Robert Krovetz

AN INVESTIGATION OF SENSE ORDERING ACROSS DICTIONARIES WITH RESPECT TO LEXICAL SEMANTIC RELATIONSHIPS

Abstract   This paper discusses an investigation of how senses are ordered across eight dictionaries. A dataset of 75 words was used for this purpose, and two senses were examined for each word. The words are divided into three groups of 25 words each according to the relationship between the senses: Homonymy, Metaphor, and Systematic Polysemy. The primary finding is that WordNet differs from the other dictionaries in terms of Metaphor. The order of the senses was more often figurative/literal, and it had the highest percentage of figurative senses that were not found. We discuss leveraging another dictionary, COBUILD, to re-order the senses according to frequency.

Keywords   Lexical semantics; word senses; corpus analysis

1.   Introduction

The order of senses in a dictionary is an important problem in lexicography (Hiorth 1954; Kipfer 1983; Lew 2013). The literature usually discusses three orderings: 1) historical, 2) frequency, and 3) logical. The first order is used by the *Oxford English Dictionary* and the *Merriam-Webster* dictionaries.[1] The second order is used by dictionaries for learners of English as a second language, such as the *Longman Dictionary of Contemporary English*, the *Cambridge International Dictionary of English*, the *Oxford Advanced Learner's Dictionary*, and *COBUILD*. There are differences about what is meant by the logical order, but literal/figurative, concrete/abstract, and general/specific are some of the distinctions that are mentioned. All dictionaries use a logical ordering to some extent.

There are several reasons why it is important to look at sense ordering. From the perspective of the user, we generally want the most frequent sense to be listed first. From the perspective of Computational Linguistics, we want to know sense order because it is important for word sense disambiguation. Many systems rely on the Most Frequent Sense (MFS) heuristic for classifying a word's sense in context (Agirre/Edmonds 2007). The skewed nature of word sense distributions means that the most frequent sense is not only more frequent, but often much more frequent than a secondary sense. From the perspective of lexicography, we want to get a better understanding of lexicographic judgment and how it relates to cognition. The frequency order in a corpus-based dictionary is not hard-and-fast. Learner's dictionaries will order senses in a way that is best for the user. There are also differences between historical order compared with logical order. For example, the figurative meaning for a word is sometimes older than the literal meaning.[2]

This paper discusses an investigation of how senses are ordered across eight dictionaries: *Longman Dictionary of Contemporary English* (LDOCE), *Collins English Dictionary*

---

[1]   We note that the ordering is changing for the Merriam-Webster dictionaries so that the order would be what is most useful for the user.

[2]   http://www.merriam-webster.com/words-at-play/6-words-whose-abstract-meanings-came-first/ engine.

(COLLINS), *WordNet* (WN), *Cambridge International Dictionary of English* (CIDE), *Webster's New World Dictionary* (WNW), *Oxford Advanced Learner's Dictionary* (OALD), *Collins COBUILD English Language Dictionary* (COBUILD) and *Merriam-Webster's 7th Collegiate Dictionary* (MW7). There were 75 words in the dataset, and two senses were examined for each word. There were 25 words for which the two senses were homonymous (e. g., *driver/car* and *driver/golf-club*, *draft/paper* and *draft/army*, *train/educate* and *train/locomotive*). Similarly, there were 25 words for which the relationship was a literal/metaphor difference (e. g., *shrimp/crustacean*, *shrimp/person*). There were instances within as well as across part-of-speech for these two subsets. For example, *train* was in the Homonymy dataset, and *parrot* was in the Metaphor dataset. Finally, there were 25 words that exhibited systematic polysemy between the senses. They were divided into different classes: animal/food, music/dance, language/people, tree/wood-of-tree, animal/hide, and natural-kind/color. All of these words were nouns. The dataset was created by the author for a variety of purposes, and it is a subset of a larger sense inventory. For words that differed in part-of-speech, we looked at the order of homographs.

The study had the following aims:

1) Compare four different learner's dictionaries to see where they agree, and where they differ, in the ordering of the senses in the three datasets. We used the *Longman Dictionary of Contemporary English* (LDOCE), the *Cambridge International Dictionary of English* (CIDE), the *Oxford Advanced Learner's Dictionary* (OALD), and the *Collins COBUILD English Language Dictionary* (COBUILD).

