

# Monolingual and bilingual dictionary approaches to the enrichment of the Spanish WordNet with adjectives

Irina Chugur

Anselmo Peñas

Julio Gonzalo

Felisa Verdejo

Departamento de Lenguajes y Sistemas Informáticos, UNED  
Ciudad Universitaria s/n,  
28080 Madrid, Spain  
{irina,anselmo,julio,felisa}@lsi.uned.es

## Abstract

We report on two different approaches to the incorporation of adjectives in Spanish WordNet based on automatic extraction techniques using EuroWordNet and machine-readable dictionaries. We show that a monolingual dictionary approach enables to exploit relations between different parts of speech and enrich the internal structure of the Spanish WordNet, while the methods based on bilingual dictionaries allow to connect the new items with the rest of the wordnets through the EuroWordNet InterLingual Index.

## Introduction

The EuroWordNet (EWN) project (Vossen, 1998) aimed at creating a semantic network for 8 European languages.<sup>1</sup> Each wordnet is organised around the notion of synset, a set of synonyms representing a concept. Basic semantic relations are encoded between synsets, including relations between different parts of speech (XPOS). Different surface realisations of similar concepts within and across languages can be matched via equivalence links between language-specific synsets and an “InterLingual Index” (ILI) which is a superset of all concepts occurring in monolingual wordnets.

This paper reports an effort to enrich the Spanish subset of EWN with adjectival information, by linking Spanish adjectives the ILI and also by obtaining internal relationships between Spanish adjectives, nouns and verbs.

Previously, the encoding of information on adjectives within Italian WordNet had been treated in (Alonge, 2000). This research focused on enriching the set of lexical-semantic relations and revising the EWN Top Ontology adapting it to include adjectives and adverbs. We did not, however, modify the set of EWN relations, forming on the development of effective techniques for (semi) automatic attachment of adjectives into SpWN, maintaining the existing EWN structure.

In a first approach, described in section 1, we tried to introduce adjectives as related to Spanish nouns and verbs, extracting derivational information from a monolingual dictionary. Nevertheless, it turned out rather problematic to link automatically the resulting adjectives to the ILI. In order to overcome this difficulty and to integrate our new candidates in the multilingual structure of the EWN, we followed another method, based on bilingual dictionaries. We created links between Spanish adjectives and ILI records. The criteria and heuristics used in this approach are reported in section 2. Finally, we draw some conclusions and future extensions of this work.

## 1 Monolingual approach. XPOS adjective relations

The EWN database includes a broad range of XPOS relations. We decided to attach new adjectives to the SpWN exploiting lexical-semantic relations between adjectives extracted from a monolingual dictionary and nouns or verbs that already belonged to the SpWN. The next step would be creating synsets with these adjectives.

### 1.1 Lexicographic definitions and EWN relations

First of all, we studied definitions of adjectives in the monolingual dictionary (Vox, 1992). Some of the recurring patterns offered the possibility of automatically extracting some relations considered in the EWN. For example, adjectives defined as “that causes + N/V” or “able to cause + N/V” are likely to capture the EWN relation “CAUSES / IS CAUSED\_BY”. While patterns such as “that involves + N/V” point to the relation “ROLE / INVOLVED”. Adjectives defined in terms of other adjectives can provide synonym or antonym clusters.

However, after a manual study of a sample we realised that this method of identification of EWN relations, first, had low productivity and, second, required a considerable manual filter, since there is not a

---

<sup>1</sup> The final database is available via ELRA: <http://www.icp.grenet.fr/ELRA/home.html>

one-to-one correspondence between the lexicographic patterns and the semantic relations.

## 1.2 Definition Patterns and Morphological Criteria

This preliminary study led us to restrict the set of candidates for linking. We decided to consider morphological criteria and exploit the relationship between denominal/deverbal adjectives and their derivational basis.

From Spanish suffixes used in derivational processes involving adjectives (Lang, 1992; Rainer, 1999) we selected a preliminary subset for a first exploration. It was composed of the following morphemes: *-able / -ible, -dor, -ense, -ista, -torio*. According to our estimations, these suffixes would permit us to obtain a large number of adjective/noun and adjective/verb pairs.

The work consisted of the following steps:

- Generating candidates applying the definition patterns to a subset of denominal and deverbal adjectives.
- Automatically extracting three types of relations: synonymy, antonymy and generic adjective-noun, adjective-verb relation.

