Quality Assessment of Large Scale Knowledge Resources

Montse Cuadros
IXA NLP Group
EHU/UPV
Donostia, Basque Country
`mcuadros001@ikasle.ehu.es`

German Rigau
IXA NLP Group
EHU/UPV
Donostia, Basque Country
`german.rigau@ehu.es`

Abstract

This paper presents an empirical evaluation of the quality of publicly available large-scale knowledge resources. The study includes a wide range of manually and automatically derived large-scale knowledge resources. In order to establish a fair and neutral comparison, the quality of each knowledge resource is indirectly evaluated using the same method on a Word Sense Disambiguation task. The evaluation framework selected has been the Senseval-3 English Lexical Sample Task. The study empirically demonstrates that automatically acquired knowledge resources surpass both in terms of precision and recall the knowledge resources derived manually, and that the combination of the knowledge contained in these resources is very close to the most frequent sense classifier. As far as we know, this is the first time that such a quality assessment has been performed showing a clear picture of the current state-of-the-art of publicly available wide coverage semantic resources.

1 Introduction

Using large-scale semantic knowledge bases, such as WordNet (Fellbaum, 1998), has become a usual, often necessary, practice for most current Natural Language Processing systems. Even now, building large and rich enough knowledge bases for broad–coverage semantic processing takes a great deal of expensive manual effort involving large research groups during long periods of development. This fact has severely hampered the state-of-the-art of current Natural Language Processing (NLP) applications. For example, dozens of person-years have been invested in the development of wordnets for various languages (Vossen, 1998), but the data in these resources seems not to be rich enough to support advanced concept-based NLP applications directly. It seems that applications will not scale up to working in open domains without more detailed and rich general-purpose (and also domain-specific) linguistic knowledge built by automatic means.

For instance, in more than eight years of manual construction (from version 1.5 to 2.0), WordNet passed from 103,445 semantic relations to 204,074 semantic relations[1]. That is, around twelve thousand semantic relations per year. However, during the last years the research community has devised a large set of innovative processes and tools for large-scale automatic acquisition of lexical knowledge from structured or unstructured corpora. Among others we can mention eXtended WordNet (Mihalcea and Moldovan, 2001), large collections of semantic preferences acquired from SemCor (Agirre and Martinez, 2001; Agirre and Martinez, 2002) or acquired from British National Corpus (BNC) (McCarthy, 2001), large-scale Topic Signatures for each synset acquired from the web (Agirre and de la Calle, 2004) or acquired from the BNC (Cuadros et al., 2005).

Obviously, all these semantic resources have been acquired using a very different set of methods, tools and corpora, resulting on a different set of new semantic relations between synsets. In fact, each resource has different volume and accuracy figures. Although isolated evaluations have been performed by their developers in different experi-

---

[1]Symmetric relations are counted only once.

mental settings, to date no comparable evaluation has been carried out in a common and controlled framework.

This work tries to establish the relative quality of these semantic resources in a neutral environment. The quality of each large-scale knowledge resource is indirectly evaluated on a Word Sense Disambiguation (WSD) task. In particular, we use a well defined WSD evaluation benchmark (Senseval-3 English Lexical Sample task) to evaluate the quality of each resource.

Furthermore, this work studies how these resources complement each other. That is, to which extent each knowledge base provides new knowledge not provided by the others.

This paper is organized as follows: after this introduction, section 2 describes the large-scale knowledge resources studied in this work. Section 3 describes the evaluation framework. Section 4 presents the evaluation results of the different semantic resources considered. Section 5 provides a qualitative assessment of this empirical study and finally, the conclusions and future work are presented in section 6.

2 Large Scale Knowledge Resources

This study covers a wide range of large-scale knowledge resources: WordNet (WN) (Fellbaum, 1998), eXtended WordNet (Mihalcea and Moldovan, 2001), large collections of semantic preferences acquired from SemCor (Agirre and Martinez, 2001; Agirre and Martinez, 2002) or acquired from the BNC (McCarthy, 2001), large-scale Topic Signatures for each synset acquired from the web (Agirre and de la Calle, 2004) or acquired from the BNC (Cuadros et al., 2005).

