

Methodology and evaluation of the Galician WordNet expansion with the WN-Toolkit*

Metodología y evaluación de la expansión del WordNet del gallego con WN-Toolkit

Xavier Gómez Guinovart

Universidade de Vigo
Vigo, Galiza
xgg@uvigo.es

Antoni Oliver

Universitat Oberta de Catalunya
Barcelona, Catalunya
aoliverg@uoc.edu

Resumen: En este artículo se presenta la metodología utilizada en la expansión del WordNet del gallego mediante el WN-Toolkit, así como una evaluación detallada de los resultados obtenidos. El conjunto de herramientas incluido en el WN-Toolkit permite la creación o expansión de wordnets siguiendo la estrategia de expansión. En los experimentos presentados en este artículo se han utilizado estrategias basadas en diccionarios y en corpus paralelos. La evaluación de los resultados se ha realizado de manera tanto automática como manual, permitiendo así la comparación de los valores de precisión obtenidos. La evaluación manual también detalla la fuente de los errores, lo que ha sido de utilidad tanto para mejorar el propio WN-Toolkit, como para corregir los errores del WordNet de referencia para el gallego.

Palabras clave: WordNet, adquisición de información léxica, corpus paralelos, recursos plurilingües

Abstract: In this paper the methodology and a detailed evaluation of the results of the expansion of the Galician WordNet using the WN-Toolkit are presented. This toolkit allows the creation and expansion of wordnets using the *expand* model. In our experiments we have used methodologies based on dictionaries and parallel corpora. The evaluation of the results has been performed both in an automatic and in a manual way, allowing a comparison of the precision values obtained with both evaluation procedures. The manual evaluation provides details about the source of the errors. This information has been very useful for the improvement of the toolkit and for the correction of some errors in the reference WordNet for Galician.

Keywords: WordNet, lexical acquisition, parallel corpora, multilingual resources

1 Introduction

WordNet (Fellbaum, 1998) is a lexical resource where nouns, verbs, adjectives and adverbs are organised in sets of synonyms called *synsets*. In this resource, synsets are connected by semantic relations as hiponymy, antonymy, meronymy, troponymy, etc. The original WordNet was created for English in the Princeton University and nowadays there are WordNet versions for several languages. In the website of the Global WordNet Association¹ a list of existing wordnets are available. Some of these wordnets hold a free li-

cense and in the Open Multilingual WordNet project (Bond and Kyonghee, 2012) they are published under a common format.

Two general methodologies are available for WordNet creation (Vossen, 1998):

- The *merge model*, that implies the creation of a new ontology for the target language.
- The *expand model*, where the variants associated with the Princeton WordNet synsets are translated using different strategies.

In our experiments we are using a set of tools based on the *expand* model, that is, we are translating English variants using several strategies.

* This research has been carried out thanks to the Project SKATeR (TIN2012-38584-C06-01 and TIN2012-38584-C06-04) supported by the Ministry of Economy and Competitiveness of the Spanish Government

¹<http://www.globalwordnet.org>

Several other wordnets have been developed using the expand model. The Spanish WordNet (Atserias et al., 1997) and the Catalan WordNet (Benítez et al., 1998) were constructed using this model and they have also been expanded using the WN-Toolkit, which is described below in section 3. The expand model has been also used in the MultiWordNet project for Italian (Pianta, Bentivogli, and Girardi, 2002), the Indonesian WordNet (Putra, Arfan, and Manurung, 2008), the Hungarian WordNet (Miháلتz et al., 2008), the Croatian WordNet (Raffaelli et al., 2014), the French WOLF WordNet (Sagot and Fišer, 2008) and more recently for the French WoNeF WordNet (Pradet, de Chalendar, and Desormeaux, 2014) and the KurdNet for Kurdish (Aliabadi, Ahmadi, and Salavati, 2014).

