

---

# Word-sense disambiguation

---

In computational linguistics, **word-sense disambiguation** (WSD) is an open problem of natural language processing, which governs the process of identifying which sense of a word (i.e. meaning) is used in a sentence, when the word has multiple meanings (polysemy). The solution to this problem impacts other computer-related writing, such as discourse, improving relevance of search engines, anaphora resolution, coherence, inference *et cetera*.

Research has progressed steadily to the point where WSD systems achieve sufficiently high levels of accuracy on a variety of word types and ambiguities. A rich variety of techniques have been researched, from dictionary-based methods that use the knowledge encoded in lexical resources, to supervised machine learning methods in which a classifier is trained for each distinct word on a corpus of manually sense-annotated examples, to completely unsupervised methods that cluster occurrences of words, thereby inducing word senses. Among these, supervised learning approaches have been the most successful algorithms to date.

Current accuracy is difficult to state without a host of caveats. In English, accuracy at the coarse-grained (homograph) level is routinely above 90%, with some methods on particular homographs achieving over 96%. On finer-grained sense distinctions, top accuracies from 59.1% to 69.0% have been reported in recent evaluation exercises (SemEval-2007, Senseval-2), where the baseline accuracy of the simplest possible algorithm of always choosing the most frequent sense was 51.4% and 57%, respectively.

## About

A disambiguation process requires two strict things: a dictionary to specify the senses which are to be disambiguated and a corpus of language data to be disambiguated (in some methods, a training corpus of language examples is also required). WSD task has two variants: "lexical sample" and "all words" task. The former comprises disambiguating the occurrences of a small sample of target words which were previously selected, while in the latter all the words in a piece of running text need to be disambiguated. The latter is deemed a more realistic form of evaluation, but the corpus is more expensive to produce because human annotators have to read the definitions for each word in the sequence every time they need to make a tagging judgement, rather than once for a block of instances for the same target word.

To give a hint how all this works, consider two examples of the distinct senses that exist for the (written) word "*bass*":

1. a type of fish
2. tones of low frequency

and the sentences:

1. *I went fishing for some sea bass.*
2. *The bass line of the song is too weak.*

To a human, it is obvious that the first sentence is using the word "*bass (fish)*", as in the former sense above and in the second sentence, the word "*bass (instrument)*" is being used as in the latter sense below. Developing algorithms to replicate this human ability can often be a difficult task, as is further exemplified by the implicit equivocation between "*bass (sound)*" and "*bass*" (musical instrument).

---

## History

WSD was first formulated as a distinct computational task during the early days of machine translation in the 1940s, making it one of the oldest problems in computational linguistics. Warren Weaver, in his famous 1949 memorandum on translation,<sup>[1]</sup> first introduced the problem in a computational context. Early researchers understood the significance and difficulty of WSD well. In fact, Bar-Hillel (1960) used the above example to argue<sup>[2]</sup> that WSD could not be solved by "electronic computer" because of the need in general to model all world knowledge.

In the 1970s, WSD was a subtask of semantic interpretation systems developed within the field of artificial intelligence, but since WSD systems were largely rule-based and hand-coded they were prone to a knowledge acquisition bottleneck.

By the 1980s large-scale lexical resources, such as the Oxford Advanced Learner's Dictionary of Current English (OALD), became available: hand-coding was replaced with knowledge automatically extracted from these resources, but disambiguation was still knowledge-based or dictionary-based.

In the 1990s, the statistical revolution swept through computational linguistics, and WSD became a paradigm problem on which to apply supervised machine learning techniques.

The 2000s saw supervised techniques reach a plateau in accuracy, and so attention has shifted to coarser-grained senses, domain adaptation, semi-supervised and unsupervised corpus-based systems, combinations of different methods, and the return of knowledge-based systems via graph-based methods. Still, supervised systems continue to perform best.

## Difficulties

### Differences between dictionaries

One problem with word sense disambiguation is deciding what the senses are. In cases like the word *bass* above, at least some senses are obviously different. In other cases, however, the different senses can be closely related (one meaning being a metaphorical or metonymic extension of another), and in such cases division of words into senses becomes much more difficult. Different dictionaries and thesauruses will provide different divisions of words into senses. One solution some researchers have used is to choose a particular dictionary, and just use its set of senses. Generally, however, research results using broad distinctions in senses have been much better than those using narrow ones.<sup>[3][4]</sup> However, given the lack of a full-fledged coarse-grained sense inventory, most researchers continue to work on fine-grained WSD.

Most research in the field of WSD is performed by using WordNet as a reference sense inventory for English. WordNet is a computational lexicon that encodes concepts as synonym sets (e.g. the concept of car is encoded as { car, auto, automobile, machine, motorcar }). Other resources used for disambiguation purposes include Roget's Thesaurus<sup>[5]</sup> and Wikipedia.<sup>[6]</sup>

### Part-of-speech tagging

In any real test, part-of-speech tagging and sense tagging are very closely related with each potentially making constraints to the other. And the question whether these tasks should be kept together or decoupled is still not unanimously resolved, but recently scientists incline to test these things separately (e.g. in the Senseval/SemEval competitions parts of speech are provided as input for the text to disambiguate).

