

# A Broad-Coverage Word Sense Tagger

Dekang Lin

Department of Computer Science  
University of Manitoba  
Winnipeg, Manitoba, Canada R3T 2N2  
lindek@cs.umanitoba.ca

Previous corpus-based Word Sense Disambiguation (WSD) algorithms (Hearst, 1991; Bruce and Wiebe, 1994; Leacock et al., 1996; Ng and Lee, 1996; Yarowsky, 1992; Yarowsky, 1995) determine the meanings of polysemous words by exploiting their **local contexts**. A basic intuition that underlies those algorithms is the following:

- (1) Two occurrences of the *same* word have *identical* meanings if they have *similar* local contexts.

In other words, previous corpus-based WSD algorithms learn to disambiguate a polysemous word from previous usages of the same word. This has several undesirable consequences. Firstly, a word must occur thousands of times before a good classifier can be trained. There are thousands of polysemous words, e.g., 11,562 polysemous nouns in WordNet (Miller, 1990). For every polysemous word to occur thousands of times each, the corpus must contain billions of words. Secondly, learning to disambiguate a word from the previous usages of the *same* word means that whatever was learned for one word is not used on other words, which obviously missed generality in natural languages. Thirdly, these algorithms cannot deal with words for which classifiers have not been trained, which explains why most previous WSD algorithms only deal with a dozen of polysemous words.

We demonstrate a new WSD algorithm that relies on a different intuition:

- (2) Two *different* words are likely to have *similar* meanings if they occur in *identical* local contexts.

The local context of a word is defined in our algorithm as a syntactic dependency relationship that the word participates in. To disambiguate a polysemous word, we search a local context database to retrieve the list of words (called **selectors**) that appeared in the same local context as the polysemous word in the training corpus. The meaning of the polysemous word is determined by maximizing its similarity to the selectors.

For example, consider the sentence:

- (3) The new facility will employ 500 of the existing 600 employees

The word “facility” has 5 possible meanings in WordNet 1.5:

1. installation
2. proficiency/technique
3. adeptness
4. readiness
5. toilet/bathroom

Since the word “facility” is the subject of “employ” and is modified by “new” in (3), we retrieve other words that appeared in the same contexts and obtain the following two groups of selectors (the  $\log \lambda$  column shows the likelihood ratios (Dunning, 1993) of these words in the local contexts):

- Subjects of “employ” with top-20 highest likelihood ratios:

word	freq	$\log \lambda$	word	freq	$\log \lambda$
ORG*	64	50.4	machine	3	6.56
plant	14	31.0	corporation	3	6.47
company	27	28.6	manufacturer	3	6.21
operation	8	23.0	insurance company	2	6.06
industry	9	14.6	aerospace	2	5.81
firm	8	13.5	memory device	1	5.79
pirate	2	12.1	department	3	5.55
unit	9	9.32	foreign office	1	5.41
shift	3	8.48	enterprise	2	5.39
postal service	2	7.73	pilot	2	5.37

\*ORG includes all proper names recognized as organizations

- Modifiers of “new” with top-20 highest likelihood ratios:

word	freq	log $\lambda$	word	freq	log $\lambda$
post	432	952.9	bonds	223	245.4
issue	805	902.8	capital	178	241.8
product	675	888.6	order	228	236.5
rule	459	875.8	version	158	223.7
law	356	541.5	position	236	207.3
technology	237	382.7	high	152	201.2
generation	150	323.2	contract	279	198.1
model	207	319.3	bill	208	194.9
job	260	269.2	venture	123	193.7
system	318	251.8	program	283	183.8

Since the similarity between Sense 1 of “facility” and the selectors is greater than that of other senses, the word “facility” in (3) is tagged “Sense 1”.

The key innovation of our algorithm is that a polysemous word is disambiguated with past usages of other words. Whether or not it appears in the training corpus is irrelevant.

Compared with previous corpus-based algorithms, our approach offers several advantages:

- The same knowledge sources are used for all words, as opposed to using a separate classifier for each individual word. For example, the same set of selectors can also be used to disambiguate “school” in “the new school employed 100 people”.
- It requires a much smaller training corpus that needs not be sense-tagged.
- It is able to deal with words that are infrequent or do not even appear in the training corpus.
- The same mechanism can also be used to infer the semantic categories of unknown words.

In the demonstrated system, the local context database is constructed with 8,665,362 dependency relationships that are extracted from a 25-million-word Wall Street Journal corpus. The corpus is parsed with a broad-coverage parser, PRINCEPAR, in 126 hours on a SPARC-Ultra 1/140 with 96MB of memory. The nouns in the input text are tagged with their senses in WordNet 1.5. Proper nouns that do not contain simple markers (e.g., Mr., Inc.) to indicate their categories are treated as 3-way ambiguous and are tagged as “group”, “person”, or “location” by the system.

## References

- Rebecca Bruce and Janyce Wiebe. 1994. Word-sense disambiguation using decomposable models. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 139–145, Las Cruces, New Mexico.
- Ted Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74, March.
- Marti Hearst. 1991. noun homograph disambiguation using local context in large text corpora. In *Conference on Research and Development in Information Retrieval ACM/SIGIR*, pages 36–47, Pittsburgh, PA.
- Claudia Leacock, Geoffrey Towwell, and Ellen M. Voorhees. 1996. Towards building contextual representations of word senses using statistical models. In *Corpus Processing for Lexical Acquisition*, chapter 6, pages 97–113. The MIT Press.
- George A. Miller. 1990. WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–312.
- Hwee Tow Ng and Hian Beng Lee. 1996. Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach. In *Proceedings of 34th Annual Meeting of the Association for Computational Linguistics*, pages 40–47, Santa Cruz, California.
- David Yarowsky. 1992. Word-sense disambiguation using statistical models of Roget’s categories trained on large corpora. In *Proceedings of COLING-92*, Nantes, France.
- David Yarowsky. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of 33rd Annual Meeting of the Association for Computational Linguistics*, pages 189–196, Cambridge, Massachusetts, June.