Some Strategies for the Improvement of a Spanish WordNet

Matías Herrera, Javier González, Luis Chiruzzo, Dina Wonsever

Facultad de Ingeniería, Universidad de la República
Montevideo, Uruguay

Abstract

Although there are currently several versions of Princeton WordNet for different languages, the lack of development of some of these versions does not make it possible to use them in different Natural Language Processing applications. So is the case of the Spanish Wordnet contained in the Multilingual Central Repository (MCR), which we tried unsuccessfully to incorporate into an anaphora resolution application and also in search terms expansion. In this situation, different strategies to improve MCR Spanish WordNet coverage were put forward and tested, obtaining encouraging results. A specific process was conducted to increase the number of adverbs, and a few simple processes were applied which made it possible to increase, at a very low cost, the number of terms in the Spanish WordNet. Finally, a more complex method based on distributional semantics was proposed, using the relations between English Wordnet synsets, also returning positive results.

1 Introduction

The Multilingual Central Repository (González-Agirre, Laparra, Rigau, & Donostia, 2012) follows the model proposed by the EuroWordNet project. EuroWordNet (Vossen, 1998) is a multilingual lexical database with wordnets for several European languages, structured in the same way as Princeton's WordNet. The MCR comprises five different languages: English, Spanish, Catalan, Basque and Galician. The Inter-Lingual-Index (ILI) allows us to link the words in one language with their equivalent translation in any of the other languages, thanks to the automatically generated mappings among WordNet versions. For example: the ILI identifier "ili-30-02084071-n" corresponds both to the English synset "eng-30-02084071-n" with lemmas "dog, domestic dog", and to the Spanish synset "spa-30-02084071-n" with lemmas "can, perro". In addition, it corresponds to the Basque synset "eus-30-02084071-n" with lemmas "zakur, or, txakur", to the synset "cat-30-02084071-n" for Catalan with lemmas "ca, canis familiaris", and also "glg-30-02084071-n" for Galician with lemmas "can, Canis familiaris". The current ILI version corresponds to WordNet 3.0. All identifiers stem from the original synset in English. In the previous example there is a translation for each one of the languages, however, this is not the most common scenario. The MCR is incomplete, at least for the Spanish version. This document presents several strategies to extend the coverage of the Spanish version. An in-depth analysis of the different problems of the Spanish MCR is presented in section 2, and section 3 describes several processes to enhance it. Section 4 presents the evaluations carried out for the strategies proposed and section 5 presents final observations on the general results and the possibility to launch an enhanced version on line.

2 Problems on the MCR Spanish WordNet

2.1 Deficiencies of the current Spanish MCR: first evaluation

For the purpose of finding the deficiencies of the MCR WordNet, our initial approach was to use it and test it out. Version 3.0 was used, since this is the latest version currently available. The web interface provided by the MCR (Benítez et al., 1998) was used to fulfill this stage. The MCR was requested to provide the results both in English and Spanish for all the searches made, in order to be able to compare them. Below we provide some ex-

amples of this initial informal evaluation and the following section presents a quantitative evaluation:

- Lack of common words

  Some common words such as "cargador" and the adverb "no" were found to be missing.

- Empty synsets

  Some Spanish synsets were available through the web interface but they were empty. For example, the synset "spa-30-00396699-r" did not contain any variants, but its English equivalent "eng-30-00396699-r" did. This shows that there were no Spanish translations in the MCR for the lemmas "meagerly", "sparingly", "slenderly" and "meagrely". When searching for the adverb "escasamente", which is a possible translation for "sparingly", it was not found.

- Very few entries for the grammatical category adverbs

  Once evaluated, it was concluded that adverb coverage of the MCR was very low. We have already mentioned the example for the adverb "no". It was also found that the adverbs "recién" (just) and "rápidamente" (quickly) were not present, although these are very commonly used in Spanish.