2 The Galnet project

The aim of the Galnet project (Gómez Guinovart et al., 2011) is building a WordNet for Galician aligned with the ILI (the inter-lingual index –namely, a list of meanings that allows a mapping of concepts of different languages) (Vossen, 1998) generated from the English WordNet 3.0 and with a lexical coverage similar to the English WordNet. The development of Galnet is integrated in the framework of the Multilingual Central Repository² (MCR) (González Agirre and Rigau, 2013). The MCR integrates in the same EuroWordNet framework wordnets from five languages (English, Spanish, Catalan, Basque and Galician) interlingually connected by the ILI and semantically categorized with several ontologies and taxonomies –IRST-Domains (Bentivogli et al., 2004), Suggested Upper Model Ontology (Pease, Niles, and Li, 2002), and Top Concept Ontology (Alvez et al., 2008). Thus, the MCR is a multilingual semantic resource of broad range suitable for use in language processing tasks that require large amounts of multilingual knowledge.

Galnet is distributed under a Creative Commons license CC BY 3.0³ as part of the MCR. The version of Galnet included in that distribution, reaches a lexical coverage of about one fifth of the English WordNet 3.0 (WN30) variants, as shown in detail in Table 1.

²<http://adimen.si.ehu.es/web/MCR/>

³<http://creativecommons.org/>

	WN30		Galnet	
	Vars	Syns	Vars	Syns
N	117798	82115	18949	14285
V	11529	13767	1416	612
Adj	21479	18156	6773	4415
Adv	4481	3621	0	0
Total	155287	117659	27138	19312

Table 1: Galnet current distribution

This early version of Galnet includes the Galician translation of the nominal and verbal synsets belonging to a set of basic concepts defined for WordNet, the Basic Level Concepts (BLC) (Izquierdo, Suárez, and Rigau, 2007), namely, 649 nominal synsets and 616 verbal synsets grouped in the `freqmin20/all`⁴ folder in the official distribution of the BLC for WordNet 3.0.⁵ This version of Galnet also includes the Galician entries for the WordNet lexicographer files (Fellbaum, 1998) corresponding to the names denoting body parts (`noun.body`) and substances (`noun.substance`), and the Galician equivalents for the adjectives of general type (`adj.all`).⁶ Finally, we extended the lexical coverage of this early version of Galnet using the WN-Toolkit (Oliver, 2012) to expand Galnet from two existing bilingual English–Galician resources, the Wikipedia and the English–Galician CLUVI Dictionary,⁷ reaching the final coverage shown in Table 1 (Gómez Guinovart et al., 2013). From this first distribution of Galnet we apply new expansions by means of lexical extraction from a Galician thesaurus (Gómez Guinovart and Simões, 2013) and using the WN-Toolkit, which is the aim of this paper.

Both the current distribution of Galnet and its development version can be explored through a specific web query interface for Galnet,⁸ or together with other lexical and textual resources for Galician through the RILG (Integrated Language Resources for Galician) platform.⁹

⁴The BLC in that set represent at least a number of synsets equal than 20, and have been obtained getting into account all relations of the synsets.

⁵<http://adimen.si.ehu.es/web/BLC/>

⁶<http://wordnet.princeton.edu/wordnet/man/lexnames.5WN.html>

⁷<http://sli.uvigo.es/diccionario/>

⁸<http://sli.uvigo.es/galnet/>

⁹<http://sli.uvigo.es/RILG/>

3 *The WN-Toolkit*

The WN-Toolkit (Oliver, 2014) is a set of programs written in Python for the creation of wordnets following the expand model. At the moment no user interface is provided so all programs must be run in a command line. The toolkit also provides some free language resources. This language resources are pre-processed so they can be easily used with the toolkit.

The toolkit is divided in the following parts:

- Dictionary-based strategies
- Babelnet-based strategies (Navigli and Ponzetto, 2012)
- Parallel corpus-based strategies
- Resources, such as freely available lexical resources, pre-processed corpora, etc.

The toolkit is distributed under the GNU-GPL license version 3.0 and can be freely downloaded from <http://lpg.uoc.edu/wn-toolkit>.

This toolkit has been developed under the SKATeR project and it has been previously successfully used for the expansion of Catalan and Spanish wordnets.