We applied patterns that permitted to identify the conceptual core of the definition to the list of adjective entries with the selected suffixes. A morphosyntactic tagger (Carmona et al, 1998) allowed obtaining the lemma and the category of the extracted word and, in this way, recognising inflected noun and verb forms. Example 1 shows two of the 35 definition patterns used and dictionary entries that match them.

<p><b>“que + Verbo 3ª p. S”</b>  <b>(that + Verb, 3 p. S form)</b></p> <p>abrumador_i_1 &lt;adj.&gt; que abruma</p>
<p><b>“que implica + (Det) + Nombre”</b>  <b>(that involves + (Determiner) + Noun)</b></p> <p>usurpatorio_i_1 &lt;adj.&gt; que implica usurpación</p>

Example 1. Definition patterns

Then we filtered those nouns and verbs that did not belong to the SpWN. Finally we obtained adjective-noun, adjective-verb pairs and synonym and antonym clusters (adjective-adjective groupings).

## 1.3 Creating synsets

Some of the nouns and verbs identified as a linking term are polysemous. Polysemy raises the problem of selecting the appropriate noun/verb sense that the adjective should be linked to.

The strong morphological and semantic relationship that holds between denominal/deverbal adjectives and their derivational basis in most cases led us to link each

derived adjective to all the senses of the noun/verb extracted from the appropriate kind of definition. For example, “traducible - que se puede traducir” (*translatable – which can be translated*). According to our hypothesis, the adjective “traducible” should be linked to every sense of the verb “traducir”.

There may be two ways to proceed:

- to create one adjective synset for each sense of the corresponding noun or verb,
- to build a unique synset for a denominal or deverbal adjective which would stand for all the adjective meanings derived from the senses of the noun/verb that the adjective is linked to.

Following (Murphy and Andrew, 1993) we assumed that adjectives could be considered monosemous, producing different extensions in combination with the meanings of the nouns that they modify. Thus, we adopted the second option.

## 1.4 Results

Table 1 reports the results obtained for each suffix. 47% of the studied adjectives can be linked to some verb. The percentage for the resulting adjective-noun relations is 17%. This difference is mainly due to the selection of suffixes. *-Able, -ible, -dor* are used in Spanish to derive adjectives from verbs, while *-ense* and *-ista* apply to nouns and generate denominal adjectives.

We obtained 2191 synsets not linked to the ILI that yielded 2282 links between adjectives and nouns or verbs. The defined patterns covered almost 80% of the selected adjectives. Cases of more complex definitions - such as “contradictorio → que se hace en presencia de los interesados” (“*that is hold in the presence of the interested parties*”) - that did not match any of the elaborated patterns correspond to the last column, entitled “no pattern”.

Suffixes	Adj - Noun	Adj - Verb	Synonyms	Antonyms	No pattern
-able	49	382	86	153	194
1055	5%	36%	8%	15%	37%
-dor	50	1251	112	0	203
1616	3%	77%	7%		13%
-ense	30	1	56	0	176
263	11%	0.4%	21%		67%
-ista	378	3	53	3	175
612	62%	0.5%	9%	0.5%	29%
-torio	84	49	24	3	57
217	9%	23%	11%	1%	26%
-ible	0	5	1	1	1
8		63%	13%	13%	13%
Total	591	1691	332	160	806
<b>3580</b>	17%	47%	9%	5%	23%

Table 1. Morphological relations.

In order to evaluate the number of erroneous connections between adjectives and inappropriate senses of corresponding nouns or verbs, we verified manually if the original adjective had any semantic

relation with every SpWN sense of the noun or verb that appeared in its lexicographic definition. We verified 830 links out of 2774 and only 84 (10%) of them resulted erroneous. We found that error margin acceptable.

### 1.5 Discussion

By means of a joint application of morphological and lexicographic patterns, we extracted synonym and antonym clusters and, on the other hand, adjective-noun, adjective-verb pairs.

This first approach allowed

- obtaining a considerable number of adjectives that are likely to be attached into the wordnet,
- relating these adjectives with nouns and/or verbs and enriching the internal structure of a particular wordnet,
- creating synonym and antonym clusters for adjectives.

There is a possibility of enlarging the number of extracted links combining different kinds of definitions. For example, “audible” is defined by means of a synonym “oíble”. “Oíble” in its turn is defined as “que se puede oír” (*that can be heard*). So, via the first synonym definition, it is possible to relate the adjective “audible” with the verb “oír”.

The next step was to link the obtained terms to the ILI, assigning them a number of ILI-record. In this way, the new elements would take part in the multilingual structure of the EWN. However, finding the adequate combination of Spanish and English adjectives under this method required a considerable manual filter. This made us search for other techniques to guarantee correct attachment to the ILI.

