**Word Sense Disambiguation (WSD): A Survey in General and for Arabic Language in Specific**

**Dr. Ahmed H. Aliwy Dr. KadhimAljanabi**

**Dept. of Computer Science Dept. of Computer Science**

**Faculty of CS and Mathematics Faculty of CS and Mathematics**

**University of Kufa University of Kufa**

**Abstract**

Word Sense Disambiguation (WSD) is the task of selecting the right sense for the word in the context which is part of Natural Language Processing (NLP) field in the computer science. Many words, in any natural languages, have more than one sense according to the context of the sentence. Selecting the correct sense for this sentence is an open problem for all natural languages and especially for Arabic language. In other words “WSD are described as an AI-complete problem that is, by analogy to NP-completeness in complexity theory”.

WSD is a very important intermediate task because many of NLP applications are improved using WSD. These tasks are like information retrieval (IR), machine translation (MT), and question answering (QA). We can use, in IR, bag of concepts instead of bag of words which, surly, improve retrieval of information. Also, in MT, we can’t get the correct translation without selecting the correct sense for the word in the sentence.

The work in this paper presents a survey for WSD problems including listing the levels of the intermediate tasks for NLP with their definitions in order to differentiate WSD task from them. Then the survey focuses on the approaches of WSD in general, their classifications and their evaluations. Also,the work presents asurvey of the approaches used for Arabic language. The cautions of limitation of works in WSD and the problems facing the researchers in Arabic language are recorded.

**Introduction:**

Natural Language Processing (NLP) is considered as the cruicial field in AI area because of the huge development and the wide applications in this field such asMachine Translation(MT), Man-Machine Interaction (using speech),Question Answering and Information Retrieval (IR). Almost all these applications have problems which must be resolved.These problems became as subfield from NLP like tokenization,stemming، part of speech (POS) tagging, parsing, Word Sense Disambiguation (WSD) and others. They can be considered as preprocessing or intermediate task for a NLP application.

WSD is the selection of the correct sense for the word in the context from list of senses for this word.It is one of the important intermediate fields. For example, we can’t get correct MT without extracting the correct sense for the word in the context.The importance of WSD arises from the fact that the human language is ambiguous where many words have more than one sense (meaning) according to the context.

Before starting with WSD, we should have **inventory** of senses as a list of all senses for each ambiguous word. For example, the word “bank” can be a “financial organization” or a “river side” or other sense. The inventory can be independent or extracted from the corpus.

**Related Work**

Pal and Saha‎1) have gone through a survey regarding the different approaches adopted in different research works, the State of the Art in the performance in this domain, works in different Indian languages and finally a survey in Bengali language.

Navigli‎2)discussed the assessment of WSD systems in the context of the Senseval/Semeval campaigns, aiming at the objective evaluation of systems participating in several differentdisambiguation tasks. Finally, applications and open problems are discussed.

Bakx‎3),in his work, studies the possible application of the algorithms and techniques of the Machine Learning field in order to handle the WSD task.

Navigli‎4) shows, in his paper, a quick tour on how to start doing research in this exciting field of WSD and suggests the hottest topics to focus on.

Zhou and Han‎5) summarize the various knowledge sources used for WSD and classifies existing WSD algorithms according to their techniques. They discussed the used knowledge sourcesand suitable applications for each class of WSD algorithms.

None of these works discussed using WSD in Arabic languages or surveyed th used techniques for Arabic language. Some of these works did not show the accuracy of each approach used by the researchers.

**Common Intermediate Task for NLP**

The common intermediate tasks for NLP application are segmentation, tokenization, POS tagging, Parsing, stemming and WSD. The following sections explain these terms briefly.

**Segmentation and Tokenization:**

Segmentation and tokenization are interchangeable in most cases. Segmentation in NLP refers to segmenting the text into paragraphs and sentences. Tokenization is the process of breaking a stream of a text into words, phrases, symbols, or other meaningful elements called tokens.This task seems to be easy task but it is really not.

