Building WordNets by machine translation of sense tagged corpora*

Antoni Oliver, Salvador Climent
Universitat Oberta de Catalunya (UOC)
Avda. Tibidabo 39-43 08035 Barcelona
aoliverg,scliment@uoc.edu

Abstract

This paper describes a methodology for the construction of WordNets based on machine translation of an English sense tagged corpus. For the construction of such a corpus we used two freely available resources: the Semcor Corpus and the Princeton WordNet Gloss Corpus. This methodology is applied on the construction of Spanish and Catalan WordNet 3.0. In our first experiments we used a simple word alignment algorithm obtaining precision results comparable to those obtained by methods based on bilingual dictionaries. After that, we used a freely available statistical word alignment algorithm (Berkeley Aligner) obtaining better results. This methodology can be suitable for those languages with an available statistical machine translation system and can be used for either constructing WordNets from the scratch or for enlarging existing WordNets.

1 Introduction

WordNet (Fellbaum, 1998) is a lexical knowledge database that organizes nouns, verbs, adjectives and adverbs in sets of synsets. Each synset represents a lexicalized concept in English. Synsets are connected to other synsets by semantic relations (hiponymy, antonymy, meronymy, troponymy, etc.). Each synset has a gloss or definition. WordNet has become a standard resource in all kind of researches and applications in Natural Language Processing.

The English WordNet (PWN *Princeton WordNet* henceforth) is a free resource and it's available for download from its web page of the University of Princeton[1]. The current version is 3.1 but at the moment of writing this paper the version available for downloading is 3.0.

Several projects have developed WordNets for other languages: EuroWordNet (Vossen, 1998) initially for Dutch, Italian, Spanish and for the enlargement and improvement of English, and in an extension of the project for German, French and Estonian; Balkanet (Tufis et al., 2004) for Bulgarian, Greek, Romanian, Serbian and Turkish and RusNet (Azarova et al., 2002) for Russian, among others. On the Global WordNet Association[2] website we can find a comprehensive list WordNets available for different languages.

Many of the available WordNets are subject to proprietary licenses. Some of the WordNets other than PWN holding a free license are: Catalan, Danish, French WOLF WordNet, Hebrew, Hindi, Japanese, Russian and Tamil WordNets among others. Our goal is to improve Catalan WordNet 3.0 and develop Spanish WordNet 3.0 and distribute both under free licenses.

2 WordNet building strategies

In this section we review some techniques that have been used to build WordNets for various languages other than English. We can distinguish two general approaches for building WordNets (Vossen, 1998):

- **Merge model**: a new ontology is constructed for the target language, with its own levels and relations. After that, relations between PWN and this local WordNet are generated.
- **Expand model**: English variants associated with PWN synsets are translated, using bilingual dictionaries or other strategies. Subsequently, relations imposed by the structure

*This research has been carried out thanks to the Project MICINN, TIN2009-14715-C04-04 of the Spanish Ministry of Science and Innovation

[1] http://wordnet.princeton.edu
[2] http://www.globalwordnet.org

of PWN are assumed to be valid for the target language. It is not necessary to establish interlingual relations because local WordNet and PWN are parallel.

Each of these strategies has a number of pros and cons (Vossen, 1996). On the one hand, the expand model is simpler from a technical point of view and ensures a greater degree of compatibility between WordNets for different languages. In contrast, WordNets developed using this technique are overly influenced by PWN and thus retain their mistakes and structural drawbacks. The merge model strategy is more complex from a technical point of view but allows a more direct use of existing ontologies and thesauri.

In the EuroWordNet project (Vossen, 1999) Dutch and Italian used the merge model, whereas Spanish chose the expand model. Catalan WordNet used the same methodology as Spanish WordNet (Benítez et al., 1998). Other projects using the merge model are RussNet (Azarova et al., 2002) and BalkaNet (Tufis et al., 2004). The expand model has been used as well in the MultiWordNet project for Italian (Pianta et al., 2002), the French WOLF WordNet (Sagot and Fišer, 2008), the Indonesian WordNet (Putra et al., 2008) and the Hungarian WordNet (Miháltz et al., 2008), among others.

