A Quick Tour of Word Sense Disambiguation, Induction and Related Approaches

Roberto Navigli

Dipartimento di Informatica
Sapienza Università di Roma,
Via Salaria 113, 00198 Roma, Italy
navigli@di.uniroma1.it
http://www.users.di.uniroma1.it/~navigli

Abstract. Word Sense Disambiguation (WSD) and Word Sense Induction (WSI) are two fundamental tasks in Natural Language Processing (NLP), i.e., those of, respectively, automatically assigning meaning to words in context from a predefined sense inventory and discovering senses from text for a given input word. The two tasks have generally been hard to perform with high accuracy. However, today innovations in approach to WSD and WSI are promising to open up many interesting new horizons in NLP and Information Retrieval applications. This paper is a quick tour on how to start doing research in this exciting field and suggests the hottest topics to focus on.

Keywords: computational lexical semantics, Word Sense Disambiguation, Word Sense Induction, text understanding.

1 Introduction

Word Sense Disambiguation (WSD) – the task of computationally determining the correct sense of a word in context – is a core research topic in computational linguistics and natural language processing. The reason for its importance lies in the ambiguity of human language, which is so pervasive that huge numbers of words can be interpreted in multiple ways depending on the context in which they occur. For example, consider the following sentences:

(a) I can hear *bass* sounds;
(b) They like grilled *bass*.

It is evident that the occurrences of the word *bass* in the two sentences denote different meanings: low-frequency tones and a type of fish, respectively. Unfortunately, identifying the specific meaning that a word assumes in context is only apparently simple, especially for a machine. In fact, while most of the time humans do not even think about the ambiguities of language, machines need to process unstructured textual information and transform it into data structures which must then be analyzed in order to determine the underlying meaning. The computational identification of meaning for words in context

is called **Word Sense Disambiguation**. For instance, as a result of disambiguation, sentence (b) above should ideally be sense-tagged as "They like/ENJOY grilled/COOKED bass/FISH". However, in order to perform WSD, a sense inventory must be available for the word of interest (*bass* in the above example). A **sense inventory** is an artifact we are all very familiar with. In fact, it is merely a list of senses of a given word, which is nothing more than what is available in traditional dictionaries – knowledge resources we have been accustomed to using since our earliest schooldays. However, sense inventories need to be updated continuously and, unfortunately, tend to take into account only lexicographic meanings (e.g., bass as a fish), and largely ignore named entities (e.g., Bass as a town in Australia). As a result, automated text understanding is hindered if real-world instances occur in text, e.g.:

(c) I just arrived in Bass, Victoria.

To account for the above-mentioned issues, one can resort to techniques aimed at the automatic discovery of word senses from text, a task called **Word Sense Induction** (WSI).

This paper is a quick tour on starting to do research in the exciting fields of WSD and WSI (see [53] for a complete survey). The paper is organized as follows: in Sections 2 and 3 we illustrate the main topics in WSD and WSI, respectively; in Section 4 we introduce the lexical substitution approach; in Section 5 we discuss other techniques for text understanding; finally we give concluding remarks in Section 6.

2   Word Sense Disambiguation

WSD is believed by many to be the first brick on the road to automated text understanding and a potentially crucial resource for applications such as information retrieval and machine translation. For this reason, much research has been devoted to it over the last few decades. However, several key issues have arisen over the years which need to be considered if a good grasp of the field is to be attained. We overview the techniques, together with the main issues in WSD, in the following subsections.

2.1   Sense Representation

Word senses are the lifeblood of WSD, regardless of the approach we intend to use. A **word sense** is a commonly-accepted meaning of a word. For instance, in the two sentences (a) and (b) from Section 1, the word *bass* can have two senses, i.e., the sound one and the fish one. These two senses are called **homonyms**, in that they are completely different (even etymologically) and are actually two completely different words which just happen to be written in the same way. However, more subtle sense distinctions can be made for a word. For instance, dictionaries can distinguish between *bass* as the lowest part of the musical range and *bass* as the lowest part in polyphonic music. This kind of ambiguity is called

polysemy. Unfortunately, polysemous senses can be created at any level of granularity, thus leading to possibly very fine-grained sense distinctions. This is in fact the case with the most widely-used machine-readable dictionary and computational lexicon of English, namely WordNet [49,19], which has been in use for almost two decades. To cope with the fine granularity problem, researchers have come up with different ideas aimed at creating coarser sense distinctions in WordNet. Two of these efforts led to the organization of WSD evaluation competitions at Semeval-2007 (the reference international semantic evaluation competition), namely:

– Manual approach: the OntoNotes project [27,67] aims at creating sense distinctions by iteratively submitting new partitions of senses for the given word to sense annotators, until the latter achieve a 90% agreement in a text annotation task.
– Automatic approach: another project [52,57] aimed at the creation of coarser-grained senses by clustering semantically similar senses using WSD techniques. The clustering is obtained by automatically mapping WordNet senses to a reference machine-readable dictionary with hierarchical sense distinctions, namely the Oxford Dictionary of English.

