

Word Senses

- ▶ Polysemy: many meanings
- ▶ The book uses “aspect” in these senses

| | |
|--------------------|-------------------------------------|
| Informal | One aspect of the meaning of a word |
| Sentiment analysis | Aspect of an entity for sentiment |
| Formal linguistics | Aspect of a verb |

- ▶ Word sense: a discrete representation of one meaning
 - ▶ Notationally, just a new term: bank¹ versus bank²
 - ▶ How can we express the semantics or content of a word sense?
- ▶ Gloss: dictionary meaning of a sense
 - ▶ Informal
 - ▶ Notoriously circular as a set, sometimes even individually
 - ▶ Red \Rightarrow blood; blood \Rightarrow red
 - ▶ Often accompanied by precedents, sentences indicating usage
 - ▶ Can be mined for understanding

Splitting Senses

Delta serves breakfast

Delta serves Atlanta

Delta serves Atlanta and breakfast

- ▶ Zeugma: conjunction of *antagonistic* readings
 - ▶ Sound anomalous since conjunction forces alignment
- ▶ Conjoin two readings
 - ▶ Anomaly (zeugma) is evidence for the senses being distinct
- ▶ Syntactic variation
 - ▶ Noun versus verb: black mark or mark time
 - ▶ Within syntactic category: serve food or serve as editor
- ▶ Dictionaries may split senses to too fine a grain
 - ▶ Clustering similar senses can be useful for NLP

Word Embeddings

- ▶ Basic embeddings, as in word2vec
 - ▶ Disregard context
 - ▶ Don't separate out senses but provide a single vector that aggregates all the occurrence contexts of a word
 - ▶ Interestingly, place a word and its antonyms close together
- ▶ Contextual embeddings, e.g., ELMo and BERT, are superior
 - ▶ They too don't separate out senses discretely

Relations Between Senses

- ▶ Homonymy: Unrelated senses of a word
- ▶ Polysemy (early in the chapter): Unrelated senses, synonym of homonymy
- ▶ Polysemy (later in the chapter): Related senses
- ▶ Synonymy: Two senses of different words are nearly identical
 - ▶ Specific to senses: big \approx large but big sister \neq large sister
- ▶ Antonymy: opposite with respect to some scale or axis
 - ▶ Differ on that axis
 - ▶ Highly similar otherwise
 - ▶ Confound word embeddings
- ▶ Hypernymy (antonym of hyponymy)
 - ▶ Superclass (sometimes called superordinate)
- ▶ Meronymy: part-whole
 - ▶ Leg as a meronym of chair
 - ▶ Chair as a holonym of leg

Structured Polysemy

Relations between senses of the same word

- ▶ Metonymy: using one aspect of an entity to refer to other aspects of (or to the entire) entity
 - ▶ These are captured in the same word
 - ▶ Organization \approx Organization \rightsquigarrow Address \rightsquigarrow Component
 - ▶ Downing Street (\approx UK Prime Minister's office) is making plans for Brexit
 - ▶ Work \approx Author (read Shakespeare)
- ▶ Synecdoche: subcategory of metonymy broadly
 - ▶ Part for a whole
 - ▶ Whole for a part
- ▶ These terms do not have set meanings
 - ▶ Sometimes defined not as subclasses

WordNet: Lexical Relations

Called a sense inventory

- ▶ Focuses on nouns (~118k), verbs (~11k), adjectives(~22k), adverbs (~4k)
- ▶ Provides a lemma entry for each included word, e.g., for “view”
 - ▶ Senses: nouns (9) and verbs (3)—ordered by decreasing popularity
 - ▶ Glosses
 - ▶ Examples
- ▶ Synset: near-synonyms of a WordNet sense
 - ▶ {view², aspect³, prospect⁴, scene³, vista¹, panorama¹}
 - ▶ Each member points to all others
 - ▶ Each member has the same synset gloss
- ▶ Synsets induce an equivalence relation: synsets are disjoint or equal

Supersenses: High-Level Conceptual Categories

Each synset identifies one supersense or *lexname*

Supersenses for nouns:

| Category | Example | Category | Example | Category | Example |
|-----------------|----------------|-----------------|----------------|-----------------|----------------|
| ACT | service | GROUP | place | PLANT | tree |
| ANIMAL | dog | POSSESSION | price | LOCATION | area |
| 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 |
| PERSON | people | COMMUNICATION | review | STATE | pain |
| FEELING | discomfort | PHENOMENON | result | TIME | day |
| FOOD | food | SUBSTANCE | oil | | |