Some WordNets have been constructed using a hybrid approach. In the Portuguese WordNet (Marrafa, 2002), for instance the vocabulary and the set of internal relations were developed independently from PWN but PWN was taken as model and source. Basque WordNet (Agirre et al., 2002) used also a hybrid approach since PWN was taken as a starting point but new hierarchies and lexical-semantic relations extracted from dictionaries were also used to enrich, and in some cases, modify those of PWN.

3 Using machine translation in the construction of WordNets

In this section we will briefly present two projects related to WordNet that use machine translation systems to translate a semantically tagged corpus: the construction of the Macedonian WordNet and the Babelnet project.

3.1 The Macedonian WordNet

In the construction of the Macedonian WordNet (Saveski and Trajkovski, 2010) a bilingual English-Macedonian dictionary has been used as a main source of information. For monosemic entries the PWN synset - Macedonian word relation can be established directly. For polysemic entries the task of establishing relations between PWN synsets and Macedonian words can be seen as a Word Sense Disambiguation problem. In this paper the authors used the proposal of (Dagan and Itai, 1994) based in the fact than a given word tends to co-occur more times in a big corpus with the open class words from its definition than with other words. The authors had to face two main problems: (i) the available English-Macedonian dictionary didn't have definitions; and (ii) no big Macedonian corpora was available. The first problem was solved by translating the PWN glosses into Macedonian using Google Translate; in this way, translated glosses were used as definitions. The second problem was solved using the web as a corpus through the *Google Similarity Distance* (Cilibrasi and Vitanyi, 2007). This distance between a word or a phrase x and a word or a phrase y is calculated using the Google search results for x, for y and form a query including x and y.

3.2 The Babelnet project

The main goal of the Babelnet project (Navigli and Ponzetto, 2010) is the creation of a big semantic network by linking the lexicographic knowledge from WordNet to the encyclopedic knowledge from Wikipedia. This is done by assigning WordNet synsets to Wikipedia entries. Since Wikipedia entries bear interlingual relations, a variant in several languages can be assigned to some of the WordNet synsets. Therefore, if a relation between a synset s and an English Wikipedia entry w_{eng} has been set, using the Wikipedia interlingual links the same relation can be set for all languages having the corresponding Wikipedia entry: w_{spa}, w_{fra}, etc. For those languages lacking the corresponding Wikipedia entry, the authors propose the use of Google Translate to translate a set of English sentences containing the synset s. This set of sentences is built using two sources:

- The Semcor corpus (Miller et al., 1993).
- Sentences from Wikipedia containing a link to the English Wikipedia page w_{eng}.

Once the automatic translation is done the most frequent translation is detected and included as a variant for the synset s in the given language.

4 Our approach

4.1 Goal

In this paper we propose an approach for the construction of WordNets based on machine translation of sense tagged corpora. As it is known, it is relatively easy to extract a bilingual dictionary from a parallel corpus (Och and Ney, 2003). If the sentences of the source language are semantically tagged with PWN synsets, we can treat these synsets as words so that we can get their corresponding variants for the equivalent synsets in the target language. To our knowledge it doesn't exist a big semantically tagged bilingual corpus for English and either Spanish or Catalan -or at least tagged for the English source. For this reason we use machine translation systems to get such a parallel corpus from an English monolingual corpus.

4.2 Sense tagged corpus

We have used two sense tagged corpora for English (see 4.6 below) where the tags are the PWN 3.0 synsets. For each sentence we can get two versions: the sentence itself, and the sentence with all tagged words substituted by its PWN 3.0 synset. We can see an example in the following lines:

```
Then he noticed that the dry wood of the
wheels had swollen .
00117620r he 02154508v that the 02551380a
15098161n of the 04574999n had 00256507v .
```

4.3 Use of Machine Translation

The machine translation system to be used in this work must be capable to perform good lexical selection, that is, it should select the correct target words for the source English sentence. For ambiguous words the system must be able to disambiguate and choose the correct translation. Other translation errors are less important for our experiments. For our experiments we use two statistical machine translation systems: Google Translate and Microsoft Bing Translator. We didn't assess in deep the ability of these systems to do a correct lexical selection, but we performed some succesful tests, as explained now on. Consider first sentences containing the English word *wood*. This word is a variant of both the PWN synset 15098161-n (*the hard fibrous lignified substance under the bark of trees*; translated into Spanish as *madera* and into Catalan as *fusta*) and 08438533-n (*the trees and other plants in a large densely wooded area*; Spanish *bosque* and Catalan *bosc*. If we take a sentence corresponding to the first sense and we translate it using the given MT systems we get:

```
This house is made of wood.
*Google Translate:
Esta casa es de madera.
Aquesta casa és de fusta.
*Microsoft Bing Translator:
Esta casa está hecha de madera.
Aquesta casa és fa de fusta
```