The difficulties arise in Arabic languages result from the nature of the language. One Arabic word can be coded many English words, for example the word “أتسألني” may represent four words “areyou asking me”. See (Aliwy-2012)‎6)for more details.

**POS Tagging:**

Many words in any language can be having more than one Part Of Speech (POS). Selecting the correct POS for these words according to the context is called POS tagging. The list of parts of speech in any language is called Tagset where each part of speech has one symbol called tag. Figure 1 shows POS tagging where the tagset is {**N**oun, **V**erb, **ADJ**ective, **P**article, **CONJ**unction}

****

Figure 1.Three different POSs for the same word

Arabic POS tagging is non-trivial task. There are many works in this fields for Arabic language but the comparison among them is very difficult because each one has its owntagset and annotated corpus see (Aliwy-2013)‎7)for more details.

**Parsing:**

It is the process of building syntactic tree for the sentence, phrases or chunks. Each parse tree can give different syntactical meaning for the sentence. Parsing process is done using algorithm and Grammar like Context Free Grammar (CFG). Figure 2show two parse trees for “يطمئن المتظاهرين في بغداد” which has two syntactic meaning where the first meaning is that the speaker is in Baghdad but not the demonstrators, the second meaning demonstratorsare in Baghdad. Statistical parser can be used for selecting the best parse tree for a given sentence.



Figure 2 Different Parse Trees for the same sentence.

**WSD**

WSD is selecting the correct sense for the word in the context from a list of senses for this word‎8)‎3). Word Sense Disambiguation (WSD) is a core research topic in computational linguistics and natural language processing‎4).

We can understand that it is different from POS tagging or parsing. Annotating the word as verb or noun is POS tagging but selecting the correct sense for the Noun is WSD. Many researchers, especially for Arabic, do not differentiate between them. We can understand the situation by an example like the word “سيف” tagged as Noun then what is the sense of this noun? It may be a tool “Sword” or a human-name “Saif”.

**Techniques used for WSD:**

There are many classifications for the approaches used in WSD. A WSD approach can be one of **corpus-based approaches** or **knowledge-based** ones. We select the classification of **Knowledge Based**, **Machine Learning Based**and **Hybrid** Approaches:

**Knowledge Based Approaches:**:

**Selectional Preferences** (or restrictions)

**Overlap Based** Approaches

**Machine Learning Based Approaches:**

**Supervised Approaches**

**Semi-supervised Algorithms**

**Unsupervised Algorithms**

**Hybrid Approaches**

**Methods Combination**

**Knowledge Based Approaches:**

**Knowledge Based Approaches** is the first used types of approaches for WSD. Selectional preference, LESK algorithm and WALKER’Salgorithm are examples for this class of approaches. Classically can be divided to two type classes Selectional Preferences and Overlap-based approaches.

**Selectional Preferences:**

Pal&Saha‎1)defines it as “Selectional preferences is finding information of the likely relations of word types, and denote common sense using the knowledge source. For example, Modeling-dress, Walk-shoes are the words with semantic relationship”. The senses of the word can be known from the counts of these pairs of words.

Ye‎9)shows that“Selectional preferences (p) are verb-sense specific. It is possible for a particular sense of a verb to have more than one selectional preference”. The verb eat has senses as **take a food**, **use up resources** and so on. The first sense requires, as selectional preference, subject of animate type and object as food. But the second sense not required a restriction for the subject.Resnik‎10)recorded 44.3 as accuracy for selectional preference.

**Overlap Based Approaches**

Overlapbased approachesdepend on checking the overlap between senses features (sense definitions or examples etc) for the ambiguous word and words features in the context. The maximum overlap sense will be selected.**LESK’S algorithm** and **WALKER’S algorithm**are examples for these approaches.

Lesk (1986) ‎11)is one of Machine-readable-Dictionary(MRD) based method‎11). It estimates the correct sense for the word by checking the maximum overlap between the definition and the context. Practically cannot be applied therefore simplified Lesk is used. There are many test for Lesk algorithm for many languages like English and Arabic. Banerjee‎12)recorded an accuracy of 31.7% for SENSEVAL-2 using adaptive Lesk.