Another way to deal with the granularity issue is to use a probabilistic generative model, e.g., based on the well-known noisy channel model [77], which can estimate the distribution of coarse-grained semantic classes, as opposed to fine-grained WordNet senses. Recently, it has also been suggested that one could abandon the adoption of the WordNet sense inventory. Examples include the use of Wikipedia [46], Wiktionary [45], or even newly-conceived machine-readable resources, such as DANTE [42]. Thus, an important open research question on the topic of sense representation is: what sense representation, and inventory, is best (and for which application [63])?

2.2   Techniques

There are three mainstream approaches to WSD, namely:

– **Supervised WSD:** these approaches use machine learning methods to learn a classifier for the target word from labeled training sets, i.e., sets of examples encoded as vectors whose elements represent features, with a special element representing the appropriate sense label (or class). Among supervised methods, memory-based learning and SVM approaches proved to be among the best systems in several international competitions [26,12,47,23,10,79].
– **Knowledge-based WSD:** these methods exploit knowledge resources (such as dictionaries, thesauri, ontologies, etc.) to determine the senses of words in context. They have the advantage of a wider coverage, thanks to the use of large amounts of structured knowledge. The best knowledge-based systems in the literature, such as Degree [56,66] or Personalized PageRank [4], exploit WordNet or other resources (e.g., BabelNet [58]) to build a semantic graph

and exploit the structural properties of the graph (either locally to the input sentence or globally) in order to choose the appropriate senses of words in context.

– **Unsupervised WSD:** these are Word Sense Induction techniques aimed at discovering senses automatically based on unlabeled corpora. They do not exploit any manually sense-tagged corpus to provide a sense choice for a word in context (see Section 3).

The question of which approach is best in general, and in which application, is still very much open. In fact, until recently, the general belief was that supervised WSD performed better than knowledge-based WSD. However, recent results show that, in the presence of enough knowledge (see Section 2.4) or within a domain (see Section 2.6), knowledge-rich systems can beat supervised approaches [66,5,1], while providing at the same time much wider coverage.

2.3   Performance

A well-known issue in Word Sense Disambiguation is performance. As recently as 2004, one of the major arguments against continuing research in the field was that state-of-the-art accuracy was around 65% in an all-words setting, i.e., when all content words must be disambiguated[1]. In other words, on average only 65 out of 100 content words could be automatically associated with an appropriate meaning [72]. It was remarked that the representation of senses was one of the main obstacles to high-performance WSD.

In recent years progress has been made that has led to a considerable boost in disambiguation performance, from about 65% to 82-83% accuracy [10,66] in an all-words setting and when the adopted sense distinctions are less fine-grained (see Section 2.1). As an alternative solution to reducing the granularity of word senses, it has been proposed to calculate the correlation between the graded sense assignments output by a fine-grained WSD system and those obtained as a result of the manual graded judgments produced by human annotators [16].

2.4   Knowledge

Another key factor in WSD, and one which is strongly linked to performance, is **knowledge**. It has been shown that the higher the amount of high-quality knowledge, the higher the performance. Specifically:

– In supervised WSD, performance can be increased considerably when hundreds of training examples are available for each target word [41].
– Knowledge-based WSD – which relies on machine-readable dictionaries or computational lexicons viewed as semantic networks – has been shown to benefit greatly from the addition of relatedness edges [56], i.e., relations of a

