

# Choosing Sense Distinctions for WSD: Psycholinguistic Evidence

**Susan Windisch Brown**

Department of Linguistics  
Institute of Cognitive Science  
University of Colorado  
Hellems 295 UCB  
Boulder, CO 80309  
susan.brown@colorado.edu

## Abstract

Supervised word sense disambiguation requires training corpora that have been tagged with word senses, which begs the question of which word senses to tag with. The default choice has been WordNet, with its broad coverage and easy accessibility. However, concerns have been raised about the appropriateness of its fine-grained word senses for WSD. WSD systems have been far more successful in distinguishing coarse-grained senses than fine-grained ones (Navigli, 2006), but does that approach neglect necessary meaning differences? Recent psycholinguistic evidence seems to indicate that closely related word senses may be represented in the mental lexicon much like a single sense, whereas distantly related senses may be represented more like discrete entities. These results suggest that, for the purposes of WSD, closely related word senses can be clustered together into a more general sense with little meaning loss. The current paper will describe this psycholinguistic research and its implications for automatic word sense disambiguation.

## 1 Introduction\*

The problem of creating a successful word sense disambiguation system begins, or should begin, well before methods or algorithms are considered. The first question should be, “Which senses do we want to be able to distinguish?” Dictionaries en-

courage us to consider words as having a discrete set of senses, yet any comparison between dictionaries quickly reveals how differently a word’s meaning can be divided into separate senses. Rather than having a finite list of senses, many words seem to have senses that shade from one into another.

One could assume that dictionaries make broadly similar divisions and the exact point of division is only a minor detail. Simply picking one resource and sticking with it should solve the problem. In fact, WordNet, with its broad coverage and easy accessibility, has become the resource of choice for WSD. However, some have questioned whether WordNet’s fine-grained sense distinctions are appropriate for the task (Ide & Wilks, 2007; Palmer et al., 2007). Some are concerned about feasibility: Is WSD at this level an unattainable goal? Others with practicality: Is this level of detail really needed for most NLP tasks, such as machine translation or question-answering? Finally, some wonder whether such fine-grained distinctions even reflect how human beings represent word meaning.

Human annotators have trouble distinguishing such fine-grained senses reliably. Interannotator agreement with WordNet senses is around 70% (Snyder & Palmer, 2004; Chklovski & Mihalcea, 2002), and it’s understandable that WSD systems would have difficulty surpassing this upper bound.

Researchers have responded to these concerns by developing various ways to cluster WordNet senses. Mihalcea & Moldovan (2001) created an unsupervised approach that uses rules to cluster senses. Navigli (2006) has induced clusters by mapping WordNet senses to a more coarse-grained lexical resource. OntoNotes (Hovy et al., 2006) is manually grouping WordNet senses and creating a corpus tagged with these sense groups. Using On-

---

\* I gratefully acknowledge the support of the National Science Foundation Grant NSF-0415923, Word Sense Disambiguation.

toNotes and another set of manually tagged data, Snow et al. (2007) have developed a supervised method of clustering WordNet senses.

Although ITA rates and system performance both significantly improve with coarse-grained senses (Duffield et al., 2007; Navigli, 2006), the question about what level of granularity is needed remains. Palmer et al. (2007) state, “If too much information is being lost by failing to make the more fine-grained distinctions, the [sense] groups will avail us little.”

Ides and Wilks (2007) drew on psycholinguistic research to help establish an appropriate level of sense granularity. However, there is no consensus in the psycholinguistics field on how lexical meaning is represented in the mind (Klein & Murphy, 2001; Pytkänen et al., 2006; Rodd et al., 2002), and, as the Ide and Wilks (2007) state, “research in this area has been focused on developing psychological models of language processing and has not directly addressed the problem of identifying senses that are distinct enough to warrant, in psychological terms, a separate representation in the mental lexicon.”

Our experiment looked directly at sense distinctions of varying degrees of meaning relatedness and found indications that the mental lexicon does not consist of separate representations of discrete senses for each word. Rather, word senses may share a greater or smaller portion of a semantic representation depending on the how closely related the senses are. Because closely related senses may share a large portion of their semantic representation, clustering such senses together would result in very little meaning loss. The remainder of this paper will describe the experiment and its implications for WSD in more detail.

## 2 Experiment

The goal of this experiment was to determine whether each sense of a word has a completely separate mental representation or not. If so, we also hoped to discover what types of sense distinctions seem to have separate mental representations.

### 2.1 Materials

Four groups of materials were prepared using the fine-grained sense distinctions found in WordNet 2.1. Each group consisted of 11 pairs of phrases. The groups comprised (1) homonymy, (2) distantly

related senses, (3) closely related senses, and (4) same senses (see Table 1 for examples). Placement in these groups depended both on the classification of the usages by WordNet and the Oxford English Dictionary and on the ratings given to pairs of phrases by a group of undergraduates. They rated the relatedness of the verb in each pair on a scale of 0 to 3, with 0 being completely unrelated and 3 being the same sense.