It is instructive to compare the word sense disambiguation problem with the problem of part-of-speech tagging. Both involve disambiguating or tagging with words, be it with senses or parts of speech. However, algorithms used for one do not tend to work well for the other, mainly because the part of speech of a word is primarily determined by the immediately adjacent one to three words, whereas the sense of a word may be determined by words further away. The success rate for part-of-speech tagging algorithms is at present much higher than that for WSD, state-of-the-art

being around 95% accuracy or better, as compared to less than 75% accuracy in word sense disambiguation with supervised learning. These figures are typical for English, and may be very different from those for other languages.

### Inter-judge variance

Another problem is inter-judge variance. WSD systems are normally tested by having their results on a task compared against those of a human. However, while it is relatively easy to assign parts of speech to text, training people to tag senses is far more difficult.<sup>[7]</sup> While users can memorize all of the possible parts of speech a word can take, it is often impossible for individuals to memorize all of the senses a word can take. Moreover, humans do not agree on the task at hand – give a list of senses and sentences, and humans will not always agree on which word belongs in which sense.<sup>[8]</sup>

Thus, a computer cannot be expected to give better performance on such a task than a human (indeed, since the human serves as the standard, the computer being better than the human is incoherent), so the human performance serves as an upper bound. Human performance, however, is much better on coarse-grained than fine-grained distinctions, so this again is why research on coarse-grained distinctions<sup>[9][10]</sup> has been put to test in recent WSD evaluation exercises.<sup>[3][4]</sup>

### Common sense

Some AI researchers like Douglas Lenat argue that one cannot parse meanings from words without some form of common sense ontology. For example, comparing these two sentences:

- "Jill and Mary are sisters." – (they are sisters of each other).
- "Jill and Mary are mothers." – (each is independently a mother).

To properly identify senses of words one must know common sense facts.<sup>[11]</sup> Moreover, sometimes the common sense is needed to disambiguate such words like pronouns in case of having anaphoras or cataphoras in the text.

### Sense inventory and algorithms' task-dependency

A task-independent sense inventory is not a coherent concept<sup>[12]</sup>: each task requires its own division of word meaning into senses relevant to the task. For example, the ambiguity of 'mouse' (animal or device) is not relevant in English-French machine translation, but is relevant in information retrieval. The opposite is true of 'river', which requires a choice in French (fleuve 'flows into the sea', or rivière 'flows into a river').

Also, completely different algorithms might be required by different applications. In machine translation, the problem takes the form of target word selection. Here the "senses" are words in the target language, which often correspond to significant meaning distinctions in the source language (bank could translate to French banque 'financial bank' or rive 'edge of river'). In information retrieval, a sense inventory is not necessarily required, because it is enough to know that a word is used in the same sense in the query and a retrieved document; what sense that is, is unimportant.

### Discreteness of senses

Finally, the very notion of "word sense" is slippery and controversial. Most people can agree in distinctions at the coarse-grained homograph level (e.g., pen as writing instrument or enclosure), but go down one level to fine-grained polysemy, and disagreements arise. For example, in Senseval-2, which used fine-grained sense distinctions, human annotators agreed in only 85% of word occurrences.<sup>[13]</sup> Word meaning is in principle infinitely variable and context sensitive. It does not divide up easily into distinct or discrete sub-meanings.<sup>[14]</sup> Lexicographers frequently discover in corpora loose and overlapping word meanings, and standard or conventional meanings extended, modulated, and exploited in a bewildering variety of ways. The art of lexicography is to generalize from the corpus to definitions that evoke and explain the full range of meaning of a word, making it seem like words are well-behaved

semantically. However, it is not at all clear if these same meaning distinctions are applicable in computational applications, as the decisions of lexicographers are usually driven by other considerations. Recently, a task – named lexical substitution – has been proposed as a possible solution to the sense discreteness problem.<sup>[15]</sup> The task consists of providing a substitute for a word in context that preserves the meaning of the original word (potentially, substitutes can be chosen from the full lexicon of the target language, thus overcoming discreteness).

## Approaches and methods

As in all natural language processing, there are two main approaches to WSD – deep approaches and shallow approaches.

Deep approaches presume access to a comprehensive body of world knowledge. Knowledge, such as "you can go fishing for a type of fish, but not for low frequency sounds" and "songs have low frequency sounds as parts, but not types of fish", is then used to determine in which sense the word is used. These approaches are not very successful in practice, mainly because such a body of knowledge does not exist in a computer-readable format, outside of very limited domains.<sup>[16]</sup> However, if such knowledge did exist, then deep approaches would be much more accurate than the shallow approaches. Also, there is a long tradition in computational linguistics, of trying such approaches in terms of coded knowledge and in some cases, it is hard to say clearly whether the knowledge involved is linguistic or world knowledge. The first attempt was that by Margaret Masterman and her colleagues, at the Cambridge Language Research Unit in England, in the 1950s. This attempt used as data a punched-card version of Roget's Thesaurus and its numbered "heads", as an indicator of topics and looked for repetitions in text, using a set intersection algorithm. It was not very successful,<sup>[17]</sup> but had strong relationships to later work, especially Yarowsky's machine learning optimisation of a thesaurus method in the 1990s.