2) Compare WordNet[3] against the other dictionaries in the same regard. Because of the importance of this dictionary as the basis for the Most Frequent Sense (MFS) heuristic for word-sense disambiguation, we wanted to see how easy it is to leverage the other dictionaries to re-order the senses in WordNet when senses are in the wrong order.

3) Compare a Merriam-Webster dictionary (MW7) against the other dictionaries to see how historical order differs.

4) Examine cases where there is disagreement and assess which senses are most frequent. This was done by using ngrams that occur in different corpora: Project Gutenberg (gutenberg.org), the Wikipedia (en.wikipedia.org), and the Internet, as represented by the Google n-grams dataset (Brants/Franz 2006). A sample of the ngrams is given in Table 1. These ngrams contained the target word (or an inflected form), and they were manually selected. Multiple ngrams were reviewed for each word in order to make the assessment. These ngrams were also used in assessing the order of senses in WordNet. The use of ngrams for this purpose is based on the *One Sense per Collocation* hypothesis (Yarowsky 1990).

The next section will discuss the results of the comparison, and a small experiment to re-order the senses in WordNet.

Word	Semantic Relation	Ngram pair
boom	homonymy	voices boomed/economy boomed
discount	homonymy	discount the price/discount the idea
draft	homonymy	draft a bill/I was drafted
entitle	homonymy	entitled to vote/entitle an act

---

3   We used version 3.0

Word	Semantic Relation	Ngram pair
frisk	homonymy	frisk and play/stop and frisk
gag	homonymy	running gags/mouth gag
john	homonymy	prostitutes and johns/use the john
mortar	homonymy	mortar shells/mortar and pestle
stigma	homonymy	social stigma/stigma and pollen
bask	metaphor	bask in the sun/bask in the glory
defuse	metaphor	defuse the bomb/defuse tension
postmortem	metaphor	postmortem exam/postmortem analysis
conceive	metaphor	children conceived/ill-conceived
purgatory	metaphor	souls in purgatory/kind of purgatory
shrimp	metaphor	brine shrimp/bully and the shrimp
underline	metaphor	words underlined/underline the importance
uproot	metaphor	uproot the trees/uproot themselves

Table 1: A sample of the ngrams used to assess sense ordering

2. Results

Table 2 gives the results of the comparison, and which sense was found to be most frequent according to the ngram analysis.

The literature mentions difficulty in making comparisons between the senses in different dictionaries (Atkins/Levin 1991), (Kilgarriff 1997). In contrast, we found it was fairly straightforward to identify the senses in the dataset and how they were ordered in the different dictionaries. This is because we are looking for specific senses rather than a general mapping.

The main problematic dictionary was CIDE. It did not enumerate the senses like the other dictionaries, but rather provided bullets, which were also used for example sentences. This made it difficult to determine when a sense was being distinguished. The main sense mapping problem was with the word *sandwich*. We were looking for a figurative sense involving time (e. g., *I can sandwich you in between 2 and 3 PM*). Most of the dictionaries defined the figurative sense only in terms of space.

Part of speech was sometimes a problem. Dictionaries differed in whether a word sense was attested as a noun or as an adjective (e. g., *turquoise*). We allowed such differences in matching the sense we were looking for.

Morphological variation was a factor, and sometimes the sense was found only under a variant form (e. g., *inflect* vs. *inflection*). In addition, sometimes the sense associated with a word form was not found, such as *plastered*, which can either mean "apply plaster" or "drunk".

The results for the different lexical semantic classes are given next. This is followed by a discussion of WordNet and a small experiment at re-ordering the senses.