## 2 Bilingual approach. Linking adjectives to the InterLingual Index

In the second approach, we tried to link automatically as many Spanish adjectives as possible to the EWN ILI (see Figure 1).

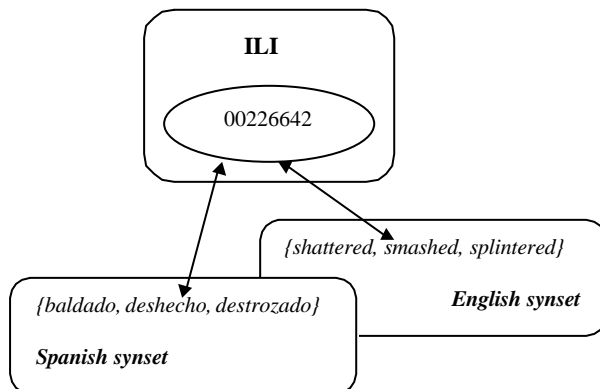


Figure 1

Some of the heuristics described below derive from (Atserias et al, 1997; Farreres et al, 1998) and have been used to build the nominal and verbal section of the SpWN. We have extended this work in order to improve precision and recall for adjective linking.

There are two complementary ways of linking a new category to the ILI:

1. translating English terms extracted from WordNet 1.5 synsets to the target language, in our case into Spanish, and
2. extracting from bilingual dictionaries Spanish - English pairs of terms and then identifying the corresponding English synsets.

### 2.1 Translating English adjectives into Spanish

#### 2.1.1 English monosemous adjectives

The first heuristic consists in extracting all the monosemous adjectives from WN. Since these adjectives appear only in one synset, their Spanish translations can be linked to them directly, as at least one of the meanings of the Spanish term should correspond to the English adjective.

```

    Cynical → 01865165
    Cynical → cínico
                link: cínico / 01865165

    Damp → 01935444
    Damp → húmedo, mojado

                links: húmedo / 01935444
                    mojado / 01935444
  
```

Example 2

#### 2.1.2 Translating variants of English synsets into Spanish

An English synset is a set of synonyms or synset variants. If we translate each of these variants into Spanish, it may happen that some of the variants have some Spanish translation. This coincidence indicates that there is a relation between the original synset and the target Spanish term. We consider that one coincidence of this kind is enough to establish a link (see Example 3).

```

    00235058 gruff hoarse husky
           gruff: bronco malhumorado rudo
           hoarse: ronco áspero
           husky: ronco
    Link: ronco => 00235058
  
```

Example 3

## 2.2 Translating Spanish adjectives into English

We extracted, from the Spanish-English dictionary, all the adjective entries and their translations. Since English adjective may be polysemous, special heuristics are needed to identify the correct ILI record which the Spanish term can be attached to. The criteria we used for this task reflect four possible situations:

### 2.2.1 Unique English translation

Certain Spanish adjectives have only one English translation, which can be either monosemous, or polysemous, being associated to only one WN synset or to several WN synsets.

#### (a) Unique monosemous English translation

If the English translation is monosemous, then the Spanish adjective can be linked to it automatically. For example, Spanish “abochornado” is translated into English by “ashamed” and “ashamed” is linked only to the index 00125050. Then “abochornado” is linked automatically to the same index.

#### (b) Unique polysemous English translation

If the unique English term that translates a Spanish adjective is polysemous, it is necessary to find the appropriate WN sense, which the Spanish adjective can be linked to.

In order to determine the correct link for the original Spanish adjective, we used a method based on inverse translation. The variants of the English synset are translated back into Spanish. Every occurrence of the original Spanish adjective among the inverse translations of the English synset variants should increase the possibility of connecting the Spanish adjective and the concept represented by the English synset.

According to this hypothesis, we introduced a simple measure  $b_r$  (*binding\_ratio*) that indicates the possibility of relating a Spanish adjective to an English adjective synset:

$$b_r(\text{adj, syn}) = \frac{\text{[Occurrences of adj among inverse translations of the synset variants]}}{\text{Number of inverse translations}}$$

The variants of English synsets that are not in the bilingual dictionary and therefore have no inverse translation (only 4%), have not been taken into account for the estimates of the *binding\_ratio*. Example 4 illustrates how the binding ratio was applied.

After we defined the measure indicating the possibility of linking, it was necessary to establish the

appropriate threshold. We examined a sample of 78 adjectives with unique polysemous translations and linked them manually to the correspondent WN synsets, obtaining 100 links. We found that if the threshold equaled 0 approximately 23% of the links generated by the *binding\_ratio* were incorrect. In order to decrease the number of incorrect or uncertain links we raised the threshold to 0,3. Thus only those links that had the *binding\_ratio* higher than 0,3 would be accepted.