Shallow approaches don't try to understand the text. They just consider the surrounding words, using information such as "if *bass* has words *sea* or *fishing* nearby, it probably is in the fish sense; if *bass* has the words *music* or *song* nearby, it is probably in the music sense." These rules can be automatically derived by the computer, using a training corpus of words tagged with their word senses. This approach, while theoretically not as powerful as deep approaches, gives superior results in practice, due to the computer's limited world knowledge. However, it can be confused by sentences like *The dogs bark at the tree* which contains the word *bark* near both *tree* and *dogs*.

There are four conventional approaches to WSD:

- Dictionary- and knowledge-based methods: These rely primarily on dictionaries, thesauri, and lexical knowledge bases, without using any corpus evidence.
- Supervised methods: These make use of sense-annotated corpora to train from.
- Semi-supervised or minimally supervised methods: These make use of a secondary source of knowledge such as a small annotated corpus as seed data in a bootstrapping process, or a word-aligned bilingual corpus.
- Unsupervised methods: These eschew (almost) completely external information and work directly from raw unannotated corpora. These methods are also known under the name of word sense discrimination.

Almost all these approaches normally work by defining a window of  $n$  content words around each word to be disambiguated in the corpus, and statistically analyzing those  $n$  surrounding words. Two shallow approaches used to train and then disambiguate are Naïve Bayes classifiers and decision trees. In recent research, kernel-based methods such as support vector machines have shown superior performance in supervised learning. Graph-based approaches have also gained much attention from the research community, and currently achieve performance close to the state of the art.

## Dictionary- and knowledge-based methods

The Lesk algorithm<sup>[18]</sup> is the seminal dictionary-based method. It is based on the hypothesis that words used together in text are related to each other and that the relation can be observed in the definitions of the words and their senses. Two (or more) words are disambiguated by finding the pair of dictionary senses with the greatest word overlap in their dictionary definitions. For example, when disambiguating the words in "pine cone", the definitions of the appropriate senses both include the words evergreen and tree (at least in one dictionary).

An alternative to the use of the definitions is to consider general word-sense relatedness and to compute the semantic similarity of each pair of word senses based on a given lexical knowledge base such as WordNet. Graph-based methods reminiscent of spreading activation research of the early days of AI research have been applied with some success. More complex graph-based approaches have been shown to perform almost as well as supervised methods<sup>[19]</sup> or even outperforming them on specific domains.<sup>[3][20]</sup> Recently, it has been reported that simple graph connectivity measures, such as degree, perform state-of-the-art WSD in the presence of a sufficiently rich lexical knowledge base.<sup>[21]</sup> Also, automatically transferring knowledge in the form of semantic relations from Wikipedia to WordNet has been shown to boost simple knowledge-based methods, enabling them to rival the best supervised systems and even outperform them in a domain-specific setting.<sup>[22]</sup>

The use of selectional preferences (or selectional restrictions) is also useful, for example, knowing that one typically cooks food, one can disambiguate the word bass in "I am cooking basses" (i.e., it's not a musical instrument).

## Supervised methods

Supervised methods are based on the assumption that the context can provide enough evidence on its own to disambiguate words (hence, world knowledge and reasoning are deemed unnecessary). Probably every machine learning algorithm going has been applied to WSD, including associated techniques such as feature selection, parameter optimization, and ensemble learning. Support Vector Machines and memory-based learning have been shown to be the most successful approaches, to date, probably because they can cope with the high-dimensionality of the feature space. However, these supervised methods are subject to a new knowledge acquisition bottleneck since they rely on substantial amounts of manually sense-tagged corpora for training, which are laborious and expensive to create.

## Semi-supervised methods

Because of the lack of training data, many word sense disambiguation algorithms use semi-supervised learning, which allows both labeled and unlabeled data. The Yarowsky algorithm was an early example of such an algorithm.<sup>[23]</sup> It uses the 'One sense per collocation' and the 'One sense per discourse' properties of human languages for word sense disambiguation. From observation, words tend to exhibit only one sense in most given discourse and in a given collocation.

The bootstrapping approach starts from a small amount of seed data for each word: either manually tagged training examples or a small number of surefire decision rules (e.g., 'play' in the context of 'bass' almost always indicates the musical instrument). The seeds are used to train an initial classifier, using any supervised method. This classifier is then used on the untagged portion of the corpus to extract a larger training set, in which only the most confident classifications are included. The process repeats, each new classifier being trained on a successively larger training corpus, until the whole corpus is consumed, or until a given maximum number of iterations is reached.

Other semi-supervised techniques use large quantities of untagged corpora to provide co-occurrence information that supplements the tagged corpora. These techniques have the potential to help in the adaptation of supervised models to different domains.

