Chapter NLP:V

V. Semantics

- Semantic Phenomena
- **Symbolic Semantics**
- Distributional Semantics
- Compositional Semantics
- Frame Semantics
Lexical semantics describes the relation between meaning and form of words:

**Semasiology** Which meaning can be assumed from word form. **Semantic relations** describe difference in meaning with identical form.

- **Polysemy:** mouse \(\text{animal} \text{ vs. } \text{input device}\)
- **Specialization:** corn \(\text{wheat} \text{ vs. } \text{oats}\)
- **Generalization:** moon \(\text{of earth} \text{ vs. } \text{any satellite}\)
- **Metaphor:** desktop \(\text{of a desk} \text{ vs. } \text{on screen}\)

**Onomasiology** Which **relations** exist between the concepts. Which forms exists for the concepts.

- **Lexical Relations:** mouse vs. rodent
- **Lexical Fields**
- **Frames**
- **Distributional relations**
Lexical Semantics

Word Senses

Meaning is usually represented discretely as a word sense.

- Senses are often denominated by superscript\(^1\).
- The informal description of a sense is its gloss,
- Senses can be modeled explicitly with their lexical relations or by componential analysis, or intrinsically via distributional semantics.

Sense description by gloss:

<table>
<thead>
<tr>
<th>Sense</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank(^1):</td>
<td>financial institution that accepts, deposits, and lends money.</td>
</tr>
<tr>
<td>Bank(^2):</td>
<td>sloping land (especially the slope beside a body of water).</td>
</tr>
<tr>
<td>Red(^1):</td>
<td>the color of blood or a ruby.</td>
</tr>
<tr>
<td>Blood(^1):</td>
<td>the red liquid that circulates in the vains of animals.</td>
</tr>
</tbody>
</table>
Lexical Semantics
Lexical Relations (selection)

Polysemy  Same lexeme, different sense.

The semantic relations are subtypes of polysemy.

Synonym  Different sense but similar meaning.

couch ←→ sofa  big ←→ large

Antonymy  Opposite meaning.

long ←→ short

Hyponymy/Hypernymy  One sense is less/more specific. Also called IS-A

car  \(\text{Hyponym}\)  vehicle

car  \(\text{Hypernym}\)  ID.3

Meronym/Holonym  The part-whole relation.

wheel  \(\text{Meronym}\)  car

car  \(\text{Holonym}\)  wheel

Relations are defined over senses, not lexemes:

Synonymous  \(\text{big}^1\)  plane  ←→  \(\text{large}^1\)  plane

Not synonymous  \(\text{big}^2\)  sister  ←→  \(\text{large}^1\)  sister
Lexical Semantics
WordNet

The largest English database of word senses is WordNet. [Fellbaum, 1998]

- WordNet has entries for lemmas.
- An entry has 1 or more synsets: sets of near-synonymous senses. Synsets represent concepts of meaning.
- Synsets and lemmas are split by word class: Noun, Verb, Adjective, Adverb.
- Each synset has a topic (supersense), assigned from a per word class fixed set.

Entry for the lemma *bass* (Noun):

<table>
<thead>
<tr>
<th>Synset</th>
<th>Topic</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>bass¹</td>
<td>attribute</td>
<td>the lowest part of the musical range</td>
</tr>
<tr>
<td>bass²</td>
<td>animal</td>
<td>edible marine and freshwater spiny-finned fishes</td>
</tr>
<tr>
<td>sea bass, bass²</td>
<td>food</td>
<td>the lean flesh of a saltwater fish of the family Serranidae</td>
</tr>
<tr>
<td>bass,¹</td>
<td>communication</td>
<td>part the lowest part in polyphonic music</td>
</tr>
<tr>
<td>basso,¹</td>
<td>person</td>
<td>an adult male singer with the lowest voice</td>
</tr>
</tbody>
</table>
Lexical Semantics

WordNet

The largest English database of word senses is WordNet. [Fellbaum, 1998]

- WordNet has entries for lemmas.
- An entry has 1 or more synsets: sets of near-synonymous senses. Synsets represent concepts of meaning.
- Synsets and lemmas are split by word class: Noun, Verb, Adjective, Adverb.
- Each synset has a topic (supersense), assigned from a per word class fixed set.

Entry for the lemma *ride* (Verb):

<table>
<thead>
<tr>
<th>Synset</th>
<th>Supersense</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>sit, ride</td>
<td>motion</td>
<td>sit and travel on the back of animal</td>
</tr>
<tr>
<td>sit, ride</td>
<td>motion</td>
<td>be carried or travel on or in a vehicle</td>
</tr>
<tr>
<td>tease, ride,</td>
<td>communication</td>
<td>harass with persistent criticism or carping</td>
</tr>
<tr>
<td>rally, bait,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rag, twit,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tantalize,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>razz, taunt,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cod</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ride</td>
<td>stative</td>
<td>continue undisturbed and without interference ?Let it ride?</td>
</tr>
</tbody>
</table>
Remarks:

- WordNet synsets are separated by word class and do not overlap:
  - Nouns: 117,798 lemmas (avg. 1.23 senses)
  - Verbs: 11,529 lemmas (avg. 2.16 senses)
  - Adjectives: 22,479 lemmas
  - Adverbs: 4,481 lemmas

- In WordNet, semantic relations are encoded as one lexical relation (Polysemy) with 3 additional subrelations:
  - Constructional/structured polysemy: Same sense entry refers to different entities
    (The) Times printed paper vs. the news contained in it vs. the organization
  - Sense extension polysemy: Derives a new synset from an old sense
    chicken animal vs. meat of animal
  - Homonymy: Same sense, very different meaning
    bank river bank vs. financial bank
Lexical Semantics
Word Sense Disambiguation

Word Sense Disambiguation (WSD) is the task of assigning to each word in a text the correct sense from a sense lexicon.