However, although these resources have been derived using different WN versions, the research community has the technology for the automatic alignment of wordnets (Daudé et al., 2003). This technology provides a mapping among synsets of different WN versions, maintaining the compatibility to all the knowledge resources which use a particular WN version as a sense repository. Furthermore, this technology allows to port the knowledge associated to a particular WN version to the rest of WN versions already connected.

Using this technology, most of these resources are integrated into a common resource called Multilingual Central Repository (MCR) (Atserias et al., 2004). In particular, all WordNet versions, eX-

tended WordNet, and the semantic preferences acquired from SemCor and BNC.

2.1 Multilingual Central Repository

The Multilingual Central Repository (MCR)[2] follows the model proposed by the EuroWordNet project. EuroWordNet (Vossen, 1998) is a multilingual lexical database with wordnets for several European languages, which are structured as the Princeton WordNet. The Princeton WordNet contains information about nouns, verbs, adjectives and adverbs in English and is organized around the notion of a *synset*. A synset is a set of words with the same part-of-speech that can be interchanged in a certain context. For example, <*party*, *political_party*> form a synset because they can be used to refer to the same concept. A synset is often further described by a gloss, in this case: "an organization to gain political power". Finally, synsets can be related to each other by semantic relations, such as hyponymy (between specific and more general concepts), meronymy (between parts and wholes), cause, etc.

The current version of the MCR (Atserias et al., 2004) is a result of the 5th Framework MEANING project. The MCR integrates into the same EuroWordNet framework wordnets from five different languages (together with four English WordNet versions). The MCR also integrates WordNet Domains (Magnini and Cavaglià, 2000) and new versions of the Base Concepts and Top Concept Ontology. The final version of the MCR contains 1,642,389 semantic relations between synsets, most of them acquired by automatic means. This represents almost one order of magnitude larger than the Princeton WordNet (204,074 unique semantic relations in WordNet 2.0). Table 1 summarizes the main sources for semantic relations integrated into the MCR.

Table 2 shows the number of semantic relations between synsets pairs in the MCR and its overlappings. Note that, most of the relations in the MCR between synsets-pairs are unique.

Hereinafter we will refer to each semantic resource as follows:

- **WN** (Fellbaum, 1998): This knowledge resource uses the direct relations encoded in WordNet 1.6 or 2.0. We also tested WN-2 (using relations at distance 1 and 2) and WN-3 (using relations at distance 1, 2 and 3).

Source	#relations
Princeton WN1.6	138,091
Selectional Preferences from SemCor	203,546
Selectional Preferences from the BNC	707,618
New relations from Princeton WN2.0	42,212
Gold relations from eXtended WN	17,185
Silver relations from eXtended WN	239,249
Normal relations from eXtended WN	294,488
Total	1,642,389

Table 1: Main sources of semantic relations

Type of Relations	#relations
Total Relations	1,642,389
Different Relations	1,531,380
Unique Relations	1,390,181
Non-unique relations (>1)	70,425
Non-unique relations (>2)	341
Non-unique relations (>3)	8

Table 2: Overlapping relations in the MCR

- **XWN** (Mihalcea and Moldovan, 2001): This knowledge resource uses the direct relations encoded in eXtended WordNet.

- **XWN+WN**: This knowledge resource uses the direct relations included in WN and XWN.

- **spBNC** (McCarthy, 2001): This knowledge resource contains the selectional preferences acquired from the BNC.

- **spSemCor** (Agirre and Martinez, 2001; Agirre and Martinez, 2002): This knowledge resource contains the selectional preferences acquired from SemCor.

- **spBNC+spSemCor**: This knowledge resource uses the selectional preferences acquired from the BNC and SemCor.