Also, an ambiguous word in one language is often translated into different words in a second language depending on the sense of the word. Word-aligned bilingual corpora have been used to infer cross-lingual sense distinctions, a kind

of semi-supervised system.

## Unsupervised methods

Unsupervised learning is the greatest challenge for WSD researchers. The underlying assumption is that similar senses occur in similar contexts, and thus senses can be induced from text by clustering word occurrences using some measure of similarity of context,<sup>[24]</sup> a task referred to as word sense induction or discrimination. Then, new occurrences of the word can be classified into the closest induced clusters/senses. Performance has been lower than other methods, above, but comparisons are difficult since senses induced must be mapped to a known dictionary of word senses. If a mapping to a set of dictionary senses is not desired, cluster-based evaluations (including measures of entropy and purity) can be performed. Alternatively, word sense induction methods can be tested and compared within an application. For instance, it has been shown that word sense induction improves Web search result clustering by increasing the quality of result clusters and the degree diversification of result lists.<sup>[25]</sup> It is hoped that unsupervised learning will overcome the knowledge acquisition bottleneck because they are not dependent on manual effort.

## Other approaches

Other approaches may vary differently in their methods:

- Identification of dominant word senses;<sup>[26][27][28]</sup>
- Domain-driven disambiguation;<sup>[29][30]</sup>
- WSD using Cross-Lingual Evidence.<sup>[31][32]</sup>

## Local impediments and summary

The knowledge acquisition bottleneck is perhaps the major impediment to solving the WSD problem. Unsupervised methods rely on knowledge about word senses, which is barely formulated in dictionaries and lexical databases. Supervised methods depend crucially on the existence of manually annotated examples for every word sense, a requisite that can so far be met only for a handful of words for testing purposes, as it is done in the Senseval exercises.

Therefore, one of the most promising trends in WSD research is using the largest corpus ever accessible, the World Wide Web, to acquire lexical information automatically.<sup>[33]</sup> WSD has been traditionally understood as an intermediate language engineering technology which could improve applications such as information retrieval (IR). In this case, however, the reverse is also true: Web search engines implement simple and robust IR techniques that can be successfully used when mining the Web for information to be employed in WSD. Therefore, the lack of training data provoked appearing some new algorithms and techniques described here:

## External knowledge sources

Knowledge is a fundamental component of WSD. Knowledge sources provide data which are essential to associate senses with words. They can vary from corpora of texts, either unlabeled or annotated with word senses, to machine-readable dictionaries, thesauri, glossaries, ontologies, etc. They can be<sup>[34][35]</sup> classified as follows:

1. Structured:
  - Thesauri
  - Machine-readable dictionaries (MRDs)
  - Ontologies
2. Unstructured:
  - Corpora: raw corpora and sense-annotated corpora
  - Collocation resources

- Other resources (such as word frequency lists, stoplists, domain labels,<sup>[36]</sup> etc.)

## Evaluation

Comparing and evaluating different WSD systems is extremely difficult, because of the different test sets, sense inventories, and knowledge resources adopted. Before the organization of specific evaluation campaigns most systems were assessed on in-house, often small-scale, data sets. In order to test one's algorithm, developers should spend their time to annotate all word occurrences. And comparing methods even on the same corpus is not eligible if there is different sense inventories.

In order to define common evaluation datasets and procedures, public evaluation campaigns have been organized. Senseval (now renamed SemEval) is an international word sense disambiguation competition, held every three years since 1998: Senseval-1<sup>[37]</sup> (1998), Senseval-2<sup>[38]</sup> (2001), Senseval-3<sup>[39]</sup> (2004), and its successor, SemEval<sup>[40]</sup> (2007). The objective of the competition is to organize different lectures, preparing and hand-annotating corpus for testing systems, perform a comparative evaluation of WSD systems in several kinds of tasks, including all-words and lexical sample WSD for different languages, and, more recently, new tasks such as semantic role labeling, gloss WSD, lexical substitution, etc. The systems submitted for evaluation to these competitions usually integrate different techniques and often combine supervised and knowledge-based methods (especially for avoiding bad performance in lack of training examples).

## Task Design Choices

**Sense Inventories.** During the first Senseval workshop the HECTOR sense inventory was adopted. The reason for adopting a previously unknown sense inventory was mainly to avoid the use of popular fine-grained word senses (such as WordNet), which could make the experiments unfair or biased. However, given the lack of coverage of such inventories, since the second Senseval workshop the WordNet sense inventory has been adopted.

**A set of testing words.** Comparison of methods can be divided in 2 groups by amount of words to test. The difference consists in the amount of analysis and processing:

- *all-words* task implies disambiguating all the words of the text
- *lexical sample* consists in disambiguating some previously chosen target words.

It is assumed that the former one is more realistic evaluation, although with very laborious testing of results. Initially only the latter was used in evaluation but later the former was included.