- WSD is similar to tagging, but more difficult.
  There are several classes for each word.
- To disambiguate a small set of words, classification works.
- Using the most frequent sense every time is a strong baseline.
  Disambiguating medical terms in lab reports
- There are several datasets (semantic concordance) where each word is annotated with its sense.
  - SEMCOR (Brown Corpus, English), SENSEVAL-3 (English), [Vossen et al., 2011] (Dutch), [Heinrich et al., 2012] (German)
  - Annotated are word class and sense id for open-class words.

Example annotations from SEMCOR:
You will find\textsuperscript{9} that avocado\textsuperscript{1} is\textsuperscript{1} unlike\textsuperscript{1} other\textsuperscript{1} fruit\textsuperscript{1} you have ever\textsuperscript{1} \textsuperscript{r} tasted\textsuperscript{2}
Lexical Semantics
Word Sense Disambiguation: Lesk

Idea: The context of a word should overlap with the words in the gloss of its sense.

- Does not need training data.
- Easy to apply to new or low-resource languages.
- Glosses can easily be extended with (annotated) examples.

<table>
<thead>
<tr>
<th>Sense</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>bank(^1)</td>
<td>A financial institution that accepts deposits and channels the money into lending activities</td>
</tr>
<tr>
<td>bank(^2)</td>
<td>sloping land (especially the slope beside a body of water)</td>
</tr>
</tbody>
</table>

We guarantee that your bank deposits will cover future costs.
Simplified Lesk:
For the word \( w_i \) in a sequence \( w_1, \ldots, w_{i-k}, \ldots, w_i, \ldots, w_{i+k}, \ldots, w_n \) with window size \( k \) and glosses \( G_{w_i} = \{ g_{w_i,1}, \ldots, g_{w_i,j} \} \):

1. Remove stopwords from and lemmatize the context window
   \[ v_i := (w_{i-k}, \ldots, w_{i-1}, w_{i+1}, \ldots, w_{i+k}) \]
   and all glosses \( g_j \in G_{w_i} \).
2. Vectorize \( v_i \) and all \( g_j \in G_{w_i} \).
3. Disambiguate \( w_i \) by the lowest cosine between context and gloss vectors.

- LESK can be improved using tf·idf-weighted vectors or any other (semantic) similarity measure.
- Gloss vectors can be pre-computed.
Lexical Semantics
Word Sense Disambiguation: Classification

**Idea:** Classify the sense with sliding-window features (cf. sequence tagging).

*Example features:*

1. Words (lemmas/stems) in the context window
2. Part-of-speech tags for each word in the window
3. $n$-grams
4. Weighted average of the word embeddings

*Example:*

\[
\begin{align*}
w_{i-4} & \quad w_{i-3} & \quad w_{i-2} & \quad w_{i-1} & \quad w_i & \quad w_{i+1} & \quad w_{i+2} & \quad w_{i+3} & \quad w_{i+4} & \quad w_{i+5} \\
\text{We guarantee that your bank deposits will cover future costs.}
\end{align*}
\]

*Features for $w_i$ with $k = 2$:*

<table>
<thead>
<tr>
<th>$w_{i-2}$</th>
<th>$\text{POS}_{i-2}$</th>
<th>$w_{i-1}$</th>
<th>$\text{POS}_{i-1}$</th>
<th>$w_{i+1}$</th>
<th>$\text{POS}_{i+1}$</th>
<th>$w_{i+2}$</th>
<th>$\text{POS}_{i+2}$</th>
<th>$w_{i+3}$</th>
<th>$w_{i+4}$</th>
<th>$w_{i+5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>that</td>
<td>IN</td>
<td>your</td>
<td>PRP</td>
<td>deposits</td>
<td>NN</td>
<td>will</td>
<td>MD</td>
<td>that your</td>
<td>deposits will</td>
<td></td>
</tr>
</tbody>
</table>
Remarks:

- The state of the art for WSD uses contextualized word embeddings.
- Word embeddings are identical for each sense of the same word form. They are not intrinsically useful for WSD.
- However, since word vector spaces embed semantic similarity, they can solve tasks like lexical substitution with a simple nearest neighbor search.
- A large language model like BERT produces contextualized word embeddings more or less as a by-product. Here, the same lexeme can have different vectors, depending on its context words. BERT solves WSD extremely well if there are vectors for each sense.
- Word Sense Induction tries to create lexicons like WordNet automatically by clustering the embedding space. This produces a synset collection with context vectors for each sense (the mean vector of each cluster) in an unsupervised fashion.
Lexical Semantics

Lexical Substitution

Lexical substitution tasks are subtask of WSD.

- Classic lexical substitution looks for one or more semantically similar replacement for certain words.
  
  My favorite thing about her is her straightforward honesty
  → My favorite thing about her is her sincere/genuine/frank honesty

- Lexical simplification looks for a easier to understand byt semantically similar replacement.
  
  John composed these verses → John wrote these poems

- Lexical substitution often uses the same techniques as WSD but does not require a lexicon.
Multi-word expressions (MWE) can function as singular lexical units.

MWE semantics can be

1. compositional,
   - driving instructor
   - argumentation quality assessment

2. idiomatic,
   - vice versa
   - kick the bucket

3. or in-between.
   - Long time no see