Improving WSD using ISR-WN with Relevant Semantic Trees and SemCor Senses Frequency

Yoan Gutiérrez
Department of Informatics
University of Matanzas, Cuba.
{yoan.gutierrez}@umcc.cu

Sonia Vázquez and Andrés Montoyo
Department of Software and Computing Systems
University of Alicante, Spain.
{svazquez, montoyo}@dlsi.ua.es

Abstract

In this paper we concentrate on the resolution of the semantic ambiguity that arises when a given word has several meanings. This specific task is commonly referred to as Word Sense Disambiguation (WSD). We propose a method that obtains the appropriate senses from a multidimensional analysis (using Relevant Semantic Trees). Our method uses different resources WordNet, WordNet Domains, WordNet-Affects and SUMO, combined with senses frequency obtained from SemCor. Our hypothesis is that in WSD it is important to obtain the most frequent senses depending on the type of analyzed context to achieve better results. Finally, in order to evaluate and compare our results, it is presented a comprehensive study and experimental work using the Senseval-2 and Semeval-2 data set, demonstrating that our system obtains better results than other unsupervised systems.

1 Introduction

The main goal of knowledge technologies is to provide meaning to the huge quantity of information that our multilingual societies generate day to day. A wide range of advanced techniques are required to progressively automate the knowledge lifecycle. For that, after performing an analysis to large data collections it is necessary to develop different approaches to automatically represent and manage a high-level of meaningful concepts (Montoyo *et al.*, 2005). Moreover, to be able to create efficient systems of Natural Language Processing (NLP) it is necessary to turn the information extracted from words in plain text into a Concept Level or meaningful word senses. This representation allows to group words with similar meanings according to the context where they appear.

In order to determine the right meanings of words in different contexts WSD systems have been developed. Furthermore, it has been proved that applications such as Machine Translation, Information Extraction, Question Answering, Information Retrieval, Text Classification, and Text Summarization require knowledge about word meanings to obtain better results. So, WSD is considered an essential task for all these applications (Ide and Véronis, 1998). For this reason many research groups are working on WSD, using a wide range of approaches.

Due to the need of evaluating different approaches to show the improvements of NLP tasks it was created the Senseval[1] competition. The first Senseval was in 1998 at Herstmonceux Castle, Succex (England) and after that every three years a new competition takes place. In Senseval, different NLP tasks are defined in order to evaluate systems using the same repositories and corpus. At present, the results obtained in WSD have been going poorer, because the requirements of each corpus are getting more complex. For example, in Senseval-2 (Cotton *et al.*, 2001) the best system obtained a 69% of accuracy in WSD, three years later in Senseval-3 (Snyder and Palmer, 2004) the best results were around 65.2% of accuracy, next in Semeval-1 (Pradhan *et al.*, 2007) a 59.1% of accuracy was obtained and in Semeval-2 (Agirre *et al.*, 2010) was 55.5%.

Due to the fact that the results are still very low in accuracy we want to go deeply in the resolution of semantic ambiguity. Firstly, we have studied the behavior of the baseline Most Frequent Sense (MFS) in each competition. This baseline has been placed among the top places of the rank; for example, in Senseval-2 a system applying this baseline could have been located on the 2nd place with a 64.58% of accuracy (Preiss, 2006). In Senseval-3 Denys Yuret of Koc University computed a 60.9% and for the same competition Bart Decadt of University of Antwerp provided a baseline of 62.4%, these results could have been located on 7th and 5th positions respectively (Snyder and Palmer,

[1] http://www.senseval.org

Proceedings of Recent Advances in Natural Language Processing, pages 233–239,
Hissar, Bulgaria, 12-14 September 2011.

2004). In Semeval-1 the baseline was positioned on 9th place of fourteen systems and for the Semeval-2 competition the MFS baseline was located on 6th place. As we can see, this probabilistic procedure can obtain effective results on WSD task, but notice that it does not take into account context information.

Taking into account these facts our hypothesis is that for WSD it is important to obtain the most frequent senses combined with contextual information.

After these considerations, a new question arises: How will we be able to develop a procedure that uses the sense frequencies combined with a technique that takes into account the context information and improves the MFS results?

