Integrating and Porting Knowledge across Languages

Jordi Atserias and **Luís Villarejo** and **German Rigau**
TALP Research Center
Jordi Girona Salgado, 1-3.
08034 Barcelona
{batalla, luisv}@talp.upc.es

IXA Group
Euskal Herriko Unibertsitatea
Donostia.
{rigau}@si.ehu.es

Abstract

This paper describes the first version of the Multilingual Central Repository. Currently, MCR integrates into the same EuroWordNet framework, five local wordnets (including three versions of the English WordNet from Princeton), the EuroWordNet Top Concept ontology, MultiWordNet Domains, and hundreds of thousand of new semantic relations and properties automatically acquired from corpora. In fact, the resulting MCR is going to constitute the largest and richest multilingual lexical–knowledge ever build.

1 Introduction

In order to develop a trustable semantic web infrastructure and a multilingual ontology framework to support open–domain knowledge management, a wide range of techniques are required to progressively automate the knowledge lifecycle. This involves extracting high-level meaning from large collections of content data and its representation and management in a common knowledge base.

However, even now, building large and rich knowledge bases takes a great deal of expensive manual effort; this has severely hampered Knowledge-Technologies and HLT application development. For example, dozens of person-years have been invested into the development of wordnets (Fellbaum 98) for various languages (Vossen 98; Bentivogli *et al.* 02), but the data in these resources is still not sufficiently rich to support advanced multilingual concept-based HLT applications directly. Furthermore, resources produced by introspection usually fail to register what really occurs in texts.

The MEANING project (Rigau *et al.* 02) [1] identifies two complementary and intermediate tasks which are crucial in order to enable the next generation of intelligent open domain HLT application systems: Word Sense Disambiguation (WSD) and large-scale enrichment of Lexical Knowledge Bases.

The advance in these two areas will allow large-scale acquisition of shallow meaning from texts, in the form of relations among concepts. WSD provides the technology to convert relations between words into relations between concepts.

However, progress is difficult due to the following interdependence:

- In order to achieve accurate WSD, we need far more linguistic and semantic knowledge than is available in current lexical knowledge bases (e.g. current wordnets).
- In order to enrich existing Knowledge Bases we need to acquire information from corpora accurately tagged with word senses.

MEANING proposes an innovative bootstrapping process to deal with this inter–dependency between WSD and knowledge acquisition exploiting a multilingual architecture based on EuroWordNet(Vossen 98).

MEANING plans to perform three consecutive cycles of large-scale WSD and acquisition processes in five European languages including Basque, Catalan, English, Italian and Spanish. As languages realize the meaning in different ways, some semantic relations that can be difficult to acquire in one language can be easy to capture in other languages.

In MEANING, the knowledge acquired for each language during the three consecutive cycles will be consistently upload and integrated into the respective local wordnets, and then ported and distributed across the rest of wordnets, balancing resources and technological advances across languages.

The Multilingual Central Repository (MCR) will grant the consistency and integrity of all the semantic knowledge produced by MEANING.

This paper describes the first version of the Multilingual Central Repository produced after

[1] http://www.lsi.upc.es/~nlp/meaning/meaning.html

the first MEANING cycle. After this introduction, section 2 presents the MCR structure, content and associated software tools. While section 3 describes the first uploading process, section 4 is devoted to the porting process. Finally section 5 draws some conclusions and future work.

2 Multilingual Central Repository

The MCR acts as a multilingual interface for integrating and distributing all the knowledge acquired in MEANING.

2.1 MCR structure

The MCR follows the model proposed by the EuroWordNet project. EuroWordNet is a multilingual lexical database with wordnets for several European languages, which are structured as the Princeton WordNet (Fellbaum 98).

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. Synsets are related to each other by semantic relations, such as hyponymy (between specific and more general concepts), meronymy (between parts and wholes), cause, entailment, etc.

The EuroWordNet architecture includes the **Inter–Lingual–Index** (ILI), a **Domain ontology** and a **Top Concept ontology** (Vossen 98). The ILI consists of a list ILI-records which interconnects word meanings in the local wordnets. During the EuroWordNet project, around 1,000 ILI-records were selected as **Base Concepts** and consistently assigned to the **Top Concept ontology**[2].

Using the **Inter–Lingual–Index**, wordnets are interconnected so that it is possible to go from the words in one language to similar words in the other languages.