4 *Experimental settings and automatic evaluation*

4.1 *Experimental settings*

In the experiments with Galnet presented in this paper we have used the following strategies and resources:

- Dictionary-based strategies
 - Apertium English–Galician dictionary¹⁰
 - English–Galician Wiktionary¹¹
- Babelnet-based strategies
 - Using Babelnet 2.0¹²
- Parallel corpus-based strategies
 - Machine translation of sense-tagged corpora (Oliver and Climent, 2012)

- * English–Galician Semcor Corpus.¹³
The translation has been performed using Google Translate.¹⁴
- Automatic sense-tagging of parallel corpora (Oliver and Climent, 2014). Only the English part of the parallel corpora has been sense-tagged using Freeling with the UKB word sense disambiguator (Padró et al., 2010).
 - * Unesco CLUVI¹⁵ Corpus of Spanish–Galician scientific-technical texts
 - * Lega CLUVI Corpus of Galician–Spanish legal texts
 - * Consumer Eroski CLUVI Corpus of Spanish-Galician texts
 - * Tectra CLUVI Corpus of English–Galician literary texts

4.2 *Automatic evaluation*

In Table 2 we can observe the precision and number of new variants obtained with each method. The evaluation has been performed in an automatic way, comparing the obtained variants with the existing variants in the current distribution of Galnet. If the variant obtained for a given synset is one of the variants in the same synset of the existing Galnet, the result is evaluated as correct. If we do not have any Galician variant for a given synset in the reference Galnet, this result is not evaluated. The automatically obtained precision values tend to be lower than real values. The reason is that sometimes we have one or more for a given synset in the reference Galnet, but the obtained variant is not present. If the obtained variant turns out to be correct, it will be evaluated as incorrect anyway.

4.3 *Getting variants from several sources*

After this first analysis of the results we have evaluated the precision of the extracted variants taking into account the number the number of sources contributing the same entry. In Table 3 we can observe the precision of the results (from automatic evaluation) for the variants obtained in several experiments. We can also observe the precision for those variants obtained in a single experiment.

¹³http://www.gabormelli.com/RKB/SemCor_Corpus/

¹⁴Thanks to the University Research Program for Google Translate.

¹⁵<http://sli.uvigo.es/CLUVI/>

¹⁰<http://sourceforge.net/projects/apertium/>

¹¹<http://www.wiktionary.org>

¹²<http://babelnet.org>

	Precision	New variants
Apertium	78,83	1.230
Wiktionary	80,91	744
Babelnet	83,29	4.794
Semcor	78,13	2.053
Unesco	80,84	2.150
Lega	77,42	1.172
Eroski	80,28	1.777
Tectra	82,74	948

Table 2: Precision values and number of new variants obtained in the experiments

	Precision	New variants
7	100	1
6	97,22	5
5	96,26	41
4	97,71	178
3	83,68	646
2	81,44	1.159
1	77,02	9.650

Table 3: Precision values and number of new variants obtained in relation with the number of experiments leading to the same variant

As expected, with very few exceptions, the higher the number of experiment leading to a variant the higher the value of precision. It is also important to keep in mind that the value of precision here is calculated in an automatic way, comparing the results with the reference WordNet for Galician.

5 Errors analysis and revision

Human revision of the results of the experiment have been done not only for the new variants found for synsets without previous Galician variants in the reference Galnet (candidates automatically not evaluated), but also for the variants found for synsets which already have some Galician variant in the Galnet (candidates automatically evaluated as incorrect). The next two subsections show the details in both these cases.

5.1 New variants for empty synsets

For the human revision of the data we have a text file (noeva-1, noeva-2, noeva-3, etc.) for each set of candidate variants, grouped by the number of experiments which lead to them, as shown in Table 3. Each line contains a proposal of new candidate variant as a sequence of offset, Galician proposed variant, English

variants, and gloss, as in the following example:

02936714-n gaiola —cage, coop —an enclosure made or wire or metal bars in which birds or animals can be kept

Lexicographers have reviewed the files line by line from the noeva-1 to the noeva-7. When there were doubts about the correctness of the proposed Galician variant, the review method involved (1) commenting with hashes the line in the noeva file and (2) generating a report containing (a) the wrong line commented, (b) another commented line with the hypothetical cause of the error, and (c) the line with the error corrected. In this way we can also profit the suggestion to extend the Galnet, and also we justify and explain the cause of the error. The reports are saved in a separate file (noeva-0).