Word	Label	Dictionaries	Corpus Results
boom	sound:grow-rapidly	COBUILD, CIDE	CORPUS-SPECIFIC
discount	bargain:opinion	WN, OALD, COLLINS, LDOCE	opinion
draft	army:paper	all except WNW	paper
entitle	book:permitted	COLLINS, COBUILD, WN, CIDE	permitted
frisk	playful:search	COBUILD, OALD	CORPUS-SPECIFIC
gag	mouth:joke	WN	CORPUS-SPECIFIC
inflect	word:voice	COBUILD, WNW	word
john	prostitute-client:toilet	(LDOCE, CIDE)/(WN, WNW)	toilet
mortar	pestle:gun	all except WNW	gun
stigma	shame:plant	WN	shame
bask	sun:approval	WN	sun
conceive	baby:imagine	COLLINS, WN, COBUILD, CIDE	imagine
defuse	bomb:situation	(LDOCE, OALD, COLLINS, WNW)/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

Table 2: Words where there was disagreement about the order of the senses compared with historical order. The table shows the sense labels in historical order (where available), the dictionaries where the order was different, and the sense that was most frequent according to corpus analysis. The first part of the table illustrates words that are in the Homonymy dataset, and the second part shows words in the Metaphor dataset

2.1 Homonymy and Metaphor Datasets

There were 5 words each in the Homonymy and Metaphor datasets that differ in part-of-speech between the senses. These were usually represented as different homographs in the dictionaries and almost all were ordered the same way. The only exception was *novel*, which was ordered adjective first except for COBUILD and Collins.

Of the 20 remaining words in the Homonymy dataset, there were differences in ordering for 10 words.[4] For the Metaphor dataset, 8 of the words differed in sense order. The four learner's dictionaries (OALD, CIDE, LDOCE, COBUILD) differed in their ordering for 5 out of 20 words in the Homonymy dataset, and for 4 out of 20 in the Metaphor dataset. These words are given in Table 2.

---

[4] There was one word, *abstract*, that we did not include due to problems with part of speech, morphological variation, and identifying a good set of n-grams for analysis. There were also two words in the Metaphor dataset that were not included: *uplift*, because of difficulty in judging which sense was most frequent, and *digest*, because the historical order differed from the order in all the other dictionaries; *digest an idea* is an older usage than *digest a meal*.

There was one word, *john*, which was only listed in one sense (toilet) in MW7, so the disagreement is given for the other dictionaries. Except for *john* and *defuse*, the order of senses in the Label column is the historical order.[5]

2.2 Systematic Polysemy Dataset

For the Systematic Polysemy dataset there was inconsistency within as well as between dictionaries. Words in each of the subsets (animal/food, music/dance, language/people, tree/wood-of-tree, animal/hide, natural-kind/color) were sometimes in that order, and sometimes not, depending on the dictionary and the word.

It is not surprising that the order differs for this set. Nor do we feel that the order is necessarily important for the users of the dictionary. It is more important in Computational Linguistics, where the distinctions are needed for natural language understanding. Our main interest in this set is from a cognitive perspective. (Panman 1982) observed that when word senses are homonymous, people will agree that they are different. But when the senses are related, people will disagree about whether the senses are distinct. We wanted to look at this question from the perspective of lexicographic judgment. How often are senses in this set distinguished compared with the other datasets? That is, to what extent does (Panman 1982)'s comment apply to lexicographers, especially since they are considered to be splitters rather than lumpers (Bejoint 1988).

The Homonymy class was indeed individuated more often than the Systematic Polysemy class. However, there were significant differences between dictionaries and between words within a class:

1) The Cambridge Dictionary did not include the Language/People class, and distinguished the senses of only 6 out of 25 words that were systematically polysemous. Webster's New World Dictionary distinguished all 25. The rest of the dictionaries distinguished 18 or more.

2) The Natural-Kind/Color class (*gold*, *silver*, *jade*, *rust*, *turquoise*) was distinguished most often, and the Music/Dance class (*waltz*, *tango*, *foxtrot*, *rumba*, *polka*) was distinguished least often amongst the Systematic Polysemy classes. The Tree/Wood-of-tree group was distinguished most often for *oak* and *pine*, and less often for *maple* and *chestnut*. This is in accord with (Hanks 1979) and many others, who have noted that dictionaries contain senses that are stereotypical of usage.

3) The Metaphor class was in between homonymy and systematic polysemy. Dictionaries differed in whether the distinction was made in a separate homograph, in a separate sense within a homograph, or as a subsense. They also differed on whether the sense was labeled as figurative.