The ILI is enhanced, enriched and structured by two separate ontologies:

- The **Top Concept ontology**, which is a hierarchy of language-independent concepts, reflecting important semantic distinctions, e.g. Object and Substance, Location, Dynamic.
- The **Domain ontology**, which is a hierarchy of domain labels, which are knowledge structures grouping meanings in terms of topics or scripts, e.g. Transport, Sports, Medicine, Gastronomy.

The main purpose of the **Top Concept ontology** is to provide a common framework for all the wordnets. It consists of 63 basic semantic distinctions that classify the Base Concepts. The Base concepts are a set of ILI-records connected to WordNet which represents the most important concepts in the different wordnets.

The **Domain ontology** groups concepts in a different way, based on scripts rather than classification. MCR uses the MultiWordNet Domains (Magnini & Cavagli 00) which were partially derived from the Dewey Decimal Classification [3]. WordNet Domains is a hierarchy of 165 Domain Labels associated to WordNet 1.6.

Information brought by Domain Labels is complementary to what is already in WordNet. First of all Domain Labels may include synsets of different syntactic categories: for instance MEDICINE groups together senses from nouns, such as *doctor* and *hospital*, and from verbs such as *to operate*. Second, a Domain Label may also contain senses from different WordNet subhierarchies. For example, SPORT contains senses such as **athlete**, deriving from *person*, **game equipment**, from *artifact*, **sport** from *act*, and **playing field**, from *location*.

The knowledge acquired locally is uploaded and ported across the rest of languages via the EuroWordNet ILI, maintaining the compatibility among them. In that way, the ILI structure (including the **Top Concept ontology** and the **Domain ontology**) will act as a natural backbone to transfer the different knowledge acquired from each local wordnet to the rest of wordnets.

2.2 MCR content

The first version of the MCR includes only conceptual knowledge. This means that only semantic relations between synsets will be acquired, upload and ported across local wordnets. However, when necessary, the relations acquired can be underspecified. For instance:

INVOLVED ← INVOLVED-PATIENT ← INVOLVED-RESULT

[2] http://www.illc.uva.nl/EuroWordNet/corebcs/topont.html [3] `http://www.oclc.org/dewey`

Although these relations are not completely specified, they will be uploaded and ported to be ready useful for other acquisition processes and languages. For instance, consider the following relation <*gain*> INVOLVED <*money*> captured as typical object. Although, this relation may be further refined into <*gain*> INVOLVED-PATIENT <*money*> in posterior cycles, other processes (like those that locate Spanish examples from large text collections) can take profit from a ported relation <*ganar*> INVOLVED <*dinero*>.

The first version of the MCR integrates:

- ILI
  - WordNet 1.6
  - EuroWordNet Base Concepts
  - EuroWordNet Top Concept ontology
  - MultiWordNet Domains
- Local wordnets
  - English WordNet 1.5, 1.6, 1.7.1
  - Basque, Catalan, Italian and Spanish wordnets
- Large collections of semantic preferences
  - Acquired from SemCor
  - Acquired from BNC
- Instances
  - Named Entities

2.3 MCR Access

The MCR also provides a web interface to the database based on Web EuroWordNet Interface[4]. Three different APIs have been also developed to provide flexible access to the MCR: first, a SOAP API to allow users to interact with the MCR, an extension of WNQUERY perl API to the MCR and a C++ API for high performance software.

3 Uploading Process

Once finished the first part of uploading all the data (checking errors and inconsistencies), a more complex second part must be performed. This second part consists of the correct integration of every piece of information into the MCR. That is, linking correctly all this knowledge to the ILI. This second part involves a complex cross–checking validation process and some complex expansion of large amounts of semantic properties and relations through the semantic structure.

The current version of the MCR uses Princeton WordNet 1.6 as ILI. Initially most of the knowledge to be uploaded into the MCR has been derived from WordNet 1.6 (automatic selectional preferences acquired from SemCor and BNC) and the Italian WordNet and the MultiWordNet Domains, both developed at IRST are using WordNet 1.6 as ILI (Bentivogli *et al.* 02; Magnini & Cavagli 00). This option also minimises side effects with other European initiatives (Balkanet, EuroTerm, etc.) and wordnet developments around Global WordNet Association. However, the ILI for Spanish, Catalan and Basque wordnets was WordNet 1.5(Atserias *et al.* 97; Benítez *et al.* 98), as well as the EuroWordNet Top Concept ontology and the associated Base Concepts.