With the aim to answer this question and to demonstrate our hypothesis we present the following contributions:

- A method that combines MFS with a multidimensional analysis of the context. It uses several semantic resources combined with Relevant Semantic Trees.
- An analysis of how the MFS influences on the Relevant Semantic Trees method.
- An analysis of the behavior of Relevant Semantic Trees and Most Frequent Senses in each one of the semantic dimensions.
- A voting process between MFS and the results of different semantic dimensions.
- An exhaustive evaluation of the proposal.
- A comparison between our results and the systems in the Senseval-2 and Semeval-2 competitions.

In Section 2 we show some related works. Our approach is described in Section 3. The evaluations and analysis are provided in Section 4. Finally, we conclude in Section 5 adding further works.

2 Motivation and related work

Many efforts have been focused on the idea of building semantic networks to help NLP systems such as: MultiWordNet[2] (MWN), Multilingual Central Repository[3] (MCR), Integration of Semantic Resources based on WordNet (ISR-WN) (Gutiérrez *et al.*, 2010b) among others. Each resource has different semantic characteristics and their usage allows to analyze the tasks of NLP from different dimensions. Among all of these resources, ISR-WN has the highest quantity of semantic dimensions aligned, so it is a suitable resource to run our proposal. Next, we present a brief description of ISR-WN.

2.1 Integration of Semantic Resources based on WordNet (ISR-WN)

Integration of Semantic Resources based on WordNet (ISR-WN) (Gutiérrez *et al.*, 2010b) is a new resource that allows the integration of several semantic resources mapped to WN. In ISR-WN, WordNet is used as a core to link several resources such as: SUMO (Niles, 2001), WordNet Domains (WND) (Magnini and Cavaglia, 2000) and WordNet Affect (WNA) (Strapparava and Valitutti, 2004). As (Gutiérrez *et al.*, 2010a) describe, the integrator resource provides a software capable to navigate inside the semantic network.

In order to apply the multidimensionality that this resource provides, we have analyzed related NLP approaches that take into account semantic dimensionality. Addressed to context analysis we have studied (Magnini *et al.*, 2008), (Vázquez *et al.*, 2004) and (Buscaldi *et al.*, 2005). In these works WSD is performed using the WND resource (domain dimension). (Zouaq *et al.*, 2009), (Villarejo *et al.*, 2005) among others, conducted a semantic analysis using SUMO ontology (category dimension), and the Relevant Semantic Trees (RST) (Gutiérrez *et al.*, 2010a) apply several dimensions at once.

Next, we present the RST method which is able to work with different resources based on WordNet.

2.2 Relevant Semantic Trees (RST)

The RST method is able to find the correct senses of each word using Relevant Semantic Trees from different resources. This approach can be used with many resources mapped to WN as we have mentioned above.

In order to measure the association between concepts according to a multidimensional perspective in each sentence, RST uses an Association Ratio (*AR*) modification based on the proposal presented by Vázquez et al. (2004).

3 WSD Method

We propose an unsupervised knowledge-based method that uses the original RST technique including senses frequency of SemCor[4] corpus

[2] http://multiwordnet.fbk.eu/
3 http://www.lsi.upc.es/~nlp/meaning/meaning.html

[4] http://www.cse.unt.edu/~rada/downloads.html#semcor

and using a voting process to find the right senses. The voting process involves MFS (Most Frequent Sense), RST over WND, WNA, WN taxonomy and SUMO. Adding this new information we are able to improve the previous results obtained by the original RST and we also improve the MFS results in Semeval-2 competition. Specifically, we provide a sort of supervised aid (i.e. MFS) to the RST method of Gutiérrez *et al.*(2010a). Our proposal consists of two phases:

- Phase 1. Obtaining the Relevant Semantic Trees.
- Phase 2. Selecting the correct senses:
  - Step 1. Obtaining the RST from candidate senses.
  - Step 2. Obtaining accumulated values of relevance for each resource and frequency sense.
  - Step 3. Voting process to obtain the final senses.

Next, we present how these phases have been developed.