**WALKER’S Algorithm** is a Thesaurus Based approach consists of two distinct steps:(i) the thesaurus category for each sense of the word is found.(ii) Calculating the score of each sense by using the context words.Kalita‎13)used 40 sentences as a test where the precision and recall were 86.66% and 61.09% respectively.

**WSD USING CONCEPTUAL DENSITY** Select a sense based on the relatedness of that word-sense to the context. The conceptual distance is founded by using WordNet.Buscaldi‎14)used GeoSemCor corpus as data set and get F-measure of 0.850

**Supervised Approaches:**

**Naïve Bayes:**

Naive Bayes has been used in most classical setting, in which, assuming independence of features, it classifies a new example x = (x1,...،xm) by assigning the sense k that maximises the conditional probability of the sense given the observed sequence of feature of that example‎3). That is:



Where *P(k)* and *P(xi|k)* are calculated from the corpus.

A test for naïve Bayes was done by Escudero‎15) using corpus contains about 192,800 examples of 191 ambiguous words (nouns and verbs). The average of accuracy was 66.95% for two sets of test.

**Exemplar Based WSD:**

Exemplar based learning ‎16)depend on estimating the distance(similarity) between two examples. The two examples are similar if the distance is small. Mooney used Hamming distance in KNearestNeighbor algorithm.Escudero‎15)gets the average of accuracy of 68 % for two sets of test.

**Decision list**

‎3)“A decision list consists of a set of ordered rules of the form (feature-value, sense, weight)”. Features and weights are estimated from training data.

The weight can be estimated by:



Escuderorestrict that “The list of all rules is sorted by decreasing values of this weight; when testing new examples, the decision list is checked, and the feature with highest weight that is matching the test example selects the winning word sense”‎3). Yarowsky‎17)reported 73.40%accuracy for 36 trainable words.

**SVM:**

Support Vector Machines (SVM) is one of the well-known machine learning algorithms. It was used in many arias in computer science and engineering problems. SVM has excellent experiment in NLP tasks as POS tagging, text categorization and many other tasks.

Palmer‎18)used SVM for WSD. She chose 21 ambiguous words. The number of examples per word was 16823 where the accuracy was 76.06%.

**HMM:**

HMM is the most used probabilistic model used in NLP applications and tasks as POS tagging, text classification, tokenization and many other tasks. The idea behind using HMM for WSD is by selecting the best sequence of senses for a given sentence‎19):



Where P(si|si-1) and P(wi|si) is the transition and emission probabilities respectively. Molino‎19)used senseval-2 of 2473 words of taggedsense. The precision was 52.30 and 58.20 for unigram and bigram with POS tagging supported.

**Other Approaches:**

There are many other approaches used for WSD as Neural networks‎20), Memory-Based Learning‎21), Genetic Algorithm‎22), Fuzzy set theory‎23), etc.

**Unsupervised WSD Approaches**

Unsupervised WSD approaches do not use an annotated data or inventories. Most of these approaches depend on the fact that the words can be similar if they appear in the same context ‎24). One of the well-known unsupervised WSD approach is**clustering based WSD** ‎24)‎25).**it**can be **Context clustering**‎24)where “thecontext will be grouped into clusters to identify the meaning of theword” or**word clustering**which cluster the words semantically identical. There are other unsupervised WSD methods as**Co-occurrenceGraph**and **Spanning tree based‎1)**.

Heng used **clustering based WSD**for 6 ambiguous words with average accuracy 85.23% ‎24).

**Semi-Supervised WSD Approaches**

It is called minimally supervised and weakly supervisedapproaches. Semi-supervised WSD methods need to small amount of annotated data. **Bootstrap** and **Monosemous Relatives** are examples for this kind of methods. **Bootstrapping**‎2)usually starts from fewannotated data A, a large corpus of unannotated data U, and a set of one or more basicclassifiers. As a result of iterative applications of a bootstrapping algorithm, the annotated corpus A grows increasingly and the untagged data set U shrinks until somethreshold is reached for the remaining examples in U.Yarowsky‎26) used **Bootstrapping**with accuracy ranged from**97% to 92%.** Leacock‎27) used **Monosemous Relatives**with precision of 96%. And Recall 100%.