3.1 Uploading local wordnets based on WordNet1.5

Although the technology to provide compatibility across wordnets exists (Daudé *et al.* 99; Daudé *et al.* 00; Daudé *et al.* 01)[5], uploading local wordnets based on WordNet1.5 to the MCR is a complex process, because between different wordnet versions, synsets can be split (1:N), joined (N:1), added (0:1) or deleted (1:0) through mapping. Thus, even if we perform manual checking of these connections, for those remaining cases of splitting or joining synsets the information inside the synsets should be verified accordingly.

The whole process of the porting wordnets using ILI based on WordNet1.5 to the new ILI based on WordNet1.6 consist of:

1. For all split synsets, all information of synset 1.5, including variants, is copied to each of the equivalent synsets in 1.6
2. For all joined synsets, all information of synsets 1.5, including variants, is copied to the equivalent synset in 1.6
3. Manual revision to validate the information assigned to split and joined synsets.

3.2 Conceptual coverage

Table 1 shows the overlapping nouns, verbs and adjectives between each wordnet pair.

At a synset level, noun overlapping is quite high and homogeneous between wordnet pairs. The maximum overlapping occurs between English and Spanish (29,502) and the lowest between Italian and Catalan (14,462).

[4] `http://nipadio.lsi.upc.es/wei.html`

[5] `http://www.lsi.upc.es/~tools/mapping.html`

NOUN	en16	spwn	itwn	cawn	bawn
en16	66,025	29,502	22,634	26,197	22,722
spwn	-	31,241	16,355	24,582	19,020
itwn	-	-	25,402	14,462	15,000
cawn	-	-	-	26,947	16,763
bawn	-	-	-	-	24,461
VERB	**en16**	**spwn**	**itwn**	**cawn**	**bawn**
en16	12,127	7,464	4,281	4,952	3,138
spwn	-	7,563	3,071	3,789	2,809
itwn	-	-	4,312	2,358	1,844
cawn	-	-	-	5,051	2,333
bawn	-	-	-	-	3,237
ADJ	**en16**	**spwn**	**itwn**	**cawn**	**bawn**
en16	17,915	11,087	2,658	4,028	0
spwn	-	11,135	1,700	3,932	0
itwn	-	-	2,686	611	0
cawn	-	-	-	4,076	0
bawn	-	-	-	-	0

Table 1: Overlapping between wordnet pairs

For verbs, at a synset level, the overlapping is also quite high but less uniform between wordnet pairs. The maximum overlapping occurs also between English and Spanish (7,464) and the lowest between Italian and Basque (1,844).

At a synset level, adjective overlapping is not high because some wordnets provide poor coverage on adjectives. While Spanish provides good overlapping with English (the maximum overlapping with 11,087 synsets), Basque wordnet do not provide adjectives at all.

3.3 Uploading Base Concepts

The original set of **Base Concepts** from EuroWordNet based on WordNet 1.5 totalized 1,030 ILI-records. Now, the Base Concepts from WordNet 1.5 has been mapped to Wordnet 1.6. After a manual revision and expansion to all WordNet 1.6 top beginners, the resulting Base Concepts for WordNet 1.6 totalized 1,535 ILI-records. In that way, the new version of Base Concepts covers the complete hierarchy of ILI-records.

3.4 Uploading the Top Ontology

The purpose of the EuroWordNet **Top Concept ontology** was to enforce more uniformity and compatibility of the different wordnet developments.

We performed also an automatic expansion of the **Top Concept** properties assigned to the Base Concepts. That is, we enriched the complete ILI structure with features coming from the Base Concepts by inheriting the Top Concept features following the hyponymy relationship.

The Eurowordnet project only performed a complete validation of the consistency of the **Top Concept ontology** of the Base Concepts. However, the classification of WordNet is not always consistent with the **Top Concept ontology**.

The following incompatibilities defined inside the **Top Concept ontology** have been used to impede the top–down propagation of the **Top Concept** properties:

- substance - object
- plant - animal - human - creature
- natural - artifact
- solid - liquid - gas

Now, we also plan to cross–check the **Top Concept ontology** expansion and the **Domain ontology** with the SUMO ontology (Niles & Pease 01).