- **MCR** (Atserias et al., 2004): This knowledge resource uses the direct relations included in MCR.

2.2 Automatically retrieved Topic Signatures

Topic Signatures (TS) are word vectors related to a particular topic (Lin and Hovy, 2000). Topic Signatures are built by retrieving context words of a target topic from large volumes of text. In our case, we consider word senses as topics. Basically, the acquisition of TS consists of A) acquiring the best possible corpus examples for a particular word sense (usually characterizing each word sense as a query and performing a search on the corpus for those examples that best match the queries), and then, B) building the TS by deriving the context words that best represent the word sense from the selected corpora.

For this study, we use the large-scale Topic Signatures acquired from the web (Agirre and de la Calle, 2004) and those acquired from the BNC (Cuadros et al., 2005).

- **TSWEB**[3]: Inspired by the work of (Leacock et al., 1998), these Topic Signatures were constructed using monosemous relatives from WordNet (synonyms, hypernyms, direct and indirect hyponyms, and siblings), querying Google and retrieving up to one thousand snippets per query (that is, a word sense). In particular, the method was as follows:

  - Organizing the retrieved examples from the web in collections, one collection per word sense.
  - Extracting the words and their frequencies for each collection.
  - Comparing these frequencies with those pertaining to other word senses using TFIDF (see formula 1).
  - Gathering in an ordered list, the words with distinctive frequency for one of the collections, which constitutes the Topic Signature for the respective word sense.

  This constitutes the largest available semantic resource with around 100 million relations (between synsets and words).

- **TSBNC**: These Topic Signatures have been constructed using ExRetriever[4], a flexible tool to perform sense queries on large corpora.

  - This tool characterizes each sense of a word as a specific query using a declarative language.
  - This is automatically done by using a particular query construction strategy, defined *a priori*, and using information from a knowledge base.

  In this study, ExRetriever has been evaluated using the BNC, WN as a knowledge base and

---

[3]http://ixa.si.ehu.es/Ixa/resources/sensecorpus
[4]http://www.lsi.upc.es/~nlp/meaning/downloads.html

TFIDF (as shown in formula 1) (Agirre and de la Calle, 2004)[5].

$$TFIDF(w, C) = \frac{wf_w}{max_w wf_w} \times log\frac{N}{Cf_w} \qquad (1)$$

Where w stands for word context, wf for the word frecuency, C for Collection (all the corpus gathered for a particular word sense), and Cf stands for the Collection frecuency.

In this study we consider two different query strategies:

- Monosemous A (**queryA**): (OR monosemous-words). That is, the union set of all synonym, hyponym and hyperonym words of a WordNet synset which are monosemous nouns (these words can have other senses as verbs, adjectives or adverbs).

- Monosemous W (**queryW**): (OR monosemous-words). That is, the union set of all words appearing as synonyms, direct hyponyms, hypernyms indirect hyponyms (distance 2 and 3) and siblings. In this case, the nouns collected are monosemous having no other senses as verbs, adjectives or adverbs.

While TSWEB use the query construction queryW, ExRetriever use both.

3   Indirect Evaluation on Word Sense Disambiguation

In order to measure the quality of the knowledge resources described in the previous section, we performed an indirect evaluation by using all these resources as Topic Signatures (TS). That is, word vectors with weights associated to a particular synset which are obtained by collecting those word senses appearing in the synsets directly related to them [6]. This simple representation tries to be as neutral as possible with respect to the evaluation framework.

All knowledge resources are indirectly evaluated on a WSD task. In particular, the noun-set

of Senseval-3 English Lexical Sample task which consists of 20 nouns. All performances are evaluated on the test data using the fine-grained scoring system provided by the organizers.

Furthermore, trying to be as neutral as possible with respect to the semantic resources studied, we applied systematically the same disambiguation method to all of them. Recall that our main goal is to establish a fair comparison of the knowledge resources rather than providing the best disambiguation technique for a particular semantic knowledge base.