Lexical sample organizers had to choose samples on which the systems were to be tested. A criticism of earlier forays into lexical-sample WSD evaluation is that the lexical sample had been chosen according to the whim of the experimenter (or, to coincide with earlier experimenters' selections). For English Senseval, a sampling frame was devised in which words were classified according to their frequency (in the BNC) and their polysemy level (in WordNet). Also, inclusion POS-tagging problem was a matter of discussion and it was decided that samples should be words with known part of speech and some indeterminants (for ex. 15 noun tasks, 13 verb tasks, 8 adjectives, and 5 indeterminates).

**Baselines.** For comparison purposes, known, yet simple, algorithms named baselines are used. These include different variants of Lesk algorithm or most frequent sense algorithm.

**Sense inventory.** WSD exercises require a dictionary, to specify the word senses which are to be disambiguated, and a corpus of language data to be disambiguated. WordNet is the most popular example of sense inventory. The reason for adopting the HECTOR database during Senseval-1 was that the WordNet inventory was already publicly available.<sup>[41]</sup>

**Evaluation measures.** During the evaluation of WSD systems two main performance measures are used:

- **Precision:** the fraction of system assignments made that are correct
- **Recall:** the fraction of total word instances correctly assigned by a system

If a system makes an assignment for every word, then precision and recall are the same, and can be called **accuracy**. This model has been extended to take into account systems that return a set of senses with weights for each occurrence.

## Software

- WordNet::SenseRelate<sup>[42]</sup>, is a project that includes free, open source systems for word sense disambiguation and lexical sample sense disambiguation.
- UKB: Graph Base WSD<sup>[43]</sup>, is a collection of programs for performing graph-based Word Sense Disambiguation and lexical similarity/relatedness using a pre-existing Lexical Knowledge Base (LKB)<sup>[44]</sup>.

## Notes

- [1] Weaver 1949.
- [2] Bar-Hillel 1964, pp. 174-179.
- [3] Navigli, Litkowski & Hargraves 2007, pp. 30-35.
- [4] Pradhan et al. 2007, pp. 87-92.
- [5] Yarowsky 1992, pp. 454-460.
- [6] Mihalcea 2007.
- [7] Fellbaum 1997.
- [8] Snyder & Palmer 2004, pp. 41-43.
- [9] Navigli 2006, pp. 105-112.
- [10] Snow et al. 2007, pp. 1005-1014.
- [11] Lenat.
- [12] Palmer, Babko-Malaya & Dang 2004, pp. 49-56.
- [13] Edmonds 2000.
- [14] Kilgarrif 1997, pp. 91-113.
- [15] McCarthy & Navigli 2009, pp. 139-159.
- [16] Lenat & Guha 1989.
- [17] Wilks, Slator & Guthrie 1996.
- [18] Lesk 1986, pp. 24-26.
- [19] Navigli & Velardi 2005, pp. 1063-1074.
- [20] Agirre, Lopez de Lacalle & Soroa 2009, pp. 1501-1506.
- [21] Navigli & Lapata 2010, pp. 678-692.
- [22] Ponzetto & Navigli 2010, pp. 1522-1531.
- [23] Yarowsky 1995, pp. 189-196.
- [24] Schütze 1998, pp. 97-123.
- [25] Navigli & Crisafulli 2010.
- [26] McCarthy et al. 2007, pp. 553-590.
- [27] Mohammad & Hirst 2006, pp. 121-128.
- [28] Lapata & Keller 2007, pp. 348-355.
- [29] Gliozzo, Magnini & Strapparava 2004, pp. 380-387.
- [30] Buitelaar et al. 2006, pp. 275-298.
- [31] Ide, Erjavec & Tufis 2002, pp. 54-60.
- [32] Chan & Ng 2005, pp. 1037-1042.
- [33] Kilgarrif & Grefenstette 2003, pp. 333-347.
- [34] Litkowski 2005, pp. 753-761.
- [35] Agirre & Stevenson 2006, pp. 217-251.
- [36] Magnini & Cavaglia 2000, pp. 1413-1418.
- [37] <http://www.itri.brighton.ac.uk/events/senseval/ARCHIVE/index.html>
- [38] <http://193.133.140.102/senseval2/>
- [39] <http://www.senseval.org/senseval3>
- [40] <http://nlp.cs.swarthmore.edu/semEval/>
- [41] Kilgarrif & Rosenzweig 2000.
- [42] <http://senserelate.sourceforge.net>
- [43] <http://ixa2.si.ehu.es/ukb/>
- [44] <http://moin.delph-in.net/LkbTopI>