3.5 Uploading Selectional Preferences

The first version of the MCR has been also enriched by a large amount of new relations. A total of 390,549 weighted Selectional Preferences (SPs) (see Table 2) obtained from two different corpora and using different approaches has been uploaded into the MCR. The first set of weighted SPs was obtained by means of probability distributions over the noun hierarchy of WordNet1.6 using the parsed trees generated by RASP (Carroll *et al.* 98) from the BNC (McCarthy 01). The second set was obtained from generalizations of the grammatical relations extracted using MiniPar (Lin 98) from Semcor (Agirre & Martinez 01; Agirre & Martinez 02).

The SPs have been included in MCR as noun–verb relations (ROLE) [6]. Although we can distinguish subjects and objects in the database, all of them have been included as a more general ROLE relation, and in fact, most of them overlap.

	#verbal synsets	#nominal synsets	#relations
Semcor SUBJ	2,490	5,398	69,840
Semcor DOBJ	3,423	6,964	110,102
BNC SUBJ	6,151	2,588	95,065
BNC DOBJ	6,125	4,185	115,542

Table 2: Selectional Preferences

[6] In EuroWordNet, INVOLVED and ROLE relationships were defined symmetric

	Spanish		English		Italian	
Relations	**UPLOAD**	**PORT0**	**UPLOAD**	**PORT0**	**UPLOAD**	**PORT0**
be_in_state	1,302	=	1,300	+2	364	+2
causes	240	=	224	+19	117	+15
near_antonym	7,444	=	7,449	+221	3,266	=
near_synonym	10,965	=	21,858	+19	4,887	+54
role	106	=	0	+106	0	+46
role_agent	516	=	0	+516	0	+227
role_instrument	291	=	0	+291	0	+151
role_location	83	=	0	+83	0	+39
role_patient	6	=	0	+6	0	+3
xpos_fuzzynym	37	=	0	+37	0	+23
xpos_near_synonym	319	=	0	+319	0	+181
Other relations	31,644	=	29,120	+2,627	9,541	+22
Total	53,272	=	59,951	+4,246	18,175	+763
role_agent-semcor	0	+52,394	69,840	=	0	+41,910
role_agent-bnc	0	+67,109	95,065	=	0	+40,853
role_patient-semcor	0	+80,378	110,102	=	0	+41,910
role_patient-bnc	0	+79,443	115,102	=	0	+50,264
Role	0	+279,324	390,109	=	0	+174,937
Instances	0	+1,599	0	+2,198	791	=
Proper Nouns	1,806	=	17,842	=	2,161	=
Base Concepts	1,169	=	1,535	=	0	+935
Domains Links	0	+55,239	109,621	=	35,174	=
Domains Synsets	0	+48,053	96,067	=	30,607	=
Top Ontology Links	3,438	=	0	+4,148	0	+2,544
Top Ontology Synsets	1,290	=	0	+1,554	0	+946

Table 3: PORT0 Main figures for Spanish, English and Italian

4 Porting Process

Having all these types of different knowledge and properties coming from different sources and methods, and completely expanded through the whole MCR, a new set of inference mechanisms can be devised to further infer new relations and knowledge inside the MCR. For instance, new relations could be generated when detecting particular *semantic patterns* occurring for some synsets having certain ontological properties, for a particular Domain, etc. That is, new relations could be generated when combining different methods and knowledge. For instance, creating new explicit relations (regular polisemy, nominalizations, etc.) when several relations derived in the integration process have confidence scores greater than certain thresholds, occurring between certain ontological properties, etc.

Obviously, new research is also needed for porting the various types of knowledge across languages. For instance, new ways to validate the ported knowledge in the target languages.

4.1 First Porting Process Results

In the first porting process all the knowledge integrated into the MCR has been ported (distributed) directly to the local wordnets (no extra semantic knowledge has been inferred).

However, by means of the first porting process, all wordnets have gained some kind of new knowledge coming from other wordnets. A direct result of the upload/integration/porting effort is that all information associated to the ILIs has been automatically ported to the other wordnets. Thus, MultiWordNet Domains are now available to the rest of local wordnets, the EuroWordnet Top Concept ontology is also available for Italian MultiWordNet and for English WordNet 1.6. Moreover, local relations have been also ported to the rest of wordnets. Thus, Italian and English WordNet has been enriched with all the new set of relations coming from EuroWordNet. In turn, Basque, Catalan, Italian and Spanish wordnets has been extensively enriched with large amounts of Selectional Preferences acquired automatically from English.

