

# WNTERM: Enriching the MCR with a terminological dictionary

**Eli Pociello, Antton Gurrutxaga**

Elhuyar R&D

Zelai Haundi kalea, 3. Osinalde Industrialdea, 20170 Usurbil. Basque Country

E-mail: {agurrutxaga, eli}@elhuyar.com

**Eneko Agirre, Izaskun Aldezabal, German Rigau**

IXA NLP Research Group

649 pk. 20.080 – Donostia. Basque Country

E-mail: {e.agirre, izaskun.aldezabal, german.rigau}@ehu.es

## Abstract

In this paper we describe the methodology and the first steps for the creation of WNTERM (from WordNet and Terminology), a specialized lexicon produced from the merger of the EuroWordNet-based Multilingual Central Repository (MCR) and the Basic Encyclopaedic Dictionary of Science and Technology (BDST). As an example, the ecology domain has been used. The final result is a multilingual (Basque and English) light-weight domain ontology, including taxonomic and other semantic relations among its concepts, which is tightly connected to other wordnets.

## 1. Introduction

In the last seven years the IXA research group has been working on the Basque WordNet (Agirre et al., 2006). The focus has been on the representation of common vocabulary (we have incorporated around 27,000 words), but now, we have turned our attention to specialized language.

Due to the unstoppable development of specialized language and terminology, it is becoming increasingly difficult to capture and organize terminological information. Domain ontologies are helpful to handle this kind of information, and have been shown to be useful in knowledge representation, management and exchange. For instance, there are several representative ontologies in the domains of e-commerce (UNSPSC<sup>1</sup>, NAICS<sup>2</sup>), medicine (GALEN<sup>3</sup>, UMLS), engineering – EngMath (Gruber & Olsen, 1994), PhysSys (Borst, 1997)–, enterprise –Enterprise Ontology (Uschold et al., 1998)–, and knowledge management –KA (Decker et al. 1999). Moreover, in the NLP community domain ontologies are being used to develop and evaluate different computational systems and applications (Navigli et al. 2003; Sagri et al., 2003; Stamou et al., 2002; Roventini & Marinelli, 2003).

The aim of WNTERM is to create a light-weight ontology belonging to the science and technology domains for Basque and English. WNTERM (i) stores the domain terminology, (ii) fixes relations among all the domain terms, and (iii) is connected to the Basque and English wordnets through the Multilingual Central Repository (MCR) (Atserias et al., 2004).

In order to have WNTERM linked to the wordnets currently available, we decided to structure it following

the MCR framework, which we also used to build the Basque WordNet. The MCR model provides English-Basque equivalence links across the English and Basque wordnets, and it also offers the possibility of enriching both the Basque WordNet and WNTERM simultaneously.

Domain ontology terms are imported into WNTERM from the MCR itself (more precisely, from the Basque and English wordnets) and the in-house *Basic Encyclopaedic Dictionary of Science and Technology* (BDST) data-base<sup>4</sup>. Taking this sources into account, there are additional objectives which we want to address with WNTERM: (i) to extend the Basque WordNet with terminology, (ii) to organize hierarchically all the terms of the BDST according to the MCR architecture, in order to provide to lexicographers more specific information about a term in their navigation through the data-base, and (iii) to link all the sources in the project: the MCR, the BDST and the domain ontology (WNTERM).

The final result of this project will be a domain ontology on science and technology, which will show the taxonomic and semantic relations among all its domain concepts, and which is also linked to the MCR and the BDST. In this paper we focus on the overall methodology for the construction of WNTERM, which will be illustrated on the ecology domain, a subset of science and technology.

This paper is organized as follows. The resources used in this project are introduced in Section 2 (MCR and the Basque WordNet) and Section 3 (BDST). In Section 4, we present the methodology followed to create a domain ontology, illustrated on the ecology domain. Finally, future work is presented in Section 5.

<sup>1</sup> <http://www.unspsc.org>

<sup>2</sup> <http://www.naics.com>

<sup>3</sup> <http://opengalen.org>

<sup>4</sup> <http://www.zientzia.net/hiztegia/index.asp>

## 2. The Multilingual Central Repository

The Multilingual Central Repository (MCR)<sup>5</sup> follows the model proposed by the EuroWordNet project<sup>6</sup>. EuroWordNet (Vossen, 1998) is a multilingual semantic lexicon with wordnets for several European languages, which are structured as the Princeton WordNet (Fellbaum, 1998).

