L90: Overview of Natural Language Processing
Lecture 10: Lexical Semantics

Weiwei Sun (materials mostly by Ann Copestake)

Department of Computer Science and Technology
University of Cambridge

Michaelmas 2020/21
Recap: formal meaning representations

- **natural language**
  - representation
  - formal language
    - formal semantics (logic)
      - models
        - representation
          - real world

- **sentences**

- **formulas**
  - entail
  - truth conditions
    - properties
      - follow
        - property

- **sentence**
  - truth condition

- **formula**
  - facts

from Yanjing Wang
Recap: formal meaning representations

natural language
  representation
  formal language
    formal semantics (logic)
      models
  real world

sentences
  sentence
    formulas
      entail
        truth conditions
          properties
            follow
                fact
Recap: formal meaning representations

natural language
  ↓
representation
  ↓
formal language
  ↓
formal semantics (logic)
  ↓
models
  ↓
representation
  ↓
real world

sentences
  ↓
formulas —— entail —— formula

truth conditions
  ↓
logic

properties —— follow —— property

facts

fact

from Yanjing Wang
Recap: formal meaning representations

natural language → representation → formal language → formal semantics (logic) → models → representation → real world

sentences

formulas → entail → formula

truth conditions

properties → follow → property

facts

fact

from Yanjing Wang
Lecture 10: Lexical Semantics

1. Lexical semantics: semantic relations
2. Polysemy
3. Word sense disambiguation
4. Word sense induction
what makes soup, soup?

www.youtube.com/watch?v=Y1HVTNxwt7w&t=22s
What makes soup, soup? (1)

Formal semantics: **extension** — what words denote e.g., soup’: the set of all soups.
What makes soup, soup? (1)

Formal semantics: **extension** — what words denote e.g., soup’: the set of all soups.

if $a$ and $b$ designate the same object, there would be no difference.

*Boris Johnson* $=$ *Prime Minister*

A sign has both a reference and a “sense”
What makes soup, soup? (2)

- Limited domain: mapping to some knowledge base term(s).
- Knowledge base constrains possible meanings. e.g. BabelNet (babelnet.org)

**EN soup**

Liquid food especially of meat or fish or vegetable stock often containing pieces of solid food **WordNet**

- Soup is a primarily liquid food, generally served warm or hot, that is made by combining ingredients such as meat and vegetables with stock, juice, water, or another liquid. **Wikipedia**

- A liquidy food **Wikipedia (disambiguation)**

- Primarily liquid food **Wikidata**

- A cooked, liquid dish (made from meat or vegetables that are mixed with broth in a pot) that is often sold in tins. **OmegaWiki**

- The liquid part of such a dish; the broth. **Wiktionary**
Issues for broad coverage systems

- Boundary between lexical meaning and world knowledge.
- Representing lexical meaning.
- Acquiring representations.
- Polysemy and multiword expressions.
Approaches to lexical meaning

- Formal semantics: **extension** — what words denote

- Semantic primitives
  e.g., *kill* means \( \text{CAUSE}(\text{NOT}(\text{ALIVE})) \)

- Meaning postulates:
  \[
  \forall e_1, x, y [\text{kill}'(e_1, x, y) \rightarrow \exists e_2 [\text{cause}'(e_1, x, e_2) \land \text{die}'(e_2, y)]]
  \]
Approaches to lexical meaning

- **Formal semantics**: extension — what words denote

- **Semantic primitives**
  - e.g., *kill* means \( \text{CAUSE}(\text{NOT}(\text{ALIVE})) \)

- **Meaning postulates**:
  \[
  \forall e_1, x, y[\text{kill}'(e_1, x, y) \rightarrow \exists e_2[\text{cause}'(e_1, x, e_2) \land \text{die}'(e_2, y)]]
  \]

  But . . .
Approaches to lexical meaning

- Formal semantics: extension — what words denote

- Semantic primitives
e.g., kill means \( \text{CAUSE}(\text{NOT}(\text{ALIVE})) \)

- Meaning postulates:
  \( \forall e_1, x, y[\text{kill}'(e_1, x, y) \rightarrow \exists e_2[\text{cause}'(e_1, x, e_2) \land \text{die}'(e_2, y)]] \)
  
  But . . .