---

[1] We use accuracy in the general sense of recall or F1 measure, depending on the task (see [53] for details on the evaluation measures).
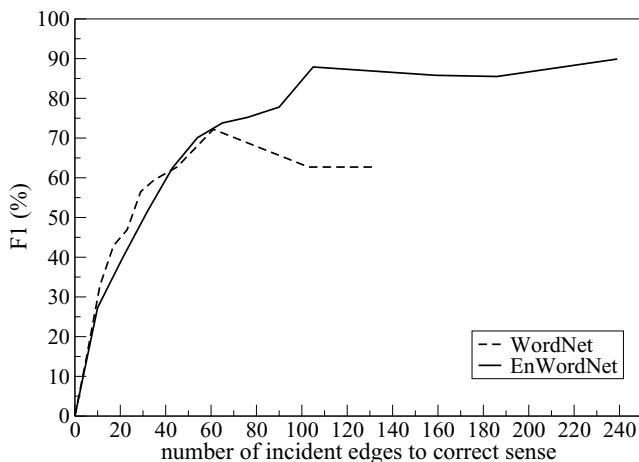
Fig. 1. The more structural knowledge (i.e., semantic relations), the better the performance (from [56])

syntagmatic nature (such as *bus* is related to *driver*), as opposed to paradigmatic relations such as hypernymy (e.g., *bus* is a *public transport*). When more than 100 semantic relations (i.e., edges in the semantic network) are available for each word sense, performance can achieve up to 90% accuracy. We report this result in Figure 1 (taken from [56]), where the performance of a simple knowledge-based WSD, based on the degree of each candidate sense in the semantic network, depends on the richness of the graph itself (EnWordNet is an enriched version of WordNet with about 60,000 additional relatedness edges).

However, these results lead to another question, that of the **knowledge acquisition bottleneck**, i.e., how can we acquire enough knowledge to:

  i) provide wide coverage of the lexicon,
 ii) obtain high performance?

It has been estimated that, at a throughput of one word instance per minute, dozens (if not hundreds) of person-years would be required to provide enough training data [15]. Moreover, this effort would have to be repeated for each language. Possible solutions to the knowledge acquisition bottleneck include the exploitation of bilingual corpora to create large and reliable training sets [78], the automatic sense labeling of corpora [14,8] and the use and integration of large-scale collaborative resources such as Wikipedia [65,66].

2.5   Multilinguality

Currently, most research in WSD is conducted on the English language, because this is the language for which the vast majority of resources are available.

However, recently there has been an increasing interest towards other languages, such as Italian [38], Spanish [40], Chinese [30], Japanese [60], Turkish [61], etc.

More importantly, attention is increasingly being paid to performing WSD across languages, a task referred to as **cross-lingual WSD** [34]. In this setting, an input sentence is provided in a source language (e.g., English) and the WSD system has to provide word senses encoded in a target language (e.g., Italian). The sense inventory is obtained using translations harvested from a parallel corpus, instead of using predefined sense labels. The basic assumption – supported by several studies [21,28,59] – is that the sense distinctions of a word in a source language are determined by the different translations of the word in other languages. This approach has several benefits over the traditional monolingual WSD task: it can be applied to any language of interest, it can easily be integrated into real applications such as machine translation, and it inherently addresses the sense granularity problem. However, it requires a wide-coverage bilingual corpus, a requirement which is not easy to satisfy.

A challenging research direction is performing truly **multilingual WSD**, i.e., returning senses lexicalized in many languages. Recent efforts in this directions include the use of contextual cues or filters from other languages [24,35] and the recently-released multilingual semantic network called BabelNet [58].

2.6   Domain WSD

In everyday life the use of natural language is often confined to specific fields of knowledge. Examples include newspaper sections, blogs and Web sites on specific topics, business documents, etc. Performing a semantic analysis of domain text is thus a crucial task that can boost applications such as Question Answering, Information Retrieval and Information Extraction.

Domain WSD is often performed in a **type-based** fashion, i.e., by assigning a single sense per word, rather than considering each single context the target word occurs in (i.e., **token-based** WSD). Distributionally similar neighbors in raw text can be used as cues to determine the predominant sense of a target word by means of a semantic similarity measure [32,43]. Other distributional methods include the use of a word-category cooccurrence matrix, where categories are coarse senses obtained from an existing thesaurus [51], and synonym-based word occurrence counts [33]. Domain-informed methods have also been proposed which make use of domain labels as cues for disambiguation purposes [22].

Domain-driven approaches have been shown to obtain the best performance among the unsupervised alternatives [73]. Their performance is typically lower than supervised systems. On the other hand, supervised systems need sense-tagged corpora to perform accurate WSD, a requirement that cannot be satisfied for most domains unless a general-purpose corpus is mixed with a smaller domain-specific training set [31].

In the presence of large amounts of structured knowledge for all domains, knowledge-based approaches obtain state-of-the-art performance [66]. A recent approach of this kind [55] with little human intervention has been proposed that proceeds in two steps: first semantic model vectors are learned for many domains,

next, a word context is classified and the selected semantic model vector is used to perform high-quality domain WSD. As a result, the text classification and disambiguation steps are effectively combined.

