SURVEY ON WORD SENSE DISAMBIGUATION

Dr. D.K. Tayal[1], Riya Goyal[2], B. Indra Thannaya[3], Hunny Gaur[4]

[1] *Professor, Computer Science and Engineering, IGDTUW, Delhi, India*
[2] *PG Student, Computer Science and Engineering, IGDTUW, Delhi, India*
[3] *Assistant Professor, Computer Science and Engineering, IGDTUW, Delhi, India*
[4] *Assistant Professor, Department of Computer Science, St. Stephen's College, Delhi, India*

ABSTRACT

The technique of determining a word's sense based on the sentence's context is known as word sense disambiguation. One of the most considerable and complex areas of natural language processing is this one. The area of word sense disambiguation has been surveyed over the last four decades. For word sense disambiguation, a variety of techniques have been proposed, including knowledge-based, supervised, unsupervised, and semi-supervised techniques. In this study, we explore the word sense disambiguation problem's previous solutions and discuss some potential directions for this branch of natural language processing going forward.

Keyword: - *Word sense disambiguation (WSD), natural language processing (NLP), supervised, unsupervised, knowledge-based (KB), semi-supervised*

---

1. INTRODUCTION

Natural language is ambiguous, which means that a lot of words have several meanings and can be perceived differently according to the context. In order to address this issue, word sense disambiguation is a branch of NLP. Determining the best appropriate meaning for an ambiguous word in the context of a sentence is the process known as WSD [1-5]. Take the term "bank" as an illustration. It may signify a variety of things, including "riverside," "banking institution," "reservoir," etc. With word sense disambiguation, the ambiguous phrase is replaced with the proper one based on the context of the sentence.
Since computers have been employed to address issues involving human language, researchers have been examining the subject of WSD, with some of the early work on machine translation occurring in the 1940s [6].

It is sometimes referred to as an "AI-complete" computational task, meaning that finding a solution assumes that human language understanding and common-sense reasoning be solved as well [7].
WSD is often set up to execute an intermediary role, either independently or when correctly embedded into a framework (thus performing disambiguation implicitly). It is yet to be demonstrated if WSD works well in practical situations. Despite the fact that a variety of articles and ideas have been published on the subject, application-oriented assessment of WSD is still an unexplored field [5].

As a result, we present a thorough summary of the literature in this survey study, outlining the most important contributions that have been made thus far and providing potential growth for the future.

2. APPLICATIONS OF WSD

Some applications of word sense disambiguation are discussed below:

1. **Machine translation***:* The accurate translation of words whose meanings rely on context requires an intuitive knowledge of word sense [8-11].

2. **Information retrieval***:* It is preferable to get only those articles where particular keywords are utilized in the required context when retrieving specific keywords [12]. The most important problem with the IR system is ambiguity resolution in a query [13-17]. The assumption made by contemporary information retrieval systems is that the user will provide sufficient extra context information to obtain appropriate results without the need for specialized word sense disambiguation algorithms.

3. **Speech processing***:* WSD is required for word segmentation and homophone distinction in voice recognition, as well as for the accurate replication of words in text synthesis [18].

4. **Text processing***:* In order to correct complex spelling errors [20], improve text classification and clustering procedures [19], and analyze texts [21,22], word sense disambiguation is utilized.

3. CHALLENGES OF WSD

There are some challenges associated with WSD, that are described below:

1. Different applications have different approaches:

Certain applications may call for completely different approaches. The issue arises while selecting the target words in machine translation. Here, words from the target language serve as the "senses," which frequently correlate to key differences in meaning from the original language [24]. Since it suffices to understand that a term is used with similar meaning in both the query and the returned document, a sensory inventory is not always necessary in information retrieval. The specific sense is irrelevant.

2. Task dependence of a sensory inventory is not possible:

A sense inventory that is independent of the task being performed is incoherent since each occurrence asks for a different analysis of word meaning in task-specific senses [24].

3. Word meaning doesn't divide into separate senses:

In theory, a word's meaning might vary endlessly and depend on its context. It is hard to break down into separate or distinctive sub-meanings [24]. It is not at all relevant, though, whether these same definition variations apply in computational applications, as lexicographers' choices are frequently influenced by certain factors.

4. APPROACHES OF WSD

Knowledge-based, Supervised, Unsupervised, and Semi-Supervised methods are the main approaches of WSD.