**Hybrid approaches:**

Hybrid approach is combining Knowledge-based and machine-learning-based features but not combination of the independent classifiers. **SenseLearner‎28)**, **Iterative Approach‎29)**and**Structural Semantic Interconnections (SSI) ‎30)**are examples for hybrid methods. For example,**Sense-learner**uses some tagged data to build a semantic language model for words seen in the training corpus. It, also, uses WordNet to derive semantic generalizations for words which are not observed in the corpus.**Mihalcea**and **Faruque‎28)**used **SenseLearner**and obtained an average accuracy of 64.6% using WordNet as knowledge-based and SENSEVAL-3 task. Mihalceaand Moldovan ‎29)used**Iterative Approach**with the 6 randomly selected files from SemCore as data set. They get an accuracy of **92.2% of 55%** ofthe disambiguated nouns and verbs.Navigli and Velardi‎30)used **Structural Semantic Interconnections** (SSI)approach on Senseval-3 and getting (precision, Recall) results (86.0%, 44.7%), (69.4%, 13.5%) and (78.6%, 26.2%) for Noun Verbs and adjectives respectively.

**Methods Combination**

We can combinemore than one classifierof WSD in many strategies. One of the simplest strategy is by using voting:

Voting‎1): number of classifier running independently and the final result is done by doing the voting between them. There are many experiments was done for voting but the most famous can be one of Un-weighted-Voting, weighted-voting and Probability Mixture.

Linear combination (AdaBoost)‎1): combining weak classifiers linearly and the disambiguation is done in levels.

Master-slaves [my thesis&Bushra Paper]: using one classifier as master and the other classifier as slaves which change the probabilities of the master according to its results.

**Arabic WSD**

Very little studies were done for Arabic WSD results from the nature of Arabic language and the lack of annotated corpora. Little Arabic researchers, as we see from our survey, did not differentiate between WSD and POS tagging. We will discuss here the used techniques for Arabic WSD and the problems facing the researcher in this field.

**The used Techniques**

All the used techniques for Arabic WSD are taken from these applied to English language. As we will see, the accuracy of the used methods can not be compared with these used for English language because each own data and the ambiguous words.

Elmougy‎31), uses Naïve Bayes algorithm and achieves a rate of precision of 73%. The used data is small private samples.

Eid and Al-said ‎32) used Rocchio Classifier, Most Frequent Sense, Naïve Bayesian classifier and support vector machine using Arabic lexical samples. They get accuracies of 88.3%,57.5%, 86.14% and 82%for the used methodsrespectively where the set was done for five Nouns.

Merhben and Zouaghi‎33) used Naïve Bayes algorithm, the decision lists and the exemplar based as case study. They get accuracies of 48.23%, 43.86 and 52.02 for these methods respectively. The same researchers ‎34) have other test using Voting where the accuracy was 83.0% and the data set was 42,316 sentences as test with 127 ambiguous words.

Pento and Rosso‎35)used unsupervised clustering approach. The data set was SemEval containing 509 ambiguous words.

Zouaghi and Merhbene‎36) used variants of LESk algorithm. The best accuracy was 78%.

**Problems in Arabic WSD**

There many problems face Arabic WSD results from the nature of Arabic language. These problems limited the Arabic researcher for working on Arabic language. Some of these problems are:

Lake of diacritics: so the same word can have many different senses and many POSs.

Arabic word order is tricky and not restricted.

“a single word in Arabic may have dozens of meanings, some of them are closely related, and interestingly there are some words thatcan mean something and its opposite” ‎37).

Many Metaphors “التعبير المجازي”: where the word is used in other meaning not it’s meaning set.

Lake of resources for Arabic language.

Arabic word can be sentence in other language.

Arabic language has many levels of complexity for tokenization, segmentation, lemmatization, POS tagging which are important for WSD.

Most of these problems are solved by separate works around the world. But the big one is by constructing a large annotated corpus for Arabic language. Tell now the researcher work to build Arabic WordNet which equivalent to English WordNet but not the same size.

