An Investigation of
Sense Ordering across Dictionaries
with Respect to
Lexical Semantic Relationships

Robert Krovetz

Lexical Research

Hillsborough, NJ

Motivation: Lexicography

- Different notions of how senses can be ordered: historical, frequency, and logical.

  – Logical order might mean *literal/figurative, concrete/abstract,* or *general/specific.*

  – To what extent do dictionaries differ in their ordering, and in the senses that are distinguished, for different lexical semantic relationships? [Homonymy, Metaphor, Systematic Polysemy]

Motivation: Computational Linguistics

- WordNet is the most widely used lexical resource for WSD.

- The Most-Frequent-Sense (MFS) heuristic is widely used to improve accuracy. It is a fall-back if we do not have enough evidence to make a different decision.

- The ordering in WordNet is partially based on SemCor, a tagged subset of the Brown corpus. It only has about 200K words.

- Can we leverage the ordering in corpus-based dictionaries to see when the senses are not in frequency-order?

Motivation: Cognition

- Follow up on Lehrer and Panman's observation: when senses are unrelated, people agree they are different, but when senses are related, people will disagree about whether they are different.

- To what extent is this true for lexicographer's judgments?

Issues with Sense Ordering

- What is the basis for frequency-ordering? To what extent do corpora differ in this, and to what extent does it differ with regard to lexical semantic relationships?

- Figurative senses can be older than literal senses, and they can also be more frequent.

- The most salient sense is not always the most frequent. For example, *keep* most frequently means to maintain: *keep still/keep warm/keep quiet.*

Methodology

- Sense Data: a Sample of 75 ambiguous words. Two senses were used for each.

- The set was divided according to lexical semantic relationships: homonymy, metaphor, and systematic polysemy (*Language/People, Substance/Color, Animal/Food, Animal/Hide, Tree/Wood,* and *Music/Dance*). The first two sets had words within as well as across part-of-speech. All of the third set were nouns.

Methodology

- Eight dictionaries were examined to assess the ordering of the 150 word senses.

- I compared historical order (based on *Merriam-Webster's 7th Collegiate Dictionary)* against the order in the other dictionaries.

- Focus on differences with learner's dictionaries (Oxford, Cambridge, Longman, and COBUILD). Where do the dictionaries agree and where do they differ?

Sample Sense Inventory

WORD	LABEL	Part-of-Speech	Semantic Class
ABSTRACT	type-of-painting	Noun	Homonymy
ABSTRACT	draft-of-paper	Noun	Homonymy
BABE	baby	Noun	Homonymy
BABE	woman	Noun	Homonymy
DRIVER	car	Noun	Homonymy
DRIVER	golf-club	Noun	Homonymy
INFLECT	word	Verb	Homonymy
INFLECT	voice	Verb	Homonymy

Sample Sense Inventory

WORD	LABEL	Part-of-Speech	Semantic Class
TRAIN	Locomotive	Noun	Homonymy
TRAIN	Educate	Verb	Homonymy
BOOM	Grow-rapidly	Verb	Homonymy
BOOM	Make-Noise	Verb	Homonymy
SANDWICH	Peanut-butter	Noun	Metaphor
SANDWICH	Find-time-for	Verb	Metaphor
CONCEIVE	Pregnancy	Verb	Metaphor
CONCEIVE	Imagine	Verb	Metaphor

Sample Sense Inventory

WORD	LABEL	Part-of-Speech	Semantic Class
GOLD	NaturalKind/Color	Noun	Metonymy
JADE	NaturalKind/Color	Noun	Metonymy
CHICKEN	Animal/Food	Noun	Metonymy
TUNA	Animal/Food	Noun	Metonymy
ALLIGATOR	Animal/Hide	Noun	Metonymy
MINK	Animal/Hide	Noun	Metonymy
OAK	Tree/Wood	Noun	Metonymy
MAPLE	Tree/Wood	Noun	Metonymy

Methodology

- Where there is a disagreement about the order of senses between WordNet and other dictionaries, how easy is it to identify ngrams that will support the frequency ordering?

- The frequency of the ngrams in *Project Gutenberg,* the *Wikipedia,* and the Internet were used as a proxy for the order of the word senses and whether the ordering is corpus-specific.

Resources: Corpora and N-grams

- **Project Gutenberg**: more than 30,000 books downloaded from gutenberg.org. It contains 2.35 billion words.

- **Wikipedia**: a download from 2021 that contains 3.57 billion words.

- **Internet**: the Google n-grams dataset. It contains the frequency of unigrams to 5-grams that occur 40 times or more in 1-Trillion words from the Internet.