1. Knowledge-based Method:

The knowledge-based method relies on information sources like corpora, WorldNet, etc. that include machine-readable dictionaries [23]. They may apply grammatical rules for disambiguation in this method. Utilizing knowledge resources to extract the context-based meanings of words is the goal of the knowledge-based method (dictionary-based approach) WSD. Dictionary, thesaurus, ontologies, collocations, and other knowledge resources are examples. Some methods used in the knowledge-based WSD are as follows:

1.1 LESK Algorithm:

It is the first-word sense disambiguation method that is machine-readable and built on a dictionary. The overlap between the definitions of the terms in a phrase in dictionaries is what this method is based on. This method [25, 26] begins by choosing from the text a precise phrase that contains an unclear term. Then, using an online Dictionary, definitions (glosses) are gathered for the phrase's other meaningful terms as well as the ambiguous word's several meanings. The meaning of every keyword is then matched to the meaning of all other terms. The intended sense of an ambiguous term is one for which the greatest number of overlaps occur.

1.2 Semantic Similarity:

According to some sources, words that are connected have a shared context, hence those meanings that are closest in semantic proximity are picked as the suitable sense [27–29]. This semantic characteristic has the power to provide harmony to the entire article. How closely two words are connected semantically may be determined using a variety of similarity metrics. This method also requires a lot of calculation when there are more than two words.

2. Supervised Method:

To create a model from explicitly sense-annotated data sets, machine learning techniques are used [5]. In order to give each occurrence of a word the correct sense, the classifier—often referred to as a term expert—concerned with a particular term conducts a classification job. A number of occurrences where a target term is explicitly reserve a sense from a reference dictionary's sense inventory often makes up the training set used to build the model. Some methods used in the supervised WSD are as follows:

2.1 Decision List:

If-then-else rules make up a decision list [30–32]. To infer the collection of characteristics for a particular word, decision lists employ training sets. These rules are used to construct a few parameters, such as feature value, sense, and score. The decision list is produced by generating the final order of rules based on the descending scores. When a word is taken into consideration, its occurrence is first computed, and from there, the decision list and score are derived using the word's representation as a feature vector. The sense of a vector is represented by its greatest score.

2.2 Decision Tree:

A tree structure splits the training dataset recursively, classification rules are represented by decision trees [33, 34]. Each branch and internal node of a tree represent an outcome of the test that will be run on a certain feature value. The sense of the word is expressed whenever a terminal node is achieved (if possible).

2.3 Naïve Bayes:

A probabilistic classifier that is based on the Bayes Theorem is the Naïve Bayes classifier [35–37]. The conditional probability of every syllable of a word (Si) and context's characteristics (fj) are two parameters used in this method to categorize text documents. The context-appropriate sense is represented by the formula's greatest value, which is its maximum evaluation.

$$\hat{S} = \underset{S_i \in Sense_D(w)}{\text{argmax}} P(S_i \mid f_1, \dots, f_m) = \underset{S_i \in Senses_D(w)}{\text{argmax}} \frac{P(f_1, \dots, f_m \mid S_i) P(S_i)}{P(f_1, \dots, f_m)}$$

$$= \underset{S_i \in Senses_D(w)}{argmax} P(S_i) \prod_{j=1}^{m} P(f_j \mid S_i)$$

2.4 Support Vector Machine:

The principle of structural risk minimization is applied in support vector machine-based algorithms [38–40]. The objective of this method is to distinguish between true and false instances with the greatest possible margin, where the margin is the space from the plane to the next true or false example. Support vectors are true and false instances that are most closely related to the hyperplane.

3. Unsupervised Method:

These methods for analyzing WSD are founded on concept that terms with the same sense will typically have term with a similar meaning nearby [5]. By grouping word occurrences and placing newly occurred into the generated clusters, which may infer word senses from the input text. In their purest form, they don't employ any machine-readable resources, don't rely on labeled training material, and rely only on human interaction. Some methods used in the unsupervised WSD are as follows:

3.1 Context Clustering:

To find the meaning of a word, context vectors first constructed and then aggregated into clusters using the Context Clustering method [41, 42]. This approach only has word-based dimensions and utilizes vector space as its word space. A word from a corpus will also be designated as a vector in this manner, and the number of times it appears in context will be evaluated. The creation of a co-occurrence matrix and the use of similarity metrics follow. After that, discrimination is carried out using any clustering method.