3   Word Sense Induction

Given the above-mentioned difficulties, Word Sense Induction is an attractive alternative to WSD. In this setting we give up on the traditional notion of a pre-defined word sense and use unsupervised techniques to automatically identify the set of senses denoted by a word. As a result we shift our focus away from how to select the most suitable senses from an inventory towards how to auto-matically discover senses from text. In fact, these methods induce word senses from raw text by clustering word occurrences on the basis of the **distributional hypothesis**, i.e., the idea that a given word – used in a specific sense – tends to co-occur with the same neighbouring words [25].

3.1   Techniques

The main approaches to WSI proposed in the literature are the following:

- **Context clustering:** the underlying hypothesis of this approach is that the distributional profile of words implicitly expresses their semantics (see also Section 5). Each occurrence of a target word in a corpus is represented here as a **context vector**. These context vectors can be either first-order vec-tors, which directly represent the context at hand, or second-order vectors, i.e., the contexts of the target word are similar if their words tend to co-occur together. The vectors are then clustered into groups, each identifying a sense of the target word. A well-known approach to context clustering is the context-group discrimination algorithm [70].
- **Word clustering:** a different approach to the induction of word senses consists of clustering words which are semantically similar and can thus convey a specific meaning. A prototypical example is Lin's algorithm [36], which exploits syntactic dependencies which occur in a corpus to produce sets of words for each discovered sense of a target word. The Clustering By Committee [64] also uses syntactic contexts, but exploits a similarity matrix to encode the similarities between words and relies on the notion of committees to output different senses of the word of interest.
- **Co-occurrence Graphs:** these methods are related to word clustering ap-proaches, but build and analyse graphs of word co-occurrence to identify the set of senses of a given word. Co-occurrences between words can be obtained on the basis of grammatical [76] or collocational relations [75]. However, suc-cessful approaches such as HyperLex [75] – a graph algorithm based on the identification of hubs in co-occurrence graphs – have to cope with the need to tune a large number of parameters [2]. To deal with this issue a graph-based algorithm has recently been proposed which is based on simple graph patterns, namely triangles and squares [54]. The patterns aim at identifying meanings using the local structural properties of the co-occurrence graph.

- **Probabilistic clustering:** another option is to adopt a probabilistic approach, e.g., a Bayesian framework [9], and formalize WSI in a generative model. First, for each ambiguous word a distribution of senses is drawn. Then, context words are generated according to this distribution. Different senses can thus be obtained which have different word distributions.

Finally, we mention a recent successful approach based on latent semantic word spaces [11], which finds latent dimensions of meaning using non-negative matrix factorization, uses these dimensions to distinguish between different senses of a target word, and then proceeds to disambiguate each given instance of that word.

3.2   Evaluation

One key issue in WSI is that of evaluation. WSI is actually a specific instance of the clustering problem, so it is just as hard to evaluate as any clustering algorithm. Unfortunately, evaluating the output of a clustering algorithm is hard even for humans. The main difficulty lies in the fact that there is no single gold-standard clustering on which human annotators can agree. Nonetheless, different approaches have been proposed in the literature to assess the output quality of a WSI algorithm. Three main evaluation techniques have been developed:

- **Unsupervised evaluation:** the clusters of sentences corresponding to the induced senses are evaluated against a gold-standard annotation. The quality of a clustering can be determined by means of measures such as the V-Measure [69], which calculates its homogeneity and completeness, Paired F-Score [39], the RandIndex [68,54], etc.
- **Supervised evaluation:** in this setting the output of WSI is evaluated in a WSD task [3]. To do this, the induced senses are mapped to gold-standard senses for the target word. Each test sentence is then annotated with the best gold standard sense, and precision and recall are used to determine the quality of the resulting WSD.
- **Within an application:** a further way to evaluate the output of WSI is within an application. A paradigmatic example of this kind is the application of WSI to Web search result clustering [54,13], where WSI techniques have been shown to consistently surpass non-semantic state-of-the-art systems.

3.3   Coverage

A second issue with WSI is that of coverage. How do we know the coverage of the induced senses? This question is strongly related to the evaluation issue above. In fact, we might answer that it depends on the application. Alternatively, we might answer that it depends on the results of a supervised evaluation based on the mapping to a gold standard sense inventory. On the other hand, as mentioned in Section 1, WSI has the potential to harvest even more senses than those available in a traditional predefined sense inventory, like the one used in WSD (e.g., WordNet).