It groups each languages' words into sets of synonyms called synsets, and records various semantic relations (such as hypernymy, hyponymy, meronymy, holonymy) between these synonym sets forming a hierarchy. Each of these synsets corresponds to a lexical concept and many have a textual gloss which often provides an explanation of what this concept represents.

The MCR (Atserias et al., 2004) is a result of the 5th Framework Meaning project<sup>7</sup> (Rigau et al., 2003). The MCR integrates in the same EuroWordNet framework wordnets from five different languages, including Spanish, Italian, Catalan and Basque (together with six English WordNet versions).

The wordnets are currently linked via an Inter-Lingual-Index (ILI) allowing the connection from words in one language to translation equivalent words in any of the other languages. In that way, the MCR constitutes a natural multilingual large-scale linguistic resource for a number of semantic processes that need large amount of multilingual knowledge to be effective tools. For instance, the English synset {*party*, *political\_party*} is linked through the ILI to the Basque synset {*partidu*, *partidu\_politiko*, *alderdi\_politiko*, *alderdi*}.

The MCR also integrates the latest version of the WordNet Domains (Magnini & Cavaglià, 2000), new versions of the Base Concepts and the Top Concept Ontology (Álvarez et al., 2008), and the SUMO ontology (Niles & Pease, 2001). The current version of the MCR contains 934,771 semantic relations between synsets, most of them acquired by automatic means. This represents almost four times larger than the Princeton WordNet (235,402 unique semantic relations in WordNet 3.0).

Although these resources have been derived using different WordNet versions, using the technology for the automatic alignment of wordnets (Daudé et al., 2001)<sup>8</sup>, most of these resources have been integrated in the MCR maintaining the compatibility among all the knowledge resources which use a particular WordNet version as a sense repository.

MCR was developed based on WordNet 1.6 version. However, for this project we have moved the MCR to the latest WordNet version (3.0), because WordNet 3.0 has a higher amount of terminology than the 1.6. The result of the automatic mapping was manually corrected.

<sup>5</sup> <http://adimen.si.ehu.es/cgi-bin/wei5/public/wei.consult.perl>

<sup>6</sup> <http://www.ilc.uva.nl/EuroWordNet/>

<sup>7</sup> <http://www.lsi.upc.edu/~nlp/meaning>

<sup>8</sup> [http://www.lsi.upc.edu/~nlp/web/index.php?option=com\\_content&task=view&id=21&Itemid=57](http://www.lsi.upc.edu/~nlp/web/index.php?option=com_content&task=view&id=21&Itemid=57)

We will now present WordNet domains and the Basque WordNet.

### 2.1 WordNet Domains

WordNet Domains<sup>9</sup> (Magnini & Cavaglià, 2000) is a lexical resource where the synsets have been annotated semi automatically with one or more domain labels. These domain labels are organized hierarchically. These labels group meanings in terms of topics or scripts, e.g. *Transport*, *Sports*, *Medicine*, *Gastronomy*. which were partially derived from the Dewey Decimal Classification. The version we used in these experiments is a hierarchy of 171 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*.

### 2.2 The Basque WordNet

The Basque WordNet<sup>10</sup> was developed by the IXA research group<sup>11</sup> developed following the framework of WordNet (Fellbaum, 1998)<sup>12</sup>, as well as on its multilingual counterparts EuroWordNet and the MCR.

The Basque WordNet has been constructed with the expand approach (Vossen, 1998), which means that the English synsets have been enriched with Basque variants. Besides, we also incorporate new synsets that exist for Basque but not for English.

Due to EuroWordNet and the MCR frameworks, the Basque WordNet is already linked to the Spanish, Catalan, English and Italian wordnets, and it can also be linked to any other wordnet tightly linked to the English WordNet.

Up to now, the Basque WordNet has been focused on general vocabulary leaving aside specialized language and terminology, and one of the goals of WINTERM is to enrich the Basque WordNet with terminological information. It currently contains 26,999 headwords, 33,302 synsets and 50,841 senses.