3.2 Co-occurrence Graph:

This method creates a graph with a vertex (V) for the text's term and an edge (E) if the terms co-occur in a relationship established by syntax inside the same sentence or phrase [43]. The graph and adjacency matrix for the graph are produced for a particular target word. The definition of the term is then determined using the Markov clustering approach. The vertex with the highest relative degree is then chosen as the hub after graph has been subjected to an iterative procedure. The target word's hubs with zero weight are linked, and graph is used to generate a least spanning tree. This spanning tree is employed to clarify the target word's precise meaning.

4. Semi-Supervised Method:

There are times when it is difficult to distinguish between supervised and unsupervised disambiguation. In fact, we may categorize approach as semi-supervised, which develop sense classifiers from annotated data with minimum or limited human supervision.

4. LITERATURE REVIEW

Despite the fact that WSD is a problem that is still not fully solved, a lot of work has been done throughout the years to address this issue. A summary of the development of WSD is provided in this paper.

In 2005, [44] present the disambiguation approach as compactness-based disambiguation which is an unsupervised WSD method. The assumption that adjacent phrases collected from a text source should be semantically related to one another forms the basis of their WSD technique. Consequently, the formulation of a lexical compactness measure for clusters of senses serves as the foundation of their WSD method. They implemented their WSD algorithm to have great accuracy.

In the paper [45], they create latent Dirichlet allocation using the unsupervised probabilistic topic model WORDNET (LDAWN), which incorporates word meaning as a unknown factor. They create a probabilistic posterior inference system that can interpret and understand a corpus of words while also figuring out the many contexts in which each word should be taken into account. They incorporate [46] structure in the topic model using the WORDNET hierarchy.

The paper [47] presented a WSD system for biological papers. All terms in the UMLS Metathesaurus that are ambiguous can be resolved by the unsupervised method. In order to do disambiguation, data from the UMLS Metathesaurus are converted into graph, and optimal meaning for each ambiguous term was chosen using the Personalized PageRank method. Experiments showed that the best results were obtained by combining all the vocabularies in the MRREL data of the Metathesaurus.

Based on word overlap between sense explanation and the target word context, [48] suggested a unique technique for word sense disambiguation. The linguistic expansions from distributional thesaurus that were used in their technique were calculated using the similarity of dependency contexts to a sizable baseline corpus. They discovered that applying their conceptually straightforward extension to two established KB techniques effectively closed the vocabulary gap, yielding quality improvements that were superior to those of advanced knowledge-based models that don’t use sensory frequency knowledge and that were on par with or even exceeded the MFS baseline.

To understand the chronological and structural behavior of the text, they examine WSD using a sequence learning neural net, or LSTM [49]. They utilize the same LSTM in semi-supervised label propagation model to solve the lack of training data in all-words WSD. They show models' best outcomes, particularly with verbs.

With no help from outside sources or custom-made features, [50] provided a BLSTM-based model for WSD that could successfully exploit word order and produce satisfactory outcomes. The experiment shows that the concept was mainly cross-linguistic and suitable for low-resource languages. Additionally, by sharing the majority of the characteristics across words, the system was created to generalize to complete vocabulary WSD.

In [51], with the use of word embeddings, their research developed supervised and unsupervised methods for WSD. On corpora containing manually annotated senses of the word from the plWordnet, they evaluated both techniques' performance. The outcomes demonstrate the unsupervised method's poor performance and point to the advantages of the supervised approach.

[52] shows that to fully utilize both labeled data and lexical knowledge, combine the target word's context and gloss information into a single framework. As a result, they suggest GAS, a gloss-augmented WSD neural network that simultaneously stores the target word's context and glosses. The experimental findings on several English all-words WSD benchmark datasets demonstrate that their model performs better than the most advanced systems.

[53] had suggested inventive methods for implementing BERT, a contextualised word representation that had been trained and successfully incorporates context into its hidden vectors. On several benchmark WSD datasets, their best technique produces accuracy levels that were significantly higher than the best previously disclosed accuracy levels.

[54] shows that lemmatization is not required for English, and the experiments revealed that this was not the case for Russian. Utilizing lemmatized training and testing data seems to produce modest but consistent gains for languages with rich morphology. This suggests that judgments regarding text pre-processing before training ELMo should consider the linguistic traits of the language being considered.