4   Lexical Substitution

Another solution to the sense representation problem – closely related to WSI – is to cast WSD as a lexical substitution task [44], i.e., by asking systems to replace a target word in context with a valid alternative substitute which preserves the meaning of the target word as much as possible. For instance, consider the sentence:

(d) They *like* grilled bass.

Suitable replacements for *like* are *love* or *are fond of*, but – in this context – not *wish*. The advantages of this approach to text understanding lie in the lack of a predefined sense inventory and, as opposed to WSI in which senses must be discovered in advance, in the freedom to choose any appropriate substitute dynamically depending on the context. Such a large degree of freedom, however, in a sense, is also a disadvantage in that, similarly to WSI, it makes it difficult to evaluate and compare lexical substitution systems.

The introduction of the lexical substitution task has generated considerable interest within the research community, especially in terms of evaluation measures [29] and multilinguality [48].

5   Other Techniques

Recently other techniques for the unsupervised semantic processing of text have been revamped, the main one being vector space models of semantics [74]. Such models are obtained by exploiting the distributional statistics of words and they are typically based on matrices whose cells encode the relationship between terms and documents, words and contexts, or word pairs and patterns. In contrast to WSI in which the meanings of a word are modeled by a set of induced senses, a word is here represented by a vector whose elements are obtained from grammatical dependencies [36,62], selectional preferences [17] or pattern-based co-occurrences [37]. The resulting matrices, which are usually of very large dimensions, can then be smoothed using truncated Singular Value Decomposition or related techniques for noise and sparsity reduction.

5.1   Compositionality and Meaning

The vectors, acquired by means of distributional semantic techniques, can be compared (e.g., using cosine similarity or geometric measures of distance) or combined by means of vector sum, or multiplication [50], or even tensor products [71]. Solutions to specific kinds of composition, such as adjective-noun, have also been proposed (e.g., by composing matrices with vectors [7]).

Recently it has been reported that state-of-the-art results can be obtained when examplar models are adopted, i.e., when the word meaning is modeled by using relevant occurrences only, rather than merging all the occurrences in a single word vector [18]. The use of more structured vectors, which take into account

syntactic relations, has also been proposed [17]. However, while a general framework has been studied [6], basic questions remain concerning the scalability and applicability to real scenarios of current approaches to distributional semantics and, similarly to WSI, their *in vitro* evaluation, i.e., outside an application.

6   Conclusions

So, how to start in Word Sense Disambiguation and Induction? In this paper we have tried to provide a smattering of the essentials of the field. While a much more detailed account of the field can be found in [53], this paper touches on topics and issues which are very fresh and deserving of the attention of researchers right now.

We can already see the first evidence that text understanding techniques improve the state of the art in fields such as search result clustering [54,13] and lexicography [20]. However, much work is still needed to prove that a proper injection of semantics into real-world applications is always beneficial. The immediate next step is to focus on highly relevant applications, including machine translation (a historical application for WSD, in fact) and semantic search.

References

 1. Agirre, E., de Lacalle, O.L., Soroa, A.: Knowledge-based WSD on specific domains: performing better than generic supervised WSD. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI). pp. 1501–1506. Pasadena, California (2009)
 2. Agirre, E., Martínez, D., de Lacalle, O.L., Soroa, A.: Evaluating and optimizing the parameters of an unsupervised graph-based WSD algorithm. In: Proceedings of TextGraphs '06. pp. 89–96. New York, USA (2006)
 3. Agirre, E., Soroa, A.: SemEval-2007 task 2: Evaluating word sense induction and discrimination systems. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), Prague, Czech Republic. pp. 7–12 (2007)
 4. Agirre, E., Soroa, A.: Personalizing PageRank for Word Sense Disambiguation. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Athens, Greece. pp. 33–41 (2009)
 5. Baldwin, T., Kim, S., Bond, F., Fujita, S., Martinez, D., Tanaka, T.: A reexamination of MRD-based Word Sense Disambiguation. ACM Transactions on Asian Language Information Processing (TALIP) 9, 4:1–4:21 (2010)
 6. Baroni, M., Lenci, A.: Distributional memory: A general framework for corpus-based semantics. Computational Linguistics 36(4), 673–721 (2010)
 7. Baroni, M., Zamparelli, R.: Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1183–1193. MIT Stata Center, Massachusetts, USA (2010)

8. Brody, S., Lapata, M.: Good neighbors make good senses: Exploiting distributional similarity for unsupervised WSD. In: Proceedings of the 22nd International Conference on Computational Linguistics (COLING). pp. 65–72. Manchester, UK (2008)