## 3. The Elhuyar Basic Dictionary of Science and Technology (BDST)

The BDST is a specialized dictionary published on line by Elhuyar Foundation<sup>13</sup>. The BDST is designed as a terminological dictionary; thus, each concept is represented in a terminological record, which includes all the information relating to that concept: the terms

<sup>9</sup> <http://wndomains.itc.it/wordnetdomains.html>

<sup>10</sup> [ixa2.si.ehu.es/mcr/wei.html](http://ixa2.si.ehu.es/mcr/wei.html)

<sup>11</sup> <http://ixa.si.ehu.es>

<sup>12</sup> <http://wordnet.princeton.edu/>

<sup>13</sup> [www.elhuyar.org](http://www.elhuyar.org)

(descriptors) that convey the concept, the definition, the domain(s), etc. One concept can be related to several domains. The BDST includes concepts of several areas of Science and Technology. The current amount of concepts is 15,627 belonging to the following knowledge areas or domains. For instance, there are 353 concepts in the ecology domain.

Aeronautics	333	Geography	124
Agriculture	474	Geology	661
Anatomy	705	Mathematics	603
Anthropology	20	Medicine	2,207
Architecture	112	Metallurgy	356
Astronomy	196	Meteorology	435
Astronautics	31	Mycology	138
Biochemistry	698	Microbiology	109
Biology	711	Mineralogy	260
Botanic	1,606	Motoring	81
Building Industry	212	Palaeontology	36
Chemistry	1,490	Photography	51
Computer Science	392	Physics	827
Ecology	353	Technology	790
Electricity	220	Telecommunications	127
Electronics	55	Zoology	2,152
General	40	Not classified	267
Genetics	156	<b>TOTAL</b>	<b>17,028</b>

Table 1: Current figures and domain classification of the BDST concepts.

The BDST includes terms in four languages: English, Spanish, French, and Basque. There is no semantic relations between concepts.

The BDST is an ongoing project. The objectives for 2008 are to include 10,000 new concepts in the dictionary. One of the most relevant term sources is the Corpus of Science and Technology (Alegria et al., 2007), a 7,6 million words corpus tagged by hand with morphosyntactic information<sup>14</sup>. Erauzterm (Alegria et al., 2004) was used to extract automatically the terms from the corpus. Further developments of the BDST are directly related to the results of WNTERM, as far as the establishment of relations between concepts is a promising task that would enrich the dictionary and enhance its value for users.

#### 4. A methodology for the construction of WNTERM

In this section, we will describe the steps for the construction of WNTERM. We first compare the two resources, with special attention to the relation between the domain labels used in each resource, and an automatic analysis of the different cases found. In the following steps we select and structure the concepts and terms to be added to WNTERM. We will illustrate the steps with data from the ecology domain.

#### 4.1 Comparison of both sources

The first step has been the comparison between both sources (the MCR and the BDST) in order to measure the amount of terms in each resource (and their overlap) taking into account the domain information across both resources. For this comparison an automatic procedure has been used, and the result have been qualitatively analyzed.

In order to carry out this comparison, the first step has been the manual mapping of the domain labels of both resources.

##### 4.1.1 Manual mapping of domain labels

As we have mentioned in Sections 2 and 3, concepts of the MCR and the BDST have a domain label. However, each of these sources has a different domain classification. In the MCR there are 171 domains labels which are organized hierarchically (for instance, *cinema*, *radio*, *post*, *tv*, *telegraphy* and *telephony* are subdomains of *telecommunications*). BDST has 34 flat domain labels. As domain information is required to measure the overlap of concepts between both resources, we have manually mapped the 171 domain labels of the MCR to the 34 domain labels of the BDST. From the BDST domains 18 have been mapped to a single MCR domain, and 16 have been related to one or more domain labels of the MCR (e.g. *telecommunications* to *cinema*, *radio*, *post*, *tv*, *telegraphy*, *telephony*). Regarding MCR domains, 64 have been mapped to a single BDST domain, but 20 have been mapped to more than one (e.g. *topography* to *geography*, *geology* and *town planning* to *architecture*, *building*). 87 of the MCR domains have not been mapped to any BDST domain because of their high degree of specification (*fencing*, *paranormal*, *numismatics*), or because the BDST have not been enriched with these fields yet (*religion*, *sport*, *gastronomy*).

##### 4.1.2 Automatic comparison and the resulting casuistry

We performed an automatic analysis of the relation between the MCR and the BDST. We checked for overlaps of terms in each resource (both in English and Basque). If the term is found in both resources, we analyze whether the domain labels can be mapped (according to the domain mapping we described in the previous sections).

The results of this comparison have been tagged using the codes described in Table 2.