[55] they fine-tune pre-trained language models like BERT on WSD tasks. They discover that when put through a conventional examination, their BERT-based models produce findings that were state-of-the-art. Additionally, they make use of sense definitions to improve the model for uncommon polysemes.

[56] focus their research on enhancing the usage of gloss information in a supervised neural WSD system. They create context-gloss pairings and suggest three WSD models based on BERT. Their pre-trained BERT model was adjusted using SemCor3.0, and outcomes obtained on numerous English all-words standard datasets demonstrate that their method exceeds the previous best systems.

As a relevance ranking challenge based on context-gloss pairings, [57] suggested the gloss selection target for supervised WSD. On five English all-words benchmark datasets, their models optimized for this aim surpass previous non-ensemble systems. Using pre-existing WordNet sample sentences, they also offer a data augmentation strategy for WSD.

[58] provide a bi-encoder model for WSD that has been jointly tuned in order to perform better on all English words. They demonstrate that without losing accuracy on the most basic intuitions, their model's advantages derive from enhanced efficiency on LFS and zero-shot cases.

In the paper [59], the problem was modeled using contextual space and a contextual path concealed in a particular text. The technique was relying on the widely used WordNet Knowledge Base (KB), and it uses Latent Semantic Analysis (LSA) and PageRank to model the contextual space and contextual path, resp. The approach has been tested in experiments, and it has achieved the best results in numerous WSD datasets.

SENSEMBERT, a knowledge-based approach for creating sense embeddings in several languages, was introduced in [60]. SENSEMBERT, which does not rely on annotated data, competes favorably against the top machine learning models on English while achieving great performance on the multilingual WSD tasks.

The data imbalance problem in WSD is addressed by [61], with the introduction of MetricWSD, a few-shot non-parametric method. The model was trained to pass information from frequently occurring words to rare ones using metric space learning and episodic training. MetricWSD performs better than earlier techniques when simply the usual annotated sense supervision was used, and it exhibits notable benefits for low-frequency words and senses.

In a supervised similarity-based WSD architecture, [62] suggests an interactive context exploitation approach from both the term and sense aspects. They also demonstrate how the suggested method has a clear advantage for learning few-shot and zero-shot WSD abilities.

In the paper [63], they provide a unique dictionary-based word vector form which captures the chronological order of the terms. The surrounding terms of the desired term were chosen via graph-based WSD algorithm based on text correlation across the word vector representations. Their WSD model performs much better on numerous benchmark datasets when compared to a reference knowledge-based WSD model. In particular, their WordNet WSD system performed better than other graph-based WSD systems that are already in use.

To correct the training on the unbalanced dataset, [64] suggests the Z-reweighting approach at the word level. The results of the studies demonstrate that the Z-reweighting technique improves performance on the WSD benchmark for standard English all words.

Using topological data analysis, [65] define and evaluate a novel unsupervised method for WSD. The problem was often approached using clustering, which was based on straightforward, low-level spatial information in word embeddings. They demonstrate that their method provides minimal error percentage on word sense induction using the SemCor dataset and a persistent homology barcode methodology.

[66] presented a new performance standard for Polish WSD jobs. They analyzed recent language models using their benchmark data and obtained the latest best performance in Polish WSD.

[67] provide a technique for condensing the sense knowledge without the aid of a glossary. For this, sentence vectors created from dictionary sense definitions were grouped into compressed senses. Through this approach, it was shown that performance was significantly better than with the uncompressed sense model and was on par with the thesaurus-based model.

On the basis of the review that we have done, we conclude that the supervised approach of WSD is used most and also gives better results as compared with different approaches.