- Lack of glosses or phrases that show the usage of the terms in Spanish.

  No Spanish gloss was found for many of the words searched. For example, we found that the result for the noun "cuchillo", "spa-30-03623556-n" and "spa-30-03624134- n" did not include a Spanish gloss for these synsets. Additionally, a generalized lack of phrases that illustrate the use of the lemmas and synsets was found.

2.2 Deficiencies in the current MCR: evaluation on a corpus

Several MCR WordNet coverage measures were applied taking Corin corpus (Grassi, Malcuori, Couto, Prada, & Wonsever, 2001) as a baseline. Corin corpus is a synchronous corpus that comprises the years 1996-2000 and contains literary-type texts by Uruguayan authors (essays and fiction) and journalistic texts published in Montevideo (articles and interviews). Several other language processing tools were used in addition to Corin, such as Freeling (Carreras, Chao, Padró, & Padró, 2004) and the dictionaries Apertium (Armentano Oller et al., 2007) and Wiktionary(*Wikimedia Foundation. 2008b. Wiktionary.*, 2008).

The following aspects were studied:

1. The percentage of available lemmas in the Spanish version of WordNet.
2. The percentage of corpus lemmas for which there was a translation available.
3. The percentage of these lemmas that was not present in the Spanish MCR but did have an available translation in the English MCR.

The results obtained are presented as follows:

2.2.1 Percentage of Corin lemmas available in the Spanish version of WordNet

POS	Lemmas found		Lemmas not found		Processed lemmas
N	69,29%	2780	30,71%	1232	4012
A	51,00%	840	49,00%	807	1647
V	75,35%	1235	24,65%	404	1639
R	32,79%	121	67,21%	248	369
Total	48,70%	3734	51,30%	3933	7667

The previous chart shows the total number of lemmas processed, their Parts Of Speech and how many of them were number found on WordNet . We can see that adverbs are the grammatical category with the lowest coverage at less than 33%. The remaining POS show a higher coverage, with verbs showing the highest one.

2.2.2 Percentage of corpus lemmas for which there was a translation available

POS	Untranslated		Translated		Lemmas
N	12,04%	483	87,96%	3529	4012
A	21,07%	347	78,93%	1300	1647
V	18,00%	295	82,00%	1344	1639
R	16,26%	60	83,74%	309	369
Total	15,46%	1185	84,54%	6482	7667

Using the two mentioned dictionaries we were able to cover a large percentage of the lemmas present in the corpus. Even so, the results do not ensure the quality of the translations. Therefore, it is necessary to improve the resources used for this purpose.

2.2.3 Lemmas not found in the Spanish MCR but with a translation available in the English MCR

Out of the 6482 lemmas translated into English, we focused on those found in the English MCR, so it was possible to compare the lemmas which were not found in the Spanish MCR but did have a translation available in the English MCR.

POS	Lemmas not in Spanish MCR		Lemmas in Spanish MCR		Total
N	43,40%	1349	56,60%	1759	3108
A	46,37%	492	53,63%	569	1061
V	15,02%	176	84,98%	996	1172
R	69,00%	187	31,00%	84	271
Total	39,27%	2204	60,73%	3408	5612

We can conclude that verbs are the grammatical category with the widest coverage, and adverbs are the most incomplete. In addition, nouns and adjectives present a coverage of just over 50%.

3 Strategies to improve WordNet

To improve the existing Spanish WordNet we conducted tests with processes that we have called "selectors", following the terminology already used in the field (WoNeF). A selector is a mechanism that, when applied to an English synset, will choose the translation or translations for the Spanish synset based on the original in English. Previously defined selectors were tested, supported by Apertium and Wiktionary translators, and in addition, two new selectors were defined, one based on morphology and the other based on the exploitation of semantic relations between synsets, with frequentist criteria used in distributional semantics. Selectors are applied in two differentiated stages, which are separately evaluated.

