Exploring the Automatic Selection of Basic Level Concepts*

Rubén Izquierdo & Armando Suárez
GPLSI. Departament de LSI. UA.
Alacant, Spain
{ruben,armando}@dlsi.ua.es

German Rigau
IXA NLP Group. EHU.
Donostia, Spain
german.rigau@ehu.es

Abstract

We present a very simple method for selecting Base Level Concepts using basic structural properties of WordNet. We also empirically demonstrate that these automatically derived set of Base Level Concepts group senses into an adequate level of abstraction in order to perform class-based Word Sense Disambiguation. In fact a very naive Most Frequent classifier using the classes selected is able to perform a semantic tagging with accuracy figures over 75%.

Keywords

WordNet, word-senses, levels of abstraction, Word Sense Disambiguation

1 Introduction

Word Sense Disambiguation (WSD) is an intermediate Natural Language Processing (NLP) task which consists in assigning the correct semantic interpretation to ambiguous words in context. One of the most successful approaches in the last years is the *supervised learning from examples*, in which statistical or Machine Learning classification models are induced from semantically annotated corpora [11]. Generally, supervised systems have obtained better results than the unsupervised ones, as shown by experimental work and international evaluation exercises such as Senseval[1]. These annotated corpora are usually manually tagged by lexicographers with word senses taken from a particular lexical semantic resource –most commonly WordNet[2] (WN) [7].

WN has been widely criticised for being a sense repository that often offers too fine–grained sense distinctions for higher level applications like Machine Translation or Question & Answering. In fact, WSD at this level of granularity, has resisted all attempts of infering robust broad-coverage models. It seems that many word–sense distinctions are too subtle to be captured by automatic systems with the current small volumes of word–sense annotated examples. Possibly, building class-based classifiers would allow to avoid the data sparseness problem of the word-based approach.

Recently, using WN as a sense repository, the organizers of the English all-words task at SensEval-3 reported an inter-annotation agreement of 72.5% [17]. Interestingly, this result is difficult to outperform by state-of-the-art fine-grained WSD systems.

Thus, some research has been focused on deriving different sense groupings to overcome the fine–grained distinctions of WN [8] [14] [12] [1] and on using predefined sets of sense-groupings for learning class-based classifiers for WSD [16] [4] [18] [5] [3]. However, most of the later approaches used the original Lexicographical Files of WN (more recently called Supersenses) as very coarse–grained sense distinctions. However, not so much attention has been paid on learning class-based classifiers from other available sense–groupings such as WordNet Domains [10], SUMO labels [13], EuroWordNet Base Concepts [19] or Top Concept Ontology labels [2]. Obviously, these resources relate senses at some level of abstraction using different semantic criteria and properties that could be of interest for WSD. Possibly, their combination could improve the overall results since they offer different semantic perspectives of the data. Furthermore, to our knowledge, to date no comparative evaluation have been performed exploring different sense–groupings.

We present a very simple method for selecting Base Level Concepts [15] using basic structural properties of WN. We also empirically demonstrate that these automatically derived set of Base Level Concepts group senses into an adequate level of abstraction in order to perform class-based WSD.

This paper is organized as follows. Section 2 introduce the different levels of abstraction that are relevant for this study, and the available sets of semi-automatically derived Base Concepts. In section 3, we present the method for deriving fully automatically a number of Base Level Concepts from any WN version. Section 4 reports the resulting figures of a direct comparison of the resources studied. Section 5 provides an empirical evaluation of the performance of the different levels of abstraction. In section 6 we provide further insights of the results obtained and finally, in section 7 some concluding remarks are provided.

2 Levels of abstraction

The notion of Base Concepts (hereinafter BC) was introduced in EuroWordNet[3] [19]. The BC are supposed to be the concepts that play the most important role in the various wordnets of different languages. This role

*This paper has been supported by the European Union under the project QALL-ME (FP6 IST-033860) and the Spanish Government under the project Text-Mess (TIN2006-15265-C06-01) and KNOW (TIN2006-15049-C03-01)

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

[2] http://wordnet.princeton.edu

[3] http://www.illc.uva.nl/EuroWordNet/

was measured in terms of two main criteria: a high position in the semantic hierarchy and having many relations to other concepts. Thus, the BC are the fundamental building blocks for establishing the relations in a wordnet. In that sense, the Lexicografic Files (or Supersenses) of WN could be considered the most basic set of BC.