- Distributional approaches (information vs knowledge)
Approaches to lexical meaning

• Formal semantics: extension — what words denote

• Semantic primitives
e.g., *kill* means $\text{CAUSE(\neg(\text{ALIVE}))}$

• Meaning postulates:
  $\forall e_1, x, y [\text{kill}'(e_1, x, y) \rightarrow \exists e_2 [\text{cause}'(e_1, x, e_2) \wedge \text{die}'(e_2, y)]]$

But . . .

• Distributional approaches (information vs knowledge)

• Ontological relationships: informal or formal
  this lecture (informal approaches).
Examples to think about

- tomato
- table
- thought
- democracy
- push
- sticky
Lexical Semantics: Semantic Relations
Taxonomic relations

Hyponymy: IS-A

- (a sense of) *dog* is a hyponym of (a sense of) *animal*; *animal* is a *hypernym* of *dog*
- *dog* is more specific and belongs to a subclass of *animal*.
- *entailment* / IS-A: a sense *A* is a hyponym of a sense *B* if everything that is *A* is also *B*, and hence being an *A* entails being a *B*.
- hyponymy relationships form a *taxonomy*
- works best for concrete nouns
Textonomic relations

Hyponymy: IS-A

- (a sense of) *dog* is a hyponym of (a sense of) *animal*; *animal* is a hypernym of *dog*
- *dog* is more specific and belongs to a subclass of *animal*.
- entailment/IS-A: a sense $A$ is a hyponym of a sense $B$ if everything that is $A$ is also $B$, and hence being an $A$ entails being a $B$.
- hyponymy relationships form a taxonomy
- works best for concrete nouns

Some issues concerning hyponymy

- not useful for all words: *thought*, *democracy*, *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
- **Near-synonymy/similarity** e.g., exciting/thrilling
  e.g., slim/slender/thin/skinny

The word *synonym* is commonly used to describe a relationship of approximate or rough synonymy.

- *craft, skill*
- *apple, fruit*
WordNet  

- large-scale, open source resource for English
- wordnets being built for other languages, e.g. Open Multilingual Wordnet (compling.hss.ntu.edu.sg/omw)
- hand-constructed
- organized into *synsets*: synonym sets (near-synonyms)
WordNet

- large-scale, open source resource for English
- wordnets being built for other languages, e.g. Open Multilingual Wordnet (compling.hss.ntu.edu.sg/omw)
- hand-constructed
- organized into synsets: synonym sets (near-synonyms)

Overview of adj red

S: (adj) red, reddish, ruddy, blood-red, carmine, cerise, cherry, cherry-red, crimson, ruby, ruby-red, scarlet (of a color at the end of the color spectrum (next to orange); resembling the color of blood or cherries or tomatoes or rubies)
- similar to
  - S: (adj) chromatic (being or having or characterized by hue)
- derivationally related form
- antonym
  - W: (adj) achromatic [Indirect via chromatic] (having no hue) “neutral colors like black or white”
WordNet labels each synset with a lexicographic category/supersenses.

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
<th>Category</th>
<th>Example</th>
<th>Category</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>service</td>
<td>GROUP</td>
<td>place</td>
<td>PLANT</td>
<td>tree</td>
</tr>
<tr>
<td>ANIMAL</td>
<td>dog</td>
<td>LOCATION</td>
<td>area</td>
<td>POSSESSION</td>
<td>price</td>
</tr>
<tr>
<td>ARTIFACT</td>
<td>car</td>
<td>MOTIVE</td>
<td>reason</td>
<td>PROCESS</td>
<td>process</td>
</tr>
<tr>
<td>ATTRIBUTE</td>
<td>quality</td>
<td>NATURAL EVENT</td>
<td>experience</td>
<td>QUANTITY</td>
<td>amount</td>
</tr>
<tr>
<td>BODY</td>
<td>hair</td>
<td>NATURAL OBJECT</td>
<td>flower</td>
<td>RELATION</td>
<td>portion</td>
</tr>
<tr>
<td>COGNITION</td>
<td>way</td>
<td>OTHER</td>
<td>stuff</td>
<td>SHAPE</td>
<td>square</td>
</tr>
<tr>
<td>COMMUNICATION</td>
<td>review</td>
<td>PERSON</td>
<td>people</td>
<td>STATE</td>
<td>pain</td>
</tr>
<tr>
<td>FEELING</td>
<td>discomfort</td>
<td>PHENOMENON</td>
<td>result</td>
<td>SUBSTANCE</td>
<td>oil</td>
</tr>
<tr>
<td>FOOD</td>
<td>food</td>
<td></td>
<td></td>
<td>TIME</td>
<td>day</td>
</tr>
</tbody>
</table>
Hyponymy in WordNet

search wordnet

Using hyponymy

- Semantic classification: e.g., for named entity recognition. e.g., *JJ Thomson Avenue* is a place.
- RTE style inference: *find*/*discover*
- Word sense disambiguation
- Query expansion in search
Collocation

- two or more words that occur together more often than expected by chance (informal description — there are others)
- some collocations are multi-word expressions
Polysemy
Polysemy

*homonymy*: unrelated word senses.