3.1 Obtaining the Relevant Semantic Trees

In this section, we describe how we have used a fragment of the original RST method with the aim to obtain Relevant Semantic Trees from the sentences. Equation 1 is used to measure and obtain the values of Relevant Concepts:

$$AR(C,s) = \sum_{i=1}^{n} AR(C,s_i)\,; \qquad (1)$$

Where

$$AR(C,w) = P(C,w) * \log_2 \frac{P(C,\ w)}{P(C)}; \qquad (2)$$

Where C is a concept; s is a sentence or a set of words (w); s_i is the *i-th* word (w) of the sentence s; $P(C,\ w)$ is joint probability distribution; and $P(C)$ is marginal probability.

The first stage is to Pre-process the sentence to obtain all lemmas. For instance, in the sentence "*But it is unfair to dump on teachers as distinct from the educational establishment.*" the lemmas are: [*unfair, dump, teacher, distinct, educational, establishment*]

Vector			
AR	Domain	*AR*	Domain
0.90	Pedagogy	0.36	Commerce
0.90	Administration	0.36	Quality
0.36	Buildings	0.36	Psychoanalysis
0.36	Politics	0.36	Economy
0.36	Environment		

Table 1. Initial Vector of Domain

Next, each lemma is searched through ISR-WN resource and it is correlated with concepts of WND (the dimension used in this example). Table 1 shows the results after applying Equation 1 over the sentence.

After obtaining the Initial Concept Vector of Domains we apply the Equation 3 in order to build the Relevant Semantic Tree related to the sentence.

$$AR(PC,s) = AR(ChC,s) - ND(IC,PC) \qquad ;(3)$$

Where:

$$ND(IC,PC) = \frac{MP(IC,PC)}{TD} \qquad ;(4)$$

Where *AR(PC, s)* represents the *AR* value of *PC* related to the sentence *s*; $AR(ChC,s)$ is the *AR* value calculated with Equation 1 in case of Child Concept (*ChC)* was included in the Initial Vector, otherwise is calculated with the Equation 3; *ND* is a Normalized Distance; *IC* is the Initial Concept from we have to add the ancestors; *PC* is Parent Concept; *TD* is Depth of the hierarchic tree of the resource to use; and *MP* is Minimal Path.

Applying the Equation 3, the algorithm to decide which parent concept will be added to the vector is shown here:

if (AR(PC, s) > 0){
if (PC had not been added to vector)
PC is added to the vector with AR(PC, s) value;
else PC value = PC value + AR(PC, s) value; }

This bottom-up process is applied for each Concept of the Initial Vector to add each Relevant Parent to the vector. After reproducing the process to each Concept of the Initial Vector, the Relevant Semantic Tree is built. As a result, the Table 2 is obtained. This vector represents the Domain tree associated to the sentence such as Figure 1 shows. As we can see, the Relevant Semantic Tree of domains in Figure 1 has associated a color intensity related to the *AR* value of each domain. The more intense the color is the more related *AR* is.

Vector			
AR	Domain	*AR*	Domain
1.63	Social_Science	0.36	Buildings
0.90	Administration	0.36	Commerce
0.90	Pedagogy	0.36	Environment
0.80	RootDomain	0.11	Factotum
0.36	Psychoanalysis	0.11	Psychology
0.36	Economy	0.11	Architecture
0.36	Quality	0.11	Pure_Science
0.36	Politics		

Table 2. Final Domain Vector based on WND

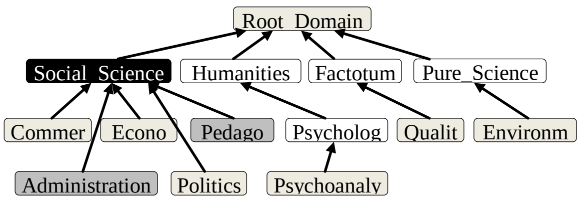

Figure 1. Relevant Semantic Tree from WND

Once the Relevant Semantic Tree is obtained, in case of the Domain dimension the Factotum category is eliminated from the tree. Due to the fact that Factotum is a generic Domain associated to words that appear in general contexts it does not provide useful information (Magnini and Cavaglia, 2000). Moreover, after conducting several experiments we have confirmed that it introduced errors.

3.2 Selecting the correct senses

To select the correct senses, three steps are applied:

Step 1. Obtaining the RST from candidate senses

In this step we associate to each possible sense of each lemma a RST based on each semantic dimension. At this stage the aim of RST is to measure the relation between each Concept and each sense. To do this we use the Equation 2 where we have substituted the variable *w* (word) with the variable sw_i, (sense) where sw_i indicates the *i-th* sense of word *w*. As a result, we convert each RST in a vector. Next, we continue with the complete process adding the parent concepts.

