Compositional and lexical semantics

- Compositional semantics: the construction of meaning (generally expressed as logic) based on syntax.
  This lecture:
  - Semantics with FS grammars

- Lexical semantics: the meaning of individual words.
  This lecture:
  - lexical semantic relations and WordNet
  - one technique for word sense disambiguation
Simple compositional semantics in feature structures

- Semantics is built up along with syntax
- Subcategorization ‘slot’ filling instantiates syntax
- Formally equivalent to logical representations (below: predicate calculus with no quantifiers)
- Alternative FS encodings possible
Objective: obtain the following semantics for
they like fish:
pron(x) \land (like_v(x, y) \land fish_n(y))

Feature structure encoding:

\[
\begin{align*}
\text{PRED} & \quad \text{and} \\
\text{ARG1} & \quad \text{PRED} \quad \text{pron} \\
\text{ARG1} & \quad \text{PRED} \quad \text{like}_v \\
\text{ARG1} & \quad \text{PRED} \quad \text{fish}_n \\
\text{ARG2} & \quad \text{PRED} \quad \text{pron} \\
\text{ARG2} & \quad \text{PRED} \quad \text{like}_v \\
\text{ARG2} & \quad \text{PRED} \quad \text{fish}_n
\end{align*}
\]
• Corresponds to \( \text{fish}(x) \) where the INDEX points to the characteristic variable of the noun (that is \( x \)).

The INDEX is unambiguous here, but e.g., \( \text{picture}(x, y) \land \text{sheep}(y) \)

\textit{picture of sheep}
like

Verb entry

- Linking syntax and semantics: ARG1 linked to the SPR’s index; ARG2 linked to the COMP index.
COMP filling rule

\[
\begin{bmatrix}
\text{HEAD} & 1 \\
\text{COMP} & \text{filled} \\
\text{SPR} & 3 \\
\text{SEM} & \text{PRE} \text{D} \text{ and} \\
\text{ARG}1 & 4 \\
\text{ARG}2 & 5
\end{bmatrix}
\rightarrow
\begin{bmatrix}
\text{HEAD} & 1 \\
\text{COMP} & 2 \\
\text{SPR} & 3 \\
\text{SEM} & 4 \\
\text{COMP} & \text{filled} \\
\text{SEM} & 5
\end{bmatrix}
\]

- As last time: object of the verb (DTR2) ‘fills’ the COMP slot
- New: semantics on the mother is the ‘and’ of the semantics of the dtrs
Apply to *like*
Apply to like fish

\[ \text{like}_v(x, y) \land \text{fish}_n(y) \]
Logic in semantic representation

- Meaning representation for a sentence is called the *logical form*
- Standard approach to composition in theoretical linguistics is lambda calculus, building FOPC or higher order representation
- Representation above is impoverished but can build FOPC in FSs
- Theorem proving
- Generation: starting point is logical form, not string.
Meaning postulates

• e.g.,
  \[ \forall x [\text{bachelor}'(x) \rightarrow \text{man}'(x) \land \text{unmarried}'(x)] \]

• usable with compositional semantics and theorem provers

• e.g. from ‘Kim is a bachelor’, we can construct the LF
  \[ \text{bachelor}'(\text{Kim}) \]
  and then deduce
  \[ \text{unmarried}'(\text{Kim}) \]

• OK for narrow domains, but ‘classical’ lexical semantic relations are more generally useful
Lexical semantic relations

Hyponymy: IS-A:

- (a sense of) *dog* is a *hyponym* of (a sense of) *animal*
- *animal* is a *hypernym* of *dog*
- hyponymy relationships form a *taxonomy*
- works best for concrete nouns

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*
WordNet

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

Overview of adj red:

1. (43) red, reddish, ruddy, blood-red, carmine, cerise, cherry, cherry-red, crimson, ruby, ruby-red, scarlet --
   (having any of numerous bright or strong colors reminiscent of the color of blood or cherries or tomatoes or rubies)
2. (8) red, reddish --
   ((used of hair or fur) of a reddish brown color; "red deer"; reddish hair)
Hyponymy in WordNet

Sense 6

big cat, cat

=> leopard, Panthera pardus
    => leoparedess
    => panther

=> snow leopard, ounce, Panthera uncia

=> jaguar, panther, Panthera onca,
    Felis onca

=> lion, king of beasts, Panthera leo
    => lioness
    => lionet

=> tiger, Panthera tigris
    => Bengal tiger
    => tigress

=> liger

=> tiglon, tigon

=> cheetah, chetah, Acinonyx jubatus

=> saber-toothed tiger, sabertooth
    => Smiledon californicus
    => false saber-toothed tiger
Some uses of lexical semantics

- Semantic classification: e.g., for selectional restrictions (e.g., the object of *eat* has to be something edible) and for named entity recognition
- Shallow inference: ‘X murdered Y’ implies ‘X killed Y’ etc
- Back-off to semantic classes in some statistical approaches
- Word-sense disambiguation
- Machine Translation: if you can’t translate a term, substitute a hypernym
- Query expansion: if a search doesn’t return enough results, one option is to replace an over-specific term with a hypernym
Word sense disambiguation

Needed for many applications, problematic for large domains. May depend on:

- 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’.
- selectional restrictions/preferences (e.g., *Kim eats bass*, must refer to fish
WSD techniques

- supervised learning: cf. POS tagging from lecture 3. But sense-tagged corpora are difficult to construct, algorithms need far more data than POS tagging
- unsupervised learning (see below)
- Machine readable dictionaries (MRDs)
- selectional preferences: don’t work very well by themselves, useful in combination with other techniques
WSD by (almost) unsupervised learning

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 plant is still operating</td>
</tr>
<tr>
<td>?</td>
<td>although thousands of plant and animal species</td>
</tr>
<tr>
<td>?</td>
<td>zonal distribution of plant life</td>
</tr>
<tr>
<td>?</td>
<td>company manufacturing plant is in Orlando etc</td>
</tr>
</tbody>
</table>

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 plant is still operating</td>
</tr>
<tr>
<td>?</td>
<td>although thousands of plant and animal species</td>
</tr>
<tr>
<td>A</td>
<td>zonal distribution of plant life</td>
</tr>
<tr>
<td>B</td>
<td>company manufacturing plant is in Orlando etc</td>
</tr>
</tbody>
</table>

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>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</td>
<td>A</td>
</tr>
<tr>
<td>etc</td>
<td></td>
<td></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 plant is still operating</td>
</tr>
<tr>
<td>A</td>
<td>although thousands of plant and animal species</td>
</tr>
<tr>
<td>A</td>
<td>zonal distribution of plant life</td>
</tr>
<tr>
<td>B</td>
<td>company manufacturing plant is in Orlando</td>
</tr>
<tr>
<td>etc</td>
<td></td>
</tr>
</tbody>
</table>

5. Iterate the previous steps 3 and 4 until convergence

6. Apply the classifier to the unseen test data

‘one sense per discourse’: can be used as an additional refinement
Yarowsky (1995): schematically

Initial state

[Diagram of initial state with symbols and question marks]

Seeds

[Diagram of seeds with symbols and question marks]

life

manu.
Iterating:

Final:
Evaluation of WSD

- SENSEVAL and SENSEVAL-2 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