By performing the same task to a sentence corresponding to the second sense we get:

```
He got lost in the wood beyond Seattle.
*Google Translate:
Se perdió en el bosque más allá de Seattle.
Es va perdre en el bosc més enllà de Seattle.
*Microsoft Bing Translator:
Se perdió la madera más allá de Seattle.
Es va perdre en la fusta més enllà de Seattle.
```

Consider now the English word *bank*. As a noun it has 10 meanings according PWN. We will concentrate on two of them: 09213565n (*sloping land (especially the slope beside a body of water)*) and 08420278n (*a financial institution that accepts deposits and channels the money into lending activities*). The first meaning has three possible variants in Spanish (*margen, orilla, vera*) and two in Catalan (*marge, vora*). The second meaning has one variant both in Spanish (*banco*) and Catalan (*banc*)[3].

If we take a sentence corresponding to the first sense and we translate it with the given MT systems we get:

```
She waits on the bank of the river.
*Google Translate:
Ella espera en la orilla del río.
Ella espera a la vora del riu.
*Microsoft Bing Translator:
Ella espera en la orilla del río.
Ella espera a la riba del riu.
```

and if we perform the same task to a sentence corresponding to the second sense we get:

```
She puts money into the bank.
*Google Translate:
Ella pone el dinero en el banco.
Ella posa els diners al banc.
*Microsoft Bing Translator:
Ella pone dinero en el Banco.
Ella posa diners en el Banc.
```

As we can see, these systems can do, at least in certain situations, a good lexical selection (note that Microsoft translator failed to select the correct Spanish and Catalan translation for *wood* as a forest, using *madera* instead of *bosque* and *fusta* instead of *bosc*). In the second example we can observe an interesting result. While in Spanish both

[3] All these variants were taken from the preliminary versions of Spanish and Catalan WordNet 3.0.

systems propose a translation for bank (of a river) that coincides with one of the variants of Spanish WordNet, in Catalan Microsoft Translator proposes a translation (*riba*) not registered as a possible variant in the Catalan WordNet, but that can be considered correct. In some situations, the use of machine translation systems can be useful for enriching existing local WordNet versions.

The machine translation systems fail to choose the correct word in some cases, thus this will undoubtedly lead to errors. Nevertheless, our evaluation methodology, described in section 5, gives us an idea of the influence of the translation quality in the results.

4.4 Bilingual corpus

Using a machine translation system we can get a bilingual corpus. For example, from the sentence used in 4.2 above we can get the pair English - Spanish:

```
Then he noticed that the dry wood of the
wheels had swollen.
Entonces se dio cuenta de que la madera
seca de las ruedas se había hinchado.
```

But in fact we are not really interested in this pair of languages. Instead, we are interested in the pair Sense Tagged English - Spanish:

```
00117620r he 02154508v that the 02551380a
15098161n of the 04574999n had 00256507v .
Entonces se dio cuenta de que la madera
seca de las ruedas se había hinchado.
```

4.5 Word alignment algorithms

So, having a parallel corpus Sense Tagged English - Target Language, the task of deriving the local WordNet can be viewed as a word alignment problem. We need an algorithm capable to select the following relations (from the example in 4.4)

```
Synset    - Spanish      - Catalan
00117620r - entonces     - llavors
02154508v - darse cuenta - adonar-se
02551380a - seco         - sec
15098161n - madera       - fusta
```

Fortunately, word alignment is a well-known task and there are several algorithms available to solve it. In this work we will test two algorithms:

- MFT: A very simple algorithm based on the detection of the most used target word of the same part-of-speech in the translation of the sentences containing the PWN synset.
- Berkeley Aligner[4] (Liang et al., 2006): A well known word alignment algorithm.

[4]http://code.google.com/p/berkeleyaligner/

In both algorithms we force two restrictions: (i) we only detect as a variant for a given synset simple lexical units, that is, no multiwords; and (ii) we only detect one variant for each synset. In a future work we will try to overcome such restrictions.