9. Brody, S., Lapata, M.: Bayesian Word Sense Induction. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL). pp. 103–111. Athens, Greece (2009)

10. Chan, Y.S., Ng, H.T., Zhong, Z.: NUS-PT: Exploiting parallel texts for Word Sense Disambiguation in the English all-words tasks. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), Prague, Czech Republic. pp. 253–256 (2007)

11. de Cruys, T.V., Apidianaki, M.: Latent semantic word sense induction and disambiguation. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT). pp. 1476–1485. Portland, Oregon, USA (2011)

12. Decadt, B., Hoste, V., Daelemans, W., van den Bosch, A.: Gambl, genetic algorithm optimization of memory-based WSD. In: Proceedings of the 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3), Barcelona, Spain. pp. 108–112 (2004)

13. Di Marco, A., Navigli, R.: Clustering web search results with maximum spanning trees. In: Proceedings of 12th International Conference of the Italian Association for Artificial Intelligence (AI*IA). pp. 201–212. Palermo, Italy (2011)

14. Diab, M.: Relieving the data acquisition bottleneck in Word Sense Disambiguation. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL). pp. 303–310. Barcelona, Spain (2004)

15. Edmonds, P.: Designing a task for SENSEVAL-2. Tech. rep., University of Brighton, U.K. (2000)

16. Erk, K., McCarthy, D.: Graded word sense assignment. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 440–449. Singapore (2009)

17. Erk, K., Padó, S.: A structured vector space model for word meaning in context. In: Proceedings of the 2008 Conference on Empirical Methods in Natural Language (EMNLP). pp. 897–906. Edinburgh, UK (2008)

18. Erk, K., Padó, S.: Exemplar-based models for word meaning in context. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL). pp. 92–97. Uppsala, Sweden (2010)

19. Fellbaum, C. (ed.): WordNet: An Electronic Database. MIT Press, Cambridge, MA (1998)

20. Flati, T., Navigli, R.: The CQC Algorithm: Cycling in graphs to semantically enrich and enhance a bilingual dictionary. Journal of Artificial Intelligence Research (JAIR), to appear (2012)

21. Gale, W.A., Church, K., Yarowsky, D.: Using bilingual materials to develop Word Sense Disambiguation methods. In: Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation. pp. 101–112. Montreal, Canada (1992)

22. Gliozzo, A., Strapparava, C., Dagan, I.: Unsupervised and supervised exploitation of semantic domains in lexical disambiguation. Computer Speech and Language 18(3), 275–299 (2004)

23. Grozea, C.: Finding optimal parameter settings for high performance Word Sense Disambiguation. In: Proceedings of the 3rd International Workshop on the Eval-

uation of Systems for the Semantic Analysis of Text (SENSEVAL-3), Barcelona, Spain. pp. 125–128 (2004)

24. Guo, W., Diab, M.T.: Combining orthogonal monolingual and multilingual sources of evidence for all words WSD. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL). pp. 1542–1551. Uppsala, Sweden (2010)

25. Harris, Z.: Distributional structure. Word 10, 146–162 (1954)

26. Hoste, V., Hendrickx, I., Daelemans, W., van den Bosch, A.: Parameter optimization for machine-learning of Word Sense Disambiguation. Natural Language Engineering 8(4), 311–325 (2002)

27. Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., Weischedel, R.: OntoNotes: The 90% solution. In: Companion Volume to the Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, New York, N.Y. pp. 57–60 (2006)

28. Ide, N., Erjavec, T., Tufiş, D.: Sense discrimination with parallel corpora. In: Proceedings of ACL-02 Workshop on WSD: Recent Successes and Future Directions. pp. 54–60. Philadelphia, USA (2002)

29. Jabbari, S., Hepple, M., Guthrie, L.: Evaluation metrics for the lexical substitution task. In: Proceedings of the Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL), Los Angeles, California. pp. 289–292 (2010)

30. Jin, P., Wu, Y., Yu, S.: SemEval-2007 task 05: Multilingual Chinese-English lexical sample. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007). pp. 19–23. Prague, Czech Republic (2007)

31. Khapra, M., Kulkarni, A., Sohoney, S., Bhattacharyya, P.: All words domain adapted WSD: Finding a middle ground between supervision and unsupervision. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL). pp. 1532–1541. Uppsala, Sweden (2010)

32. Koeling, R., McCarthy, D., Carroll, J.: Domain-specific sense distributions and predominant sense acquisition. In: Proceedings of the Human Language Technology Conference and the 2005 Conference on Empirical Methods in Natural Language Processing (EMNLP), Vancouver, B.C., Canada. pp. 419–426 (2005)