Additionally, 15 for verbs, 2 for adjectives, 1 for adverbs

Sense Relations in WordNet

Noun relations:

| Relation | Definition | Example |
|-------------------|------------------------------------|--|
| Hypernym | From concepts to superordinates | breakfast ¹ → meal ¹ |
| Hyponym | From concepts to subtypes | meal ¹ → lunch ¹ |
| Instance Hypernym | From instances to their concepts | Austen ¹ → author ¹ |
| Instance Hyponym | From concepts to their instances | composer ¹ → Bach ¹ |
| Part Meronym | From wholes to parts | table ² → leg ³ |
| Part Holonym | From parts to wholes | course ⁷ → meal ¹ |
| Antonym | Semantic opposition between lemmas | leader ¹ ↔ follower ¹ |
| Derivation | Lemmas w/same morphological root | destruction ¹ ↔ destroy ¹ |

Verb relations:

| Relation | Definition | Example |
|----------|---|---|
| Hypernym | From events to superordinate events | fly ⁹ → travel ⁵ |
| Troponym | From events to subordinate event | walk ¹ → stroll ¹ |
| Entails | From verbs (events) to the verbs (events) they entail | snore ¹ → sleep ¹ |
| Antonym | Semantic opposition between lemmas | increase ¹ ↔ decrease ¹ |

WSD: Word Sense Disambiguation

- ▶ Lexical sample task: map
 - ▶ Small, fixed set of target words
 - ▶ Senses for each word from a lexicon
 - ▶ Supervised classification works well
- ▶ Semantic concordance = Corpus, each word labeled with its sense
 - ▶ SemCor \subseteq Brown Corpus
 - ▶ 226k words, manually tagged using WordNet
 - ▶ Example with POS as subscript and sense as superscript

You will find_v⁹ that avocado_n¹ is_v¹ unlike_j¹ any other_j¹ fruit_n¹
you have ever_r¹ tasted_v²

- ▶ All-words task
 - ▶ Entire lexicon of words and senses
 - ▶ Data sparseness
- ▶ Choose the correct WordNet sense

Evaluation of WSD Approaches

- ▶ F_1 score on held-out corpus
- ▶ Effective baseline: Most frequent sense in WordNet
 - ▶ Also a good default
- ▶ One sense per discourse
 - ▶ A word tends to retain its sense, especially among unrelated senses (homonyms)
 - ▶ Not an effective baseline
 - ▶ Useful heuristic for disambiguation

Contextual Word Embeddings

Assumes a contextual embedding technique, such as BERT or ELMo

- ▶ Embedding of sense (synset): mean of embeddings of the words in it
 - ▶ c_i is labeled with sense s
 - ▶ There are n occurrences of words labeled s in the corpus

$$v_s = \frac{1}{n} \sum_i^n c_i$$

- ▶ Example counts: [view: 100; prospect: 111; panorama: 1,000]
 - ▶ Precompute sense embeddings
- ▶ At test time, compute the contextual embedding of a word
 - ▶ Find the nearest sense embedding of that word (same lemma)
- ▶ For words with unknown sense embeddings
 - ▶ Use most frequent sense in WordNet as default
 - ▶ Works because SemCor is a small subset of WordNet
 - ▶ Won't work for words not present in WordNet

Estimating Missing Sense Embeddings using WordNet

Loureiro and Jorge's alternative to most frequent sense

- ▶ Apply WordNet relations, in increasing order of abstraction
- ▶ Estimate each level based on estimates of lower levels
- ▶ For a given word, use the mean sense embedding of the first abstraction level that has data for that word
 - ▶ Synset: mean of other synset members with known embeddings
 - ▶ Embedding: mean of known embeddings of words in it
 - ▶ If the synset's embedding is known, use it; skip the rest
 - ▶ Else, Hypernyms: mean of hypernyms with known embeddings
 - ▶ Embedding: mean of known embeddings of synsets below it
 - ▶ If the hypernym's embedding is known, use it; skip the rest
 - ▶ Else, Lexnames: if some lexname (supersense) have known embeddings, take their mean
 - ▶ Embedding: mean of known embeddings of synsets in it
 - ▶ If the lexname's embedding is known, use it

Other Sources for Word Sense Information

- ▶ Wikipedia
 - ▶ Use the URI of a page as a sense
- ▶ Thesauruses, especially for handling antonyms
 - ▶ Modify embedding technique to use antonym relations
 - ▶ Retrofitting or counterfitting
 - ▶ Train static embeddings as usual
 - ▶ Modify those embeddings to bring synonyms closer and take antonyms farther

Word Sense Induction

Unsupervised learning

- ▶ Compute a context embedding for each occurrence of a word w
- ▶ Cluster these embeddings
 - ▶ Predefined number of clusters
- ▶ Each cluster expresses a sense of w
- ▶ The centroid of a cluster is a sense embedding
- ▶ Upon receiving the word (in some new context)
 - ▶ Compute its context embedding
 - ▶ Assign the word to the closest sense