Multilingual Central Repository version 3.0: upgrading a very large lexical knowledge base

Aitor González Agirre, Egoitz Laparra, German Rigau
Basque Country University
Donostia, Basque Country
{aitor.gonzalez, egoitz.laparra, german.rigau}@ehu.es

Abstract

This paper describes the upgrading process of the Multilingual Central Repository (MCR). The new MCR uses WordNet 3.0 as Interlingual-Index (ILI). Now, the current version of the MCR integrates in the same EuroWordNet framework wordnets from five different languages: English, Spanish, Catalan, Basque and Galician. In order to provide ontological coherence to all the integrated wordnets, the MCR has also been enriched with a disparate set of ontologies: Base Concepts, Top Ontology, WordNet Domains and Suggested Upper Merged Ontology. We also suggest a novel approach for improving some of the semantic resources integrated in the MCR, including a semi-automatic method to propagate domain information. The whole content of the MCR is freely available.

1 Introduction

Building large and rich knowledge bases is a very costly effort which involves large research groups for long periods of development. For instance, hundreds of person-years have been invested in the development of wordnets for various languages (Fellbaum, 1998; Vossen, 1998; Tufis et al., 2004; K. et al., 2010). In the case of the English WordNet, in more than ten years of manual construction (from 1995 to 2006, that is, from version 1.5 to 3.0), WordNet grew from 103,445 to 235,402 semantic relations[1], which represents a growth of around one thousand new relations per month.

The Multilingual Central Repository (MCR)[2] (Atserias et al., 2004b) follows the model proposed by the European project EuroWordNet (LE-2 4003) (Vossen, 1998). The MCR is the result of the European MEANING project (IST-2001-34460) (Rigau et al., 2002), as well as projects KNOW (TIN2006-15049-C03)[3] (Agirre et al., 2009), KNOW^2 (TIN2009-14715-C04)[4] and several complementary actions associated to the KNOW^2 project. The original MCR was aligned to the 1.6 version of WordNet. In the framework of the KNOW^2 project, we decided to upgrade the MCR to be aligned to a most recent version of WordNet.

The previous version of the MCR, aligned to the English 1.6 WordNet version, also integrated the eXtended WordNet project (Mihalcea and Moldovan, 2001), large collections of selectional preferences acquired from SemCor (Agirre and Martinez, 2001) and different sets of named entities (Alfonseca and Manandhar, 2002). It was also enriched with semantic and ontological properties as Top Ontology (Álvez et al., 2008), SUMO (Pease et al., 2002) or WordNet Domains (Magnini and Cavaglià, 2000).

The new MCR integrates wordnets of five different languages, including English, Spanish, Catalan, Basque and Galician. This paper presents the work carried out to upgrade the MCR to new versions of these resources. By using technology to automatically align wordnets (Daudé et al., 2003), we have been able to transport knowledge from different WordNet versions. Thus, we can maintain the compatibility between all the knowledge bases that use a particular version of WordNet as a sense repository. However, most of the ontological knowledge have not been directly ported from the previous version of the MCR.

Furthermore, WordNet Domains[5] was generated semi-automatically and has never been verified completely. Additionaly, it was aligned to

[1] Symmetric relations are counted only once.
[2] http://adimen.si.ehu.es/web/MCR
[3] http://ixa.si.ehu.es/know
[4] http://ixa.si.ehu.es/know2
[5] http://wndomains.fbk.eu/

WordNet 1.6. Thus, one goal of this work is the automatic construction of a new semantic resource derived from WordNet Domains and aligned to WordNet 3.0.

To assist in the correction and maintenance of the integrated resources in the MCR, we also adapted and enhanced the Web EuroWordNet Interface (WEI) in both consult and edit modes.

2 Multilingual Central Repository 3.0

The first version of the MCR was built following the model proposed by the EuroWordNet project. The EuroWordNet architecture includes the Inter-Lingual Index (ILI), a Domain Ontology and a Top Ontology (Vossen, 1998).