Table 3 summarises the main results before (UPLOAD0) and after the whole porting process (PORT0) for Spanish, English and Italian. In this table, relations do not consider hypo/hypernym relations and *links* stands for total number of Domains or Top Concept ontology properties ported (before the top-down expansion).

4.2 The *Vaso* example

When uploading and porting coherently all this semantic knowledge into the MCR a full range of new possibilities appear for improving both Acquisition and WSD processes. We will illustrate the current content of MCR by a simple example. The Spanish noun *vaso* has three possible senses in the Spanish WordNet. Now, the MCR integrates consistently a large set of explicit knowledge about the senses of *vaso* that can be used to differentiate and characterize better their particular meanings.

VASO_1 02755829-n 06-NOUN.ARTIFACT FACTOTUM	
English	drinking_glass glass
Italian	bicchiere
Basque	edontzi baso edalontzi
Catalan	got vas
Gloss	a glass container for holding liquids while drinking
Top Concept ontology	
1stOrderEntity-Form-Object 1stOrderEntity-Origin-Artifact 1stOrderEntity-Function-Container 1stOrderEntity-Function-Instrument	

Table 4: Vaso_1

VASO_3 09914390-n 23-NOUN.QUANTITY NUMBER	
English	glassful glass
Italian	bicchierata bicchiere
Basque	basocada
Catalan	got vas
Gloss	the quantity a glass will hold
Top Concept ontology	
1stOrderEntity-Composition-Part 2ndOrderEntity-SituationType-Static 2ndOrderEntity-SituationComponent-Quantity	

Table 5: Vaso_3

While sense 1 of *vaso* (see Table 4) corresponds to the container sense of *glass*, sense 3 (see Table 5) corresponds to the quantity a glass holds. This is represented in the MCR as follows. On the one hand, vaso_1 is connected to the same ILI as the English synset <drinking_glass glass>, which belongs to the Semantic File ARTIFACT and having no specific MutiWordNet Domain (FACTOTUM). On the other hand, vaso_3 is connected to <glassful glass>, and the Semantic File is QUANTITY and its corresponding domain is FACTOTUM-NUMBER. The Top Concept ontology also provides further clues about their meanings: vaso_1 includes the *Function-Container* property and vaso_3 the *SituationComponent-Quantity*. Further, coming from the weighted Selectional Preferences acquired from SemCor, we know that the typical things that somebody does with vaso_1 are for instance the corresponding equivalent translations to Spanish for <*polish, shine, smooth, smoothen*> or <*beautify, embellish, prettify*>, and with vaso_3 <*drink imbibe*> or <*consume have ingest take take_in*>. Finally, we must add that this also holds for the rest of languages connected.

5 Conclusions and Future Work

The first version of the MCR integrates now into the same EuroWordNet framework (using an upgraded release of Base Concepts and Top Concept ontology and MultiWordNet Domains) five local wordnets (with three English WordNet versions) with hundreds of thousand of new semantic relations, instances and properties fully expanded. All wordnets gained some kind of new knowledge coming from other wordnets by means of the first porting process. In fact, the resulting MCR is the largest and richest multilingual lexical knowledge base ever build.

In this way, this version of the MCR produced by MEANING is going to constitute a natural multilingual large-scale knowledge resource for a number of semantic processes that need large amounts of linguistic knowledge to be effective tools (e.g. Semantic Web ontologies).

Future versions of the MCR may include language dependent information (including syntactic information, subcategorization frames, diathetic alternations, etc), following the current standards for representing this information, i.e. the EAGLES recommendations, LCS Lexical Conceptual Structures, complex semantic relations (Lin & Pantel 01), etc.

We need to investigate new inference facilities to enhance the uploading and porting process as suggested before. Now, after full expansion (**Realization**) of the EuroWordNet Top Con-

cept ontology properties, we plan a full expansion through the nominal part of the hierarchy of the selectional preferences acquired from SemCor and BNC (and possibly other implicit semantic knowledge currently available in WordNet such as meronymy information).