Basic Level Concepts [15] (hereinafter BLC) should not be confused with **Base Concepts**. BLC are a compromise between two conflicting principles of characterization: a) to represent as many concepts as possible (abstract concepts), and b) to represent as many distinctive features as possible (concrete concepts).

As a result of this, Basic Level Concepts typically occur in the middle of hierarchies and less than the maximum number of relations. BC mostly involve the first principle of the Basic Level Concepts only. BC are generalizations of features or semantic components and thus apply to a maximum number of concepts. Our work focuses on devising simple methods for selecting automatically an accurate set of Basic Level Concepts from WN.

WordNet synsets are organized in forty five Lexicographer Files, or SuperSenses, based on syntactic categories (nouns, verbs, adjectives and adverbs) and logical groupings, such as person, phenomenon, feeling, location, etc. There are 26 basic categories for nouns, 15 for verbs, 3 for adjectives and 1 for adverbs. Within EuroWordNet, initially, a set of 1,024 Common Base Concepts was selected from WN1.5. The Balkanet project[4] selected his own list of BC extending the original set of BC of EWN to a final set of 4,698 ILI records from WN2.0[5] (3,210 nouns, 1,442 verbs and 37 adjectives). In the the Meaning project[6], the number of BC selected from WN1.6 was 1,535 (793 for nouns and 742 for verbs).

3 Automatic Selection of Base Level Concepts

This section describes a simple method for deriving a set of Base Level Concepts (BLC) from WN. The method has been applied to different WN versions for nouns and verbs. Basically, to select the appropriate BLC of a particular synset, the algorithm only considers the relative number of relations of their hypernyms. We derived two different sets of BLC depending on the type of relations considered: a) all types of relations encoded in WN (All) and b) only the hyponymy relations encoded in WN (Hypo).

The process follows a bottom-up approach using the chain of hypernym relations. For each synset in WN, the process selects as its Base Level Concept the first local maximum according to the relative number of relations. For synsets having multiple hypernyms, the path having the local maximum with higher number of relations is selected. Usually, this process finishes having a number of "fake" Base Level Concepts. That is, synsets having no descendants (or with a very small number) but being the first local maximum according to the number of relations considered. Thus, the process finishes checking if the number of concepts subsumed by the preliminary list of BLC is higher than a certain threshold. For those BLC not representing enough concepts according to a certain threshold, the process selects the next local maximum following the hypernym hierarchy. Thus, depending on the type of relations considered to be counted and the threshold established, different sets of BLC can be easily obtained for each WN version.

#rel.	synset
18	group_1,grouping_1
19	social_group_1
37	organisation_2,organization_1
10	establishment_2,institution_1
12	faith_3,religion_2
5	Christianity_2,**church_1**,Christian_church_1
#rel.	**synset**
14	entity_1,something_1
29	object_1,physical_object_1
39	artifact_1,artefact_1
63	construction_3,structure_1
79	building_1,edifice_1
11	place_of_worship_1, ...
19	**church_2**,church_building_1
#rel.	**synset**
20	act_2,human_action_1,human_activity_1
69	activity_1
5	ceremony_3
11	religious_ceremony_1,religious_ritual_1
7	service_3,religious_service_1,divine_service_1
1	**church_3**,church_service_1

Table 1: *Possible Base Level Concepts for the noun* Church *in WN1.6*

An example is provided in Table 1. This table shows the possible BLC for the noun "church" using WN1.6. The table presents the hypernym chain for each synset together with the number of relations encoded in WN for the synset. The local maxima along the hypernym chain of each synset appears in bold. Obvioulsy, different criteria will select a different set of Base Level Concepts.

Instead of highly related concepts, we also considered highly frequent concepts as possible indicator of a large set of features. Following the same basic algorithm, we also used the relative frequency of the synsets in the hypernym chain. That is, we derived two other different sets of BLC depending on the source of relative frequencies considered: a) the frequency counts in SemCor (FreqSC) and b) the frequency counts appearing in WN (FreqWN). The frequency of a synset has been obtained summing up the frequencies of its word senses. In fact, WN word-senses were ranked using SemCor and other sense-annotated corpora. Thus, the frequencies of SemCor and WN are similar, but not equal.