A common WSD method has been applied to all knowledge resources. A simple word overlapping counting (or weighting) is performed between the Topic Signature and the test example[7]. Thus, the **occurrence** evaluation measure counts the amount of overlapped words and the **weight** evaluation measure adds up the weights of the overlapped words. The synset having higher overlapping word counts (or weights) is selected for a particular test example. However, for TSWEB and TSBNC the better results have been obtained using occurrences (the weights are only used to order the words of the vector). Finally, we should remark that the results are not skewed (for instance, for resolving ties) by the most frequent sense in WN or any other statistically predicted knowledge.

Figure 3 presents an example of Topic Signature from TSWEB using queryW and the web and from TSBNC using queryA and the BNC for the first sense of the noun party. Although both automatically acquired TS seem to be closely related to the first sense of the noun party, they do not have words in common.

As an example, table 4 shows a test example of Senseval-3 corresponding to the first sense of the noun party. In bold there are the words that appear in TSBNC-queryA. There are several important words that appear in the text that also appear in the TS.

4   Evaluating the quality of knowledge resources

In order to establish a clear picture of the current state-of-the-art of publicly available wide coverage knowledge resources we also consider a number of basic baselines.

---

[5]Although other measures have been tested, such as Mutual Information or Association Ratio, the best results have been obtained using TFIDF formula.

[6]A weight of 1 is given when the resource do not has associated weight.

[7]We also consider multiword terms.

democratic	0.0126	socialist	0.0062
tammany	0.0124	organization	0.0060
alinement	0.0122	conservative	0.0059
federalist	0.0115	populist	0.0053
missionary	0.0103	dixiecrats	0.0051
whig	0.0099	know-nothing	0.0049
greenback	0.0089	constitutional	0.0045
anti-masonic	0.0083	pecking	0.0043
nazi	0.0081	democratic-republican	0.0040
republican	0.0074	republicans	0.0039
alcoholics	0.0073	labor	0.0039
bull	0.0070	salvation	0.0038

party	4.9350	trade	1.5295
political	3.7722	parties	1.4083
government	2.4129	politics	1.2703
election	2.2265	campaign	1.2551
policy	2.0795	leadership	1.2277
support	1.8537	movement	1.2156
leader	1.8280	general	1.2034
years	1.7128	public	1.1874
people	1.7044	member	1.1855
local	1.6899	opposition	1.1751
conference	1.6702	unions	1.1563
power	1.6105	national	1.1474

Table 3: Topic Signatures for party#n#1 using TSWEB (24 out of 15881 total words) and TS-BNC(queryA) with TFIDF (24 out of 9069 total words)

<instance id="party.n.bnc.00008131" docsrc="BNC"> <context> Up to the **late** 1960s , catholic nationalists were **split** between two **main political** groupings . There was the Nationalist Party , a weak **organization** for which **local** priests had to **provide** some **kind** of legitimation . As a <head>party</head> , it really only exercised a modicum of **power** in relation to the Stormont administration . Then there were the republican **parties** who focused their attention on Westminster **elections** . The disorganized **nature** of catholic nationalist **politics** was only turned round with the emergence of the **civil rights movement** of 1968 and the subsequent forming of the SDLP in 1970 . </context> </instance>

Table 4: Example of test num. 00008131 for party#n which its correct sense is 1

4.1 Baselines

We have designed several baselines in order to establish a relative comparison of the performance of each semantic resource:

- **RANDOM**: For each target word, this method selects a random sense. This baseline can be considered as a lower-bound.

- **WordNet MFS (WN-MFS)**: This method selects the most frequent sense (the first sense in WordNet) of the target word.

- **TRAIN-MFS**: This method selects the most frequent sense in the training corpus of the target word.

- **Train Topic Signatures (TRAIN)**: This baseline uses the training corpus to directly build a Topic Signature using TFIDF measure for each word sense. Note that in this case, this baseline can be considered as an upper-bound of our evaluation framework.