## Works cited

- Agirre, E.; Lopez de Lacalle, A.; Soroa, A. (2009) "Knowledge-based WSD on Specific Domains: Performing better than Generic Supervised WSD" (<http://www.ijcai.org/papers09/Papers/IJCAI09-251.pdf>) *Proc. of IJCAI*
- Agirre, E.; M. Stevenson. 2006. Knowledge sources for WSD. In *Word Sense Disambiguation: Algorithms and Applications*, E. Agirre and P. Edmonds, Eds. Springer, New York, NY.
- Bar-Hillel, Y. (1964). *Language and information*. Reading, MA: Addison-Wesley.
- Buitelaar, P.; B. Magnini, C. Strapparava and P. Vossen. 2006. Domain-specific WSD. In *Word Sense Disambiguation: Algorithms and Applications*, E. Agirre and P. Edmonds, Eds. Springer, New York, NY.
- Chan, Y. S.; H. T. Ng. 2005. Scaling up word sense disambiguation via parallel texts. In *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI, Pittsburgh, PA)*.
- Edmonds, P. 2000. Designing a task for SENSEVAL-2 (<http://www.itri.brighton.ac.uk/events/senseval/SENSEVAL2/task-design.ps>). Tech. note. University of Brighton, Brighton. U.K.
- Fellbaum, Christiane (1997) "Analysis of a handwriting task" *Proc. of ANLP-97 Workshop on Tagging Text with Lexical Semantics: Why, What, and How? Washington D.C., USA*
- Gliozzo, A.; B. Magnini and C. Strapparava. 2004. Unsupervised domain relevance estimation for word sense disambiguation (<http://www.aclweb.org/anthology/W/W04/W04-3249.pdf>). In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP, Barcelona, Spain)*.
- Ide, N.; T. Erjavec, D. Tufis. 2002. Sense discrimination with parallel corpora (<http://acl.ldc.upenn.edu/W/W02/W02-0808.pdf>). In *Proceedings of ACL Workshop on Word Sense Disambiguation: Recent Successes and Future Directions (Philadelphia, PA)*.
- Kilgarriff, A. 1997. I don't believe in word senses (<http://www.kilgarriff.co.uk/Publications/1997-K-CHum-believe.pdf>). *Comput. Human.* 31(2), pp. 91–113.
- Kilgarriff, A.; G. Grefenstette. 2003. Introduction to the special issue on the Web as corpus (<http://www.aclweb.org/anthology/J03-3001.pdf>). *Computational Linguistics* 29(3), pp. 333–347
- Kilgarriff, Adam; Joseph Rosenzweig, *English Senseval: Report and Results May–June, 2000*, University of Brighton
- Lapata, M.; and F. Keller. 2007. An information retrieval approach to sense ranking (<http://homepages.inf.ed.ac.uk/keller/papers/naacl07.pdf>). In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL, Rochester, NY)*.
- Lenat, D. "Computers versus Common Sense" ([http://www.youtube.com/watch?v=KSrUHGaUE\\_c](http://www.youtube.com/watch?v=KSrUHGaUE_c)). Retrieved 2008-12-10. (GoogleTachTalks on YouTube)
- Lenat, D.; R. V. Guha. 1989. *Building Large Knowledge-Based Systems*, Addison-Wesley
- Lesk; M. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone (<http://www.cse.ohio-state.edu/~cbrew/732/labs/p24-lesk.pdf>). In *Proc. of SIGDOC-86: 5th International Conference on Systems Documentation, Toronto, Canada*.
- Litkowski, K. C. 2005. Computational lexicons and dictionaries. In *Encyclopaedia of Language and Linguistics (2nd ed.)*, K. R. Brown, Ed. Elsevier Publishers, Oxford, U.K.
- Magnini, B; G. Cavaglià. 2000. Integrating subject field codes into WordNet. In *Proceedings of the 2nd Conference on Language Resources and Evaluation (LREC, Athens, Greece)*.
- McCarthy, D.; R. Koeling, J. Weeds, J. Carroll. 2007. Unsupervised acquisition of predominant word senses (<http://www.aclweb.org/anthology/J/J07/J07-4005.pdf>). *Computational Linguistics* 33(4): 553–590.
- McCarthy, D.; R. Navigli. 2009. The English Lexical Substitution Task ([http://www.dsi.uniroma1.it/~navigli/pubs/LRE\\_2009\\_McCarthy\\_Navigli.pdf](http://www.dsi.uniroma1.it/~navigli/pubs/LRE_2009_McCarthy_Navigli.pdf)), *Language Resources and Evaluation*, 43(2), Springer.
- Mihalcea, R. 2007. Using Wikipedia for Automatic Word Sense Disambiguation (<http://www.cs.unt.edu/~rada/papers/mihalcea.naacl07.pdf>). In *Proc. of the North American Chapter of the Association for Computational Linguistics (NAACL 2007), Rochester, April 2007*.

