Overview of Natural Language Processing
Part II & ACS L90

Lecture 10: Lexical Semantics

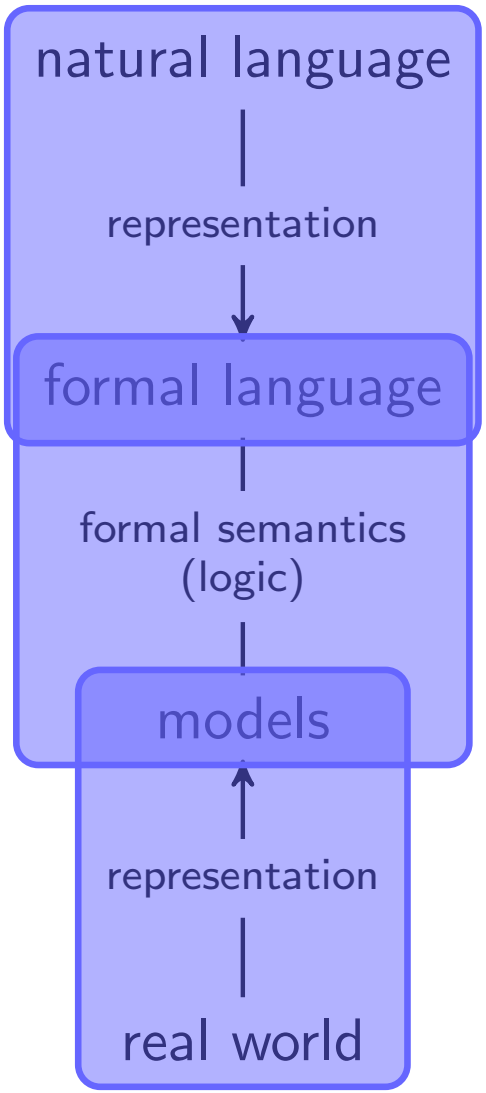
Guy Emerson
Based on slides by Weiwei Sun

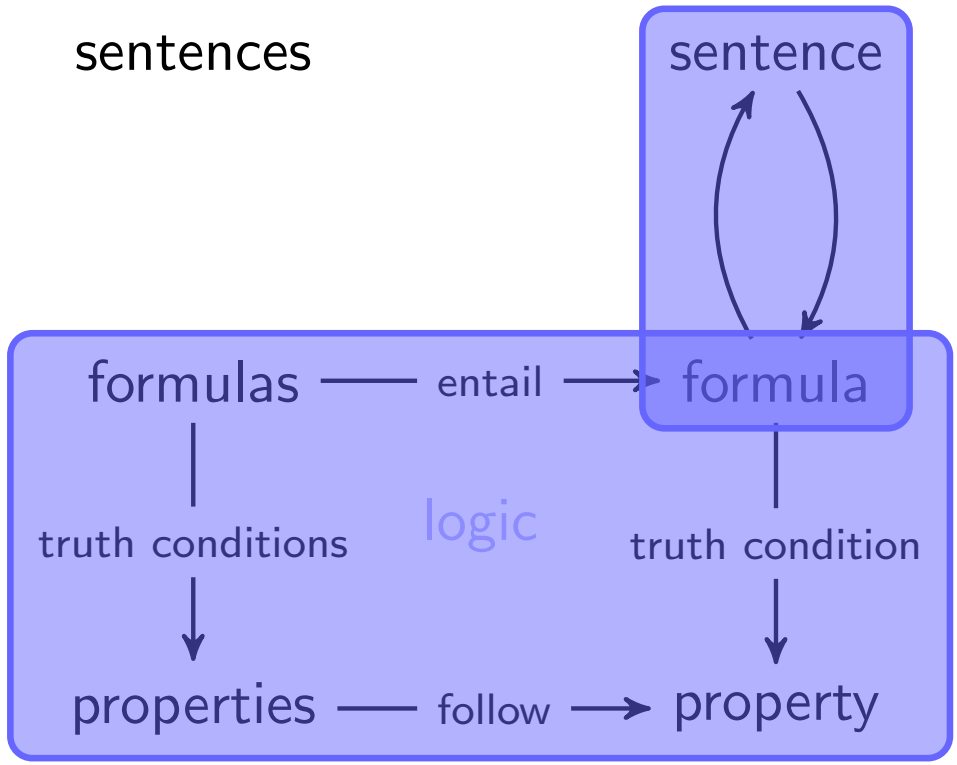
Department of Computer Science and Technology
University of Cambridge

Michaelmas 2021/22

language without meaning is meaningless

Recap: formal meaning representations





facts

fact

from Yanjing Wang

Life is Too Short…
becomingminimalist.com

Life (2017) - IMDb
imdb.com

Life (American TV series) - Wikip…
en.wikipedia.org

9 Ways to Take Responsibility …
thriveglobal.com

Pursue Meaning Instead of Happiness ...
medium.com

How Short Your Life REALLY Is - YouTube
m.youtube.com

Habits to Help You Balance a Creative Life
thriveglobal.com

What Will Life Be Like After the ...
knowledge.insead.edu

to Improve Your Personal Development ...
inc.com

Having a sense of meaning in life is ...
theconversation.com

The 10 Benefits of Knowing Your Purpos…
goalcast.com

search results by google

what is the meaning of life? — life'