Initially most of the knowledge uploaded into the MCR was aligned to WordNet 1.6 and the Spanish, Catalan, Basque and Italian WordNet and the MultiWordNet Domains, were using WordNet 1.6 as ILI (Bentivogli et al., 2002; Magnini and Cavaglià, 2000). Thus, the original MCR used Princeton WordNet 1.6 as ILI. This option also minimized side efects with other European initiatives (Balkanet, EuroTerm, etc.) and wordnet developments around Global WordNet Association. Thus, the Spanish, Catalan and Basque wordnets as well as the EuroWordNet Top Ontology and the associated Base Concepts were transported from its original WordNet 1.5 to WordNet 1.6 (Atserias et al., 1997; Benítez et al., 1998; Atserias et al., 2004a).

The release of new free versions of Spanish and Galician wordnets aligned to Princeton WordNet 3.0 (Fernández-Montraveta et al., 2008; Xavier et al., 2011) brought with it the need to update the MCR and transport all its previous content to a new version using WordNet 3.0 as ILI. Thus, as a first step, we decided to transport Catalan and Basque wordnets and the ontological knowledge: Base Concepts, SUMO, WordNet Domains and Top Ontology.

2.1 Upgrading from 1.6 to 3.0

This section describes the process carried out for adapting the MCR to ILI 3.0. Due to its size and complexity, all this process have been mainly automatic.

To perform the porting between the wordnets 1.6 and 3.0 we have followed a similar process to the one used to port the Spanish and Catalan versions from 1.5 to 1.6 (Atserias et al., 2004a).

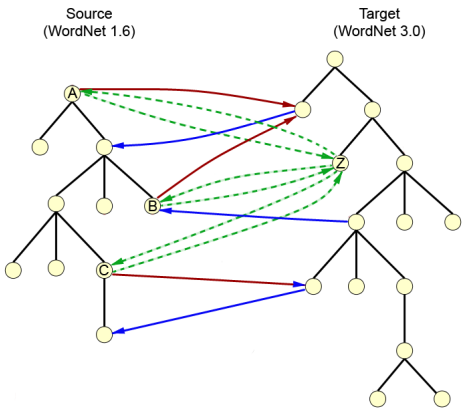

Figure 1: Example of a multiple intersection in the mapping between two versions of WordNet.

Upgrading ILI: The algorithm to align wordnets (Daudé et al., 2000; Daudé et al., 2001; Daudé et al., 2003) produces two mappings for each POS, one in each direction (from 1.6 to 3.0, and from 3.0 to 1.6). To upgrade the ILI, different approaches were applied depending on the POS.

For nouns, those synsets having multiple mappings from 1.6 to 3.0 were checked manually (Pociello et al., 2008).

For verbs, adjectives and adverbs, for those synsets having multiple mappings, we took the intersection between the two mappings (from 1.6 to 3.0, and from 3.0 to 1.6).

Upgrading WordNets: Finally, using the previous mapping, we transported from ILI 1.6 to ILI 3.0 the Basque (Pociello et al., 2008) and Catalan (Benítez et al., 1998) wordnets. The English WordNet was uploaded directly from the source files while the Spanish (Fernández-Montraveta et al., 2008) and Galician (Xavier et al., 2011) wordnets were directly uploaded from their database dumps.

It is possible to have multiple intersections for a source synset. When multiple intersections collapsed into the same target synset, we decided to join the set of variants from the source synsets to the target synset.

Figure 1 shows an example of this particular case (the intersections are displayed as dot lines). Therefore, the variants of the synsets A, B and C of WordNet 1.6 will be placed together in the synset Z of WordNet 3.0.

Upgrading Base Concepts: We used *Base Concepts* directly generated for WN 3.0[6]

[6]http://adimen.si.ehu.es/web/BLC

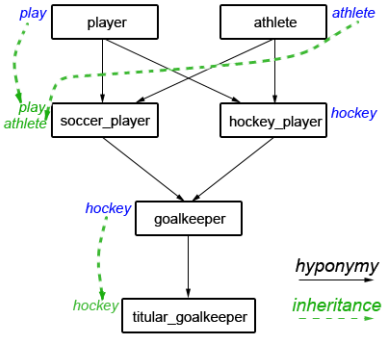

Figure 2: Example of a multiple intersection in the mapping between two versions of WordNet.