4 Comparing Base Level Concepts

Different sets of Base Level Concepts (BLC) have been generated using different WN versions, types of relations (All and Hypo), sense frequencies (FreqSC and FrecWN) and thresholds.

Table 2 presents the total number of BLC and its

[4] http://www.ceid.upatras.gr/Balkanet
[5] http://www.globalwordnet.org/gwa/5000_bc.zip
[6] http://www.lsi.upc.es/~nlp/meaning

average depth for WN1.6[7] varying the threshold and the type of relations considered (All or Hypo) and the type of frequency (WN or SemCor).

		BLC		Depth	
Threshold	Relation	Noun	Verb	Noun	Verb
0	all	3,094	1,256	7.09	3.32
	hypo	2,490	1,041	7.09	3.31
	SemCor	34,865	3,070	7.44	3.41
	WN	34,183	2,615	7.44	3.30
10	all	971	719	6.20	1.39
	hypo	993	718	6.23	1.36
	SemCor	690	731	5.74	1.38
	WN	691	738	5.77	1.40
20	all	558	673	5.81	1.25
	hypo	558	672	5.80	1.21
	SemCor	339	659	5.43	1.22
	WN	340	667	5.47	1.23
50	all	253	633	5.21	1.13
	hypo	248	633	5.21	1.10
	SemCor	94	630	4.35	1.12
	WN	99	631	4.41	1.12

Table 2: *Automatic Base Level Concepts for WN1.6 using relations or frequencies*

As expected, when increasing the threshold, the total number of automatic BLC and its average depth decrease. For instance, using all relations on the nominal part of WN, the total number of BLC ranges from 3,094 (no threshold) to 253 (threshold 50). However, although the number of total BLC for nouns decreases dramatically (around 10 times), the average depth of the synsets selected only ranges from 7.09 to 5.21 using both types of relations (All and Hypo). This fact, possibly indicates the robustness of the approach.

Also as expected, the verbal part of WN behave differently. In this case, since the verbal hierarchies are much shorter, the average depth of the synsets selected ranges from 3.32 to only 1.13 using all relations, and from 3.31 to 1.10 using hypo relations.

In general, when using the frequency criteria, we can observe a similar behaviour than when using the relation criteria. However, now the effect of the threshold is more dramatic, specially for nouns. Again, although the number of total BLC for nouns decreases dramatically, the average depth of the synsets selected only ranges from 7.44 to 4.35 and 4.41. As expected, verbs behave differently than nouns. The number of BLC (for both SemCor and WN frequencies) reaches a plateau of around 600. In fact, this number is very close to the verbal top beginners.

Table 3 summarizes the Balkanet and Meaning Base Concepts including the total number of synsets and their average depth.

Set	**PoS**	**#BC**	**Depth.**
Balkanet	Noun	3,210	5.08
	Verb	1,442	2.45
Meaning	Noun	793	4.93
	Verb	742	1.36

Table 3: Balkanet *and* Meaning *Base Concepts*

5 Sense–groupings as semantic classes

In order to study to what extend the different sense–groupings could be of the interest for class–based WSD, we present a comparative evaluation of the different sense–groupings in a controlled framework. We tested the behaviour of the different sets of sense–groupings (WN senses, Balkanet BC, Meaning BC, automatic BLC and SuperSenses) using the English all–words task of SensEval–3. Obviously, different sense–groupings would provide different abstractions of the semantic content of WN, and we expect a different behaviour when disambiguating nouns and verbs. In fact, the most common baseline used to test the performance of a WSD system, is the Most Frequent Sense Classifier. In this study, we will use this simple but robust heuristic to compare the performances of the different sense–groupings. Thus, we will use SemCor[8] [9] to train for Most Frequent Classifiers for each word and sense–grouping. We only used brown1 and brown2 parts of SemCor to train the classifiers. We used standard Precision, Recall and F1 measure (harmonic mean between Precision and Recall) to evaluate the performance of each classifier.

	Senses	**BLC-A**	**BLC-S**	**SS**
Nouns	4.93	4.07	4.00	3.06
Verbs	11.00	8.64	8.72	4.08
N + V	7.66	6.13	6.13	3.52

Table 4: *Polysemy degree over SensEval–3*

For WN senses, Meaning BC, the automatic BLC, and Lexicographic Files, we used WN1.6. For Balkanet BC we used the synset mappings provided by [6][9], translating the BC from WN2.0 to WN1.6. For testing the Most Frequent Classifiers we also used these mappings to translate the sense–groupings from WN1.6 to WN1.7.1.