*bank* (raised land) vs *bank* (financial institution)

*bank* (financial institution) vs *bank* (the building belonging to a financial institution):

(1) The bank is on the corner of Nassau and Witherspoon.

\[\text{BUILDING} \leftrightarrow \text{ORGANIZATION}\]

*bank* (N) (raised land) vs *bank* (V) (to create some raised land): *regular polysemy*. Compare *pile*, *heap* etc

Related but distinct senses

- No clearcut distinctions.
- Dictionaries are not consistent.
Word Sense Disambiguation
Word sense disambiguation

- selecting the correct sense for a word in a context.
- challenges: inventory of potential word senses, datasets
- needed for many applications, problematic for large domains.
- *lexical sample task*: to disambiguate a small pre-selected set of words. simple supervised classification approaches work very well.
- *all-words task*: to disambiguate every word in the text. similar to part-of-speech tagging.
- SemCor: a subset of the Brown Corpus; over 226,036 words; manually tagged with WordNet senses; all-words task

Assumes that we have a standard set of word senses (e.g., WordNet)

<table>
<thead>
<tr>
<th>Sense</th>
<th>Supersense</th>
<th>Target Word in Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>bass⁴</td>
<td>FOOD</td>
<td>...fish as Pacific salmon and striped <em>bass</em> and...</td>
</tr>
<tr>
<td>bass⁷</td>
<td>ARTIFACT</td>
<td>...play <em>bass</em> because he doesn’t have to solo...</td>
</tr>
</tbody>
</table>
Aspects of WSD

- **baseline**: *most frequent sense*
  frequency: e.g., *diet*: the food sense (or senses) is much more frequent than the parliament sense (*Diet of Wurms*)

- **one sense per discourse**: a word appearing multiple times in a discourse often appears with the same sense.

- **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 with neural encoders and classifiers (lecture 6–8).
- sense-tagged corpora are difficult to construct; algorithms need far more data than POS tagging
- contextual embeddings + nearest-neighbor

\[ v_s = \frac{1}{n} \sum_i c_i \]

- Feature-based algorithms for WSD are extremely simple and function almost as well as contextual language model algorithms.
  - part-of-speech tags
  - collocation features of words or \(N\)-grams
  - weighted average of embeddings of all words in a window
WSD techniques

- supervised learning: cf. POS tagging from lecture 3 with neural encoders and classifiers (lecture 6–8).
- sense-tagged corpora are difficult to construct; algorithms need far more data than POS tagging
- contextual embeddings + nearest-neighbor
- Feature-based algorithms for WSD are extremely simple and function almost as well as contextual language model algorithms.
  - part-of-speech tags
  - collocation features of words or \( N \)-grams
  - weighted average of embeddings of all words in a window
- unsupervised learning: learn from \( \{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\} \)
  - machine readable dictionaries (MRDs): e.g., look at overlap with words in definitions and example sentences
  - selectional preferences: don’t work very well by themselves, useful in combination with other techniques
Word Sense Induction
How can we get proper representations?

*word sense induction* automatically create the set of “senses” of each word.

*grammar induction* automatically create the tree of each sentence.

*semantic role induction* automatically create predicate–argument links of words.
Clustering

- For each token \(w_i\) of word \(w\) in a corpus, compute a context vector \(c\).
- Use a clustering algorithm to cluster these word-token context vectors \(c\) into a predefined number of groups or clusters. Each cluster defines a sense of \(w\).
- Compute the vector centroid of each cluster. Each vector centroid \(s_j\) is a sense vector representing that sense of \(w\).
Lexical meaning: what doesn’t work

• meaning of tomato is tomato’ or TOMATO
• meaning postulates
• dictionary definition: good dictionary definition allows reader with some familiarity with a concept to identify it
tomato: mildly acid red or yellow pulpy fruit eaten as a vegetable

Unanswered questions
• how far does distributional semantics on get us?
• grounding often claimed for systems combining vision and language: is this enough?
• are virtual worlds a possible basis for grounding?
Readings

• Ann’s notes