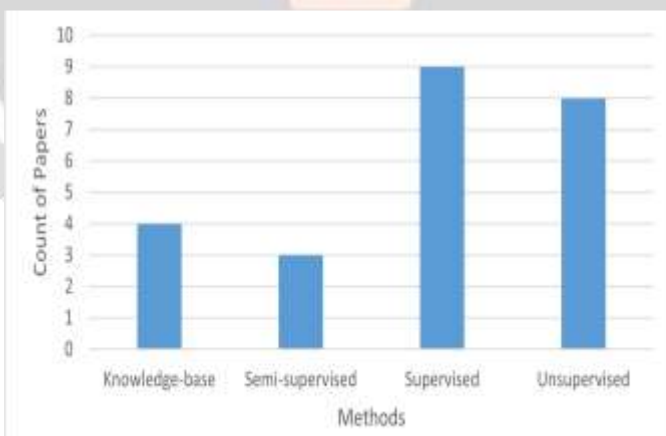

Chart-1 Comparison of WSD Approaches

4. CONCLUSION

In this study, we looked at the WSD problem, which is one of the most critical open NLP tasks. Many issues arise in WSD since it relies on information gained from many sources, some of which have been touched upon. Lexical complexity is another issue that WSD addresses. In WSD, the above discussed methodologies are all employed. Because training data is completely dependent on a given domain, the supervised technique outperforms all other approaches.

7. REFERENCES

[1]. Ide, N., & Véronis, J. (1998). Word sense disambiguation: The state of the art. Computational linguistics, 24(1), 1-40.

[2]. Florian, R., Cucerzan, S., Schafer, C., & Yarowsky, D. (2002). Combining classifiers for word sense disambiguation. Natural Language Engineering, 8(4), 327-341.

[3]. Nameh, M., Fakhrahmad, S. M., & Jahromi, M. Z. (2011, July). A new approach to word sense disambiguation based on context similarity. In Proceedings of the World Congress on Engineering (Vol. 1, pp. 6-8).

[4]. Wang, X., & Matsumoto, Y. (2003, October). Chinese word sense disambiguation by combining pseudo training data. In International Conference on Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003 (pp. 138-143). IEEE.

[5]. Navigli, R. (2009). Word Sense Disambiguation: a Survey. ACM Computing Surveys. Vol. 41 (2).

[6] Hutchins, J. (1999). Warren Weaver memorandum: 50th anniversary of machine translation. MT News International, 8(1), 5-6.

[7]. Mallery, J. C. (1988). Thinking about foreign policy: Finding an appropriate role for artificially intelligent computers. In Master's thesis, MIT Political Science Department.

[8]. Carpuat, M., & Wu, D. (2005, June). Word sense disambiguation vs. statistical machine translation. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL’05) (pp. 387-394).

[9]. http://www.ling.gu.se/~sl/Undervisning/StatMet 11/wsd-mt.pdf date: 14/05/2015

[10]. http://nlp.cs.nyu.edu/sk-symposium/note/P-28.pdf date: 14/05/2015

[11]. Chan, Y. S., Ng, H. T., & Chiang, D. (2007, June). Word sense disambiguation improves statistical machine translation. In Proceedings of the 45th annual meeting of the association of computational linguistics (pp. 33-40).

[12]. Salton, G. (1968). Automatic information organization and retrieval.

[13]. Sanderson, M. (1994). Word sense disambiguation and information retrieval. In SIGIR’94 (pp. 142-151). Springer, London..

[14]. Stokoe, C., Oakes, M. P., & Tait, J. (2003, July). Word sense disambiguation in information retrieval revisited. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval (pp. 159-166).

[15]. http://www.aclweb.org/anthology/P121029 date: 14/05/2015

[16]. https://www.comp.nus.edu.sg/~nght/pubs/esair 11.pdf date: 14/05/2015

[17]. http://cui.unige.ch/isi/reports/2008/CLEF2008-LNCS.pdf date: 14/05/2015

[18]. Seneff, S. (1992). TINA: A natural language system for spoken language applications. Computational linguistics, 18(1), 61-86.

[19]. Grineva, M., Grinev, M., Turdakov, D., Velikhov, P., & Boldakov, A. (2008). Harnessing Wikipedia for smart tags clustering. In KASW: International Workshop on «Knowledge Acquisition from the Social Web».—2008.

[20]. Yarowsky, D. (1994). Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French. arXiv preprint cmp-lg/9406034.

[21]. Grineva, M. P., Grinev, M., & Lizorkin, D. (2009). Effective Extraction of Thematically Grouped Key Terms From Text. In AAAI Spring Symposium: Social Semantic Web: Where Web 2.0 Meets Web 3.0 (pp. 39-44).

[22]. Grineva, M., Grinev, M., & Lizorkin, D. (2009, April). Extracting key terms from noisy and multitheme documents. In Proceedings of the 18th international conference on World wide web (pp. 661-670).

