense extension
  - zero-derivation, e.g. *tango* (N) vs *tango* (V), or *rabbit, turkey, halibut* (meat / animal)
  - metaphorical senses, e.g. *swallow* [food], *swallow* [information], *swallow* [anger]
  - metonymy, e.g. he played *Bach*; he drank his *glass*.

- vagueness: *nurse, lecturer, driver*

- cultural stereotypes: *nurse, lecturer, driver*

No clearcut distinctions.
Dictionaries are not consistent.
Word sense disambiguation

Needed for many applications, difficult for large domains.

- typically assumes that we have a standard set of word senses (e.g., WordNet)
- frequency: e.g., diet: the food sense (or senses) is much more frequent than the parliament sense (Diet of Wurms)
- collocations: e.g. striped bass (the fish) vs bass guitar: syntactically related or in a window of words (latter sometimes called ‘cooccurrence’). Generally ‘one sense per collocation’.
WSD techniques

- **supervised** learning: cf. POS tagging from lecture 3. Need a training corpus. But sense-tagged corpora are difficult to construct, algorithms need far more data than POS tagging

- **minimally-supervised** learning (Yarowsky, 1995)

- **unsupervised** sense induction (lecture 8)
WSD by minimally-supervised learning

Yarowsky, David (1995) *Unsupervised word sense disambiguation rivalling supervised methods*

Disambiguating *plant* (factory vs vegetation senses):
1. Find contexts in training corpus:

<table>
<thead>
<tr>
<th>sense</th>
<th>training example</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>company said that the <em>plant</em> is still operating</td>
</tr>
<tr>
<td>?</td>
<td>although thousands of <em>plant</em> and animal species</td>
</tr>
<tr>
<td>?</td>
<td>zonal distribution of <em>plant</em> life</td>
</tr>
<tr>
<td>?</td>
<td>company manufacturing <em>plant</em> is in Orlando</td>
</tr>
<tr>
<td>etc</td>
<td></td>
</tr>
</tbody>
</table>
Yarowsky (1995): schematically

Initial state

[Diagram with various question marks and placeholders]
2. Identify some seeds to disambiguate a few uses. e.g., ‘plant life’ for vegetation use (A) ‘manufacturing plant’ for factory use (B):

<table>
<thead>
<tr>
<th>sense</th>
<th>training example</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>company said that the <em>plant</em> is still operating</td>
</tr>
<tr>
<td>?</td>
<td>although thousands of <em>plant</em> and animal species</td>
</tr>
<tr>
<td>A</td>
<td>zonal distribution of <em>plant</em> life</td>
</tr>
<tr>
<td>B</td>
<td>company manufacturing <em>plant</em> is in Orlando</td>
</tr>
<tr>
<td></td>
<td>etc</td>
</tr>
</tbody>
</table>
Seeds

Word sense disambiguation
3. Train a **decision list** classifier on the Sense A/Sense B examples.

<table>
<thead>
<tr>
<th>reliability</th>
<th>criterion</th>
<th>sense</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.10</td>
<td><em>plant</em> life</td>
<td>A</td>
</tr>
<tr>
<td>7.58</td>
<td>manufacturing <em>plant</em></td>
<td>B</td>
</tr>
<tr>
<td>6.27</td>
<td><em>animal</em> within 10 words of <em>plant</em></td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>etc</td>
<td></td>
</tr>
</tbody>
</table>

Decision list classifier: automatically trained if/then statements.

Experimenter decides on classes of test by providing definitions of features of interest: system builds specific tests and provides reliability metrics.
4. Apply the classifier to the training set and add reliable examples to A and B sets.

<table>
<thead>
<tr>
<th>sense</th>
<th>training example</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>company said that the <em>plant</em> is still operating</td>
</tr>
<tr>
<td>A</td>
<td>although thousands of <em>plant</em> and animal species</td>
</tr>
<tr>
<td>A</td>
<td>zonal distribution of <em>plant</em> life</td>
</tr>
<tr>
<td>B</td>
<td>company manufacturing <em>plant</em> is in Orlando</td>
</tr>
<tr>
<td></td>
<td>etc</td>
</tr>
</tbody>
</table>

5. Iterate the previous steps 3 and 4 until convergence
Iterating:
Final:
6. Apply the classifier to the unseen test data

- ‘one sense per discourse’: can be used as an additional refinement
- Yarowsky’s experiments were nearly all on homonyms: these principles may not hold as well for sense extension.
Evaluation of WSD

- Yarowsky reported an accuracy of 95%, but on ’easy’ homonymous examples
- SENSEVAL competitions
- evaluate against WordNet
- baseline: pick most frequent sense — hard to beat (but don’t always know most frequent sense)
- human ceiling varies with words
- MT task: more objective but sometimes doesn’t correspond to polysemy in source language
Uses of WSD and lexical semantics in NLP

- any NLP application that needs access to semantics!
- e.g. sentiment analysis:
  - *feel drained* vs *drain pasta*
Metaphor and sentiment examples

He injected new life into the performance.
He added new life into the performance.
inject hydrogen into the balloon

I can't buy this story.
I can't believe this story.
This sum will buy you a ride on the train

The speech crowned the meeting.
The speech culminated the meeting.
The prince was crowned in Westminster Abbey

The police smashed the drug ring after they were tipped off.
The police arrested the drug ring after they were tipped off.
She smashed her car against the guard rail

She salts her lectures with jokes.
She complements her lectures with jokes.
people used to salt meats on ships
Uses of WSD and lexical semantics in NLP

- any NLP application that needs access to semantics!
- e.g. sentiment analysis: 
  *feel drained* vs *drain pasta*
- or information retrieval: 
  query expansion by synonymy or hyponymy