3.1 Translation methods

The translation process used was key for the application of this method to create the Spanish WordNet based on the English WordNet. We used two different methods: automatic translation and dictionaries. With regard to dictionaries, Wiktionary was used as well as a dictionary created based on the XML stem files of the Apertium dictionary. The automatic translation used was the one provided by Bing Translator (*Bing Online Translator.*, 2015). These tools were chosen mainly due to their availability, since they are either free and/or open. Wiktionary and Apertium were downloaded from their respective websites, and Bing Translator was used online through its API.

Microsoft's Bing Translator does not take into account the grammatical category of the word to be translated, therefore, there were cases where if verbs were translated, it would return nouns, or even the same verb but in a different conjugated form, instead of the infinitive form used in the search. In order to solve this problem, it was decided to use the results returned by the translator, and conduct a morphological analysis applying Freeling. The procedure entails obtaining all the possible grammatical categories of the word and its lemma, to afterwards select the words with the same grammatical category as the originally translated English word.

We decided to use a dictionary created based on the XML stem files of the Apertium dictionary rather than the already processed Apertium dictionary, since, for some reason, when making a request it would only return one possible translation, even if the XML file contained more. It was possible to obtain all the available translations for each word using the XML stem files.

3.2 Phase 1: Initial selectors

Below we present the experiments conducted with simple selectors already reported in the literature: monosemy and single translation. It is surprising that these selectors are still productive over the currently available version of WordNet, as our experiments show.

Monosemy Monosemy takes those words found in a single synset. This condition seems to show that there is no ambiguity and, therefore, all translations obtained are added to the corresponding synsets in the Spanish WordNet. For example, when applying this selector to the synset "eng-30-00048268-r" whose lemma is "currently" the three possible translations obtained by the translators "hoy", "ahora" and "actualmente" are selected since "currently" is only found in one synset in the English WordNet.

Single translation This selector takes all the words that have a single translation into Spanish and places it in all corresponding Spanish WordNet synsets. For example, when applying this selector to the

synset “eng-30-00061528-r”, whose lemma is “abruptly” and the translation returned is “abruptamente”, this will be selected since it is the single translation.

Factorization The factorization selector works at synset level. It takes all synsets from the English WordNet and returns all possible translations for each lemma. Once the set of translations for each lemma is put together, the selector selects those translations found as a common translation for all the lemmas in the synset, that is, with the intersection of the translation sets for each lemma. For example, consider the synset “eng-30-01309991-a”, whose lemmas are “artless” and “ingenuous”. The translations for “artless” are: “inocente”, “ingenuo” and “cándido” and those for “ingenuous” are: “inocente” and “ingenuo”. In this case, by applying the selector we obtained “inocente” and “ingenuo”, as a common translation.

Derived Adverb This selector obtains adverbs from the English WordNet and then the adjectives from which these derive. The property “is_derived_from” provided by the MCR was used to obtain the adjectives from which these adverbs derive. Once the adjective synsets are returned, we will obtain all the variants. These are in turn translated so as to later apply the morphological derivation rules to build adverbs in Spanish. By applying this selector to the synset “eng-30-00033562-r” whose lemma is “mildly” and is linked to the POS adjective synset “eng-30-01508719-a” whose lemma is “mild”, we will obtain “suavemente” and “levemente”. The latter are generated based on both available translations for “mild”: “suave” and “leve”, and by applying the following morphological derivation rules.

If the adjective ends in an “o”, it will be replaced by the sequence “amente”, for example, “lento” resulting in “lentamente”. If the adjective ends in an “r” or “n”, then , add the sequence “amente”, for example, “encantador” and “fanfarron” and their respective results “encantadoramente” and “fanfarronamente”. The sequence “mente” will be added to the rest of the adjectives that do not fall in the categories above mentioned, for example, “educada” and “educadamente”. Since this selector builds words by applying morphological derivation rules, we observed that sometimes it would return adverbs that do not exist in Spanish. Therefore, we decided to validate them against a corpus comprised of Spanish news text. To do so, we extracted all adverbs from said corpus to put together a list of adverbs to validate the existence of the adverbs built by the selector. The weakness of such validation method lies in the fact that it may discard adverbs which are correct as they are not found in the reference corpus. However, we considered more pertinent to ensure that accurate words were added. Moreover, it is always possible to use a longer list of known adverbs to reduce the number of false negatives.