Table 5 presents the F1 measure (harmonic mean of recall and precision) of the different baselines. In this table, TRAIN has been calculated with a fixed vector size of 450 words. As expected, RANDOM baseline obtains the poorest result while the most frequent sense of WordNet (WN-MFS) is very close to the most frequent sense of the training corpus (TRAIN-MFS), but

Baselines	F1
TRAIN	65.1
TRAIN-MFS	54.5
WN-MFS	53.0
RANDOM	19.1

Table 5: Baselines

both are far below to the Topic Signatures acquired using the training corpus (TRAIN).

4.2 Performance of the knowledge resources

Table 6 presents the performance of each knowledge resource uploaded into the MCR and the average size of its vectors. In bold appear the best results for precision, recall and F1 measures. The lowest result is obtained by the knowledge directly gathered from WN mainly because of its poor coverage (Recall of 17.6 and F1 of 25.6). Its performance is improved using words at distance 1 and 2 (F1 of 33.3), but it decreases using words at distance 1, 2 and 3 (F1 of 30.4). The best precision is obtained by WN (46.7), but the best performance is achieved by the combined knowledge of MCR-spBNC[8] (Recall of 42.9 and F1 of 44.1). This represents a recall 18.5 points higher than WN. That is, the knowledge integrated into the MCR (WordNet, eXtended WordNet and the selectional preferences acquired from SemCor) although partly derived by automatic means performs much better

---

[8]MCR without Selectional Preferences from BNC

KB	P	R	F1	Av. Size
MCR-spBNC	45.4	**42.9**	**44.1**	115
MCR	41.8	40.4	41.1	235
spSemCor	43.1	38.7	40.8	56
spBNC+spSemCor	41.4	30.1	40.7	184
WN+XWN	45.5	28.1	34.7	68
WN-2	38.0	29.7	33.3	72
XWN	45.0	25.6	32.6	55
WN-3	31.6	29.3	30.4	297
spBNC	36.3	25.4	29.9	128
WN	**46.7**	17.6	25.6	13

Table 6: P, R and F1 fine-grained results for the resources integrated into the MCR.

in terms of recall and F1 measures than using the knowledge currently present in WN alone (with a small decrease in precision). It also seems that the knowledge from spBNC always degrades the performance of their combinations[9].

Regarding the baselines, all knowledge resources integrated into the MCR surpass RANDOM, but none achieves neither WN-MFS, TRAIN-MFS nor TRAIN.

Figure 1 plots F1 results of the fine-grained evaluation on the nominal part of the English lexical sample of Senseval-3 of the baselines (including upper and lower-bounds), the knowledge bases integrated into the MCR, the best performing Topic Signatures acquired from the web and the BNC evaluated individually and in combination with others. The figure presents F1 (Y-axis) in terms of the size of the word vectors (X-axis)[10].

In order to evaluate more deeply the quality of each knowledge resource, we also provide some evaluations of the combined outcomes of several knowledge resources. The combinations are performed following a very simple voting method: first, for each knowledge resource, the scoring results obtained for each word sense are normalized, and then, for each word sense, the normalized scores are added up selecting the word sense with higher score.

Regarding Topic Signatures, as expected, in general the knowledge gathered from the web (TSWEB) is superior to the one acquired from the BNC either using queryA or queryW (TSBNC-queryA and TSBNC-queryW). Interestingly, the performance of TSBNC-queryA when using the

first two hundred words of the TS is slightly better than using queryW (both using the web or the BNC).

Although TSBNC-queryA and TSBNC-queryW perform very similar, both knowledge resources contain different knowledge. This is shown when combining the outcomes of these two different knowledge resources with TSWEB. While no improvement is obtained when combining the knowledge acquired from the web and the BNC when using the same acquisition method (queryW), the combination of TSWEB and TSBNC-queryA (TSWEB+ExRetA) obtains better F1 results than TSWEB (TSBNC-queryA have some knowledge not included into TSWEB).