Table 4 presents the polysemy degree for nouns and verbs of the different words when grouping its senses with respect the different semantic classes on SensEval–3. Senses stand for WN senses, BLC-A for automatic BLC derived using a threshold of 20 and all relations, BLC-S for automatic BLC derived using a threshold of 20 and frequencies from SemCor and SS for the SuperSenses. As expected, while increasing the abstraction level the polysemy degree decreases. Notice that the reduction is dramatic for verbs (from 11.0 to only 4.08). Notice also, that when using the Base Level Concept representations a high degree of polysemy is maintained for nouns and verbs.

Table 5 presents for polysemous words the performance in terms of F1 measure of the different sense-groupings when training the class–frequencies on SemCor and testing on SensEval–3. That is, for each polysemous word in SensEval–3 the Most Frequent Class is obtained from SemCor. Best results are marked using bold.

As expected, SuperSenses obtain very high F1 results for nouns and verbs. Comparing the BC from Balkanet and the best results seems to be achieved by Meaning BC for both nouns and verbs. Notice that the set of BC from Balkanet was larger than the ones selected in Meaning, thus indicating that the BC from Meaning provide a better level of abstraction.

Regarding the relations criteria, all sets of auto-

[7] WN1.6 have 66,025 nominal and 12,127 verbal synsets.

[8] Annotated using WN1.6.

[9] http://www.lsi.upc.edu/˜nlp/

	All		Hypo		Semcor		WN	
Class	**Nouns**	**Verbs**	**Nouns**	**Verbs**	**Nouns**	**Verbs**	**Nouns**	**Verbs**
Senses	63.69	49.78	63.69	49.78	63.69	49.78	63.69	49.78
Balkanet	65.15	50.84	65.15	50.84	65.15	50.84	65.15	50.84
Meaning	65.28	53.11	65.28	53.11	65.28	53.11	65.28	53.11
BLC–0	66.36	54.30	65.76	54.30	64.45	52.27	64.95	51.75
BLC–10	66.31	54.45	65.86	54.45	64.98	53.21	65.59	53.29
BLC–20	**67.64**	54.60	**67.28**	54.60	65.73	53.97	66.30	53.44
BLC–30	67.03	54.60	66.72	54.60	66.46	54.15	66.67	53.61
BLC–40	66.61	55.54	66.77	**55.54**	68.46	**54.63**	**69.16**	54.22
BLC–50	67.19	**55.69**	67.19	**55.54**	**68.84**	**54.63**	69.11	**54.63**
SuperSenses	**73.05**	**76.41**	**73.05**	**76.41**	**73.05**	**76.41**	**73.05**	**76.41**

Table 5: *F1 measure for polysemous words tested on SensEval–3*

matic BLC perform better than those BC provided by Balkanet or Meaning. Also in this case, for nouns, the best results are obtained when using a threshold of only 20. We should highlight this result since this set of BLC obtain better WSD performance than the rest of automatically derived BLC while maintaining more information of the original synsets. That is, BLC-20 using all relatons (558 classes) achieves an F1–score of 67.64, while SuperSenses using a much smaller set (26 classes) achieves 73.05. We can also observe that in general, using hyponymy relations we obtain slightly lower performances than using all relations. Possibly, this fact indicates that a higher number of hyponymy relations is required for a Base Level Concept to compensate minor (but richer) number of relations. These results suggest that intermediate levels of representation such as the automatically derived Base Concept Levels could be appropriate for learning class–based WSD classifiers.

Also in Table 5, we present the results of using frequencies from SemCor and frequencies from WN for selecting the BLC. In this case, not all sets of automatic BLC surpass the BC from Balkanet and Meaning. The best results are obtained when using higher thresholds. However, in this case, verbal BLC obtain slightly lower results than using the relations criteria (both all and hypo). We can also observe that in general, using SemCor frequencies we obtain slightly lower performances than using WN frequencies.

These results for polysemous words reinforce our initial observations. That is, that the method for automatically deriving intermediate levels of representation such the Base Concept Levels seems to be robust enough for learning class-based WSD classifiers. In particular, it seems that BLC could achieve high levels of accuracy while maintaining adequate levels of abstraction (with hundreds of BLC). In particular, the automatic BLC obtained using the relations criteria (All or Hypo) surpass the BC from Balkanet and Meaning. For verbs, it seems that even the unique top beginners require an extra level of abstraction (that is, the SuperSense level) to be affective.