2.3 Frequency, learner's dictionaries, and historical order

We used ngrams from three large corpora (Project Gutenberg, the Wikipedia, and the Google n-grams dataset) to make an assessment of the ordering based on corpus frequency, and the sense which is most frequent is indicated in Table 2.

---

[5] The order for *john* in the Labels column (prostitute-client:toilet) corresponds to the difference between the four dictionaries, which were the only ones in which both senses were listed. The word *defuse* was not defined in MW7.

The corpus results generally support the ordering in the learner's dictionaries for the Homonymy dataset. The results for *inflect* depended on the word form. The root was typically associated with inflecting a word (the older sense), but the derived form *inflection* was associated more often with vocal inflection, and *inflectional* was associated with inflecting a word. For all words in the Metaphor dataset (with the exception of *digest*), MW7 ordered the figurative sense after the literal sense, which is generally what we would expect, and so there is less disagreement for the Metaphor dataset. That is, the historical order and the most frequent order are generally the same. An exception is the words *conceive* and *defuse*, which were more frequent in the figurative than the literal sense in all three corpora.

2.4   WordNet and Metaphor

The results for the Metaphor dataset showed that the literal sense was usually most frequent, but not always, and the most frequent sense is sometimes corpus-specific (e.g, *underline* is used most frequently in the literal sense in the Project Gutenberg corpus, and the figurative sense is most common in the Wikipedia and the Internet-based corpus).

WordNet differed from all of the other dictionaries in terms of Metaphor. It had the most words (6 out of 25) that were only defined in the literal sense. WordNet also ordered metaphorical senses before literal senses more than the other dictionaries.

WordNet is the most widely used dictionary in Computational Linguistics, and the ordering is partially based on the frequency of the senses in SemCor, a subset of the Brown corpus that has been manually tagged with senses from WordNet (Landes/Leacock/Tengi 1998). This corpus is small (only about 200,000 word tokens), and many word senses appear infrequently.

We conducted a small experiment to leverage the ordering in COBUILD to re-order the corresponding senses in WordNet with regard to a literal/metaphor distinction. This was based on a manual mapping between the senses in our own sense inventory, and the corresponding senses (where found) in WordNet We looked at an additional 20 words that have a literal/figurative distinction: *agitator*, *avalanche*, *bankrupt*, *barrage*, *beak*, *beanpole*, *bloodsucker*, *blight*, *lamb*, *leech*, *pedestal*, *shark* (nouns), and *applaud*, *backfire*, *backtrack*, *bait*, *devour*, *nosedive*, *unmask*, *unseat* (verbs). Of these words, six were defined only in a literal or figurative sense in WordNet, and one word was not defined. Most of the remaining words were in a literal/figurative order in both WordNet and in COBUILD. We examined those senses where the figurative sense was listed before the literal sense. As Table 2 shows, there are some cases such as *conceive* and *defuse* where a figurative sense is more frequent than a literal sense, but for most of the words in our dataset the literal sense is more frequent. We then examined COBUILD to see if the order supported the ordering in WordNet, or if the ordering differed. The aim is to leverage the larger corpus frequency that COBUILD is based on. The ordering differed for *blight* and *devour*. *W*e identified ngrams in the Project Gutenberg, Wikipedia, and Google datasets for these two words. We found that they are generally used literally more often than figuratively.[6]

---

6   The primary exceptions are *urban blight*, and *suburban blight*.

3.   Discussion

The order of senses is not an easy decision. There can be a conflict between the most frequent sense and the most salient sense, or between the most frequent sense and a sense order that would follow a consistent pattern such as putting the literal sense before the figurative one.

From the perspective of Computational Linguistics though, the decision is easier – we want to identify the sense that is most frequent. The Most Frequent Sense heuristic is widely used in research on word sense disambiguation (Agirre/Edmonds 2007). It is used as a back-off method when we do not have enough information to make a more informed choice. We found it was relatively straightforward to identify corresponding senses between WordNet and other dictionaries with regard to a literal/figurative distinction, and we were able to use this information to propose a re-ordering of the WordNet senses. We were also able to use corpus ngrams from Gutenberg, the Wikipedia, and the Internet to support that re-ordering.