(Izquierdo et al., 2007).

Upgrading SUMO: SUMO has been directly ported from version 1.6 using the mapping. Those unlabelled synsets have been filled through inheritance. The ontology of the previous version is a modified version of SUMO, trimmed and polished, to allow the use of first-order theorem provers (like *E-prover* or *Vampire*) for formal reasoning, called AdimenSUMO[7]. The next step is to update AdimenSUMO using the latest version of SUMO for WordNet 3.0 (available on the website of SUMO[8]).

Upgrading WordNet Domains: As SUMO, what is currently in the MCR has been transported directly from version 1.6 using the mapping. Again, those unlabelled synsets have been filled through inheritance. In addition, new versions have been generated using graph techniques (see Section 4 for a detailed description of the process).

Upgrading the Top Ontology: Similar to SUMO and WordNet Domains, what is currently available in the MCR has been transported directly from version 1.6 using the mapping. Once more, those unlabelled synsets have been filled through inheritance. It remains to check the incompatibilities between labels following (Álvez et al., 2008).

An example of how to perform the process of inheritance used for SUMO, WordNet Domains and Top Ontology is shown in Figure 2. The example is presented for domains, but it can be applied to the other two cases.

Figure 2 shows a sample hierarchy where each node represents a synset. The domains are displayed on the sides. The inherited domain labels are highlighted using dot lines. In this specific example synset ***soccer_player*** inherits labels *play* and *athlete* from its hypernyms ***player*** and ***athlete***, respectively. Note that synset ***hockey_player*** does not inherit any label form its hypernyms because of it owns a domain (*hockey*). Similarly, synset ***goalkeeper*** does not inherit domains coming from the synset ***soccer_player***. Finally, synset ***titular_goalkeeper*** inherits *hockey* domain (but neither *play* nor *athlete* domains).

Thus, some of the current content of the MCR will require a future revision. Fortunately, by cross-checking its ontological knowledge most of these errors can be easily detected.

2.2 Web EuroWordNet Interface

WEI is a web application that allows consulting and editing the data contained in the MCR and navigating throw them. Consulting refers to exploring the content of the MCR by accessing words, a synsets, a variants or ILIs. The interface presents different searching parameters and displays the query results. The different searching parameters are:

- **Item**: a value to search for, it can be a Word, a Synset a Variant or an ILI.
- **Item type**: the type of item to search for: Word, a Synset a Variant or an ILI.
- **PoS**: the item' s gramatical cathegory or Part of Speech: Nouns, Verbs, Adjectives, Adverbs.
- **Search**: the type of search and subsearches (which are dinamically loaded from the database): Synonyms, Hyponyms, etc.
- **WordNet Source**: the WordNet from which navigate.
- **Navigation WordNet**: the WordNet to which navigate.
- **Gloss**: if selected it shows the glosses of the Synsets.
- **Score**: if selected, it shows the confidence factor.
- **Rels**: if selected, it shows information about the relations that each Synset has in all the target languages.

[7] http://adimen.si.ehu.es/web/adimenSUMO
[8] http://www.ontologyportal.org/

- **Full**: if selected, makes a recursive search.
- **Target WordNets**: the target WordNets of our search.

2.3 Automatic translations

The new version of WEI is able to use *Automatic Translation Web Services* for translating automatically the glosses and examples from other wordnets. This new feature helps users to complete and/or improve the gloss or examples of a given WordNet more quicky. Both glosses and examples are taken from the original English WordNet and translated to the target language. Suggestions for glosses and/or examples will appear below the existing ones, and may choose the most appropriate. In the current version, the translations of the glosses and examples are translated only from English (despite the possibility of translating from any available source).

2.4 Marks for synsets and variants

In the new version of WEI it is possible to assign a mark to a variant or synset to indicate special properties. We can also write a small note or comment to explain better the reason to assign that mark.