Levenshtein This selector uses Levenshtein’s edit distance, based on the assumption that, if the distance between a word in English and its translation is short, they can be considered to have the same sense. Minor modifications are made to reduce the distance between one word and its translation. One example of these transformations is the inversion of the letters “r” and “e” to be applied to the word “tiger” and corresponding translation “tigre”. After doing the transformation, Levenshtein’s distance becomes 0. When implementing the initial selectors we decided not to use it since it did not return good results during the initial experiments. A possible explanation for this is that Spanish and English do not share as many cognate terms as English and French do, as discussed in the WoNeF article.

Singular translation selectors, monosemy and single factorization Levenshtein were inspired in (Atserias, Climent, Farreres, Rigau, & Guez, 1997), while Levenshtein was used in (Pradet, de Chalendar, & Desormeaux, 2014). Derived adverbs was our own production.

3.3 Phase 2: distributional semantics

For the expansion stage we proposed a selector that would exploit the relations between synsets and frequencies of occurrence of both words within a corpus, to determine which translation is the correct one for each ambiguous synset. It

is worth noting that this selector would be used when both related lemmas in English are known, and one of them gets only translation but for the other one there are several possible translations.

A detailed explanation of the implementation of this phase is presented below:

Let's suppose that we have a synset SA associated to synset SB in WordNet through a hypernymy relation. In addition, we have two English lemmas LA and LB for SA and SB respectively. The translations for LA are TA_1 and TA_2, and the translations for LB are TB. So to decide which translation is correct for this lemma, we searched for the occurrence of each translation in a corpus. These searches are considered as a function and represented with letter Θ. This process is called disambiguation.

For example, for calculating $\Theta(TA_1, TB)$ we count all occurrences of the words TA_1 and TB that happen within the same sentence.

$$O_1 = \frac{(\Theta(TA_1, TB)}{\Theta(TA_1) + \Theta(TB)}$$

$$O_2 = \frac{(\Theta(TA_2, TB)}{\Theta(TA_2) + \Theta(TB)}$$

In case $O_1 \geq O_2 \implies TA_1$ is chosen as the translation of LA.

However, if $O_1 < O_2 \implies TA_2$ is chosen as the translation of LA.

An example of the application of this expansion phase follows:

We know that SA = "eng-30-09776346-n" and SB = "eng-30-09816771-n" are related through the hypernym relation and they have the lemmas LA = "affiliate" and LB = "associate" respectively. Furthermore, we know that TB = "asociado" and the translation candidates for "affiliate" are TA_1 = "filial" and TA_2 = "afiliado". Because $O(filial, asociado) = 0.0$ and $O(afiliado, asociado) = 8.18129755379e^{-05}$, then we know that $O(afiliado, asociado) \geq O(filial, asociado) \implies$ the word TA_2 = "afiliado" is chosen as the translation of LA.

The previous result is correct because the English gloss for SA = "eng-30-09776346-n" is: "a subordinate or subsidiary associate; a person who is affiliated with another or with an organization".

The semantic relations used for this process were hypernymy, meronimy and antonymy, and the frequency counts were performed over the Spanish news text corpus.

4 Evaluation of results

We show evaluations for the initial selectors, for the phase 2 process and a global evaluation of results within a lexical semantics effort.

4.1 Quantitative evaluation of phase 1 results

In the evaluation we randomly selected 1000 synsets for each POS (verb, adverb, noun and adjective). The translations of every lemma in all the sorted synsets were obtained and the four selectors mentioned above were applied. The results obtained were stored in a database.

