Outline of today’s lecture

Lecture 5: Introduction to semantics & lexical semantics
  Dependency structures
  Words and concepts
  Semantic relations
  Polysemy
  Supervised classification in NLP
  Word sense disambiguation
Dependency structure

- Alternative representation to syntax trees.
- Relate words to each other via labelled directed arcs – dependencies.

```
some  big  angry  dogs  bark  loudly
       MOD  SUBJ MOD  MOD
       MOD  DET
```
Why are dependencies important?

Provide an interface to semantics (‘who does what to whom’)

Example
John hit the ball.

Dependency parsing
(Subj head=hit dep=John)
(OBJ head=hit dep=ball)
(DET head=ball dep=the)
The cost of parsing errors...

Incorrect dependencies

(SUBJ head=hit dep=ball)
(OBJ head=hit dep=John)
(DET head=ball dep=the)
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
Semantic relations

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
- multiple inheritance: e.g., is *coin* a hyponym of both *metal* and *money*?
Other semantic 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
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
- Relies on context, e.g. collocations: *striped bass* (the fish) vs *bass guitar*.

Methods:
- **supervised learning:**
  - Assume a predefined set of word senses, e.g. WordNet
  - Need a large sense-tagged training corpus (difficult to construct)
- **minimally-supervised** learning (Yarowsky, 1995)
- **unsupervised** sense induction (lecture 7)
Supervised classification in NLP

Used in many NLP tasks, for instance:

- Text classification (e.g. sentiment analysis, spam filtering)
- Semantic tasks (e.g. WSD, named entity recognition)
- Discourse processing (lecture 9)

**Input:**

- a set of data points $d \in D$
- a set of classes $C = \{c_1, c_2, ..., c_K\}$

**Output:**

- a predicted class $c \in C$
Features in supervised classification

- Data points are represented in the form of features
- Features link what we observe in the data \((D)\) with the class \(c\) we want to predict
- During training we learn weights for different features

Features in semantic classification tasks:

- Internal structure of words (for some tasks, e.g. named entity recognition)
- Context: e.g. previous or next word; a window of 10 words
- Syntactic relations with words in the context
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></td>
<td>etc</td>
</tr>
</tbody>
</table>
Yarowsky (1995): schematically

Initial state
2. Identify some seeds to disambiguate a few uses:

   ‘*plant* life’ for vegetation use (A)
   ‘manufacturing *plant*’ for factory use (B)

<table>
<thead>
<tr>
<th></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 etc</td>
</tr>
</tbody>
</table>
Seeds

Word sense disambiguation
3. Train a decision list classifier on Sense A/Sense B examples. 
Rank features by log-likelihood ratio:

$$\log \left( \frac{P(\text{Sense}_A | f_i)}{P(\text{Sense}_B | f_i)} \right)$$

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

etc

5. Iterate the previous steps 3 and 4 until convergence
Iterating:
Final:

A A B B
A A
A A BB
A A
A A BB
A A
A A
A A
A B B B
B B
B B
B B
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.
Problems with WSD as supervised classification

Yarowsky reported an accuracy of 95%, but ...

- on ’easy’ homonymous examples
- real performance around 75% (in SENSEVAL)
- need to predefine word senses (not theoretically sound)
- need a very large training corpus (difficult to annotate, humans do not agree)
- learn a model for individual words — no real generalisation

Better way:

- unsupervised sense induction (but a very hard task)
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