The available marks are the following:

- Variant marks:
  - DUBLEX: For those variant with dubious lexicalization.
  - INFL: Indicates that the variant is a inflected one.
  - RARE: Old fashioned or rarely used variant.
  - SUBCAT: Subcategorization.
  - VULG: For those variants that are vulgar, rude, or offensive.
- Synset marks:
  - GENLEX: Non-lexicalized general concepts that are introduced to better organize the hierarchy.
  - HYPLEX: Indicates that the hypernym has identical lexicalization.
  - SPECLEX: Domain specific terms that should be checked.

2.5 User management

We also included a new user access control to WEI. The previous user access control to WEI was carried out using Apache, in the Operating System. This implies that the access control and user management was done outside WEI. The MCR is being edited in a distributed way. Several research groups are editing the MCR in some of the languages. Each group has different users. Thus, the responsibility of managing the users is also distributed.

3 Current state of the MCR

In this section provide some information about the current state of the MCR, including the progress over the English WordNet.

Tables 1, 2 and 3 present respectively the current number of synsets and variants, the number of glosses and the number of examples of each wordnet per PoS.

4 A proposal for upgrading WordNet Domains

WordNet Domains[9] (WND) is a lexical resource developed at ITC-IRST where synsets have been semi-automatically annotated with one or more domain labels from a set of 165 hierarchically organized domains (Magnini and Cavaglià, 2000). WND allow to reduce the polysemy degree of the words, grouping those senses that belong to the same domain (Magnini et al., 2002).

But the semi-automatic method used to develop this resource was not free of errors and inconsistencies. By cross-checking the ontological content of the MCR it is possible to find some of these problems. For instance, noun synset <diver_1 frogman_1 underwater_diver_1> defined as *someone who works underwater* has domain *history* because it inherits from its hypernym <explorer_1 adventurer_2>.

4.0.1 Domain inheritance

WND was developed using WordNet 1.6. One consequence of the automatic mapping that we used to upgrade version 1.6 to 3.0 is that many synsets were left unlabeled (because there are new synsets, changes in the structure, etc.).

One of the first tasks undertaken has been to fill these gaps. For them, we has carried out a propagation of the labels by inheritance for nominal and verbal synsets. The inherent structure of WordNet for adjectives and adverbs makes this spread

[9] http://wndomains.fbk.eu/

WordNet	Nouns	Verbs	Adjectives	Adverbs	Synsets	WN %
EngWN3.0	147,360	25,051	30,004	5,580	118,431	100%
SpaWN3.0	40,009	11,107	7,005	1,106	59,227	50%
CatWN3.0	51,598	11,577	7,679	2	46,027	39%
EusWN3.0	41,071	9,472	148	0	30,615	26%
GalWN3.0	9,114	1,413	4,866	0	9,320	8%

Table 1: Current number of synsets and variants of each WN.

WordNet	Nouns	Verbs	Adjectives	Adverbs	Synsets	WN %
EngWN3.0	82,379	13,767	18,156	3,621	117,923	100%
SpaWN3.0	13,014	3,469	1,965	687	19,135	16%
CatWN3.0	6,289	44	840	1	7,174	6%
EusWN3.0	2,854	78	0	0	2,932	2%
GalWN3.0	4,997	2	3,111	0	8,111	7%

Table 2: Current number of glosses of each WN.

not trivial. Therefore, this simple process has been carried out only for nouns and verbs.

We have worked exclusively on those synsets that had no labels at all. We inherited the labels from its hypernyms. If a synset has more than one hypernym, the domain labels are taken from all of them. We used a small list of incompatible labels to detect incompatibilities. Therefore, the same synset can not be both *factotum* and *biology*, or *animals* and *plants*.

This process increased our domain information by nearly a 18-19%, as shown in Tables 4 and 5:

PoS	Before	After	Increase
Nouns	66,595	83,286	+25%
Verbs	12,219	14,224	+16%
All	100,315	119,011	+19%

Table 4: Number of synsets with domain labels.

PoS	Before	After	Increase
Nouns	87,938	108,665	+24%
Verbs	13,026	15,051	+16%
All	124,551	146,899	+18%

Table 5: Total number of domain labels.