Step 2. Obtaining accumulated values of relevance for each resource and frequency sense

To measure the similarity between the RST of the sentences and senses, we have applied a fragment of the original method from (Gutiérrez *et al.*, 2010a) introducing sense frequency ($Freq_s$) as a new modification. Our goal is to obtain a new value to measure the Most Frequent Sense (MFS) in a given context. The *AR* value is accumulated when a matching exists between the vector elements of the sense and the vector elements of the sentence. The process is shown in the Equation 5.

$$AC(s, ARV) = \frac{\sum_k ARV[Vs_k]}{\sum_{i=1} ARV_i} + Freq_s \quad ; \qquad (5)$$

Where *AC* is the *AR* value accumulated for the analyzed elements; *ARV* is the vector of relevant concepts of the sentence with the format: *ARV*[concept1 | *AR* value,*...]*; *Vs* is the vector of relevant concepts of the sense with the format *Vs*[concepts]; Vs_k is the *k-th* concept of the vector *Vs*; *ARV* [Vs_k] represents the value of *AR* assigned to the concept Vs_k for the value *ARV; $Freq_s$* represents the normalized value of frequency sense obtained from *cntlist* file from WN 1.6; and $\sum_{i=1} VRA_i$ is the term that normalizes the result.

AC is calculated for each RST (or Relevant Vector) of each semantic dimension. In this approach we have obtained four *AC* values (for WN taxonomy, WND, WNA and SUMO).

Notice that once we have obtained *AC* values for each sense in each dimension, if the senses calculated do not match with the grammatical category that Freeling (Atserias *et al.*, 2006) suggests, we discriminate these senses adding a zero value to *AC;* in other case we add a one value. Adding these values we can maintain all the candidates in the solution despite the grammatical category is wrong.

Finally, the proposed sense will have the highest *AC* value among all senses in each lemma.

Step 3. Voting process to obtain the final senses

As we have explained above, each semantic dimension provides a possible sense. It is important to remark that the sense frequency is also included as a semantic dimension. So, in order to decide the right sense among the different semantic dimensions sense proposals we use a voting process. To apply this idea we define the next equation:

$$Ps = max_i(max_k(V[VAC]_k)_i); \qquad (6)$$

Where *VAC* corresponds to a vector composed by *AC* values of each sense for one lemma; *V [VAC]* is a vector of the *VAC*; *k* corresponds to each resource; $V\,[VAC]_k$: corresponds to *k-th* *VAC* for resource *k*; $\max_k(V\,[VAC]_k)$: determines the sense with maximum *AC* value of each *VAC*; *i*: is *i-th* sense; $\max_i$: determines the sense that was selected more times by $\max_k$ among all resources; and *Ps*: indicates proposed sense.

The *VAC* format is as follows: VAC [*AC* value sense#1, *AC* value sense#2, *AC* value sense#n]. And the *V [VAC]* format is: *V* [*VAC*-Domains, *VAC*-Emotions, *VAC*-WordNet Taxonomies, *VAC*-SUMO, *VAC*-Frequency *Senses*]

In *VAC* we also define a vector built with the frequency values of SemCor corpus for each lemma. Then we conduct a voting process with five *AC* values. If in an exceptional situation we obtain a tie or disjoin senses, the proposed sense will be the most frequent. We have chosen this option because of empirical studies have demonstrated that MFS works better than others (Molina *et al.*, 2002).

4 Evaluations and Analysis

In this section our purpose is to confirm the hypothesis presented in Section 1. We have evaluated this method with two different test corpus, Senseval-2 on "English All words" task and Semeval-2 on "English All words on Specific Domain" task. Moreover, we have compared our results with the participating systems of the aforementioned competitions. The goal of these experiments is to demonstrate how the sense frequencies combined with RST can improve the original RST results.

4.1 Evaluation with Senseval-2 corpus

First, we analyzed how the addition of the sense frequencies to accumulated value (*AC*) of each sense improved the results of the previous work published on (Gutiérrez *et al.*, 2010a). To do this we used as test corpus the file d00.txt and we conducted some experiments:

- Exp 1: Adding to *AC* value a 0% of $Freq_S$.
- Exp 2: Adding to *AC* value a 50% of $Freq_S$.
- Exp 3: Adding to *AC* value a 100% of $Freq_S$.