**Discussion**

The work in this paper starts with listing the levels of the intermediate tasks for NLP, surveying the approaches of WSD in general, survey the approaches used for Arabic language and the problems facing the researchers in Arabic.

The comparison among the WSD works is very difficult because each researcher used different data set. Little researchers try to apply little different approaches to the same data set. Some researchers reported that one approaches can give better results for one data set but the same approach give the worst results for other data set.

The survey recordedthat, some Arabic researcher did not differentiate between WSD and other NLP task like POS tagging. For example the word is Noun or Verb, or verb is active or passive are POS tagging problems but they take it as WSD problem.

Most of the Combination of independent classifiers studied for WSD were recorded that raisingaccuracy.

Finally we recommended that WSD is an open problem in all natural languages and it is AI-Complete problem. For Arabic language it can be more difficult than the other languages because of the nature of Arabic language.

**REFERENCES**

1. Pal, A. R., &Saha, D. (2015). Word sense disambiguation: a survey. arXiv preprint arXiv:1508.01346.
2. Navigli, R. (2009). Word sense disambiguation: A survey. ACM Computing Surveys (CSUR), 41(2), 10.
3. Bakx, G. E., Villodre, L. M., &Claramunt, G. R. (2006). Machine Learning Techniques for Word Sense Disambiguation. Unpublished doctoral dissertation, UniversitatPolitecnica de Catalunya.
4. Navigli, R. (2012). A quick tour of word sense disambiguation, induction and related approaches. In SOFSEM 2012: Theory and practice of computer science (pp. 115-129). Springer Berlin Heidelberg.
5. Zhou, X., & Han, H. (2005, May). Survey of Word Sense Disambiguation Approaches. In FLAIRS Conference (pp. 307-313).
6. Aliwy, A. H. (2012). Tokenization as Preprocessing for Arabic Tagging System. International Journal of Information and Education Technology, 2(4), 348.
7. Aliwy, A. H. (2013). Arabic morphosyntactic raw text part of speech tagging system (Doctoral dissertation, RepozytoriumUniwersytetuWarszawskiego).
8. Martin, D. J. J., &Jurafsky, D. (2000). Speech and Language Processing. An introduction to natural language processing, computational linguistics, and speech recognition.
9. Ye, M. (2004). SelectionalPreferenced Based Verb Sense Disambiguation Using WordNet.
10. Resnik, P. (1997, April). Selectional preference and sense disambiguation. InProceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How (pp. 52-57).
11. Lesk, M. (1986, June). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. InProceedings of the 5th annual international conference on Systems documentation (pp. 24-26). ACM.
12. Banerjee, S., & Pedersen, T. (2002). An adapted Lesk algorithm for word sense disambiguation using WordNet. In Computational linguistics and intelligent text processing (pp. 136-145). Springer Berlin Heidelberg.
13. Kalita, P., & Barman, A. K. Implementation of Walker Algorithm in Word Sense Disambiguation for Assamese language. Information retrieval, 1(2), 3.
14. Buscaldi, D., &Rosso, P. (2008). A conceptual density‐based approach for the disambiguation of toponyms. International Journal of Geographical Information Science, 22(3), 301-313.
15. Escudero, G., Màrquez, L., &Rigau, G. (2000). Naive Bayes and exemplar-based approaches to word sense disambiguation revisited. arXiv preprint cs/0007011.
16. Ng, H. T. (1997). Exemplar-based word sense disambiguation: Some recent improvements. arXiv preprint cmp-lg/9706010.
17. Yarowsky, D. (2000). Hierarchical decision lists for word sense disambiguation. Computers and the Humanities, 34(1-2), 179-186.
18. Palmer, M., Ng, H. T., & Dang, H. T. (2006). Evaluation of WSD systems.Word Sense Disambiguation: Algorithms and Applications, 75-106.