**premeditado**  
Translation: deliberate (4 senses):

1. 01016975: by conscious design or purpose: "intentional damage"; "a knowing attempt to defraud"; "a willful waste of time"
2. 00965893: "walking at the same measured pace"
3. 01834448: marked by careful consideration or reflection: "a deliberate decision"
4. 01407902: "a calculated insult"; "with measured irony"

Inverse translations and binding-ratio for each synset:

01016975 deliberate intentional knowing willful wilful  
deliberate: deliberado premeditado  
deliberate: pausado lento  
intentional: intencional  
knowing: de\_complicidad  
willful:  
wilful: voluntarioso terco  
wilful: premeditado

$b_r = 2^2 / 6 = 0.67 \Rightarrow$  link: premeditado / 01016975

00965893 careful deliberate measured  
careful: cuidadoso  
careful: prudente  
deliberate: deliberado premeditado  
deliberate: pausado lento  
measured:

$b_r = 1^2 / 4 = 0.25 \Rightarrow$  no link

01834448 deliberate  
deliberate: deliberado premeditado  
deliberate: pausado lento

$b_r = 1^2 / 2 = 0.5 \Rightarrow$  link: premeditado / 01834448

01407902 calculated deliberate measured  
calculated:  
deliberate: deliberado premeditado  
deliberate: pausado lento  
measured:

$b_r = 1^2 / 2 = 0.5 \Rightarrow$  link: premeditado / 01407902

Example 4. Links from *binding\_ratio*.

We verified that the *binding\_ratio* at this threshold yielded 46 correct links, while all incorrect links (24) were automatically discharged. There was only 1 correct link that we lost because it did not reach the required threshold. The remaining 29 candidate links were

correct connections that could not be detected due to the lack of inverse translation. At this threshold no incorrect link was accepted.

### 2.2.2 Multiple English translation

If there are two or more English translations for a Spanish adjective, we consider another two cases in which it is possible to link the adjective term to the ILI:

#### (a) Multiple and monosemous English translation

The Spanish term has several English translations. At least one of them is monosemous. Here, like in 2.2.1 (a), the Spanish adjective can be attached to every monosemous English term. For example, Spanish “chillón” can be translated into English by “screaming”, “loud” and “garish”. “Screaming” has only one meaning represented by the index 00235682. That enables us to link the Spanish adjective to this concept automatically. “Garish” (01818583) is also monosemous in WN, that is why “chillón” is also linked to 01818583.

#### (b) Multiple English translation with synsets in common

The last case is that of several monosemous and polysemous translations that share some synset(s). This time the link can be established between the original Spanish adjective and this common synset. Example 5 shows how two polysemous translations of “encharcado” (“flooded” and “swamped”) share the sense corresponding to the index 00817914, enabling us to link the adjective “encharcado” to this synset.

<b>encharcado</b> Translations: 1. flooded 2. swamped
<b>flooded</b> 00816602 (esp of a boat; covered or overflowing with water) 00817914 (filled to overflowing as if with water)
<b>swamped</b> 00064913 ("inundated farmlands"; "a swamped boat") 00817914 (filled to overflowing as if with water) 01398176 (rendered powerless esp by an excessive amount or profusion)
Link: encharcado → 00817914

Example 5

## 2.3 Results

The above heuristics produced links for 3102 different Spanish adjectives (see Table 2). This figure corresponds to 5210 links distributed between 3424 different synsets (1,53 Spanish variants per synset). This figure surpasses the WN 1.5 ratio for category 3 (adjectives excluding satellites) which is 1,26 (8102

links for 6440 different synsets). In each case a 30% sample has been verified manually, and no incorrect links have been found.

		Adjective – synset links	Number of adjectives	Number of synsets
From English to Spanish	monosemous adjectives	1768	1470	1199
	common synset	871	551	676
From Spanish to English	unique monosemous translations	592	592	530
	unique polysemous translations	1271	474	1209
	several translations (at least one monosemous term)	1166	931	730
	several translations (at least one shared synset)	606	361	415
Total		5237	3102	3424

Table 2.

## Conclusions

The reported work allowed attaching automatically 5210 adjectives organised in 3424 synsets. We also obtained links between 2774 adjectives and related noun and/or verb synsets.