We also plan further investigation to perform also full bottom–up expansion (**Generalization**), rather than merely expanding top–down the knowledge and properties represented into the MCR. In this case, different knowledge and properties can collapse on particular Base Concepts, Semantic Files, Domains and/or Top Concepts.

With respect the *porting process*, we plan to investigate also a new set of inference mechanisms in order to further infer new explicit relations and knowledge (regular polisemy, nominalizations, etc).

Finally, new research is also needed to verify the various types of semantic knowledge ported across languages.

Acknowledgments

The authors want to thank the anonymous reviewers for their useful comments and suggestions in order to prepare the final version of the paper.

This research has been partially funded by the Spanish Research Department (HERMES TIC2000-0335-C03-02) and by the European Commission (MEANING IST-2001-34460).

References

(Agirre & Martinez 01) E. Agirre and D. Martinez. Learning class-to-class selectional preferences. In *Proceedings of CoNLL01*, Toulouse, France, 2001.

(Agirre & Martinez 02) E. Agirre and D. Martinez. Integrating selectional preferences in wordnet. In *Proceedings of the first International WordNet Conference in Mysore*, India, 21-25 January 2002.

(Atserias *et al.* 97) J. Atserias, S. Climent, X. Farreres, G. Rigau, and H. Rodríguez. Combining multiple methods for the automatic construction of multilingual wordnets. In *Procceeding of RANLP'97*, pages 143–149, Bulgaria, 1997.

(Benítez *et al.* 98) L. Benítez, S. Cervell, G. Escudero, M. López, G. Rigau, and M. Taulé. Methods and tools for building the catalan wordnet. In *Proceedings of the ELRA Workshop on Language Resources for European Minority Languages, First International Conference on Language Resources & Evaluation*, Granada, Spain, 1998.

(Bentivogli *et al.* 02) L. Bentivogli, E. Pianta, and C. Girardi. Multiwordnet: developing an aligned multilingual database. In *First International Conference on Global WordNet*, Mysore, India, 2002.

(Carroll *et al.* 98) J. Carroll, G. Minnen, and E. Briscoe. Can subcategorisation probabilities help a statistical parser? In *Proceedings of the Sixth ACL/SIGDAT Workshop on Very Large Corpora*, pages 118–126, 1998.

(Daudé *et al.* 99) J. Daudé, L. Padró, and G. Rigau. Mapping Multilingual Hierarchies Using Relaxation Labeling. In *Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC'99)*, Maryland, US, 1999.

(Daudé *et al.* 00) J. Daudé, L. Padró, and G. Rigau. Mapping WordNets Using Structural Information. In *Proceedings of 38th annual meeting of the Association for Computational Linguistics (ACL'2000)*, Hong Kong, 2000.

(Daudé *et al.* 01) J. Daudé, L. Padró, and G. Rigau. A complete wn1.5 to wn1.6 mapping. In *Proceedings of NAACL Workshop "WordNet and Other Lexical Resources: Applications, Extensions and Customizations"*, Pittsburg, PA, United States, 2001.

(Fellbaum 98) C. Fellbaum, editor. *WordNet. An Electronic Lexical Database*. The MIT Press, 1998.

(Lin & Pantel 01) D. Lin and P. Pantel. Discovery of inference rules for question answering. *Natural Language Engineering*, 7(4):343–360, 2001.

(Lin 98) D. Lin. Extracting collocations from text corpora. In *Proceedigns of First Workshop on Computational Terminology*, Montreal, Canada, 1998.

(Magnini & Cavagli 00) B. Magnini and G. Cavagli. Integrating subject field codes into wordnet. In *In Proceedings of the Second Internatgional Conference on Language Resources and Evaluation LREC'2000*, Athens. Greece, 2000.

(McCarthy 01) D. McCarthy. *Lexical Acquisition at the Syntax-Semantics Interface: Diathesis Aternations, Subcategorization Frames and Selectional Preferences*. Unpublished PhD thesis, University of Sussex, 2001.

(Niles & Pease 01) I. Niles and A. Pease. Towards a standard upper ontology. In *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.

(Rigau *et al.* 02) G. Rigau, B. Magnini, E. Agirre, P. Vossen, and J. Carroll. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of COLLING Workshop*, Taipei, Taiwan, 2002.

(Vossen 98) P. Vossen, editor. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks* . Kluwer Academic Publishers , 1998.