6 Discussion

We can put the current results in context, although indirectly, by comparison with the results of the English SensEval–3 all–words task systems. In this case, the best system presented an accuracy of 65.1%, while the "WN first sense" baseline would achieve 62.4%[10].

[10] This result could be different depending on the treatment of multiwords and hyphenated words.

	Relations			Frequencies		
Class	**Noun**	**Verb**	**N+V**	**Noun**	**Verb**	**N+V**
Senses	71.79	52.89	63.24	71.79	52.89	63.24
Balkanet	73.06	53.82	64.37	73.06	53.82	64.37
Meaning	73.40	56.40	65.71	73.40	56.40	65.71
BLC–0	74.80	58.32	67.35	72.99	55.33	65.01
BLC–10	74.99	58.46	67.52	74.60	57.08	66.69
BLC–20	76.12	58.60	68.20	75.62	57.22	67.31
BLC–30	75.99	58.60	68.14	76.10	57.63	67.76
BLC–40	75.76	59.70	68.51	**78.03**	58.18	69.07
BLC–50	**76.22**	**59.83**	**68.82**	**78.03**	**58.87**	**69.38**
SuperSns	**81.87**	**79.23**	**80.68**	**81.87**	**79.23**	**80.68**

Table 6: *F1 measure for nouns and verbs using all relations and WN frequencies criteria for selecting BLC*

Furthermore, it is also worth mentioning that in this edition there were a few systems above the "WN first sense" baseline (4 out of 26 systems). Usually, this baseline is very competitive in WSD tasks, and it is extremely hard to improve upon even slightly.

Table 6 present for monosemous and polysemous nouns and verbs the F1 measures of the different sense-groupings obtained with all relations and WN frequencies criteria when training the class–frequencies on SemCor and testing on SensEval–3. Best results are marked using bold.

Obviously, higher accuracy figures are obtained when incorporating also monosemous words. Note this naive system achieves for Senses an F1 of 63.24, very similar to those reported in SensEval–3, and SuperSenses obtain a very high F1 of 80.68. Regarding the automatic BLC, the best results are obtained for BLC–50, but all of them outperform the BC from Balkanet and Meaning. However, for nouns and using all relations, BLC–20 (with 558 classes) obtain only slightly lower F1 figures than BLC–50 (with 253 classes). When using WN frequencies instead of all relations, BLC even achieve higher results but not all of them outperform the BC from Balkanet and Meaning.

Surprisingly, these naive Most frequent WSD systems trained on SemCor are able to achieve very high-levels of accuracy. For nouns, using BLC-20 (selected from all relations, 558 semantic labels) the system reaches 76.12, while using BLC-40 (selected from WN frequencies, 132 semantic labels) the system achieves 78.03. Finally, using SuperSenses for verbs (15 semantic labels) this naive system scores 79.23.

To our knowledge, the best results for class–based WSD are those reported by [3]. This system performs a sequence tagging using a perceptron–trained HMM, using SuperSenses, training on SemCor and testing on the SensEval–3. The system achieves an F1–score of 70.74, obtaining a significant improvement from a baseline system which scores only 64.09. In this case, the first sense baseline is the SuperSense of the most frequent synset for a word, according to the WN sense ranking. Possibly, the origin of the discrepancies between our results and those reported by [3] is twofold. First, because they use a BIO sequence schema for annotation, and second, the use of the brown-v part of SemCor to establish sense–frequencies.

7 Conclusions and further work

The WSD task seems to have reached its maximum accuracy figures with the usual framework. Some of its limitations could come from the sense–granularity of WordNet (WN). Moreover, it is not clear how WSD can contribute with the current result to improve other NLP tasks. Changing the set of classes could be a solution to enrich training corpora with many more examples. In fact, our most frequent naive systems are able to perform a semantic tagging with accuracy figures over 75%.