19. Molina, A., Pla, F., &Segarra, E. (2002). A hidden markov model approach to word sense disambiguation. In Advances in Artificial Intelligence—IBERAMIA 2002 (pp. 655-663). Springer Berlin Heidelberg.
20. Veronis, J., & Ide, N. M. (1990, August). Word sense disambiguation with very large neural networks extracted from machine readable dictionaries. InProceedings of the 13th conference on Computational linguistics-Volume 2(pp. 389-394). Association for Computational Linguistics.
21. Veenstra, J., Van den Bosch, A., Buchholz, S., &Daelemans, W. (2000). Memory-based word sense disambiguation. Computers and the Humanities,34(1-2), 171-177.
22. Hausman, M. (2011). A Genetic Algorithm Using Semantic Relations for Word Sense Disambiguation (Doctoral dissertation, University of Colorado at Colorado Springs).
23. Diou, C., Katsikatsos, G., &Delopoulos, A. (2006, December). Constructing Fuzzy Relations fromWordNetforWord Sense Disambiguation. In Semantic Media Adaptation and Personalization, 2006. SMAP'06. First International Workshop on (pp. 135-140). IEEE.
24. Ji, H. (2010, October). One sense per context cluster: Improving word sense disambiguation using web-scale phrase clustering. In Universal Communication Symposium (IUCS), 2010 4th International (pp. 181-184). IEEE.
25. Wanton, T. M., &LLavori, R. B. (2012). A clustering-based Approach for Unsupervised Word Sense Disambiguation. Procesamiento del lenguaje natural, 49, 49-56.
26. Yarowsky, D. (1995, June). Unsupervised word sense disambiguation rivaling supervised methods. In Proceedings of the 33rd annual meeting on Association for Computational Linguistics (pp. 189-196). Association for Computational Linguistics.
27. Leacock, C., Miller, G. A., &Chodorow, M. (1998). Using corpus statistics and WordNet relations for sense identification. Computational Linguistics,24(1), 147-165.
28. Mihalcea, R., &Faruque, E. (2004, July). Senselearner: Minimally supervised word sense disambiguation for all words in open text. InProceedings of ACL/SIGLEX Senseval (Vol. 3, pp. 155-158).
29. Mihalcea, R., & Moldovan, D. I. (2000, May). An Iterative Approach to Word Sense Disambiguation. In FLAIRS Conference (pp. 219-223).
30. Navigli, R., &Velardi, P. (2005). Structural semantic interconnections: a knowledge-based approach to word sense disambiguation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27(7), 1075-1086.
31. Elmougy, S., Taher, H., &Noaman, H. (2008, March). Naïve Bayes classifier for Arabic word sense disambiguation. In proceeding of the 6th International Conference on Informatics and Systems (pp. 16-21).
32. Eid, M. S., Al-Said, A. B., Wanas, N. M., Rashwan, M. A., &Hegazy, N. H. (2010, December). Comparative Study of Rocchio Classifier Applied to supervised WSD Using Arabic Lexical Samples. In Proceedings of the tenth conference of language engeneering (SEOLEC’2010), Cairo, Egypt.
33. Merhben, L., Zouaghi, A., &Zrigui, M. (2012). Lexical Disambiguation of Arabic Language: An Experimental Study. Polibits, (46), 49-54.
34. Merhbene, L., Zouaghi, A., &Zrigui, M. (2013). A Semi-Supervised Method for Arabic Word Sense Disambiguation Using a Weighted Directed Graph. InIJCNLP (pp. 1027-1031).
35. Pinto, D., Rosso, P., Benajiba, Y., Ahachad, A., & Jiménez-Salazar, H. (2007, December). Word sense induction in the arabic language: A self-term expansion based approach. In Proc. 7th Conference on Language Engineering of the Egyptian Society of Language Engineering-ESOLE (pp. 235-245).
36. Zouaghi, A., Merhbene, L., &Zrigui, M. (2011). Word Sense disambiguation for Arabic language using the variants of the Lesk algorithm. WORLDCOMP,11, 561-567.
37. Al-Gedawy, M., &Hegazy, O. (2012). Handling Text Mining Problems in Arabic using Domain-Specific Approach. International Journal of Computer Applications, 45(16).