POS	Translated		Untranslated	
R	82,80%	1187	17,20%	246
V	71,90%	1226	28,10%	478
A	59,50%	969	40,50%	659
N	71,20%	1036	28,80%	419
All	71,00%	4418	29,00%	1802

Table 1: Translated lemmas

As can be seen, 71 % of the lemmas processed returned a translation. When we analyze the data at grammatical category level, we see that adverbs is the category with the highest translation percentage, with over 80 %. The other categories behave in a similar way to each other, adjectives being the category with the least coverage with almost 60 % of translations returned.

The following table shows the distribution of the translation of the lemmas for each of the 4000 synsets selected. Our aim was to obtain the results returned for each selector over the total of lemmas translated, but avoiding the overlapping of results by providing an order of importance. There follows the order applied: single selector, monosemy selector, factorization selector and others. For "V", "A" and "N" POS, the others include the translations that were not selected by any selector. For "R" POS, as well as translations not selected by any selector, the translations determined by the derived adverbs selector are also included.

POS	Singulars	Monosemic and not singular	Not monosemic, not singular and factored
R	56,40%	6,10%	1,30%
V	58,00%	2,00%	0,80%
A	77,60%	5,20%	0,90%
N	72,70%	4,60%	1,40%
All	70,20%	4,70%	1,20%

Table 2: Translation by selector

As seen here, verbs and adverbs had the worst result, while adjectives had the best result: 16.3%. We must remember that these data do not consider the results of the derived adverbs selector. These were excluded from the comparison because they could not be compared with the rest of the POS.

4.2 Synsets for which the initial selectors obtained results

POS	Yes		No	
R	73,90%	739	26,10%	261
V	52,80%	528	47,20%	472
A	59,90%	599	40,10%	401
N	63,70%	637	36,30%	363
All	62,60%	2845	37,40%	1155

Table 3: Synsets for which the initial selectors obtained results

As seen here, the POS with the highest coverage by initial selectors were adverbs, with almost 74%; without distinguishing according to POS, there is a 62.60% coverage.

4.3 Comparison with current WordNet

POS	New		Existent	
R	83,80%	694	16,20%	134
V	50,40%	390	49,60%	384
A	62,50%	429	37,50%	257
N	54,80%	423	45,20%	349
All	63,30%	1936	36,70%	1124

Table 4: Comparison with current WordNet

As seen here, for each POS there was a high percentage of synsets that had translations which were not found in the current Spanish WordNet (MCR 3.0). Adverbs is the grammatical category with the highest percentage: approximately 83%. In total there were just over 63% new synsets. As only the initial selectors were applied, we concluded that we would see a significant improvement at the end of the process.

A manual qualitative evaluation was conducted to measure the accuracy of the results. We randomly selected 25 synsets for each POS (verb, adverb, noun and adjective) of the added ones, and we verified if the result was correct or not. For the selectors that work at synset level, the data in table 5 reflect the percentages of the resulting correct or incorrect synsets, and for the selectors that work at lemma level, the percentages correspond to the resulting correct or incorrect synsets.

POS	Monosemy	Single translation	Factorization	Derived adverb
V	93.48%	98.39%	100.00%	-
A	96.08%	100.00%	96.00%	-
N	93.48%	100.00%	100.00%	-
R	97.14%	94.59%	92.00%	92.00%
All	95.04%	98.25%	97.00%	-

Table 5: Accuracy for the initial selectors

Although the derived adverbs selector was the least accurate one, it returned a very good result: 92%.

As seen in the charts above, the results of the four selectors were very good: all show over 92 % of effectiveness and some reach 100 % for some POS.