<sup>14</sup> [www.ztcorpusa.net](http://www.ztcorpusa.net)

Codes	Description
1 - 0	The term is in the MCR but it is not in the BDST
1 - 1	The term is both in the MCR and the BDST
1 - 2	The term is both in the MCR and the BDST, but they have different domains
1 - 3	The term is only in the MCR, but there is one synonym of this term in the BDST and they have the same domain label
1 - 4	The term is only in the MCR, and there is one synonym of this term in the BDST but they have not the same domain label
0 - 1	The term is in the BDST but it is not in the MCR
1 - 1	The term is both in the BDST and the MCR
2 - 1	The term is both in the BDST and the MCR, but they have different domains
3 - 1	The term is only in the BDST, but there is one synonym of this term in the MCR and they have the same domain label
4 - 1	The term is only in the BDST, and there is one synonym of this term in the MCR but they have not the same domain label

Table 2: Codes used in the automatic procedure.

#### 4.1.3 First conclusions of the comparison

Table 3 shows the results of the automatic comparison. There are 51,469 overlapped terms in English and 26,379 in Basque. Therefore, as we expected, the overlap between the Basque WordNet and the BDST is smaller. As we have already mentioned, this is due to the fact that the Basque WordNet has focused on common vocabulary.

MCR-BDST overlap		
Overlap type	English terms	Basque terms
1 - 1	2,613	1,853
1 - 2	18,591	12,190
1 - 3	2,850	803
1 - 4	16,572	4,447
2 - 1	9,228	6,040
3 - 1	130	59
4 - 1	1,485	987
<b>TOTAL</b>	<b>51,469</b>	<b>26,379</b>

Table 3: Overlap between terms of the MCR and the BDST.

Note that we applied a strict string match among terms. This means that often the same term is written in similar but different ways in both resources. For instance, the BDST has *videotape*, *vanilla plant*, *goldfish* and *desertization* while the MCR contains very similar terms like *video tape*, *vanilla* and *gold-fish*, or a synonym like *desertification*. The numbers in Table 3 correspond to this strict term match, and thus, they are an underestimation of the potential overlap between the MCR and the BDST. In the next section, we will see that a manual review of the 0-1 and 1-0 pairs allows to detect us this false mismatches.

Table 3 also shows that most of the overlapped terms have different domain labels (45,876 in English and 23,664 in Basque). These terms must be previously checked by hand because some of them may not refer to the same concept. For instance, the term *horn* has been marked in the automatic comparison as 1-2, that is, this term is both in the MCR and in the BDST, but they have different domains. In the MCR the term *horn* can belong to the domain of *anatomy* and in the BDST belongs to *zoology*, both referring to the same concept (*one of the bony outgrowths on the heads of certain ungulates*). On the other hand, there is another *horn* in the MCR belonging to the *transport* domain (*a device on an automobile for making a warning noise*) which has nothing to do with the BDST term from *zoology*. Therefore, once more, these numbers are an underestimation of the potential overlap between the MCR and the BDST.

Domain	English terms	Basque terms
Medicine	1,111	1,592
Chemistry	1,085	994
Zoology	465	1,141
Biochemistry	459	228
Botanics	374	775
Geology	332	417
Computer Science	329	368
Meteorology	311	366
Metallurgy	310	321
Biology	308	482
Technology	287	386
Physics	274	460
Aeronautics	266	291
Anatomy	244	468
Ecology	211	245
Mathematics	203	325
Agriculture	197	177
Telecommunications	138	135
Mineralogy	129	206
Genetics	97	45
Electricity	74	123
Building Industry	62	97
Astronautics	37	95
Motoring	36	40
Microbiology	33	81
Architecture	29	51
Mycology	25	90
Geography	24	31
Photography	15	26
Palaeontology	9	23
General	6	7
<b>TOTAL</b>	<b>7,480</b>	<b>10,086</b>

Table 4: Number of English and Basque terms missing in the MCR with comparison to the BDST, organized by BDST domains and ordered according to the number of English missing terms.

The comparison between both resources showed that the MCR contains many English words which are missing from the BDST (80,119 in English and 7,541 thousand in Basque), and that the number of terms missing from the MCR which are covered in the BDST is smaller (7,480 in English). This is not surprising, as the BDST focuses on just science and technology related terms.

In any case, besides the construction of WNTERM itself, the cross-enrichment of both resources is significant, as both of them benefit from the information contained in the other.