4.5.1 MTF - *Most Frequent Translation*

This algorithm works as follows:

- We take all the synsets in the corpus ordered by frequency, beginning with the most frequent.
- For every synset we take all target sentences alligned to the corresponding source sentences containing this synset.
- The most frequent target lemma of the same POS from this set of sentences is the candidate variant for the synset.

For example, the synset 00393105a (*white_1*) appears 1190 times in the corpus. By applying the algorithm we get a set of candidates. If we take the most frequent adjectives in the translation of the sentences containing this synset, we obtain for Spanish:

```
blanco:1139;pequeño:212;perenne:160;
grande:105;rosa:86;rojo:83;amarillo:81;
fragante:66;púrpura:60;azul:53;
```

The most frequent is *blanco*, that is several times more frequent than the second one. So we take this word as a variant in Spanish for this synset.

4.5.2 Berkeley Aligner

In our experiments we also used the Berkeley Aligner (Liang et al., 2006). We used it with default options and only one variant for each synset was taken -the one bearing the higher probability. In these experiments we used lemma and part-of-speech information.

4.6 Linguistic resources

For our experiments we need a sense tagged corpus. The tags must be the PWN 3.0 synsets. Fortunately there are two freely available resources:

- The Semcor corpus[5] (Miller et al., 1993).
- The Princeton WordNet Gloss Corpus (PWGC) [6], consisting of the WordNet 3.0 glosses with semantic annotation.

[5]http://www.cse.unt.edu/ rada/downloads.html
[6]http://wordnet.princeton.edu/glosstag.shtml

Corpus	Sentences	Words
Semcor	37.176	721.622
PWGC	117.659	1.174.666
Total	154.835	1.896.288

Table 1: Size of the corpus

	Spanish		Catalan	
	Synsets	Variants	Synsets	Variants
T	115.989	69.829	70.856	46.027
N	80.324	48.676	51.598	36.454
V	16.805	9.010	11.577	5.424
A	17.752	11.450	7.679	4.148
R	1.108	693	2	1

Table 2: Size of the preliminar WordNets used for evaluation

In table 1 we can observe the total number of sentences and words in the corpus.

4.7 Linguistic tools

4.7.1 Machine Translation Systems

- Google Translate[7].
- Microsoft Bing Translator[8].

Both systems are statistical and allow us to translate from English to Spanish and Catalan.

4.7.2 Morphosyntactic tagger

All sentences in the corpus, both source and target sentences, have been morphosyntactically tagged whith Freeling (Padró et al., 2010a).

4.7.3 Preliminary versions of the Spanish and Catalan 3.0 WordNets

The results obtained with the different algorithms have been automatically evaluated with a preliminary version of the Spanish and Catalan 3.0 WordNets. These preliminary versions have been obtained from versions 1.6 applying a mapping and other techniques. In table 2 we can see the number of synsets and variants of these versions.

5 Evaluation

We conducted the experiments for three language pairs:

- English - English: this dummy language pair allows to evaluate the effect of the quality of the machine translation system in the results. As we are using the source language we simulate a perfect machine translation system.
- English - Spanish
- English - Catalan

[7]http://translate.google.com

[8]http://www.microsofttranslator.com/

We have tested all the algorithms for every language pair. As machine translations systems, for English-Catalan and English-Spanish we have used both Google Translate and Microsoft Translator.

5.1 Algorithm MFT

In figure 1 we can see the results of this algorithm for all the language pairs (for the translated pairs we show only the results for Google Translate). In the figure we can see the precision obtained regarding the number of translated synsets. As we can realize, precision decreases along with the number of synsets. As we start the evaluation with the most frequent synset, this means that the precision decreases with the decrease of frequency, as it was expected.

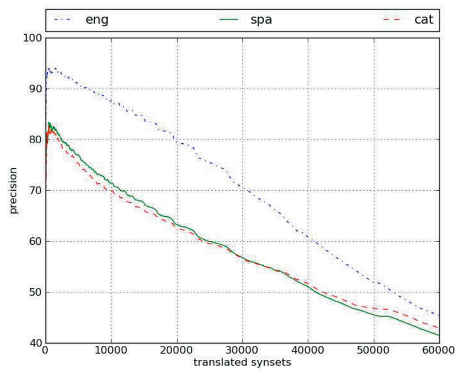

Figure 1: Precision of MFT for English, Spanish and Catalan (using Google Translate).