Surprisingly, the knowledge integrated into the MCR (MCR-spBNC) surpass the knowledge from Topic Signatures acquired from the web or the BNC, using queryA, queryW or their combinations.

Furthermore, the combination of TSWEB and MCR-spBNC (TSWEB+MCR-spBNC) outperforms both resources individually indicating that both knowledge bases contain complementary information. The maximum is achieved with TS vectors of at most 700 words (with 49.3% precision, 49.2% recall and 49.2% F1). In fact, the resulting combination is very close to the most frequent sense baselines. This fact indicates that the resulting large-scale knowledge base almost encodes the knowledge necessary to behave as a most frequent sense tagger.

4.3 Senseval-3 system performances

For sake of comparison, tables 7 and 8 present the F1 measure of the fine-grained results for nouns of the Senseval-3 lexical sample task for the best and worst unsupervised and supervised systems, respectively. We also include in these tables some of the baselines and the best performing combination of knowledge resources (including TSWEB and MCR-spBNC)[11]. Regarding the knowledge resources evaluated in this study, the best combination (including TSWEB and MCR-spBNC) achieves an F1 measure much better than some supervised and unsupervised systems and it is close to the most frequent sense of WordNet (WN-MFS) and to the most frequent sense of the training corpora (TRAIN-MFS).

---

[9]All selectional preferences acquired from SemCor or the BNC have been considered including those with very low confidence score.

[10]Only varying the size of TS for TSWEB and TSBNC.

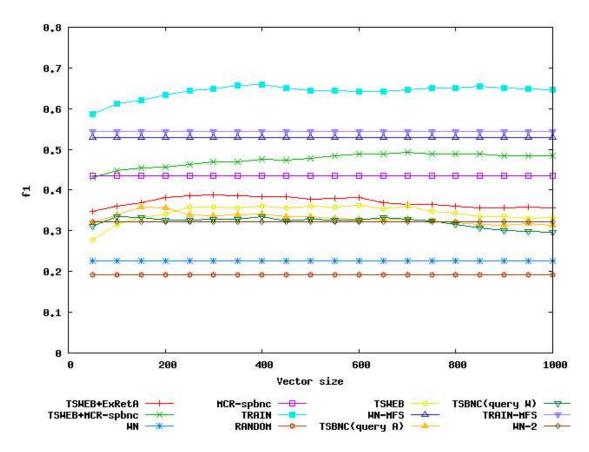[11]Although we maintain the classification of the organizers, system s3_wsdiit used the train data.

Figure 1: Fine-grained evaluation results for the knowledge resources

s3 systems	F1
s3_wsdiit	68.0
WN-MFS	53.0
Comb TSWEB MCR-spBNC	49.2
s3_DLSI	17.8

Table 7: Senseval-3 Unsupervised Systems

s3 systems	F1
htsa3 U.Bucharest (Grozea)	74.2
TRAIN	65.1
TRAIN-MFS	54.5
DLSI-UA-LS-SU U.Alicante (Vazquez)	41.0

Table 8: Senseval-3 Supervised Systems

We must recall that the main goal of this research is to establish a clear and neutral view of the relative quality of available knowledge resources, not to provide the best WSD algorithm using these resources. Obviously, much more sophisticated WSD systems using these resources could be devised.

5 Quality Assessment

Summarizing, this study provides empirical evidence for the relative quality of publicly available large-scale knowledge resources. The relative quality has been measured indirectly in terms of precision and recall on a WSD task.

The study empirically demonstrates that automatically acquired knowledge bases clearly surpass both in terms of precision and recall the knowledge manually encoded from WordNet (using relations expanded to one, two or three levels).