[23]. Krovetz, R., & Croft, W. B. (1989, May). Word sense disambiguation using machine-readable dictionaries. In Proceedings of the 12th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 127-136).

[24]. Edmonds, P., & Agirre, E. (2008). Word sense disambiguation. Scholarpedia, 3(7), 4358.

[25]. Banerjee, S., & Pedersen, T. (2002, February). An adapted Lesk algorithm for word sense disambiguation using WordNet. In International conference on intelligent text processing and computational linguistics (pp. 136-145). Springer, Berlin, Heidelberg.
[26]. Lesk, M. (1986, June). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In Proceedings of the 5th annual international conference on Systems documentation (pp. 24-26).
[27]. http://www.dlsi.ua.es/projectes/srim/publicacio nes/CICling-2002.pdf date: 14/05/2015
[28]. Mittal, K., & Jain, A. (2015). WORD SENSE DISAMBIGUATION METHOD USING SEMANTIC SIMILARITY MEASURES AND OWA OPERATOR. ICTACT Journal on Soft Computing, 5(2).
[29]. http://www.d.umn.edu/~tpederse/Pubs/cicling2 003-3.pdf date: 14/05/2015
[30]. Yarowsky, D. (2000). Hierarchical decision lists for word sense disambiguation. Computers and the Humanities, 34(1), 179-186.
[31]. Parameswarappa, S., Narayana, V. N., & Yarowsky, D. (2013). Kannada word sense disambiguation using decision list. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 2(3), 272-278.
[32]. Palanati, D. P., & Kolikipogu, R. (2013). Decision list algorithm for word sense disambiguation for TELUGU natural language processing. Int. J. Electron. Commun. Comput. Eng, 4(6), 176-180.
[33]. Singh, R. L., Ghosh, K., Nongmeikapam, K., & Bandyopadhyay, S. (2014). A decision tree based word sense disambiguation system in Manipuri language. Advanced Computing, 5(4), 17.
[34]. Pedersen, T. (2001). A decision tree of bigrams is an accurate predictor of word sense. arXiv preprint cs/0103026.
[35]. Le, C. A., & Shimazu, A. (2004, December). High WSD accuracy using Naive Bayesian classifier with rich features. In Proceedings of the 18th Pacific Asia conference on language, information and computation (pp. 105-114).
[36]. http://www.cs.upc.edu/~escudero/wsd/00-ecai.pdf date: 14/05/2015
[37]. Aung, N. T. T., Soe, K. M., & Thein, N. L. (2011). A word sense disambiguation system using Naïve Bayesian algorithm for Myanmar language. International Journal of Scientific & Engineering Research, 2(9), 1-7.
[38]. https://www.comp.nus.edu.sg/~nght/pubs/se3.p df date: 14/05/2015
[39]. Buscaldi, D., Rosso, P., Pla, F., Segarra, E., & Arnal, E. S. (2006, February). Verb sense disambiguation using support vector machines: Impact of WordNet-extracted features. In International Conference on Intelligent Text Processing and Computational Linguistics (pp. 192-195). Springer, Berlin, Heidelberg.
[40]. Joshi, M., Pedersen, T., & Maclin, R. (2005, December). A Comparative Study of Support Vector Machines Applied to the Supervised Word Sense Disambiguation Problem in the Medical Domain. In IICAI (pp. 3449-3468).
[41]. http://nlp.cs.rpi.edu/paper/wsd.pdf date: 14/05/2015
[42]. Niu, C., Li, W., Srihari, R. K., Li, H., & Crist, L. (2004, July). Context clustering for Word Sense Disambiguation based on modeling pairwise context similarities. In SENSEVAL@ ACL.
[43]. Widdows, D., & Dorow, B. (2002). A graph model for unsupervised lexical acquisition. In COLING 2002: The 19th International Conference on Computational Linguistics.
[44]. Mavroeidis, D., Tsatsaronis, G., Vazirgiannis, M., Theobald, M., & Weikum, G. (2005, October). Word sense disambiguation for exploiting hierarchical thesauri in text classification. In European Conference on Principles of Data Mining and Knowledge Discovery (pp. 181-192). Springer, Berlin, Heidelberg.
[45]. Boyd-Graber, J., Blei, D., & Zhu, X. (2007, June). A topic model for word sense disambiguation. In Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL) (pp. 1024-1033).
[46]. Abney, S., & Light, M. (1999). Hiding a semantic hierarchy in a Markov model. In Unsupervised Learning in Natural Language Processing.
[47]. Agirre, E., Soroa, A., & Stevenson, M. (2010). Graph-based word sense disambiguation of biomedical documents. Bioinformatics, 26(22), 2889-2896.
[48]. Miller, T., Biemann, C., Zesch, T., & Gurevych, I. (2012, December). Using distributional similarity for lexical expansion in knowledge-based word sense disambiguation. In Proceedings of COLING 2012 (pp. 1781-1796).
[49]. Yuan, D., Richardson, J., Doherty, R., Evans, C., & Altendorf, E. (2016). Semi-supervised word sense disambiguation with neural models. arXiv preprint arXiv:1603.07012.
[50]. Kågebäck, M., & Salomonsson, H. (2016). Word sense disambiguation using a bidirectional lstm. arXiv preprint arXiv:1606.03568.