As we can see in the figure, with this algorithm we can obtain for the pair English-English about 20.000 translated synsets with a precision of 80%, and about 30.000 with a precision score of 70%. For this language pair we don't have the negative effects of errors coming from the machine translation system. As we can observe, results for Spanish and Catalan are very similar so that we can obtain about 2.200 translated synset with 80% precision and 10.000 with 70%.

We must keep in mind that we are performing an automatic evaluation using the full PWN 3.0 for English but using preliminary versions for Spanish and Catalan. Therefore, in some cases we might be rejecting correct variants, as they are not present in the preliminary versions. For English this is less likely to happen as PWN is much more complete.

Results obtained with Microsoft Translator are very similar with a very slight decrease in precision. Figure 2 shows a comparison for Spanish.

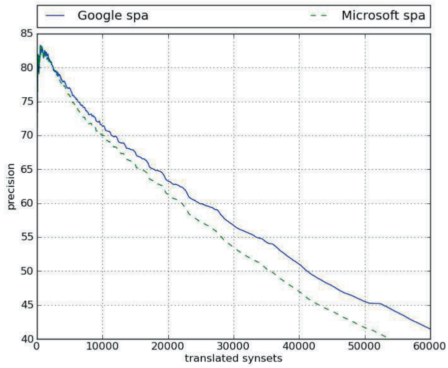

Figure 2: Comparison of MFT for Spanish using Google Translate and Microsoft Translator

5.1.1 Berkeley Aligner

Figure 3 shows the results achieved by using the Berkeley Aligner -please notice the change of x and y scales respect figure 1. In the figure we can observe some interesting facts. First of all, precision has increased significantly for the pair English-English so that it remains practically constant regardless of the frequency of the synset. For this ideal language pair we obtain precision scores higher than 96%. For real language pairs results are not so good since precision decreases with frequency. We can also observe that results for Spanish are better than for Catalan and the difference is higher for smaller frequencies.

As we are using a statistical word alignment algorithm, we can get a probability score for each synset. We can improve precision by taking only those alignments with a probability higher than a given threshold. Of course, this will have a cost on the recall but the goal is to get a subset of results indicating high precision. We have taken 0.9 as probability threshold, that is, we get the alignment if its probability is 0.9 or higher. In figure 4 we can observe the results for Spanish and a comparison with MFT and Berkeley Aligner taking the best candidate for all alignments, regardless of its precision. As we can observe, Berlkeley Aligner performs better than MFT for all frequencies -about 5% better. Taking those alignments with a precision of 0.9 or higher we get a good improvement of the results until the 15.000th synset. Until the 5.000th synset the improvement is about 7 points with respect the Berkeley Aligner.

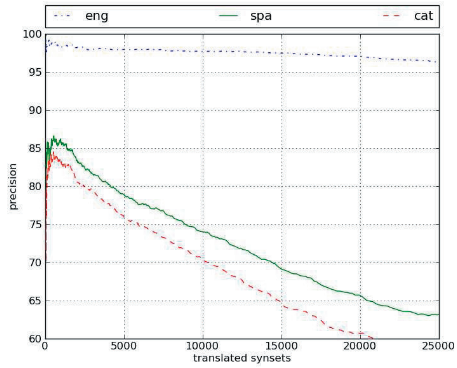

Figure 3: Precision for Berkeley Aligner for English, Spanish and Catalan.

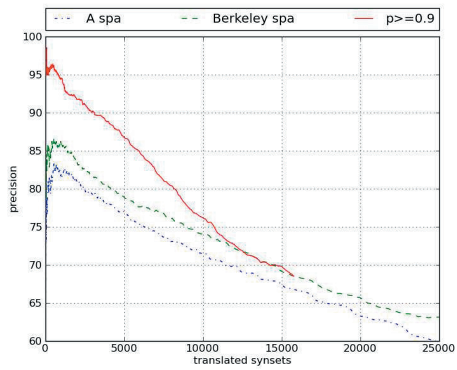

Figure 4: Comparison of MFT, Berkeley Aligner and Berkeley Aligner for $p \geq 0.9$ for Spanish

In figure 5 we can see the same comparison for Catalan. In this case Berkeley Aligner starts with better results but for a given frequency MFT starts to perform better. If we see the results for the Berkeley Aligner with probabilities higher than 0.9, we can observe that we get better results until the 10.000th synset. Until the 5.000th synset the improvement is also about 7 points with respect the Berkeley Aligner.