- Mohammad, S; G. Hirst. 2006. Determining word sense dominance using a thesaurus (<http://acl.ldc.upenn.edu/E/E06/E06-1016.pdf>). In Proceedings of the 11th Conference on European chapter of the Association for Computational Linguistics (EACL, Trento, Italy).
- Navigli, R. 2006. Meaningful Clustering of Senses Helps Boost Word Sense Disambiguation Performance (<http://acl.ldc.upenn.edu/P/P06/P06-1014.pdf>). Proc. of the 44th Annual Meeting of the Association for Computational Linguistics joint with the 21st International Conference on Computational Linguistics (COLING-ACL 2006), Sydney, Australia.
- Navigli, R.; G. Crisafulli. Inducing Word Senses to Improve Web Search Result Clustering ([http://www.dsi.uniroma1.it/~navigli/pubs/EMNLP\\_2010\\_Navigli\\_Crisafulli.pdf](http://www.dsi.uniroma1.it/~navigli/pubs/EMNLP_2010_Navigli_Crisafulli.pdf)). Proc. of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP 2010), MIT Stata Center, Massachusetts, USA.
- Navigli, R.; M. Lapata. An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation ([http://www.dsi.uniroma1.it/~navigli/pubs/PAMI\\_2010\\_Navigli\\_Lapata.pdf](http://www.dsi.uniroma1.it/~navigli/pubs/PAMI_2010_Navigli_Lapata.pdf)). IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 32(4), IEEE Press, 2010.
- Navigli, R.; K. Litkowski, O. Hargraves. 2007. SemEval-2007 Task 07: Coarse-Grained English All-Words Task (<http://www.aclweb.org/anthology/S/S07/S07-1006.pdf>). Proc. of Semeval-2007 Workshop (SemEval), in the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007), Prague, Czech Republic.
- Navigli, R.; P. Velardi. 2005. Structural Semantic Interconnections: a Knowledge-Based Approach to Word Sense Disambiguation ([http://www.dsi.uniroma1.it/~navigli/pubs/PAMI\\_2005\\_Navigli\\_Velardi.pdf](http://www.dsi.uniroma1.it/~navigli/pubs/PAMI_2005_Navigli_Velardi.pdf)). IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 27(7).
- Palmer, M.; O. Babko-Malaya and H. T. Dang. 2004. Different sense granularities for different applications (<http://aclweb.org/anthology/W/W04/W04-2807.pdf>). In Proceedings of the 2nd Workshop on Scalable Natural Language Understanding Systems in HLT/NAACL (Boston, MA).
- Ponzetto, S. P.; R. Navigli. Knowledge-rich Word Sense Disambiguation rivaling supervised systems (<http://aclweb.org/anthology-new/P/P10/P10-1154.pdf>). In Proc. of the 48th Annual Meeting of the Association for Computational Linguistics (ACL), 2010.
- Pradhan, S.; E. Loper, D. Dligach, M. Palmer. 2007. SemEval-2007 Task 17: English lexical sample, SRL and all words (<http://www.aclweb.org/anthology/S/S07/S07-1016.pdf>). Proc. of Semeval-2007 Workshop (SEMEVAL), in the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007), Prague, Czech Republic.
- Schütze, H. 1998. Automatic word sense discrimination (<http://www.aclweb.org/anthology/J/J98/J98-1004.pdf>). Computational Linguistics, 24(1): 97–123.
- Snow, R.; S. Prakash, D. Jurafsky, A. Y. Ng. 2007. Learning to Merge Word Senses ([http://www.stanford.edu/~jurafsky/wordsenses\\_emnlp07.pdf](http://www.stanford.edu/~jurafsky/wordsenses_emnlp07.pdf)), Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL).
- Snyder, B.; M. Palmer. 2004. The English all-words task (<http://acl.ldc.upenn.edu/acl2004/senseval/ps/snyder.ps>). In Proc. of the 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (Senseval-3), Barcelona, Spain.
- Weaver, Warren (1949). "Translation" (<http://www.hutchinsweb.me.uk/MTNI-22-1999.pdf>). In Locke, W.N.; Booth, A.D.. *Machine Translation of Languages: Fourteen Essays*. Cambridge, MA: MIT Press.
- Wilks, Y.; B. Slator, L. Guthrie. 1996. *Electric Words: dictionaries, computers and meanings*. Cambridge, MA: MIT Press.
- Yarowsky, D. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora (<http://www.cs.jhu.edu/~yarowsky/pubs/coling92.ps>). In Proc. of the 14th conference on Computational linguistics (COLING), 1992.
- Yarowsky, D. 1995. Unsupervised word sense disambiguation rivaling supervised methods (<http://www.cs.jhu.edu/~yarowsky/acl95.ps>). In Proc. of the 33rd Annual Meeting of the Association for Computational Linguistics.