Surprisingly, the knowledge contained into the MCR (WordNet, eXtended WordNet, Selectional Preferences acquired automatically from SemCor) is of a better quality than the automatically acquired Topic Signatures. In fact, the knowledge resulting from the combination of all these large-scale resources outperforms each resource individually indicating that these knowledge bases contain complementary information. Finally, we should remark that the resulting combination is very close to the most frequent sense classifiers.

Regarding the automatic acquisition of large-scale Topic Signatures it seems that those acquired from the web are slightly better than those acquired from smaller corpora (for instance, the BNC). It also seems that queryW performs better than queryA but that both methods (queryA and

540

queryW) also produce complementary knowledge. Finally, it seems that the weights are not useful for measuring the strength of a vote (they are only useful for ordering the words in the Topic Signature).

6   Conclusions and future work

During the last years the research community has derived a large set of semantic resources using a very different set of methods, tools and corpus, resulting on a different set of new semantic relations between synsets. In fact, each resource has different volume and accuracy figures. Although isolated evaluations have been performed by their developers in different experimental settings, to date no complete evaluation has been carried out in a common framework.

In order to establish a fair comparison, the quality of each resource has been indirectly evaluated in the same way on a WSD task. The evaluation framework selected has been the Senseval-3 English Lexical Sample Task. The study empirically demonstrates that automatically acquired knowledge bases surpass both in terms of precision and recall to the knowledge bases derived manually, and that the combination of the knowledge contained in these resources is very close to the most frequent sense classifier.

Once empirically demonstrated that the knowledge resulting from MCR and Topic Signatures acquired from the web is complementary and close to the most frequent sense classifier, we plan to integrate the Topic Signatures acquired from the web (of about 100 million relations) into the MCR. This process will be performed by disambiguating the Topic Signatures. That is, trying to obtain word sense vectors instead of word vectors. This will allow to enlarge the existing knowledge bases in several orders of magnitude by fully automatic methods. Other evaluation frameworks such as PP attachment will be also considered.

7   Acknowledgements

References

E. Agirre and O. Lopez de la Calle. 2004. Publicly available topic signatures for all wordnet nominal senses. In *Proceedings of LREC*, Lisbon, Portugal.

E. Agirre and D. Martinez. 2001. Learning class-to-class selectional preferences. In *Proceedings of CoNLL*, Toulouse, France.

E. Agirre and D. Martinez. 2002. Integrating selectional preferences in wordnet. In *Proceedings of GWC*, Mysore, India.

J. Atserias, L. Villarejo, G. Rigau, E. Agirre, J. Carroll, B. Magnini, and Piek Vossen. 2004. The meaning multilingual central repository. In *Proceedings of GWC*, Brno, Czech Republic.

M. Cuadros, L. Padró, and G. Rigau. 2005. Comparing methods for automatic acquisition of topic signatures. In *Proceedings of RANLP*, Borovets, Bulgaria.

J. Daudé, L. Padró, and G. Rigau. 2003. Validation and Tuning of Wordnet Mapping Techniques. In *Proceedings of RANLP*, Borovets, Bulgaria.

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

C. Leacock, M. Chodorow, and G. Miller. 1998. Using Corpus Statistics and WordNet Relations for Sense Identification. *Computational Linguistics*, 24(1):147–166.

C. Lin and E. Hovy. 2000. The automated acquisition of topic signatures for text summarization. In *Proceedings of COLING*. Strasbourg, France.

B. Magnini and G. Cavaglià. 2000. Integrating subject field codes into wordnet. In *Proceedings of LREC*, Athens. Greece.

D. McCarthy. 2001. *Lexical Acquisition at the Syntax-Semantics Interface: Diathesis Aternations, Subcategorization Frames and Selectional Preferences*. Ph.D. thesis, University of Sussex.

R. Mihalcea and D. Moldovan. 2001. extended wordnet: Progress report. In *Proceedings of NAACL Workshop on WordNet and Other Lexical Resources*, Pittsburgh, PA.

P. Vossen, editor. 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks* . Kluwer Academic Publishers .