As for MFT results for Berkeley Aligner for the different machine translation systems are practically identical. In figure 6 we can see a compar-

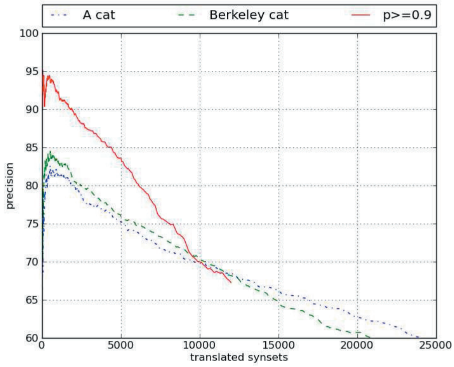

Figure 5: Comparison of MFT, Berkeley Aligner and Berkeley Aligner for $p \geqslant 0.9$ for Catalan

ison for Spanish using Google Translate and Microsoft translator. As we can see we obtain slight better results with Google Translate. The same applies for Catalan.

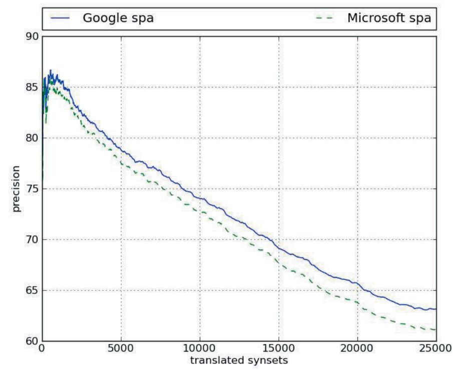

Figure 6: Comparison of Berkeley Aligner for Spanish using Google Translate and Microsoft Translator.

6 Conclusions and future work

In this paper we have presented a methodology for the construction of WordNets based on machine translation of a sense disambiguated corpus of English. This methodology has been used for the construction of the Spanish and Catalan WordNets 3.0 along with methodologies based on a mapping of previous existing versions.

This methodology achieves values of precision and number of obtained synsets comparable to methodologies based on bilingual dictionaries for Spanish (Atserias et al., 1997) and Catalan (Benítez et al., 1998). To perform a good comparison we need to further analyse our results in order to group synsets according to the degree of polisemy. For Spanish, our best algorithm (Berkeley Aligner for $p \geqslant 0.9$) performs better than all the criteria presented in (Atserias et al., 1997) except monosemic-1 criterion. Nevertheless, our proposal performs worse than their combination of criteria as they can obtain 7.131 with a precision higher than 85% whereas we can obtain only 5.890 in the same conditions. For Catalan (Benítez et al., 1998) obtained much better results and our best proposal only outperforms monosemic-3 and all polysemic criteria.

This methodology can be used for the construction of new WordNets and also for the enrichment of existing WordNets. Our proposal can be applied to those languages having an available English-source machine translation system. Such a system must be able to perform good lexical selection while other translation errors are less important. It has to be noticed that Google Translate offers machine translation from English to more than 50 languages and Microsoft Translator to more than 35 languages. Of course, not all language pairs achieve the same quality, but they can be a good starting point.

In future work we plan to overcome some of the restrictions of the algorithms presented in this paper. First of all, we will use Berkeley Aligner to try to get more of one variant for each synset. This can be done by observing the assigned probability and taking more than one candidate if probability scores are similar enough. We will aim to overcome as well the restriction of considering the possible variants only when they are simple lexical units. This can also be done by further exploring the possibilities of the alignment algorithm.

Finally, we want to cope with the limit of recall due the corpus: only those synsets present in the corpus can be retrieved. To achieve this goal we will to explore several possibilities, including the use of automatic semantic tagging (Padró et al., 2010b) and the use of Wikipedia page links as a pseudo-semantical tagging as in (Navigli and Ponzetto, 2010).

References

E. Agirre, O. Ansa, X. Arregi, J. M Arriola, A. D de Ilarraza, E. Pociello, and L. Uria. 2002. Methodological issues in the building of the basque WordNet: quantitative and qualitative analysis. In *Proceedings of the first International WordNet Conference in Mysore, India*, page 21–25.