In the next subsections we will illustrate the methodology to build WNTERM on the terms from a single domain. Table 4 shows that the terms from BDST of the medicine domain are those that could benefit the MCR most. However, the chosen domain is that of ecology, mainly because it is related to the KYOTO project<sup>15</sup>, where we take part and which is focused on environmental issues.

## 4.2 Selection of concepts

In this step, we select the concepts that are going to be added to WNTERM. We focus on English (because the Basque WordNet is a subset of the English WordNet, and the Basque and English BDST are equivalent), and on the ecology domain. The idea is to take as basis the concepts (synsets) already in the MCR, and add all of them to WNTERM. If new concepts exist in the BDST, they are also added to WNTERM and manually linked to their corresponding superclass (hypernym).

All the concepts in the MCR tagged with the ecology domain label, irrespective of their presence in the BDST, have been copied to WNTERM. In other words, those terms of the MCR tagged with the codes 1-0, 1-1, 1-3 and 3-1 have been directly added to WNTERM. Note that terms tagged with 1-2 and 1-4 have been ignored for the time being, because, as we said in section 4.1.3, these terms, having different domain labels from those in the BDST, must be previously checked by hand to confirm whether they refer to the same concept.

We then have turned our attention to those concepts in the BDST not present in the MCR (those marked as 0-1). We first have examined term matching problems (see examples in section 4.1.3). If we think two terms are the same, and if the concept behind the term was not already in WNTERM, the relevant concept is copied to WNTERM. If the term corresponds to a concept which was not in the MCR, this has been copied to WNTERM, and a link to its hypernym in the Interlingual Index of the MCR added. The result of this additions can be seen in Figure 1, where concepts in bold correspond to those coming from the BDST and those in italics to their hypernyms.

Term values	WNTERM concepts	Noun	Adj.	Verb
<b>1 - 0</b>	8	7	1	0
<b>1 - 1</b>	2	2	0	0
<b>1 - 3</b>	2	2	0	0
<b>3 - 1</b>	94	89	5	0
<b>Revised 0-1</b>	116	80	34	2
<b>TOTAL</b>	222	180	40	2

Table 5: Figures of the ecology domain concepts in the WNTERM.

Table 5 shows the final amount of concepts included in WNTERM. As it can be seen most of the concepts are nominal (222). Adjectives have been also added (40) but due to the fact they can also function as nouns (e.g. *mutant(s)*, *perennial(s)*), we have given special attention to their inclusion in WNTERM. For the time being, we have focused on the *Elhuyar Basque Dictionary* to decide their category. However, we would like to study the possibility of introducing these concepts twice (both as nominal and as adjectival) and relating them with the MCR *XPOS\_EQ\_near\_synonymy* relation. This relation is used when a meaning matches multiple ILI-records simultaneously (in this case, a noun and an adjective), or when there is some doubt about the precise mapping.

## 4.3 Selection of terms

The third step corresponds to the terms. Both English and Basque terms have been automatically added from the MCR and the BDST to their corresponding concepts of WNTERM. As we have already detected the relevant concepts in the previous section, we just look up the terms for those concepts in the two resources, and copy them to WNTERM.

Term values	MCR		BDST	
	English	Basque	English	Basque
<b>1 - 0</b>	22	4	-	-
<b>Revised 0 - 1</b>	-	-	160	143
<b>1 - 1</b>	2	3	2	4
<b>1 - 3</b>	7	3	-	-
<b>3 - 1</b>	-	-	3	3
<b>TOTAL</b>	<b>31</b>	<b>10</b>	<b>165</b>	<b>150</b>

Table 6: Figures of ecology domain terms in the WNTERM.

Table 6 shows the amount of terms added from the MCR and the BDST in WNTERM.

Finally, in the case of Basque, some gaps have been automatically detected, and therefore, Basque variants have been added by hand.

## 4.4 Structure the domain ontology

Finally, the hypernymy links have been used to structure the hierarchy of concepts in the domain ontology, i.e. we have used the hypernymy tree of WordNet to hierarchically organize the terms, plus the hypernym links for the new concepts that we added by hand in the previous section. Irrelevant splits in the hierarchy can be automatically detected and deleted (Vossen, 2001).