However, this process may also have propagated innapropriate domain labels to unlabeled synsets. It remains for future research an accurate evaluation of this new resource.

In the next section we present some examples using a new graph-based method for propagating domain labels through WordNet. Additionaly, the method can also be used to detect anomalies in the original WND labels.

4.0.2 A new graph based method

UKB[10] algorithm (Agirre and Soroa, 2009) applies personalized PageRank on a graph derived from a wordnet. This algorithm has proven to be very competitive on Word Sense Disambiguation tasks and it is easily portable to other languages that have a wordnet (Agirre et al., 2010). Now, we present a novel use of the UKB algorithm for propagating information through a wordnet structure.

Given an input context, '*ukb_ppv*' (*Personalized PageRank Vector*) algorithm outputs a ranking vector over the nodes of a graph, after applying a *Personalized PageRank* over it. We just need to use a wordnet as a knowledge base and pass to the application the contexts we want to process, performing a kind of *spreading activation* through the WordNet structure.

As context we use those synsets labelled with a particular domain. Thus, for each of the 169[11] domain labels included in the MCR we generate a context. Each file contains the list of offsets corresponding to those synsets with a particular domain label. After creating the context file, we just need to execute '*ukb_ppv*' that returns a ranking of weights for each WordNet synset with respect to that particular domain.

[10] http://ixa2.si.ehu.es/ukb/

[11] Excluding *factotum* labels.

WordNet	Nouns	Verbs	Adjectives	Adverbs	Synsets	WN %
EngWN3.0	10,433	11,583	15,615	3,674	41,305	100%
SpaWN3.0	478	27	195	967	700	2%
CatWN3.0	2,103	46	368	0	2,517	6%
EusWN3.0	2,377	0	0	0	2,377	6%
GalWN3.0	270	2	4,291	0	4,563	11%

Table 3: Current number of examples of each WN.

Once made the process for all domains we have weights for each synset and for each of the domains. Therefore, we know which are the highest weights for each domain and the highest weights for each synset. This allows us to estimate which synsets are more representative of each domain and which domains are best for each synset.

Basically, what we do is to mark some synsets with a domain (using the labels we already know from the original porting process) and use the wordnet graph to propagate the new labelling. We work on the assumption that a synset directly related to several synsets labelled with a particular domain (i.e *biology*) would itself possibly be also related somehow to that domain (i.e. *biology*). Therefore, it makes no sense to use the domain *factotum* for this technique.

Table 6 shows the first ten domains and weights resulting from the application of this method on synset <diver_1 frogman_1 underwater_diver_1>. The suggestions of the algorithm seems to improve the current labeling because it suggests *sub* (possibly the best one) and *diving* (possibly, the second best option). Moreover, the method suggests the wrong label with a much lower weight.

Weight	Domain
0.0144335:	sub
0.0015939:	diving
0.0001725:	swimming
0.0001297:	history
0.0000557:	nautical
0.0000529:	fashion
0.0000412:	jewellery
0.0000315:	ethnology
0.0000274:	archaeology
0.0000204:	gas

Table 6: PPV weight rankings for sense $diver_n^1$.

Table 7 shows the first ten domains and weights resulting from the application of this method on synset <pornography_1 porno_1 porn_1 erotica_1 smut_5> defined as *creative activity (writing or pictures or films etc.) of no literary or artistic value other than to stimulate sexual desire* and labelled with the domain *law*. The suggestions of the algorithm seems to improve the current labeling because it suggests *sexuality* (possibly the best one) and *cinema* (possibly, the second best option). Moreover, the wrong label dissapears.

Method 3	
Weight	**Domain**
0.000123453:	sexuality
0.000112444:	cinema
0.000077780:	theatre
0.000075525:	painting
0.000062377:	telecommunication
0.000060640:	publishing
0.000050370:	psychological_features
0.000047003:	photography
0.000046853:	artisanship
0.000040458:	graphic_arts

Table 7: PPV weight rankings for sense $porno_n^1$.