J. Atserias, S. Climent, X. Farreres, G. Rigau, and H. Rodriguez. 1997. Combining multiple methods for the automatic construction of multi-lingual WordNets. In *Recent Advances in Natural Language Processing II. Selected papers from RANLP*, volume 97, page 327–338.

I. Azarova, O. Mitrofanova, A. Sinopalnikova, M. Yavorskaya, and I. Oparin. 2002. Russnet: Building a lexical database for the Russian language. In *Workshop on WordNet Structures and Standarisation, and how these affect WordNet Application and Evaluation*, pages 60–64, Las Palmas de Gran Canaria (Spain).

Laura Benítez, Sergi Cervell, Gerard Escudero, Mònica López, German Rigau, and Mariona Taulé. 1998. Methods and tools for building the catalan WordNet. In *In Proceedings of the ELRA Workshop on Language Resources for European Minority Languages*.

R. L Cilibrasi and P. M.B Vitanyi. 2007. The Google similarity distance. *IEEE Transactions on Knowledge and Data Engineering*, 19(3):370–383.

I. Dagan and A. Itai. 1994. Word sense disambiguation using a second language monolingual corpus. *Computational Linguistics*, 20(4):563–596.

C. Fellbaum. 1998. *WordNet: An electronic lexical database*. The MIT press.

Percy Liang, Ben Taskar, and Dan Klein. 2006. Alignment by agreement. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, HLT-NAACL '06, page 104–111, Stroudsburg, PA, USA. Association for Computational Linguistics. ACM ID: 1220849.

Palmira Marrafa. 2002. Portuguese WordNet: general architecture and internal semantic relations. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 18(spe).

M. Miháltz, C. Hatvani, J. Kuti, G. Szarvas, J. Csirik, G. Prószéky, and T. Váradi. 2008. Methods and results of the Hungarian wordnet project. In *Proceedings of the Fourth Global WordNet Conference. GWC*, pages 387–405, Szeged, Hungary.

George A Miller, Claudia Leacock, Randee Tengi, and Ross T Bunker. 1993. A semantic concordance. In *Proceedings of the workshop on Human Language Technology*, HLT '93, page 303–308, Stroudsburg, PA, USA. Association for Computational Linguistics. ACM ID: 1075742.

Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, page 216–225, Stroudsburg, PA, USA. Association for Computational Linguistics. ACM ID: 1858704.

F. J Och and H. Ney. 2003. A systematic comparison of various statistical alignment models. *Computational linguistics*, 29(1):19–51.

L. Padró, M. Collado, S. Reese, M. Lloberes, and I. Castellón. 2010a. FreeLing 2.1: Five years of open-source language processing tools. In *LREC*, volume 10, page 931–936.

L. Padró, S. Reese, E. Agirre, and A. Soroa. 2010b. Semantic services in freeling 2.1: Wordnet and UKB. In *Proceedings of the 5th International Conference of the Global WordNet Association (GWC-2010)*.

E. Pianta, L. Bentivogli, and C. Girardi. 2002. MultiWordNet. developing an aligned multilingual database. In *1st International WordNet Conference*, pages 293–302, Mysore (India).

D. D Putra, A. Arfan, and R. Manurung. 2008. Building an Indonesian WordNet. In *Proceedings of the 2nd International MALINDO Workshop*.

B. Sagot and D. Fišer. 2008. Building a free French wordnet from multilingual resources. In *Proceedings of OntoLex*.

M. Saveski and I. Trajkovski. 2010. Automatic construction of wordnets by using machine translation and language modeling. In *13th Multiconference Information Society*, Ljubljana, Slovenia.

D. Tufis, D. Cristea, and S. Stamou. 2004. BalkaNet: aims, methods, results and perspectives: a general overview. *Science and Technology*, 7(1-2):9–43.

P. Vossen. 1996. Right or wrong. combining lexical resources in the EuroWordNet project. In *Proceedings of Euralex-96*, page 715–728, Goetheborg.

P. Vossen. 1998. Introduction to Eurowordnet. *Computers and the Humanities*, 32(2):73–89.

P. Vossen. 1999. EuroWordNet a multilingual database with lexical semantic networks. *Computational Linguistics*, 25(4).