For instance, there is a wrong proposal in noeva-1 which states:

02522399-n brincadeira —cod, codfish —major food fish of Arctic and cold-temperate waters

But in Galician a *brincadeira* is a jibe or a joke, and never a cod, as is the meaning of this synset. So we comment this line in noeva-1 and create three new lines in the report file noeva-0, including as the third line the manually corrected proposal:

#02522399-n brincadeira —cod, codfish —major food fish of Arctic and cold-temperate waters

##Polysemy: “cod” in English can be translated in Galician in another sense by “brincadeira”. The wrong sense was chosen.

02522399-n bacallau —cod, codfish —major food fish of Arctic and cold-temperate waters

In this way, we document the error of the extraction, remark its cause (English lexical polisemy), and manually create a right proposal for file export to Galnet.

While typology of errors is very varied, the errors in extraction implying capitalization are very frequent. From 173 errors identified, 89 imply an error in the use of capital letters, for instance:

#09034967-n dar_es_salaam —Dar_es_Salaam, capital_of_Tanzania —the capital and largest port city of Tanzania on the Indian Ocean

##Letter case

09034967-n Dar_es_Salaam —Dar_es_Salaam,

capital_of_Tanzania —the capital and largest port city of Tanzania on the Indian Ocean

The errors related to the Spanish source of the current distribution of MCR are another frequent cause of errors in extraction, accounting for 40 error cases, for instance:

#07410207-n cinto —knock, bash, bang, smash, belt —a vigorous blow

##Spanish bad or dubious equivalent

07410207-n golpe —knock, bash, bang, smash, belt —a vigorous blow

In this proposal, the Galician variant *cinto* has been extracted by the WN Toolkit from its alignment in the processed Spanish-Galician resources with Spanish variant *cin-turón*, which is present in the synset by error (in fact, it is an alternative translation of English *belt*).

The third most frequent cause of errors in extraction (16 cases) is the bad selection of meaning from polysemous variants in English (as in the previous example of *cod*) or Spanish, for instance:

#03365592-n solo —floor, flooring —the inside lower horizontal surface (as of a room, hallway, tent, or other structure)

##Polysemy: “suelo” in Spanish can be translated in Galician in another sense by “solo”. The wrong sense was chosen.

03365592-n chan —floor, flooring —the inside lower horizontal surface (as of a room, hallway, tent, or other structure)

That being said, the results of human evaluation of the new variants extracted for synsets without previous Galician variants in the reference Galnet, in comparison with their automatic evaluation, are shown in Table 4. Due to time limitations, human review of *noeva-1* has been limited by now to the 200 first lines.

5.2 New variants for not empty synsets

The new variants extracted for synsets which already have some Galician variant in the Galnet –and automatically evaluated as *incorrect* candidates– are stored in a set of text files (*incorrect-1*, *incorrect-2*, *incorrect-3*...) grouped by the number of experiments which lead to them, as shown in Table 5.

Each line contains a sequence of offset, Galician proposed variant, Galician variants

	AP	CV	WC	RP
7	100	1	0	100
6	97,22	5	0	100
5	96,26	41	1	97,56
4	97,71	178	0	100
3	83,68	646	18	97,21
2	81,44	1.159	94	91,89
1	77,02	9.650	60	70

Table 4: Human evaluation of candidate new variants for empty synsets (AP = automatic precision, CV = candidate variants, WC = wrong candidates, RP = real precision)

	Variants
6	1
5	4
4	7
3	80
2	187
1	2.053

Table 5: Number of candidate variants for non-empty synsets obtained in relation with the number of experiments leading to the same variant

already in the synset, English variants, and gloss, as in the following example:

14541852-n risco —perigo —hazard, jeopardy, peril, risk, endangerment —a source of danger; a possibility of incurring loss or misfortune

Lexicographers have also reviewed the files from the *incorrect-1* to the *incorrect-6*. When a wrong proposed Galician variant is found, the review method involved (1) commenting with hashes the line in the *incorrect* file, (2) generating a report in a separate file (*incorrect-0*) containing (a) the wrong line commented, (b) another commented line with the hypothetical cause of the error, and (c) if possible, the line with the error corrected, and (3) generating a report in a separate file (*modify-0*) with the corrections needed in the existing variants (or examples) of the synset.