5 Evaluation of phase 2 results

5.1 Lemmas processed

The 1040 synsets that were not translated in phase 1 because they were ambiguous were applied and evaluated in phase 2. As phase 2 can fail for various reasons, in this section we present detailed information about the results obtained to identify such reasons. As phase 2 exploits the relations between the existing synsets in WordNet up to the present, if the synsets are not related to any other synsets, or if they are, but such synsets are empty for Spanish, this method returns no results. Therefore three different groups can be observed on the following table.

POS	With relations	With relations and no trans.	With relations and trans.
R	83,10%	10,56%	6,34%
V	1,20%	10,40%	88,40%
A	1,79%	34,52%	63,69%
N	0,00%	22,17%	77,83%
Total	12,21%	16,92%	70,87%

As seen here, adverbs is the grammatical category that has the least connected synsets, which shows that our method does not return good results for this POS. The other grammatical categories have enough relations and they are sufficiently complete for phase 2 to return results.

5.2 Lemmas processed in phase 2 with relations and with translations for these relations

It is important to highlight that for lemmas corresponding to synsets associated to other already complete synsets, the method applied in phase 2 can fail if there were no occurrences in the corpus of the possible candidates for all lemmas. This is explained in the following results.

POS	With result		Without result	
R	33,33%	3	66,67%	6
V	63,80%	282	36,20%	160
A	60,75%	65	39,25%	42
N	70,95%	127	29,05%	52
Total	64,72%	477	35,28%	260

As can be seen here, there is margin for improvement: 35 %, which can be improved by increasing the size of the search corpus.

5.3 Comparison with current WordNet

In this section we compare the results obtained in phase 2 with the results of the current WordNet, as only the results that do not appear in the current WordNet will entail a real increase in the completeness of WordNet.

POS	Not present		Present	
R	66,67%	2	33,33%	1
V	73,05%	206	26,95%	76
A	52,31%	34	47,69%	31
N	62,20%	79	37,80%	48
Total	67,30%	321	32,70%	156

5.4 Manual evaluation of disambiguated synsets

A manual qualitative evaluation was conducted to measure the accuracy of the results. We randomly selected 25 synsets for each POS (verb, adverb, noun and adjective) and we verified if the result was correct or not. We must remember that for adverbs there were only two results. It is important to remember that most of the errors detected at this stage correspond to lemmas that had been accurately translated but whose translation was not the correct one for the synset in question. The lemma “cup” of synset “eng-30-03147901-n” with the sense of “trophy” is a good example of this. The translations obtained for the lemma were “taza” and “copa”, and when requesting disambiguation the process selected “taza”, which was not the correct meaning for this synset.

POS	Correct	Incorrect
R	100.00%	0.00%
V	68.00%	32.00%
A	84.00%	16.00%
N	68.00%	32.00%
Total	74.03%	25.97%

From these evaluations we can conclude that phase 2 was not as accurate as phase 1. These results could be improved by increasing the size of the corpus or by improving the method. A larger corpus would have more sentences, that is to say, more contexts where the meaning of candidates can be validated. The translations where the gender does not match in English could be discarded to improve the method. Doing this would discard cases like that of synset “spa-30-10129825-n”, whose gloss is “mujer joven”. For the lemma “girl”, which corresponds to said English synset, a possible translation obtained was “chico”. This is a clear example where the original lemma in English and the resulting translation do not match in gender. Another way to improve the method would be to prioritize some specific relations.

6 Evaluation of the results on Corin lexicon

To evaluate the results obtained in both phases we implemented a task to measure the semantic coverage on a small corpus, in this case Corin. For this task we obtained all the lemmas in the corpus, applied Freeling to know the grammatical category, and then searched WordNet. This process was first executed with the original WordNet, our starting point, and then with the resulting WordNet. The aim was to measure the improvement in the coverage of the existing lemmas in the corpus under study of the resulting WordNet regarding the current WordNet. We must remember that the process to improve WordNet was executed on a random set of 1000 synsets per POS. The results obtained must be weighed considering the percentage these synsets represent within the total number of synsets for each POS. These percentages are shown in the following table.