[51]. Wawer, A., & Mykowiecka, A. (2017, April). Supervised and unsupervised word sense disambiguation on word embedding vectors of unambigous synonyms. In Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications (pp. 120-125).
[52]. Luo, F., Liu, T., Xia, Q., Chang, B., & Sui, Z. (2018). Incorporating glosses into neural word sense disambiguation. arXiv preprint arXiv:1805.08028.
[53]. Hadiwinoto, C., Ng, H. T., & Gan, W. C. (2019). Improved word sense disambiguation using pre-trained contextualized word representations. arXiv preprint arXiv:1910.00194.
[54]. Kutuzov, A., & Kuzmenko, E. (2019). To lemmatize or not to lemmatize: how word normalisation affects ELMo performance in word sense disambiguation. arXiv preprint arXiv:1909.03135.
[55]. Du, J., Qi, F., & Sun, M. (2019). Using BERT for word sense disambiguation. arXiv preprint arXiv:1909.08358.
[56]. Huang, L., Sun, C., Qiu, X., & Huang, X. (2019). GlossBERT: BERT for word sense disambiguation with gloss knowledge. arXiv preprint arXiv:1908.07245.
[57]. Yap, B. P., Koh, A., & Chng, E. S. (2020). Adapting BERT for word sense disambiguation with gloss selection objective and example sentences. arXiv preprint arXiv:2009.11795.
[58]. Blevins, T., & Zettlemoyer, L. (2020). Moving down the long tail of word sense disambiguation with gloss-informed biencoders. arXiv preprint arXiv:2005.02590.
[59]. Wang, Y., Wang, M., & Fujita, H. (2020). Word sense disambiguation: A comprehensive knowledge exploitation framework. Knowledge-Based Systems, 190, 105030.
[60]. Scarlini, B., Pasini, T., & Navigli, R. (2020, April). Sensembert: Context-enhanced sense embeddings for multilingual word sense disambiguation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 05, pp. 8758-8765).
[61]. Chen, H., Xia, M., & Chen, D. (2021). Non-parametric few-shot learning for word sense disambiguation. arXiv preprint arXiv:2104.12677.
[62]. Wang, M., & Wang, Y. (2021, August). Word sense disambiguation: Towards interactive context exploitation from both word and sense perspectives. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 5218-5229).
[63]. Kwon, S., Oh, D., & Ko, Y. (2021). Word sense disambiguation based on context selection using knowledge-based word similarity. Information Processing & Management, 58(4), 102551.
[64]. Su, Y., Zhang, H., Song, Y., & Zhang, T. (2022, May). Rare and Zero-shot Word Sense Disambiguation using Z-Reweighting. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 4713-4723).
[65]. Rawson, M., Dooley, S., Bharadwaj, M., & Choudhary, R. (2022). Topological Data Analysis for Word Sense Disambiguation. arXiv preprint arXiv:2203.00565.
[66]. Janz, A., Dziob, A., Oleksy, M., & Baran, J. (2022). A Unified Sense Inventory for Word Sense Disambiguation in Polish. In International Conference on Computational Science (pp. 682-689). Springer, Cham.
[67]. Park, J. Y., Shin, H. J., & Lee, J. S. (2022). Word Sense Disambiguation Using Clustered Sense Labels. Applied Sciences, 12(4), 1857.