For instance, there is a wrong proposal in *incorrect-3* which states:

08586825-n sede —sé —see —the seat within a bishop’s diocese where his cathedral is located

But, differently from Spanish *sede* –a polysemous word which means see (as in ‘the Holy See’), venue or headquarters–, Galician

sede means only venue or headquarters, not see, as is the meaning of this synset. In fact, the Galician word for *see* is the yet existing Galician variant *sé*. So we comment this line in *incorrect-3* and create two new lines in the report file *incorrect-0*:

```
#08586825-n sede —sé —see —the seat within
a bishop’s diocese where his cathedral is located
##Polysemy: “sede” in Spanish can be translated
in Galician in another sense by “sede”.
The wrong sense was chosen.
```

While typology of errors is varied, the three causes most frequent are again the polysemy of the English or Spanish lexical source, the bad or dubious Spanish source, and the bad choice of letters case. Nevertheless, a characteristic feature of these errors in the new variants extracted for not empty synsets is that often they indicate an error in the existing variants of the distribution version of Galnet.

For instance, there is a correct proposal in *incorrect-3* which states:

```
05320899-n oído —orella —ear —the sense organ
for hearing and equilibrium
```

With the revision of this proposal, lexicographers can discover that there is a bad existing Galician variant *orella* for this synset in the reference Galnet (*orella* doesn’t mean ‘the sense organ for hearing and equilibrium’ but ‘the externally visible cartilaginous structure of the external ear’). In that case, the review protocol implies (1) not commenting the line in *incorrect-3*, which implies including the new Galician variant *oído* in the extended version of Galnet; and (2) copying the line, such as it is, in the file *modify-0*, which implies deleting the existing Galician variant *orella* from this synset in the extended version of Galnet.

All in all, the results of human evaluation of the new variants extracted for synsets with previous Galician variants in the reference Galnet, are shown in Table 6. Due to time limitations, human review of *incorrect-1* has been limited by now to the 100 first lines.

5.3 Galnet expansion results

After the revision of errors, we have incorporated the new variants and the required modifications of synsets to Galnet, obtaining the results shown in Table 7.

	CV	WC	RP	SM
6	1	0	100	0
5	4	0	100	0
4	7	0	100	3
3	80	14	85,10	20
2	187	22	88,23	22
1	2.053	47	53	11

Table 6: Human evaluation of candidate new variants for not empty synsets (CV = candidate variants, WC = wrong candidates, RP = real precision, SM= suggested modifications of synsets in reference Galnet)

	Syns	Vars	Unique vars
Reference GN	19.312	27.138	23.125
Extended GN	21.509	29.687	24.661
Δ	2.197	2.549	1.536

Table 7: Galnet expansion results

Both the reference Galnet (its current distribution in the MCR) and the extended Galnet (the work in process) can be explored through the interface at <http://sli.uvigo.es/galnet/>, where the results of the expansion of Galnet with the WN-Toolkit can be viewed selecting “wnt7” as experiment in the query of the development version.

6 Conclusions and future work

In this paper we have presented a practical application of multilingual resources exploitation for lexical acquisition. We have discussed the efficiency of a tool for lexical extraction as the WN-Toolkit in extending Galician WordNet coverage from bilingual resources of Galician in combination with English and Spanish, including lexical resources such as the dictionaries of Apertium, Wiktionary and Babelnet, and textual resources such as the CLUVI and SemCor corpora.

The precision of the extraction has reached high levels of efficiency in obtaining new variants coming from two or more bilingual resources: 91,89% for candidates to new variants for empty synsets and 88,23% for candidates to new variants for synsets not empty. Parallel corpora based strategies have the problem of a very low recall, mainly due to the use of a very simple alignment algorithm based on the most frequent translation. In future experiments, we will try to widen the coverage of the results im-

proving the WN-Toolkit with new alignment algorithms yielding greater coverage such as those used by Giza++ (Och and Ney, 2003).