5 Concluding Remarks and Future Directions

As a result of this work, the current version of the MCR consistently maintains new wordnet versions for five languages (English, Spanish, Catalan, Basque and Galician), and the ontological knowledge from WordNet Domains, Top Ontology and SUMO.

In particular, the main contributions of our work can be summarized as follows:

We have created a new version of the MCR using WordNet 3.0 as ILI.

We have improved the existing Web EuroWordNet Interface (WEI) (both consult and edit interfaces) to work with the new version of the MCR. Now, the interface includes automatic translation

facilities, making it easier and faster the development of the resources integrated into the MCR. We also added new editing facilities for recording new linguistic information associated to the variants and synsets.

We have uploaded into the new version of the MCR the English WordNet 3.0, the new Spanish WordNet 3.0 (Fernández-Montraveta et al., 2008) and a new Galician WordNet 3.0.

We have used a complete mapping from WordNet 1.6 to WordNet 3.0 (covering not only nouns, but verbs, adjectives and adverbs) to transport the Basque and Catalan wordnets and the ontological knowledge from the existing version of the MCR (using WordNet 1.6 as ILI) to the new MCR version (using WordNet 3.0 as ILI).

We have applied a very simple estrategy to complete the ontological information by exploiting basic inheritance mechanisms. This process has been applied to WordNet Domains, Top Ontology and SUMO.

We have also investigated a new approach for consistently propagating domain formation through the WordNet structure by exploiting a well-known graph algorithm using UKB. Although an exhaustive empirical evaluation should be addressed in a near future, a preliminary review of the new resources created using this process presents very interesting insights for future research.

Obviously, further investigation is needed to assess the quality of the new labelling of WordNet Domains. We plan to evaluate the quality of these new resources indirectly by comparing their performance on a common Word Sense Disambiguation task. We would also like to continue studing different ways for selecting the most appropriate set of domain labels per synset. We also plan to derive domain information from Wikipedia by exploiting WordNet++ (Navigli and Ponzetto, 2010).

The whole content of the MCR and the new WEI is freely available [12].

Moreover, the maintenance of this type of resources is continuous, and all the integrated knowledge should be constantly updated and revised.

Acknowledgments

We thank the IXA NLP group from the Basque Country University. This work was been possible thanks to the support of the FP7 PATHS project (ICT-2010-270082) and the Spanish national projects KNOW (TIN2006-15049-C03) and KNOW2 (TIN2009-14715-C04-04).

[12] http://adimen.si.ehu.es/web/MCR

References

Agirre, E., Cuadros, M., Rigau, G., and Soroa, A. (2010). Exploring Knowledge Bases for Similarity. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC10). European Language Resources Association (ELRA). ISBN: 2-9517408-6-7. Pages 373–377.".*

Agirre, E. and Martinez, D. (2001). Knowledge sources for Word Sense Disambiguation. In *Proceedings of the Fourth International Conference TSD 2001, Plzen (Pilsen), Czech Republic. Published in the Springer Verlag Lecture Notes in Computer Science series. Václav Matousek, Pavel Mautner, Roman Moucek, Karel Tauser (eds.) Copyright Springer-Verlag. ISBN: 3-540-42557-8. ".*

Agirre, E., Rigau, G., Castellón, I., Alonso, L., Padró, L., Cuadros, M., Climent, S., and Coll-Florit, M. (2009). KNOW: Developing large-scale multilingual technologies for language understanding. *Procesamiento del Lenguaje Natural*, (43):377–378.

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

Alfonseca, E. and Manandhar, S. (2002). Distinguishing concepts and instances in WordNet. In *Proceedings of the first International Conference of Global WordNet Association*, Mysore, India.

Álvez, J., Atserias, J., Carrera, J., Climent, S., Laparra, E., Oliver, A., and Rigau, G. (2008). Complete and Consistent Annotation of WordNet using the Top Concept Ontology. In *Proceedings of the the 6th Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech (Morocco).

Atserias, J., Climent, S., Farreres, J., Rigau, G., and Rodríguez, H. (1997). Combining Multiple Methods for the Automatic Construction of Multilingual WordNets. In *Proceedings of International Conference on Recent Advances in Natural Language Processing (RANLP'97)*, Tzigov Chark, Bulgaria.