Base Level Concepts (BLC) are concepts that are representative for a set of other concepts. In the present work, a simple method for automatically selecting BLC from WN based on the hypernym hierarchy and the number of stored frequencies or relationships between synsets have been shown. Although, some sets of Base Concepts are available at this moment (e.g. EuroWordNet, Balkanet, Meaning), a huge manual effort should be invested for its development. Other sets of Base Concepts, like WN Lexicographer Files are clearly insufficient in order to describe and distinguish between the enormous number of concepts that are used in a text. Using a very simple baseline, the Most Frequent Class, our approach empirically shows a clear improvement over such other sets. In addition, our method is capable to get a more or less detailed sets of BLC without losing semantic discrimination power.

Other selection criteria for selecting BLC should be investigated. We are also interested in the direct comparison between automatically and manually selected BLC. Finally, we plan to use BLC for supervised class–based WSD.

References

[1] E. Agirre and O. LopezDeLaCalle. Clustering wordnet word senses. In *Proceedings of RANLP'03*, Borovets, Bulgaria, 2003.

[2] M. J. Atserias J., Climent S. and R. G. A proposal for a shallow ontologization of wordnet. In *Proceedings of the 21th Annual Meeting of the Sociedad Espaola para el Procesamiento del Lenguaje Natural*, Granada, Spain, 2005.

[3] M. Ciaramita and Y. Altun. Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'06)*, pages 594–602, Sydney, Australia, 2006. ACL.

[4] M. Ciaramita and M. Johnson. Supersense tagging of unknown nouns in wordnet. In *Proceedings of the Conference on Empirical methods in natural language processing (EMNLP'03)*, pages 168–175. ACL, 2003.

[5] J. Curran. Supersense tagging of unknown nouns using semantic similarity. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL'05)*, pages 26–33. ACL, 2005.

[6] J. Daudé, L. Padró, and G. Rigau. Validation and tuning of wordnet mapping techniques. In *Proceedings of the International Conference on Recent Advances on Natural Language Processing (RANLP'03)*, Borovets, Bulgaria., 2003.

[7] C. Fellbaum, editor. *WordNet. An Electronic Lexical Database.* The MIT Press, 1998.

[8] M. Hearst and H. Schütze. Customizing a lexicon to better suit a computational task. In *Proceedingns of the ACL SIGLEX Workshop on Lexical Acquisition*, Stuttgart, Germany, 1993.

[9] H. Kuĉera and W. N. Francis. *Computational Analysis of Present-Day American English.* Brown University Press, Providence, RI, USA, 1967.

[10] B. Magnini and G. Cavaglia. Integrating subject fields codes into wordnet. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)*, 2000.

[11] L. Màrquez, G. Escudero, D. Martínez, and G. Rigau. Supervised corpus-based methods for wsd. In *E. Agirre and P. Edmonds (Eds.) Word Sense Disambiguation: Algorithms and applications.*, volume 33 of *Text, Speech and Language Technology.* Springer, 2006.

[12] R. Mihalcea and D. Moldovan. Automatic generation of coarse grained wordnet. In *Proceding of the NAACL workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations*, Pittsburg, USA, 2001.

[13] I. Niles and A. Pease. Towards a standard upper ontology. In *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, pages 17–19. Chris Welty and Barry Smith, eds, 2001.

[14] W. Peters, I. Peters, and P. Vossen. Automatic sense clustering in eurowordnet. In *First International Conference on Language Resources and Evaluation (LREC'98)*, Granada, Spain, 1998.

[15] E. Rosch. Human categorisation. *Studies in Cross-Cultural Psychology*, I(1):1–49, 1977.

[16] F. Segond, A. Schiller, G. Greffenstette, and J. Chanod. An experiment in semantic tagging using hidden markov model tagging. In *ACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 78–81. ACL, New Brunswick, New Jersey, 1997.

[17] B. Snyder and M. Palmer. The english all-words task. In R. Mihalcea and P. Edmonds, editors, *Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pages 41–43, Barcelona, Spain, July 2004. Association for Computational Linguistics.

[18] L. Villarejo, L. Màrquez, and G. Rigau. Exploring the construction of semantic class classifiers for wsd. In *Proceedings of the 21th Annual Meeting of Sociedad Espaola para el Procesamiento del Lenguaje Natural SEPLN'05*, pages 195–202, Granada, Spain, September 2005. ISSN 1136-5948.

[19] P. Vossen, L. Bloksma, H. Rodriguez, S. Climent, N. Calzolari, A. Roventini, F. Bertagna, A. Alonge, and W. Peters. The eurowordnet base concepts and top ontology. Technical report, Paris, France, France, 1998.