<sup>15</sup> <http://www.kyoto-project.eu>

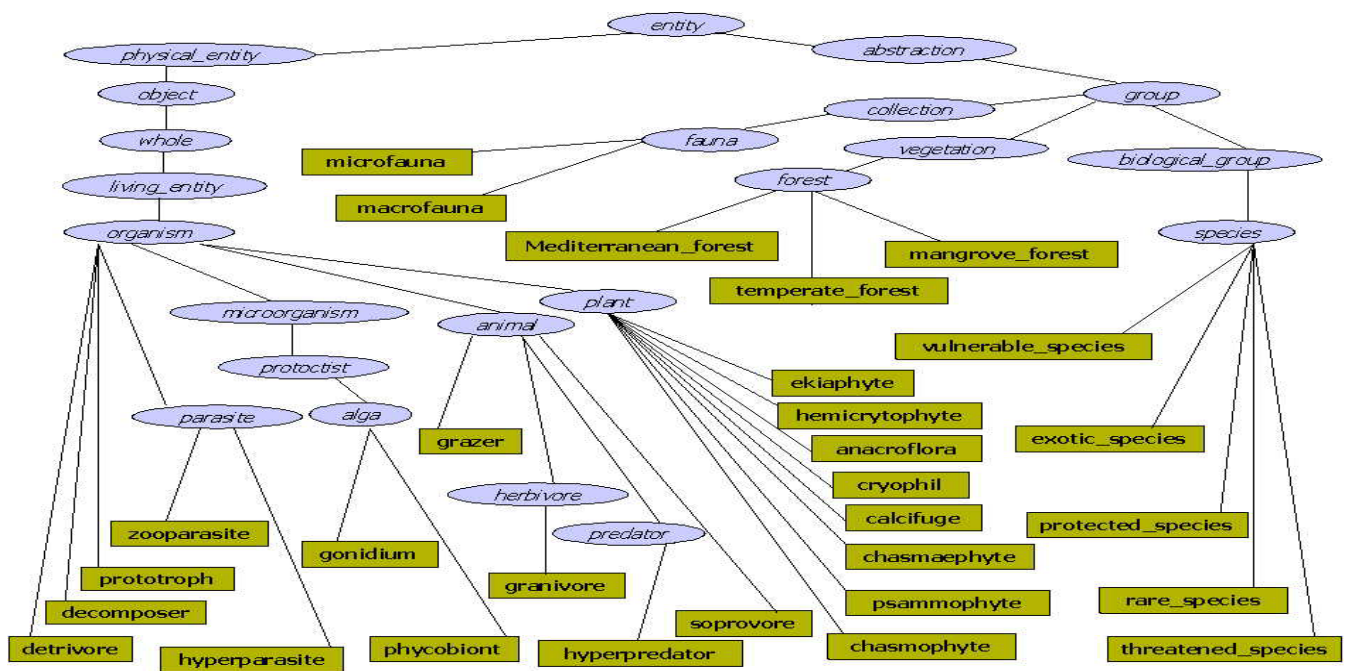


Figure 1: Subhierarchy of WNTERM, with terms added from the BDST (in bold) and their hypernym links in the MCR (in italics).

Later on, and in case the domain experts deem it necessary, the hierarchy can be changed to suit the particularity of the domain. Alternatively to WordNet hierarchy, SUMO or Top Ontology (both of which are linked to WordNet synsets in the MCR) can also be used, depending on the user needs.

As a result, a new wordnet (more precisely, an ecological domain ontology) has been constructed with its own set of concepts and relations, which is linked to the rest of wordnets via the ILLI.

## 5. Conclusions and future work

In this paper we have presented the construction methodology of WNTERM, a science and technology light-weight ontology based on the MCR architecture which has been built with the combination of the Basque and English wordnets and the BDST. The construction methodology has been illustrated on the ecology domain. The final result is a multilingual (Basque and English) light-weight domain ontology, including taxonomic and other semantic relations among its concepts, which is tightly connected to the English WordNet 3.0 and other wordnets. The study also reveals that both wordnets and the BDST are mutually enriched in the process.

In the future, we would like to reorganize the hierarchy coming from the MCR, that is, to design the terminological hierarchical organization according to the criteria of domain experts (lexicographers working on the BDST). We also want to develop a semi-automatic framework based on domain corpora (Vossen, 2001), in order to speed up the construction of other domain ontologies. WNTERM will be used as the evaluation benchmark for the automatic procedures.

## Acknowledgements

We would like to thank Mari Susperregi and Pili Lizaso who have collaborated in this research. Eli Pociello has a grant from the Provincial Council of Gipuzkoa. Part of this work has been funded by the European Commission (KYOTO, ICT-2007-211423) and the Education Ministry (KNOW, TIN2006-15049-C03-01).