Our study confirms that the use of bilingual dictionaries is more adequate for building wordnets than methods based on monolingual dictionaries, if the goal is to establish direct correspondence to the synsets of other languages. On the other hand, by means of monolingual techniques it is possible to enrich the internal structure of a particular wordnet, providing very valuable information for various NLP applications.

It would also be interesting to combine both approaches. Our first attempts in this direction showed the complexity of the task. We studied, for instance, the possibility of enriching existing synsets with synonym clusters obtained by means of definition patterns. We expected to be able to enrich Spanish synsets with synonym clusters extracted from the monolingual dictionary. Example 6 illustrates this point. The synset {0028964 bayo} was obtained applying bilingual techniques, while the synonym cluster {bayo, asustado} was extracted from the monolingual dictionary by means of a specific lexicographic pattern. Being the synonym cluster a superset of the synset, it should be possible to include “asustado” automatically in the

synset. However, this would lead to an error, since the meaning of the adjective “bayo” in the synset is completely different.

<p><u>Spanish synset:</u>          {0028964 bayo}          (of a moderate reddish-brown color)</p> <p><u>Synonym cluster from the monolingual dictionary:</u>          {bayo, asustado}          (frightened, scared)</p>
---

*Example 6*

Once again, polysemy and differences between definitions in WN and bilingual or monolingual Machine Readable Dictionaries (MRD) pose problems for automatic creation of links.

We conclude that the described work provides a valid methodology for linking not only adjectives but also any other syntactic category of any language, requiring WN and bilingual MRD.

#### **Acknowledgements**

This work has been partially supported by European Commission, EuroWordNet project (LE #4003), by Comisión Interministerial de Ciencia y Tecnología, the project TIC-96-1234-C03-01, and a PhD grant from UNED.

#### **References**

- Alonge A., Bertagna F., Calzolari N., Roventini A., and Zampolli A. (2000) Encoding Information on Adjectives in a Lexical-Semantic Net for Computational Applications. In Proceedings of NAACL'2000, Seattle.
- Atserias J., Climent S., Farreres X., Rigau G. and Rodríguez H. (1997) Combining Multiple Methods for the Automatic Construction of Multilingual WordNets. In Proceedings of International Conference “Recent Advances in NLP (RANLP'97), Bulgaria.
- Carmona J., Cervell S., Màrquez L., Martí M., Padró L., Placer R., Rodríguez H., Taulé M. and Turmo J. (1998) An Environment for Morphosyntactic Processing of Unrestricted Spanish Text. In Proceedings of the 1st International Conference on Language Researches and Evaluation (LREC'98), Granada, Spain.
- Gonzalo J., Verdejo F., Peters C., and Calzolari N. (1998<sup>a</sup>) Applying EuroWordNet to Multilingual Text Retrieval. Special Issue on EuroWordNet, Journal of Computers and the Humanities, 32 (2-3), pp. 185-207. Kluwer Academic Publisher, Dordrecht.
- Gonzalo J., Verdejo F., Chugur I., López F. and Peñas A. (1998b) Extracción de relaciones semánticas entre nombres y verbos en EuroWordNet. Sociedad Española para Procesamiento del Lenguaje Natural, 23, pp. 97-103.
- Farreres X., Rigau G., and Rodríguez H. (1998) Using WordNet for Building WordNets. Proceedings of COLING – ACL. Workshop “Usage of WordNet in Natural Language processing Systems”. Montreal, Canada.
- Lang M. F (1992) Formación de palabras en español: morfología derivativa productiva en el léxico moderno. Madrid, Cátedra.
- Miller G. (1990) WordNet: An Online Lexical Database. Special Issue of International Journal of Lexicography, 3 (4), pp. 235 - 244.
- Murphy G.L. and Andrew J.M. (1993) The Conceptual Basis of Antonymy and Synonymy in Adjectives. Journal of Memory and Language, 32, pp.301-319.
- Rainer F. (1999) La Derivación Adjetival. In “Gramática Descriptiva de la Lengua Española”, Bosque I. and Demonte V., Espasa Calpe, Madrid, vol. 3.
- Soler C. Los Adjetivos en SpanishWordNet, in preparation.
- Vossen P. (1998) Introduction to EuroWordNet. Computers and the Humanities, Special Issue on EuroWordNet.
- Vossen P., Peters W., Gonzalo J. (1999) Towards a Universal Index of Meaning. In Proceedings of ACL/SIGLEX'99: Standardizing Lexical Resources, University of Maryland.
- VOX (1992) Diccionario de la Lengua Española.
- VOX (1992) Vox Harrap, Spanish-English, English-Spanish dictionary.