Atserias, J., Rigau, G., and Villarejo, L. (2004a). Spanish WordNet 1.6: Porting the Spanish Wordnet across Princeton versions. In *LREC'04*.

Atserias, J., Villarejo, L., Rigau, G., Agirre, E., Carroll, J., Vossen, P., and Magnini, B. (2004b). The MEANING Multilingual Central Respository. In *Proceedings of the Second International Global WordNet Conference (GWC'04)*.

Bentivogli, L., Pianta, E., and Girardi, C. (2002). MultiWordNet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*, Mysore, India.

Benítez, L., Cervell, S., Escudero, G., López, M., Rigau, G., and Taulé, M. (1998). Methods and Tools for Building the Catalan WordNet. In *Proceedings of ELRA Workshop on Language Resources for European Minority Languages*, Granada, Spain.

Daudé, J., Padró, L., and Rigau, G. (2000). Mapping WordNets Using Structural Information. In *38th Annual Meeting of the Association for Computational Linguistics (ACL'2000).*, Hong Kong.

Daudé, J., Padró, L., and Rigau, G. (2001). A Complete WN1.5 to WN1.6 Mapping. In *Proceedings of NAACL Workshop 'WordNet and Other Lexical Resources: Applications, Extensions and Customizations' (NAACL'2001).*, Pittsburg, PA, USA.

Daudé, J., Padró, L., and Rigau, G. (2003). Making Wordnet Mappings Robust. In *Proceedings of the 19th Congreso de la Sociedad Española para el Procesamiento del Lenguage Natural, SEPLN*, Universidad Universidad de Alcalá de Henares. Madrid, Spain.

Fellbaum, C. (1998). *WordNet. An Electronic Lexical Database*. Language, Speech, and Communication. The MIT Press.

Fernández-Montraveta, A., Vázquez, G., and Fellbaum, C. (2008). *Text Resources and Lexical Knowledge*, volume 33 of *Text, Translation, Computational Processing*, chapter The Spanish Version of WordNet 3.0, pages 175–182. Mouton de Gruyter.

Izquierdo, R., Suárez, A., and Rigau, G. (2007). Exploring the Automatic Selection of Basic Level Concepts. In *Proceedings of RANLP*.

K., R., S., T., T., C., V., S., and H., I. (2010). WNMS: Connecting the Distributed WordNet in the Case of Asian WordNet. In *Proceedings of the 5th Global WordNet Conference (GWC'10)*, Mumbai (India).

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

Magnini, B., Satrapparava, C., Pezzulo, G., and Gliozzo, A. (2002). The Role of Domains Informations. In *In Word Sense Disambiguation*, Treto, Cambridge.

Mihalcea, R. and Moldovan, D. (2001). eXtended WordNet: Progress Report. In *Proceedings of NAACL Workshop* WordNet and Other Lexical Resources: Applications, Extensions and Customizations, pages 95–100, Pittsburg, PA, USA.

Navigli, R. and Ponzetto, S. P. (2010). Building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 216–225, Uppsala, Sweden.

Pease, A., Niles, I., and Li, J. (2002). The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In *AAAI-2002*.

Pociello, E., Gurrutxaga, A., Agirre, E., Aldezabal, I., and Rigau, G. (2008). WNTERM: Combining the Basque WordNet and a Terminological Dictionary. In *6th international conference on Language Resources and Evaluation, LREC'08*.

Rigau, G., Magnini, B., Agirre, E., Vossen, P., and Carroll, J. (2002). MEANING: A Roadmap to Knowledge Technologies. In *Proceedings of COLING Workshop A Roadmap for Computational Linguistics*, Taipei, Taiwan.

Tufis, D., Cristea, D., and Stamou, S. (2004). BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. *Romanian Journal on Science Technology of Information. Special Issue on Balkanet*, 7(3–4):9–44.

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

Xavier, G. G., Clemente, X. M. G., Pereira, A. G., and Lorenzo, V. T. (2011). Galnet: WordNet 3.0 do galego. *Linguamática*, 3(1):61–67.