The more difficult problem is the missing senses in WordNet. It stood out among the dictionaries in terms of identifying figurative senses least often. In the additional sample of 20 words that we used to assess sense order and metaphor, six of the words were missing a literal or a figurative sense in WordNet, and one word was not defined. WordNet has been criticized for being „too fine-grained" (making too many distinctions), but this is a case where additional distinctions are needed.

The results on word-sense individuation show that there is a great deal of consistency for words that are in the Homonymy class. Over the set of of 25 words, we found that almost all dictionaries distinguished the senses. The dictionaries differed in the order of the senses, but usually not in the fact that they were distinguished. The results on the Systematic Polysemy dataset show that the senses in this group are distinguished least often, and this is what we would expect. (Panman 1982) found that when two senses are homonyms, people agree that the senses are different, and when the senses are related people disagree about whether they are distinct meanings. However, we found that even amongst the sets of words that are systematically related, there is an ordering of different to similar. Dictionaries distinguished senses most often for words with a substance/color relationship (as with *gold*, *silver*, and *amber*), and least often for words with a music/dance relationship (as with *waltz*, *foxtrot*, and *tango*).

4.   Conclusion

This paper looked at sense-ordering across a number of different dictionaries. WordNet differed from all the dictionaries with respect to metaphor. A small experiment showed that the COBUILD dictionary and an ngram analysis can be leveraged to re-order those senses that were out-of-order with regard to frequency. The dataset of 75 words and the information about their ordering in the different dictionaries is available from: http://lexicalresearch.com/resources/euralex-2022-dataset.tar

References

Agirre, E./Edmonds, P. (eds.) (2007): Word sense disambiguation: algorithms and applications, Heidelberg.

Atkins, B./Levin, B. (1998): Admitting impediments. In: Zernik U. (ed.): Lexical acquisition: exploiting on-line resources to build a lexicon. Hillsdale, NJ, pp. 233–262.

Bejoint, H. (1988): Monosemy and the dictionary. In: BudaLex'88 Proceedings: Papers from the 3rd International Euralex Congress, pp. 11–26.

Brants, T./Franz, A. (2006): Web 1T 5-gram corpus version 1.1. Technical report, google research. The resource is available from the linguistic data consortium. https://catalog.ldc.upenn.edu/LDC2006T13.

Cambridge international dictionary of English (1995). Cambridge, MA.

Collins COBUILD English language dictionary (1987). London/New York.

Collins English dictionary (1979). First edition. London/New York.

Fellbaum, C. (ed.) (1998): WordNet: an electronic lexical database. Cambridge, MA.

Hanks, P. (1979): To what extent does a dictionary definition define? In: ITL International Journal of Applied Linguistics 45, pp. 32–38.

Hiorth, F. (1954): Arrangement of meanings. In: Lexicography, Lingua 4, pp. 413–424.

Kilgarriff, A. (1997): I don't believe. In: Word Senses, Computers and the Humanities 31 (2), pp. 91–113.

Kipfer, B. (1983): Methods of ordering senses within entries. In: Proceedings of Euralex, pp. 49–54.

Landes, S./Leacock, C./Tengi, R. (1998): Building semantic concordances. In: Fellbaum, C. (ed.): WordNet: an electronic lexical database. Cambridge, MA, pp. 199–216.

Lew, R. (2013): Identifying, ordering and defining senses. In: The Bloombury companion to lexicography, pp. 284–302.

Longman dictionary of contemporary English (1978). London.

Oxford advanced learner's dictionary (1989). Fourth edition. Oxford.

Panman, O. (1982): Homonymy and polysemy. In: Lingua, pp. 105–136.

Webster's seventh new collegiate dictionary (1965). Springfield, MA.

Webster's new world dictionary of the American language (1970). Second college edition. New York.

Yarowsky, D. (1993): One sense per collocation. In: Proceedings of the Workshop on Human Language Technology, pp. 266–271.

Contact information

Robert Krovetz
Lexical Research
rkrovetz@lexicalresearch.com