33. Lapata, M., Keller, F.: An information retrieval approach to sense ranking. In: Proceedings of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics, Rochester, N.Y. pp. 348–355. Rochester, USA (2007)

34. Lefever, E., Hoste, V.: SemEval-2010 task 3: Cross-lingual Word Sense Disambiguation. In: Proceedings of the 5th International Workshop on Semantic Evaluation. pp. 15–20. Uppsala, Sweden (2010)

35. Lefever, E., Hoste, V., Cock, M.D.: Parasense or how to use parallel corpora for Word Sense Disambiguation. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT). pp. 317–322. Portland, Oregon, USA (2011)

36. Lin, D.: Automatic retrieval and clustering of similar words. In: Proceedings of the 17th International Conference on Computational linguistics (COLING). pp. 768–774. Montreal, Quebec, Canada (1998)

37. Lin, D., Pantel, P.: Dirt – discovery of inference rules from text. In: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (KDD). pp. 323–328. San Francisco, CA, USA (2001)

38. Magnini, B., Giampiccolo, D., Vallin, A.: The italian lexical sample task at Senseval-3. In: Proceedings of Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text. pp. 17–20. Barcelona, Spain (2004)

39. Manandhar, S., Klapaftis, I.P., Dligach, D., Pradhan, S.S.: SemEval-2010 task 14: Word sense induction & disambiguation. In: Proceedings of the 5th International Workshop on Semantic Evaluation. pp. 63–68. Uppsala, Sweden (2010)

40. Màrquez, L., Taulé, M., Martí, A., Artigas, N., García, M., Real, F., Ferrés, D.: Senseval-3: The Spanish lexical sample task. In: Proceedings of the 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3), Barcelona, Spain. pp. 21–24 (2004)

41. Martinez, D.: Supervised Word Sense Disambiguation: Facing Current Challenges, Ph. D. Thesis. University of the Basque Country, Spain (2004)

42. McCarthy, D.: Dante: a new resource for research at the syntax-semantics interface. In: Proceedings of Interdiciplinary Workshop on Verbs. Pisa, Italy (2010)

43. McCarthy, D., Koeling, R., Weeds, J., Carroll, J.: Unsupervised acquisition of predominant word senses. Computational Linguistics 33(4), 553–590 (2007)

44. McCarthy, D., Navigli, R.: The English lexical substitution task. Language Resources and Evaluation 43(2), 139–159 (2009)

45. Meyer, C.M., Gurevych, I.: How web communities analyze human language: Word senses in Wiktionary. In: Proceedings of the Second Web Science Conference. Raleigh, NC, USA (2010)

46. Mihalcea, R., Csomai, A.: Wikify! Linking documents to encyclopedic knowledge. In: Proceedings of the Sixteenth ACM Conference on Information and Knowledge management, Lisbon, Portugal. pp. 233–242 (2007)

47. Mihalcea, R., Faruque, E.: SenseLearner: Minimally supervised Word Sense Disambiguation for all words in open text. In: Proceedings of the 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3), Barcelona, Spain. pp. 155–158. Barcelona, Spain (2004)

48. Mihalcea, R., Sinha, R., McCarthy, D.: Semeval-2010 task 2: Cross-lingual lexical substitution. In: Proceedings of the 5th International Workshop on Semantic Evaluation. pp. 9–14. Uppsala, Sweden (2010)

49. Miller, G.A., Beckwith, R., Fellbaum, C.D., Gross, D., Miller, K.: WordNet: an online lexical database. International Journal of Lexicography 3(4), 235–244 (1990)

50. Mitchell, J., Lapata, M.: Vector-based models of semantic composition. In: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL). pp. 236–244. Columbus, Ohio, USA (2008)

51. Mohammad, S., Hirst, G.: Determining word sense dominance using a thesaurus. In: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Trento, Italy. pp. 121–128 (2006)

52. Navigli, R.: Meaningful clustering of senses helps boost word sense disambiguation performance. In: Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING-ACL), Sydney, Australia. pp. 105–112 (2006)

53. Navigli, R.: Word Sense Disambiguation: A survey. ACM Computing Surveys 41(2), 1–69 (2009)

54. Navigli, R., Crisafulli, G.: Inducing word senses to improve web search result clustering. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 116–126. Boston, USA (2010)