## External links and suggested reading

- Computational Linguistics Special Issue on Word Sense Disambiguation (<http://www.aclweb.org/anthology-new/J/J98/>) (1998)
- *Evaluation Exercises for Word Sense Disambiguation* (<http://www.senseval.org>) The de-facto standard benchmarks for WSD systems.
- Roberto Navigli. *Word Sense Disambiguation: A Survey* ([http://www.dsi.uniroma1.it/~navigli/pubs/ACM\\_Survey\\_2009\\_Navigli.pdf](http://www.dsi.uniroma1.it/~navigli/pubs/ACM_Survey_2009_Navigli.pdf)), *ACM Computing Surveys*, 41(2), 2009, pp. 1–69. *An up-to-date state of the art of the field.*
- *Word Sense Disambiguation* ([http://www.scholarpedia.org/article/Word\\_sense\\_disambiguation](http://www.scholarpedia.org/article/Word_sense_disambiguation)) as defined in Scholarpedia
- *Word Sense Disambiguation: The State of the Art* (<http://www.up.univ-mrs.fr/~veronis/pdf/1998wsd.pdf>) (PDF) *A comprehensive overview By Prof. Nancy Ide & Jean Véronis (1998).*
- *Word Sense Disambiguation Tutorial* (<http://www.d.umn.edu/~tpederse/WSDTutorial.html>), by Rada Mihalcea and Ted Pedersen (2005).
- *Word Sense Disambiguation: Algorithms and Applications*, edited by Eneko Agirre and Philip Edmonds (2006), Springer. Covers the entire field with chapters contributed by leading researchers. [www.wsdbook.org](http://www.wsdbook.org) site of the book (<http://www.wsdbook.org>)
- Bar-Hillel, Yehoshua. 1964. *Language and Information*. New York: Addison-Wesley.
- Edmonds, Philip & Adam Kilgarriff. 2002. Introduction to the special issue on evaluating word sense disambiguation systems. *Journal of Natural Language Engineering*, 8(4):279-291.
- Edmonds, Philip. 2005. Lexical disambiguation. *The Elsevier Encyclopedia of Language and Linguistics*, 2nd Ed., ed. by Keith Brown, 607-23. Oxford: Elsevier.
- Ide, Nancy & Jean Véronis. 1998. Word sense disambiguation: The state of the art. *Computational Linguistics*, 24(1):1-40.
- Jurafsky, Daniel & James H. Martin. 2000. *Speech and Language Processing*. New Jersey, USA: Prentice Hall.
- Litkowski, K. C. 2005. Computational lexicons and dictionaries. In *Encyclopaedia of Language and Linguistics* (2nd ed.), K. R. Brown, Ed. Elsevier Publishers, Oxford, U.K., 753–761.
- Manning, Christopher D. & Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press. <http://nlp.stanford.edu/fsnlp/>
- Mihalcea, Rada. 2007. Word sense disambiguation. *Encyclopedia of Machine Learning*. Springer-Verlag.
- Resnik, Philip and David Yarowsky. 2000. Distinguishing systems and distinguishing senses: New evaluation methods for word sense disambiguation, *Natural Language Engineering*, 5(2):113-133. <http://www.cs.jhu.edu/~yarowsky/pubs/nle00.ps>
- Yarowsky, David. 2001. Word sense disambiguation. *Handbook of Natural Language Processing*, ed. by Dale et al., 629-654. New York: Marcel Dekker.

# Article Sources and Contributors

**Word-sense disambiguation** *Source:* <http://en.wikipedia.org/w/index.php?oldid=480593378> *Contributors:* .alyn.post., 16@r, Activelink, Aditya Kabir, Agbee, Alansohn, Alexander Wilks, Alvations, Anomalocaris, Aubrey09, Augur, B7T, BD2412, BPalles, Beaniebrad, Bkonrad, Bobo192, Bootecat, Brandon.herygers1, Bunnyfaceperson123, Burschik, Calvin 1998, CharlesGillingham, Chris Roy, Chris the speller, ClamDip, Coaltrain420, Colonies Chris, DARTH SIDIOUS 2, Dale Chock, Dandude13, David spector, Dbmag9, Dekimasu, Demnevanni, Derek Ross, DevOhm, Devoindahouse, Dikisoccer, DimaBorger, Dispenser, Drunken Pirate, Euchiasmus, Ezubaric, Foma84, Funandtrvl, Furrykef, Fvw, Fwedasd, Galoubet, Gbchaosmaster, Gbeverungen, Good Olfactory, Greenleaf, Grumpyyoungman01, Hairy Dude, Harej, IffyNikki, Iridescent, Isaaclovesyou, IvanLanin, J.delanoy, Jafeluv, Jamelan, Jhog1978, Jitse Niesen, Jodi.a.schneider, John of Reading, Juliancolton, Jusdafax, Kwamikagami, Latinatransl, Laurusnobilis, Lordandricus, Lucyin, Mattgirling, Maxim, Michael Hardy, Msbmsb, Nightkhaos, Occidentalst, OffsBlink, Oneiros, Orballo, Pearle, Peter.r.bailey, Petiatil, PierreAbbat, Powerapollo, Ppntori, PrepareToQualify, Psychonaut, Qilisis, Queenmomcat, R'n'B, Radiojon, Rajah, Rama, Robykiwi, Rror, Sandman2007, Santiazpi, Simetrical, Smjg, Soshial, Spencerk, T J Haigh, TXiKi, Tbear1234, The Interior, The Thing That Should Not Be, Thebw, Tiddly Tom, Tohd8BohaithuGh1, Tony1, Tothwolf, Tregoweth, Vacindak, Veinor, Viriditas, Visor, Vuongvina, Wiki0709, WildWildBil, Woohookitty, Wtmitchell, Wwoods, Xeno, Yevrah342, Zinjixmaggir, 147 anonymous edits

## License

---

Creative Commons Attribution-Share Alike 3.0 Unported  
[//creativecommons.org/licenses/by-sa/3.0/](https://creativecommons.org/licenses/by-sa/3.0/)