A pair was considered to represent the same sense if the usage of the verb in both phrases was categorized by WordNet as the same and if the pair received a rating greater than 2.7. Closely related senses were listed as separate senses by WordNet and received a rating between 1.8 and 2.5. Distantly related senses were listed as separate senses by WordNet and received ratings between 0.7 and 1.3. Because WordNet makes no distinction between related and unrelated senses, the Oxford English Dictionary was used to classify homonyms. Homonyms were listed as such by the OED and received ratings under 0.3.

	Prime	Target
Unrelated	banked the plane	banked the money
Distantly related	ran the track	ran the shop
Closely related	broke the glass	broke the radio
Same sense	cleaned the shirt	cleaned the cup

Table 1. Stimuli.

### 2.2 Method

The experiment used a semantic decision task (Klein & Murphy, 2001; Pytkänen et al., 2006), in which people were asked to judge whether short phrases “made sense” or not. Subjects saw a phrase, such as “posted the guard,” and would decide whether the phrase made sense as quickly and as accurately as possible. They would then see another phrase with the same verb, such as “posted the letter,” and respond to that phrase as well. The response time and accuracy were recorded for the second phrase of each pair.

### 2.3 Results and Discussion

When comparing response times between same sense pairs and different sense pairs (a combina-

tion of closely related, distantly related, and unrelated senses), we found a reliable difference (same sense mean: 1056ms, different sense mean: 1272ms;  $t_{32} = 6.33$ ;  $p < .0001$ ). We also found better accuracy for same sense pairs (same sense: 95.6% correct vs. different sense: 78% correct;  $t_{32} = 7.49$ ;  $p < .0001$ ). When moving from one phrase to another with the same meaning, subjects were faster and more accurate than when moving to a phrase with a different sense of the verb.

By itself, this result would fit with the theory that every sense of a word has a separate semantic representation. One would expect people to access the meaning of a verb quickly if they had just seen the verb used with that same meaning. One could think of the meaning as already having been “activated” by the first phrase. Accessing a completely different semantic representation when moving from one sense to another should be slower.

If all senses have separate representations, access to meaning should proceed in the same way for all. For example, if one is primed with the phrase “fixed the radio,” response time and accuracy should be the same whether the target is “fixed the vase” or “fixed the date.” Instead, we found a significant difference between these two groups, with closely related pairs accessed, on average, 173ms more quickly than the mean of the distantly and unrelated pairs ( $t_{32} = 5.85$ ;  $p < .0005$ ), and accuracy was higher (91% vs. 72%;  $t_{32} = 8.65$ ;  $p < .0001$ ).

A distinction between distantly related pairs and homonyms was found as well. Response times for distantly related pairs was faster than for homonyms (distantly related mean: 1253ms, homonym mean: 1406ms;  $t_{32} = 2.38$ ;  $p < .0001$ ). Accuracy was enhanced as well for this group (distantly related mean: 81%, unrelated mean: 62%;  $t_{32} = 5.66$ ;  $p < .0001$ ). Related meanings, even distantly related, seem to be easier to access than unrelated meanings.

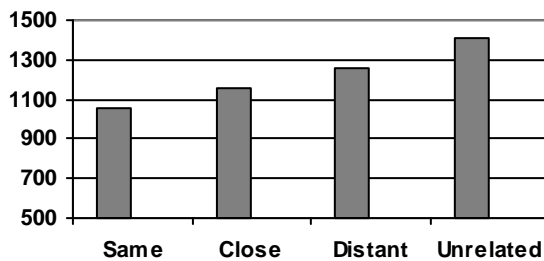


Figure 1. Mean response time (ms).

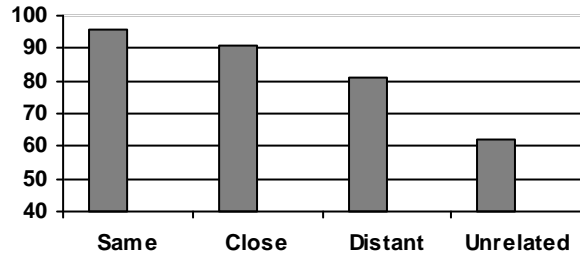


Figure 2. Mean accuracy (% correct).

A final planned comparison tested for a linear progression through the test conditions. Although somewhat redundant with the other comparisons, this test did reveal a highly significant linear progression for response time ( $F_{1,32} = 95.8$ ;  $p < .0001$ ) and for accuracy ( $F_{1,32} = 100.1$ ;  $p < .0001$ ).

People have an increasingly difficult time accessing the meaning of a word as the relatedness of the meaning in the first phrase grows more distant. They respond more slowly and their accuracy declines. However, closely related senses are almost as easy to access as same sense phrases. These results suggest that closely related word senses may be represented in the mental lexicon much like a single sense, perhaps sharing a core semantic representation.