Sample N-grams used for Assessing Frequency

- Draft (*draft a bill/I was drafted*)
  Entitle (*entitled to vote/entitle a book*)
  Mortar (*mortar shells/mortar and pestle*)

- Defuse (*defuse the bomb/defuse tension*)
  Conceive (*children conceived/ill-conceived*)

Results: Mapping Senses between Dictionaries

- It was generally fairly easy to determine which senses were distinguished, and in which order.

- Problems:
  - Morphology: *inflect, inflection, inflectional*
                    plastering (with-plaster/defeat)
  - Part-of-Speech: noun/adjective
  - Semantics: *sandwich*

Results on Sense Ordering

- Differences between part-of-speech were almost always ordered the same way. The only exception was *novel* in COBUILD and Collins.

- For the Homonymy dataset, there was a difference in 10 out of 20 words compared with historical order for at least one dictionary.

- For the Metaphor dataset, there was a difference in 8 out of 20 words compared with historical order for at least one dictionary..

Results on Sense Ordering

- There was one word in the Homonymy dataset that was listed with only one sense in MW7: *john,* and one word in the Metaphor dataset was not found at all: *defuse.*

- The OALD, CIDE, LDOCE, and COBUILD dictionaries differed in order for 5 out of the 20 words in the Homonymy dataset, and 4 out of 20 in the Metaphor dataset.

Results on Sense Ordering

- For the Metaphor dataset, frequency order and logical order were usually the same, so dictionaries were usually in agreement. The primary exception was WordNet.

- WordNet also differed from the other dictionaries in that it distinguished literal and figurative senses least often (6 out of 25 words were only defined in a literal sense).

Results on Sense Ordering

Word	Label	Dictionaries	Corpus Results
bask	sun:approval	WN	sun
conceive	baby:imagine	COLLINS,WN,COBUILD	imagine
defuse	bomb:situation	(LDOCE,OALD,COLLINS)/ COBUILD	situation
postmortem	death:final-analysis	WN	death
purgatory	hell:bad-place	WN	hell
shrimp	crustacean:person	WN	crustacean
underline	writing:emphasis	COBUILD, WN	CORPUS-SPECIFIC
uproot	plants:from-home	WN, COBUILD	CORPUS-SPECIFIC

WordNet and Metaphor

- Examined an additional 20 words with a literal/figurative distinction. Of these, 1 word was not found in WordNet, and 6 were defined in a literal, but not figurative sense, or vice-versa.

- I used COBUILD and ngrams to support a re-ordering of two senses.

Issues with N-gram Analysis

- N-grams often followed the *one sense per collocation* hypothesis, but not always. Consider *uprooted by Katrina.* Compare *has nosedived* and *plane/economy nosedived.*

- A few ngrams sometimes dominated the frequency distribution. For example, *urban/suburban blight.*

- Corpus bias sometimes influenced the relative order of senses. Out of 18 disagreements investigated, 5 were corpus-specific in the ordering.

Results on Sense Ordering

- For the Systematic Polysemy dataset, there were differences within dictionaries and between dictionaries on how the senses were ordered.

- There were also differences in how senses were distinguished. Sometimes there were different numbered senses, sometimes a related number is given (e.g., 1 vs. 1a), and sometimes a parenthetical is used (e.g., "(the music for) this dance").

- There was a range between dictionaries about how many words had different senses.   The Cambridge dictionary distinguished 6 out of 25, and Webster's New World distinguished all 25.

Results on Word Sense Individuation

- (Panman 1982) noted that when senses are homonymous, people agree that the senses are different, but when they are have related meanings people disagree about whether the senses are distinct.

- The results show that this is also reflected in lexicographic judgment.   There was an ordering between homonymy, metaphor, and systematic polysemy about whether senses were individuated.

Similarity between Dictionaries

- Almost all words in the Homonymy dataset were distinguished by most dictionaries (avg. 7.6 out of 8). Only 6 out of 25 words had a different ordering between learner's dictionaries.

- Metaphor was distinguished by fewer dictionaries (avg. 7.1 out of 8), but this is still high.   Only 4 out of 25 words had a different ordering between learner's dictionaries.

- There was an ordering between classes in the Systematic Polysemy dataset. Substance/color was distinguished most often, and Music/Dance least often.

Future Work

- Identify additional cases in WordNet where the senses are incorrectly ordered by frequency.

- Look at other measures for establishing a metric for word-sense granularity:
  - Crowdsourcing
  - How often the word senses co-occur in a discourse

Conclusion

- The ordering of senses is not an easy decision.   There can be conflicts between frequency and salience, and between frequency and logical order.

- WordNet stands out among the dictionaries that were studied with regard to Metaphor. It distinguished literal from figurative senses least often. It also ordered the senses incorrectly (according to frequency) most often.

- I was able to leverage a corpus-based learner's dictionary, COBUILD, to re-order WordNet senses, and I supported this with an ngram analysis in large corpora.