## References

- Agirre, E., Aldezabal, I., Etxeberria, J., Izagirre, E., Mendizabal, K., Quintian, M. & Pociello, E. (2006). A methodology for the joint development of the Basque WordNet and Semcor. In *Proceedings of the 5th International Conference on Language Resources and Evaluations (LREC)*. Genoa (Italy).
- Alegria, I., Areta, N., Artola, X., Díaz de Ilarraza, A., Ezeiza, N., Gurrutxaga, A., Leturia, I. & Sologaitoa, A. (2007). ZT Corpus: Annotation and tools for Basque corpora. In *Corpus Linguistics 2007*. Birmingham (England).
- Alegria, I., Gurrutxaga, A., Lizaso, P., Saralegi, X., Ugartetxea, S. & Urizar, R. (2004). A Xml-Based Term Extraction Tool for Basque. In *LREC2004: 4th International Conference On Language Resources And Evaluation*. Lisbon (Portugal).
- Álvez, J., Atserias, J., Carrera, J., Climent, S., Oliver, A. & Rigau, G. (2008). Consistent annotation of EuroWordNet with the Top Concept Ontology. In *Proceedings of The 4th Global Wordnet Association Conference*. Szeged (Hungary).
- Atserias, J., Villarejo, L., Rigau, G., Agirre, E., Carroll, J., Magnini, B. & Vossen, P. (2004). The MEANING Multilingual Central Repository. In *Proceedings of the*

- 2<sup>nd</sup> *Global WordNet Conference*. Brno (Czech Republic).
- Borst, W.N. (1997). Construction of Engineering Ontologies. Centre for Telemática and Information Technology, University of Tweety. Enschede, The Netherlands.
- Daudé, J., Padró, L. & Rigau, G. (2001). A Complete WordNet1.5 to WordNet1.6 Mapping. In *Proceedings of NAACL'01 Workshop on WordNet and Other Lexical resources*. Pittsburgh (Pennsylvania).
- Decker, S., Erdmann, M., Fensel, D. & Suder, R. (1999). Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information. In *Proceedings of Semantic Issues in Multimedia Systems (DS-8)*. Rotorua (New Zealand).
- Fellbaum, C. (1998). *WordNet: An electronic Lexical Database*. The MIT Press, Cambridge, Massachusetts. London (England).
- Gruber, T.R., & Olsen, F. (1994). An ontology for Engineering Mathematics. In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning*. Bonn (Germany).
- Magnini, B. & Cavaglià, G. (2000). Integrating subject field codes into WordNet. In *Proceedings of LREC*, pages 1413–1418. Athens (Greece).
- Navigli, R., Velardi, P. & Gangemi, A. (2003). Ontology Learning and Its Application to Automated Terminology Translation. In *IEEE Intelligent Systems*, 18 (1), pp. 22-31.
- Niles, I. & Pease, A. (2001). Towards a standard upper ontology. In *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems*, pp. 17–19.
- Rigau, G., Agirre, E. & Atserias, J. (2003). The MEANING project. In *Proceedings of the XIX Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)*. Alcalá de Henares (Madrid).
- Roventini, A., Marinelli, R. (2004). Extending the Italian WordNet with the Specialized Language of the Maritime Domain. In *Proceedings of the 2<sup>nd</sup> Global WordNet Conference*. Brno (Czech Republic).
- Sagri, M. T., Tiscornia, D. & Bertagna, F. (2004). Jur-WordNet. In Sojka, P. et al. (Eds.) In *Second International WordNet Conference*. Brno (Czech Republic).
- Stamou, S., Ntoulas, A., Hoppenbrouwers, J., Saiz-Noeda, M. & Christodoulakis, D. (2002). EUROTERM: Extending EWN using the expand and merge model. In *Proceedings of the 1<sup>st</sup> Global WordNet Conference*. Mysore (India).
- Uschold, M., King, M., Moralee, S. & Zorgios, Y. (1998) The Enterprise Ontology, *Knowledge Engineering Review*, 13(1). 31-89.
- Vossen, P. (1998). *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers.
- Vossen, P. (2001) Extending, Trimming and Fusing WordNet for Technical Documents. In *Proceedings of the NAACL Workshop on Extending Wordnet*. Pittsburgh (Pennsylvania).