The linear progression through meaning relatedness is also compatible with a theory in which the semantic representations of related senses overlap. Rather than being discrete entities attached to a main “entry”, they could share a general semantic space. Various portions of the space could be activated depending on the context in which the word occurs. This structure allows for more coarse-grained or more fine-grained distinctions to be made, depending on the needs of the moment.

A structure in which the semantic representations overlap allows for the apparently smooth progression from same sense usages to more and more distantly related usages. It also provides a simple explanation for semantically underdetermined usages of a word. Although separate senses of a word can be identified in different contexts, in some contexts, both senses (or a vague meaning indeterminate between the two) seem to be represented by the same word. For example, “newspaper” can refer to a physical object: “He tore the newspaper in half”, or to the content of a publication: “The newspaper made me mad today, suggesting that our committee is corrupt.” The sen-

tence “I really like this newspaper” makes no commitment to either sense.

### 3 Conclusions

What does this mean for WSD? Most would agree that NLP applications would benefit from the ability to distinguish homonym-level meaning differences. Similarly, most would agree that it is not necessary to make very fine distinctions, even if we can describe them. For example, the process of cleaning a cup is discernibly different from the process of cleaning a shirt, yet we would not want to have a WSD system try to distinguish between every minor variation on cleaning. The problem comes with deciding when meanings can be considered the same sense, and when they should be considered different.

The results of this study suggest that some word usages considered different by WordNet provoke similar responses as those to same sense usages. If these usages activate the same or largely overlapping meaning representations, it seems safe to assume that little meaning loss would result from clustering these closely related senses into one more general sense. Conversely, people reacted to distantly related senses much as they did to homonyms, suggesting that making distinctions between these usages would be useful in a WSD system.

A closer analysis of the study materials reveals differences between the types of distinctions made in the closely related senses and the types made in the distantly related senses. Most of the closely related senses distinguished between different concrete usages, whereas the distantly related senses distinguished between a concrete usage and a figurative or metaphorical usage. This suggests that grouping concrete usages together may result in little, if any, meaning loss. It may be more important to keep concrete senses distinct from figurative or metaphorical senses. The present study, however, divided senses only on degree of relatedness rather than type of relatedness. It would be useful in future studies to address more directly the question of distinctions based on concreteness, animacy, agency, and so on.

### References

- Chklovski, Tim, and Rada Mihalcea. 2002. Building a sense tagged corpus with open mind word expert. *Proc. of ACL 2002 Workshop on WSD: Recent Successes and Future Directions*. Philadelphia, PA.
- Duffield, Cecily Jill, Jena D. Hwang, Susan Windisch Brown, Dmitriy Dligach, Sarah E. Vieweg, Jenny Davis, Martha Palmer. 2007. Criteria for the manual grouping of verb senses. *Linguistics Annotation Workshop, ACL-2007*. Prague, Czech Republic.
- Hovy, Eduard, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes: The 90% solution. *Proc. of HLT-NAACL 2006*. New York, NY.
- Ide, Nancy, and Yorick Wilks. 2007. Making sense about sense. In *Word Sense Disambiguation: Algorithms and Applications*, E. Agirre and P. Edmonds (eds.). Dordrecht, The Netherlands: Springer.
- Klein, D., and Murphy, G. (2001). The representation of polysemous words. *J of Memory and Language* 45, 259-282.
- Mihalcea, Rada, and Dan I. Moldovan. 2001. Automatic generation of a coarse-grained WordNet. In *Proc. of NAACL Workshop on WordNet and Other Lexical Resources*. Pittsburg, PA.
- Navigli, Roberto. 2006. Meaningful clustering of word senses helps boost word sense disambiguation performance. *Proc. of the 21<sup>st</sup> International Conference on Computational Linguistics*. Sydney, Australia.
- Palmer, Martha, Hwee Tou Ng, and Hoa Trang Dang. 2007. Evaluation of WSD systems. In *Word Sense Disambiguation: Algorithms and Applications*, E. Agirre and P. Edmonds (eds.). Dordrecht, The Netherlands: Springer.
- Pylkkänen, L., Llinás, R., and Murphy, G. L. (2006). The representation of polysemy: MEG evidence. *J of Cognitive Neuroscience* 18, 97-109.
- Rodd, J., Gaskell, G., and Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *J. of Memory and Language*, 46, 245-266.
- Snow, Rion, Sushant Prakash, Dan Jurafsky and Andrew Y. Ng. 2007. Learning to merge word senses. *Proc. of EMNLP 2007*. Prague, Czech Republic.
- Snyder, Benjamin, and Martha Palmer. 2004. The English all-words task. *Proc. of ACL 2004 SENSEVAL-3 Workshop*. Barcelona, Spain.