In the original method the authors calculated an accumulated value for each resource and summed up all the values to obtain the total accumulated value to combine all resources. In this new approach we also add the $Freq_S$ to the total accumulated value. Table 3 shows how each experiment obtains better results when Sense Frequencies ($Freq_S$) parameter is increased. Notice that we do not keep increasing this weight (i.e. 150%, 200%, etc) because the proposal would become converted only in selection process of MFS.

In order to determine whether the $Freq_S$ enhances the Most Frequent Senses (MFS) baseline, we conducted new experiments.

Next, we show how we have used the original method adding to *AC* the 100% of $Freq_S$ but only using one dimension at the same time:

- Exp4: Using MFS using $Freq_S$
- Exp5: Using WND resource
- Exp6: Using SUMO resource
- Exp7: Using WNA resource
- Exp8: Using WN Taxonomy resource

After doing these experiments we were able to determine which dimension worked better. As we can see on Table 3, these five experiments obtained promising results.

Another experiment was to combine these five experiments in a voting process to obtain even better results. This idea has led us to make our main proposal.

- Exp9: Applying a voting process among Exp4, Exp5, Exp6, Exp7 and Exp8 results

Table 3 shows all the results obtained from d00.txt file of Senseval-2. The result of MFS is underlined and the approach that exceeded it is in bold. We can see that the voting process (Exp9) obtained the best results.

Exp	Precision	Recall	Exp	Precision	Recall
Exp1	0,408	0,407	Exp6	0,561	0,560
Exp2	0,490	0,490	Exp7	0,555	0,554
Exp3	0,535	0,534	**Exp8**	**0,572**	**0,572**
Exp4	0,565	0,564	**Exp9**	**0,575**	**0,575**
Exp5	**0,572**	**0,572**			

Table 3. Results over d00.txt from Senseval-2

Following, we present the results after analyzing the entire corpus of the Senseval-2 competition. For that, we applied two experiments to the entire corpus.

- Exp10: Applying WSD with MFS of $Freq_S$
- Exp11: Applying a voting process using the five dimensions

We show in Table 4 a comparison among the results of the best performances of our voting process, MFS using $Freq_s$ and MFS obtained by (Preiss, 2006). The baseline used by Preiss was based on *cntlist* file from WN 1.7 version and our Exp10 was based on *cntlist* from WN 1.6. Notice, that are different although both are based on frequency information.

English All words - Fine-grained Scoring							
Rank	Precision	Recall		Rank	Precision	Recall	
1	0.690	0.690	S	**Exp11**	**0,610**	**0,609**	U
MFS	0.669	0.646	-	Exp10	0,601	0,599	-
2	0.636	0.636	S	4	0.575	0.569	U
3	0.618	0.618	S	..	..	..	

Table 4. Senseval-2 ranking

As we can see, our approach improves the Exp10 results. These results were obtained by our system, but the baseline MFS results obtained by Preiss were better than ours. This means that we could enhance the MFS that we use. So, we need to integrate in our approach a better MFS resource to obtain better results. Table 4 shows that our proposal would have the best results of all unsupervised methods.

4.2 Evaluation with Semeval-2 corpus

Our approach was also evaluated using corpus from Semeval-2 competition. The voting process obtained 52.7% and 51.5% of precision and recall respectively, improving the MFS baseline with 1% of accuracy. The original method from (Gutiérrez *et al.*, 2010a) was improved on 19.3% of accuracy such as Table 5 shows.

Rank	Precision	Recall	Rank	Precision	Recall
1	0.570	0.555	…	….	…
2	0.554	0.540	…	…	…
3	0.534	0.528	26	0.370	0.345
4	0.522	0.516	<u>27</u>	<u>0.328</u>	<u>0.322</u>
Our	**0,527**	**0,515**	<u>28</u>	<u>0.321</u>	<u>0.315</u>
5	0.513	0.513	29	0.312	0.303
MFS	0.505	0.505	Random	0.23	0.23

Table 5. Semeval-2 ranking

The underlined results pertain to original method from (Gutiérrez *et al.*, 2010a) and the bold results pertain to our approach. As a result, we can see that we can improve the MFS proposal from Semeval-2 competition.