There follows a table with the percentages of

POS	Total	Processed Synsets	
V	13845	1000	7.22%
N	83090	1000	1.20%
R	3621	1000	27.62%
A	18156	1000	5.51%

coverage obtained according to each POS, for the two versions of WordNet: the original one and the one expanded by this method.

POS	Original WordNet		Expanded WordNet		In the Corpus
V	75.35%	1235	77.36%	1268	1639
N	69.29%	2780	70.09%	2812	4012
R	32.79%	121	62.87%	232	369
A	51.00%	840	54.34%	895	1647

We can conclude that adverbs was the category with the best results, reaching a coverage of almost 63 % over the original 33 %. Two reasons explain this: first, adverbs is the category least covered by the original WordNet, and it was also the POS where the strategy was implemented more times, which was executed on just over 27 % of its synsets. The coverage also improved for the other POS. Though it is true that the improvement was relatively small (between 1 % and 3 %), we must remember that in these cases the method was applied to a small percentage of the synsets in WordNet.

7 Conclusions

Different strategies were designed and implemented in order to enrich the current Spanish WordNet from the English WordNet within the context of the expansion model. The strategy was to use a series of selectors which were called "initial selectors" as a first step. We then applied a method based on the exploitation of the semantic relations of WordNet so as to add variants that the initial selectors had not been able to add. The results obtained show that the strategy used is effective as it entails a significant improvement of the current Spanish WordNet, thus complying with the initial expectations. One of the weaknesses lies in the translation methods and tools, as they provide the resources our proposals are based on. This is why they strongly condition the final results. Regarding the strategy implemented, the initial selectors are sufficient to significantly improve the current WordNet, with a 92 % accuracy, while there was a 74 % accuracy in phase 2.

References

Armentano Oller, C., Corbí Bellot, A. M., Forcada Zubizarreta, M. L., Ginestí Rosell, M., Montava Belda, M. A., Ortiz Rojas, S., ... others (2007). Apertium, una plataforma de código abierto para el desarrollo de sistemas de traducción automática. Universidad de Cádiz. Servicio de Publicaciones.

Atserias, J., Climent, S., Farreres, X., Rigau, G., & Guez, H. R. (1997). Combining multiple methods for the automatic construction of multilingual wordnets. In *In proceedings of international conference on recent advances in natural language processing (ranlp'97), tzigov chark* (pp. 143–149).

Benítez, L., Cervell, S., Escudero, G., López, M., Rigau, G., & Taulé, M. (1998). Methods and tools for building the catalan wordnet. In *in proceedings of elra workshop on language resources for european minority languages.*

Bing online translator. (2015). `https://www.bing.com/translator`.

Carreras, X., Chao, I., Padró, L., & Padró, M. (2004). Freeling: An open-source suite of language analyzers. In *Proceedings of the 4th international conference on language resources and evaluation (lrec'04).*

González-Agirre, A., Laparra, E., Rigau, G., & Donostia, B. C. (2012). Multilingual central repository version 3.0: upgrading a very large lexical knowledge base. In *Gwc 2012 6th international global wordnet conference* (p. 118).

Grassi, M., Malcuori, M., Couto, J., Prada, J. J., & Wonsever, D. (2001). Corpus informatizado: textos del español del uruguay (corin). In *Slplt-2-second international workshop on spanish language processing and language technologies-jaén, españa.*

Pradet, Q., de Chalendar, G., & Desormeaux, J. B. (2014). Wonef, an improved, expanded and evaluated automatic french translation of wordnet. *Volume editors*, 32.

Vossen, P. (1998). Introduction to eurowordnet. In *Eurowordnet: A multilingual database with lexical semantic networks* (pp. 1–17). Springer.

Wikimedia foundation. 2008b. wiktionary. (2008). `http://www.wiktionary.org`.