Lecture 10: Lexical Semantics

1. Semantic relations
2. Polysemy
3. Word sense disambiguation & induction
4. Grounding & open questions

materials
mostly by
Ann Copestake

what makes soup, soup?

www.youtube.com/watch?v=Y1HVTNxwt7w

What makes soup, soup? (1)

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






search results by google

? 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* **缺少定义**

- 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*

Multilinguality

<table>
<tr><th>English</th><th>German</th><th>Danish</th><th>French</th></tr>
<tr><td>tree</td><td>Baum</td><td rowspan="2">træ</td><td>arbre</td></tr>
<tr><td>wood</td><td>Holz</td><td rowspan="3">skov</td><td rowspan="2">bois</td></tr>
<tr><td>woods</td><td rowspan="2">Wald</td></tr>
<tr><td>forest</td><td>forêt</td></tr>
</table>

Hjelmslev's depiction of the tree/wood/forest semantic field.

- What would a language-independent knowledge base look like? Is it even possible?

Issues for broad coverage systems

- Boundary between lexical meaning and world knowledge.
  - Knowing the word ‘soup’ vs. knowing how to make soup
- Representing lexical meaning.
  - What is the meaning of ‘soup’?
- Acquiring representations.
  - How can the meaning of ‘soup’ be learnt?
- Polysemy and multiword expressions.
  - Different senses of ‘soup’ in different cuisines?
  - *Primordial soup* (mix of chemicals on Earth when life started)
  - *In the soup* (‘in a difficult situation’)

Approaches to lexical meaning

- Formal semantics: extension — what words denote
- Semantic primitives
  e.g., *kill* means $\textsc{cause}(\textsc{not}(\textsc{alive}))$
- Meaning postulates:
  $\forall e_1, x, y[\text{kill'}(e_1, x, y) \rightarrow \exists e_2[\text{cause'}(e_1, x, e_2) \wedge \text{die'}(e_2, y)]]$
  Problematic in the general case. . .
- Distributional semantics (corpus-based)
- Ontological relationships
  This lecture: informal approaches

Examples to think about

- tomato
- table
- thought
- democracy
- push
- sticky

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*

Some issues concerning hyponymy

- not useful for all words: *thought*, *democracy*, *push*, *sticky*?
  (works best for concrete nouns)
- individuation differences: is *table* a hyponym of *furniture*?
- multiple inheritance: is *tomato* a hyponym of both *fruit* and *vegetable*?
- what does the top of the hierarchy look like?

Other semantic relations

Classical relations

- **Synonymy** e.g., *aubergine*/*eggplant*.
- **Antonymy** e.g., *big*/*little*
- **Near-synonymy/similarity** e.g., *exciting*/*thrilling*
  e.g., *slim*/*slender*/*thin*/*skinny*
- **Meronomy**: PART-OF e.g., *arm* is a meronym of *body*,
  steering wheel is a meronym of *car*
  (is *stock* a meronym of *soup*?)
  (is *noodle* a meronym of *soup*?)

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

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

WordNet wordnet.princeton.edu

- 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 wordnet.princeton.edu

WordNet labels each synset with a lexicographic category/*supersenses*.

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

Hyponymy in WordNet

search WordNet

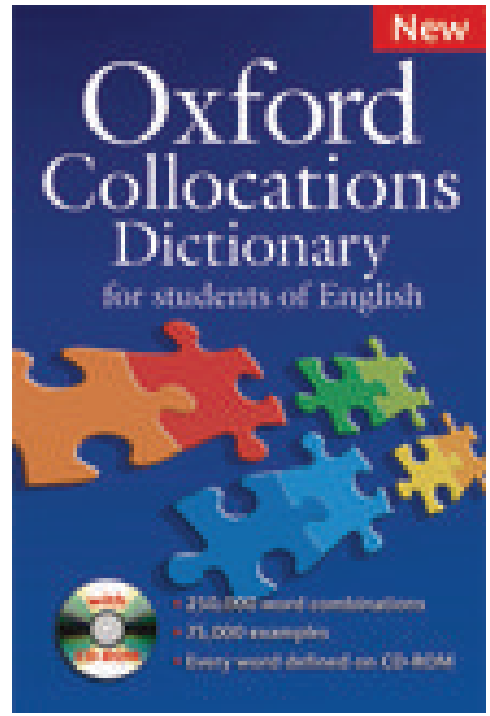
`http://wordnetweb.princeton.edu/perl/webwn`
`https://www.nltk.org/howto/wordnet.html`

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
- some collocations are multi-word expressions



Homonymy and Polysemy

Homonymy and Polysemy

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

(1) They pulled the canoe up on the bank.