In this competition only were evaluated nouns and verbs. The behavior of our approach for each category was: nouns 54.4% of precision and 53.7% of recall, and verbs 49.4% of precision and 45.4% of recall. Each category is effective in comparison with the best results obtained on this competition.

In order to determine if the annotation of grammatical categories influences on the results, we discovered that the Freeling tool introduced a noise of 2.62% when detecting nouns and for verbs 8.20%. These analyses indicate that the results would be better using another more accurate tool.

4.3 Comparison with newer works

In this section we present a comparison with some relevant WSD methods. We can mention those approaches using page-rank such as (Sinha and Mihalcea, 2007), and (Agirre and Soroa, 2009). These proposals were tested using "English All Words" task corpus from Senseval-2. In both proposals, Page-Rank method has been used to determine the centrality of structural lexical network using the semantic relations of WordNet. Then, to disambiguate each word the most weighted sense was chosen. These approaches obtained 58.6% and 56.37% of recall respectively. Other significant work is the ACL 2004 paper by (Mc.Carthy *et al.*, 2004) where the most frequent senses were obtained from a variety of resources (Reuters Corpus and SemCor Corpus), some of which provide domain information. This proposal obtained a 64% of precision in all-nouns task; this is just 3% higher than our results. However, we achieved better results than Mihalcea and Agirre exceeding them around 5%. This improvement could seem very poor but talking about WSD is a great step forward.

5 Conclusions and further works

In this paper we have presented the hypothesis that for word-sense disambiguation it is important to obtain the Most Frequent Senses depending on the kind of analyzed context. In order to demonstrate this hypothesis, we have studied how several semantic dimensions combined with sense frequencies could improve the obtained results of many approaches that only conducted the WSD analysis with one dimension. We have proposed an adaptation of an unsupervised knowledge-based method that combines the original Relevant Semantic Trees method with senses frequency in a voting process. As a result, we have been able to determine which percentage of sense frequency is needed to help the Relevant Semantic Trees method. Therefore, we have demonstrated that the WSD results are better when more percentage of sense frequency is added.

Moreover, we have conducted different experiments in order to know which semantic dimensions achieve better results. These experiments demonstrated that the Domain dimension (WND) and WordNet dimension (WN Taxonomy) worked better than MFS (Frequency dimension). Also, a voting process has been applied among all dimensions obtaining in Senseval-2 an of 60.9% and achieving the best results of all unsupervised systems. Furthermore, related to Semeval-2 our approach has improved the baseline MFS and the original RST method.

As further work we propose to use other resources on the voting process in order to add more dimensions and also, use a better frequency resource. Apart from that, we also have considered to use another grammatical categorizer, in order to reduce the noise introduced by misclassifying words.

Acknowledgements

This paper has been supported partially by Ministerio de Ciencia e Innovación - Spanish Government (grant no. TIN2009-13391-C04-01), and Conselleria d'Educación - Generalitat

Valenciana (grant no. PROMETEO/2009/119, ACOMP/2010/288 and ACOMP/2011/001).

References

Antonio Molina, Ferran Pla, Encarna Segarra and Lidia Moreno. 2002. Word Sense Disambiguation using Statistical Models and WordNet. *Proceedings of 3rd International Conference on Language Resources and Evaluation*: Las Palmas de Gran Canaria.

Andrés Montoyo, Armando Suárez, German Rigau and Manuel Palomar. 2005. Combining Knowledge- and Corpus-based Word-Sense-Disambiguation Methods. *Journal of Artificial Intelligence,* 23:299-330.

Amal Zouaq, Michel Gagnon and Benoit Ozell. 2009. A SUMO-based Semantic Analysis for Knowledge Extraction. *Proceedings of the 4th Language & Technology Conference*: Poznań, Poland.

Bernardo Magnini and Gabriela Cavaglia. 2000. Integrating Subject Field Codes into WordNet. *Proceedings of Third International Conference on Language Resources and Evaluation (LREC-2000)*: 1413--1418.