55. Navigli, R., Faralli, S., Soroa, A., de Lacalle, O., Agirre, E.: Two birds with one stone: Learning semantic models for text categorization and Word Sense Disambiguation. In: Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM). pp. 2317–2320. Glasgow, UK (2011)
56. Navigli, R., Lapata, M.: An experimental study on graph connectivity for unsupervised Word Sense Disambiguation. IEEE Transactions on Pattern Anaylsis and Machine Intelligence 32(4), 678–692 (2010)
57. Navigli, R., Litkowski, K.C., Hargraves, O.: SemEval-2007 task 07: Coarse-grained English all-words task. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), Prague, Czech Republic. pp. 30–35 (2007)
58. Navigli, R., Ponzetto, S.P.: BabelNet: Building a very large multilingual semantic network. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL), Uppsala, Sweden. pp. 216–225 (2010)
59. Ng, H.T., Wang, B., Chan, Y.S.: Exploiting parallel texts for Word Sense Disambiguation: an empirical study. In: Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL), Sapporo, Japan. pp. 455–462 (2003)
60. Okumura, M., Shirai, K., Komiya, K., Yokono, H.: SemEval-2010 task: Japanese WSD. In: Proceedings of the 5th International Workshop on Semantic Evaluation (Semeval-2010). pp. 69–74. Uppsala, Sweden (2010)
61. Orhan, Z., Çelik, E., Neslihan, D.: SemEval-2007 task 12: Turkish lexical sample task. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007). pp. 59–63. Prague, Czech Republic (2007)
62. Pado, S., Lapata, M.: Dependency-based construction of semantic space models. Computational Linguistics 33(2), 161–199 (2007)
63. Palmer, M., Babko-Malaya, O., Dang, H.T.: Different sense granularities for different applications. In: Proceedings of 2nd Workshop on Scalable Natural Language Understanding Systems at HLT-NAACL-04. Boston, MA (2004)
64. Pantel, P., Lin, D.: Discovering word senses from text. In: Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining (KDD). pp. 613–619 (2002)
65. Ponzetto, S.P., Navigli, R.: Large-scale taxonomy mapping for restructuring and integrating Wikipedia. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI), Pasadena, California, USA. pp. 2083–2088 (2009)
66. Ponzetto, S.P., Navigli, R.: Knowledge-rich Word Sense Disambiguation rivaling supervised system. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL). pp. 1522–1531. Uppsala, Sweden (2010)
67. Pradhan, S., Loper, E., Dligach, D., Palmer, M.: SemEval-2007 task-17: English lexical sample, SRL and all words. In: Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), Prague, Czech Republic. pp. 87–92 (2007)
68. Rand, W.M.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66(336), 846–850 (1971)
69. Rosenberg, A., Hirschberg, J.: V-measure: A conditional entropy-based external cluster evaluation measure. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). pp. 410–420. Prague, Czech Republic (2007)
70. Schütze, H.: Automatic word sense discrimination. Computational Linguistics 24(1), 97–124 (1998)
71. Smolensky, P.: Tensor product variable binding and the representation of symbolic structures in connectionist systems. Artificial Intelligence 46(1-2), 159–216 (1990)

72. Snyder, B., Palmer, M.: The english all-words task. In: Proceedings of the 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3), Barcelona, Spain. pp. 41–43. Barcelona, Spain (2004)
73. Strapparava, C., Gliozzo, A., Giuliano, C.: Pattern abstraction and term similarity for Word Sense Disambiguation: IRST at Senseval-3. In: Proceedings of the 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3), Barcelona, Spain. pp. 229–234. Barcelona, Spain (2004)
74. Turney, P.D., Pantel, P.: From frequency to meaning: Vector space models of semantics. Journal of Artificial Intelligence Research (JAIR) 37, 141–188 (2010)
75. Véronis, J.: Hyperlex: lexical cartography for Information Retrieval. Computer, Speech and Language 18(3), 223–252 (2004)
76. Widdows, D., Dorow, B.: A graph model for unsupervised lexical acquisition. In: Proceedings of the 19th International Conference on Computational Linguistics (COLING). pp. 1–7. Taipei, Taiwan (2002)
77. Yuret, D., Yatbaz, M.A.: The noisy channel model for unsupervised Word Sense Disambiguation. Computational Linguistics 36(1), 111–127 (2010)
78. Zhong, Z., Ng, H.T.: Word Sense Disambiguation for all words without hard labor. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI). pp. 1616–1622. Pasadena, California (2009)
79. Zhong, Z., Ng, H.T.: It makes sense: A wide-coverage Word Sense Disambiguation system for free text. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL). pp. 78–83. Uppsala, Sweden (2010)