(2) They cashed a cheque at the bank.

polysemy: related word senses.
bank (financial institution) vs *bank* (building of a financial institution)

(3) The bank granted the loan.

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

Related but distinct senses

- No clear-cut distinctions.
- Dictionaries are not consistent.

Regular Polysemy

regular polysemy: sense distinction shared by a group of words

- building $\leftrightarrow$ organisation
  - bank (*the bank on the corner...*, *the bank granted...*)
  - school
  - library
  - theatre
  - restaurant
  - ...
- physical configuration (N) $\leftrightarrow$ create a physical configuration (V)
  - bank (*a sand bank*, *to bank sand*)
  - heap
  - pile
  - can (*a can of fish*, *to can fish*)
  - box
  - ...

Word Sense Disambiguation

Word sense disambiguation

- selecting the correct sense for a word in a context.
- challenges: inventory of potential word senses, datasets.
- challenges: application in downstream tasks, scaling to 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)

Sense	**Supersense**	**Target Word in Context**
bass^4	FOOD	. . . fish as Pacific salmon and striped *bass* and. . .
bass^7	ARTIFACT	. . . play *bass* because he doesn't have to solo. . .

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 also called 'co-occurrence').
  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).
- 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
- selectional preferences: don't work very well by themselves, useful in combination with other techniques
- *distant supervision* using machine readable dictionaries: look at overlap with words in definitions and example sentences

sense-tagged corpora are difficult to construct;
algorithms need far more data than POS tagging

Word Sense Induction

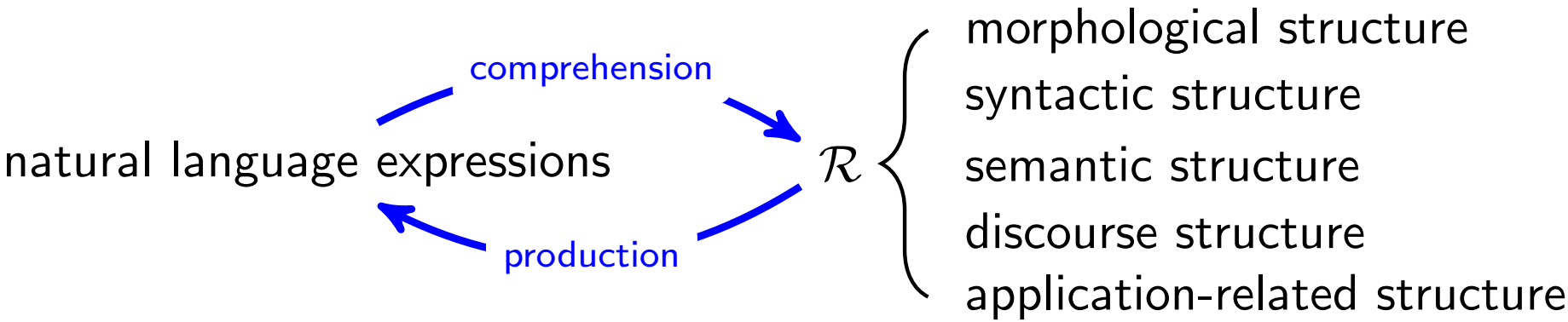
Word Sense Induction

- Word Sense Disambiguation: supervised, pre-defined senses
- Word Sense Induction: unsupervised, no pre-defined senses

Clustering

- For each token w_i of word w in a corpus, compute a context vector c_i.
- Use a clustering algorithm (e.g. k-means) to group these context vectors c_i into a number of clusters. Each cluster defines a sense of w.
- Compute the centroid (vector average) of each cluster. Each centroid s_j is a vector representing that sense of w.

The Bigger Picture



How can we get proper representations?

word sense induction automatically create a sense for each word token.

grammar induction automatically create a tree for each sentence.

semantic role induction automatically create predicate–argument links between word tokens.

Grounding and Open Questions

The Symbol Grounding Problem

Aspects of lexical semantics that we've covered:

- semantic relations (hyponymy, antonymy, etc.)
- word senses, polysemy
- distributional semantics (including language models)
- input for compositional semantics

Problem: defining symbols in terms of symbols

Beyond treating words as symbols:

- meaning isn't (just) about symbols: humans need to recognize and manipulate things in the world.
- *grounding*: relate symbols to the real world
  (often associated with Harnad, but other authors too)
  (recent work often refers to 'multimodality', which glosses over the difference between sensory modalities and language)

Unanswered Questions

- How can we develop a compositional grounded semantics which captures relations and senses?
- How far can we get with distributional semantics?
- How much grounding / what kind of grounding is necessary?
- How can machine learning models represent, learn, and manipulate semantic structure?

- *What makes soup soup?*
  `www.youtube.com/watch?v=Y1HVTNxwt7w`

Readings

- Ann's notes
- D Jurafsky and J Martin. *Speech and Language Processing* Chapter 19. `web.stanford.edu/~jurafsky/slp3/19.pdf`