Bernardo Magnini, Carlo Strapparava, Giovanni Pezzulo and Alfio Gliozzo. 2008. Using Domain Information for Word Sense Disambiguation. *Proceedings of the First International Conference on Emerging Trends in Engineering and Technology (icetet 2008)*: 1187-1191. Nagpur, India.

Benjamin Snyder and Martha Palmer. 2004. The English All Word Task. *SENSEVAL-3: Third International Workshop on the evaluation of System of the Semantic Analysis of Text*: Barcelona, Spain.

Christiane Fellbaum. 1998. WordNet. An Electronic Lexical Database. *The MIT Press*: University of Cambridge.

Carlo Strapparava and Alessandro Valitutti. 2004. WordNet-Affect: an affective extension of WordNet. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*: 1083-1086. Lisbon.

Davide Buscaldi, Paolo Rosso and Francesco Masulli. 2005. Integrating Conceptual Density with WordNet Domains and CALD Glosses for Noun Sense Disambiguation. *Proceedings of España for Natural Language Processing (ESTAL-2005)*: 267-276. Alicante, Spain.

Diana Mc.Carthy, Rob Koeling, Julie Weeds and John Carroll. 2004. Finding Predominant Word Senses in Untagged Text. *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04)*: 279--286.

Eneko Agirre, Oier López De Lacalle, Christiane Fellbaum, Shu-Kai Hsieh, Maurizio Tesconi, Monica Monachini, Piek Vossen and Roxanne Segers. 2010. SemEval-2010 task 17: All-words word sense disambiguation on a specific domain. *Proceedings of the 5th International Workshop on Semantic Evaluation*: 75-80. Los Angeles, California.

Eneko Agirre and Aitor Soroa. 2009. Personalizing PageRank for Word Sense Disambiguation. *Proceedings of the 12th conference of the European chapter of the Association for Computational Linguistics (EACL-2009)*: Athens, Greece.

Ian Niles. 2001. Mapping WordNet to the SUMO Ontology. *Teknowledge Corporation.*

Jordi Atserias, Bernardino Casas, Elisabet Comelles, Meritxell González, Lluis Padró and Muntsa Padró. 2006. FreeLing 1.3: Syntactic and semantic services in an opensource NLP library. *Proceedings of LREC'06*: Genoa, Italy.

Judita Preiss. 2006. A detailed comparison of WSD systems:an analysis of the system answers for the Senseval-2 English all words task. *Natural Language Engineering*, 12(3):209–228.

Luis Villarejo, Lluís Márquez and German Rigau. 2005. Exploring the construction of semantic class classiers for WSD. *Sociedad Española para el Procesamiento del Lenguaje Natural,* 35: 195-202.

Nancy Ide and Jean Véronis. 1998. Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art. *Computational Linguistics,* 24(1):1-40.

Ravi Sinha and Rada Mihalcea. 2007. Unsupervised Graph-based Word Sense Disambiguation Using Measures of Word Semantic Similarity. *Proceedings of the IEEE International Conference on Semantic Computing (ICSC 2007)*: Irvine, CA.

Sameer S. Pradhan, Edward Loper, Dmitriy Dligach and Martha Stone Palmer. 2007. SemEval-2007 Task-17: English Lexical Sample SRL and All Words. *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*: 87-92. Prague.

Scott Cotton, Phil Edmonds, Adam Kilgarriff and Martha Palmer. 2001. English All word. *SENSEVAL-2: Second International Workshop on Evaluating Word Sense Disambiguation Systems*: Toulouse, France.

Sonia Vázquez, Andrés Montoyo and German Rigau. 2004. Using Relevant Domains Resource for Word Sense Disambiguation. *IC-AI'04. Proceedings of the International Conference on Artificial Intelligence*: Ed: CSREA Press. Las Vegas, E.E.U.U.

Yoan Gutiérrez, Antonio Fernández, Andrés Montoyo and Sonia Vázquez. 2010a. UMCC-DLSI: Integrative resource for disambiguation task. *Proceedings of the 5th International Workshop on Semantic Evaluation*: 427-432. Uppsala, Sweden.

Yoan Gutiérrez, Antonio Fernández, Andrés Montoyo and Sonia Vázquez. 2010b. Integration of semantic resources based on WordNet. *XXVI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural,* 45: 161-168. Universidad Politécnica de Valencia, Valencia.