References

- Aliabadi, Purya, Mohamed Sina Ahmadi, and Kyumars Sheykh Salavati, Shahin adn Esmaili. 2014. Towards building kurdnet, the kurdish wordnet. In *Proceedings of the 7th Global WordNet Conference*, Tartu, Estonia.
- Alvez, Javier, Jordi Atserias, Jordi Carrera, Salvador Climent, Antoni Oliver, and German Rigau. 2008. Consistent annotation of eurowordnet with the top concept ontology. In *Proceedings of the 4th Global WordNet Conference*, Szeged, Hungary.
- Atserias, Jordi, Salvador Climent, Xavier Farreres, German Rigau, and Horacio 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, pages 327–338.
- Bentivogli, Luisa, Pamela Forner, Bernardo Magnini, and Emanuele Pianta. 2004. Revising wordnet domains hierarchy: Semantics, coverage, and balancing. In *Proceedings of COLING Workshop on Multilingual Linguistic Resources*, pages 101–108, Ginebra.
- Benítez, Laura, 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*.
- Bond, Francis and Paik Kyonghee. 2012. A survey of wordnets and their licenses. In *Proceedings of the 6th International Global WordNet Conference*, pages 64–71, Matsue, Japan.
- Fellbaum, Christiane. 1998. *WordNet: An electronic lexical database*. The MIT press.
- Gómez Guinovart, Xavier, Xosé María Gómez Clemente, Andrea González Pereira, and Verónica Taboada Lorenzo. 2011. Galnet: WordNet 3.0 do galego. *Linguamática*, 3(1):61–67.
- Gómez Guinovart, Xavier, Xosé María Gómez Clemente, Andrea González Pereira, and Verónica Taboada Lorenzo. 2013. Sinonimia e rexistros na construción do WordNet do galego. *Estudos de lingüística galega*, 5:27–42.
- Gómez Guinovart, Xavier and Alberto Simões. 2013. Retreading dictionaries for the 21st century. In José Paulo Leal, Ricardo Rocha, and Alberto Simões, editors, *2nd Symposium on Languages, Applications and Technologies*, pages 115–126, Saarbrücken. Dagstuhl Publishing.
- González Agirre, Antoni and German Rigau. 2013. Construcción de una base de conocimiento léxico multilingüe de amplia cobertura: Multilingual central repository. *Linguamática*, 5(1):13–28.
- Izquierdo, Rubén, Armando Suárez, and German Rigau. 2007. Exploring the automatic selection of basic level concepts. In *Proceedings of the International Conference on Recent Advances on Natural Language Processing (RANLP'07)*, Borovetz, Bulgaria.
- Miháltz, M., 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.
- Navigli, Roberto and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.
- Och, Franz Josef and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.
- Oliver, A. and S. Climent. 2012. Building wordnets by machine translation of sense tagged corpora. In *Proceedings of the Global WordNet Conference, Matsue, Japan*.
- Oliver, Antoni. 2012. WN-Toolkit: un toolkit per a la creació de wordnets a partir de diccionaris bilingües. *Linguamática*, 4(2):93–101.
- Oliver, Antoni. 2014. Wn-toolkit: Automatic generation of wordnets following the

- expand model. In *Proceedings of the 7th Global WordNetConference*, Tartu, Estonia.
- Oliver, Antoni and Salvador Climent. 2014. Automatic creation of wordnets from parallel corpora. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).
- Padró, L., S. Reese, E. Agirre, and A. Soroa. 2010. Semantic services in freeling 2.1: Wordnet and UKB. In *Proceedings of the 5th International Conference of the Global WordNet Association (GWC-2010)*.
- Pease, Adam, Ian Niles, and John Li. 2002. The suggested upper merged ontology: A large ontology for the semantic web and its applications. In *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*, Edmonton.
- Pianta, E., L. Bentivogli, and C. Girardi. 2002. MultiWordNet. developing an aligned multilingual database. In *1st International WordNet Conference*, pages 293–302, Mysore, India.
- Pradet, Quentin, Gaël de Chalendar, and Jaume Baguenier Desormeaux. 2014. Wonef, an improved, expanded and evaluated automatic french translation of wordnet. In *Proceedings of the 7th Global WordNetConference*, Tartu, Estonia.
- Putra, D. D, A. Arfan, and R. Manurung. 2008. Building an Indonesian WordNet. In *Proceedings of the 2nd International MALINDO Workshop*.
- Raffaelli, Ida, Bekavac Božo, Željko Agić, and Marko Tadić. 2014. Building croatian wordnet. In *Proceedings of the 4th Global WordNet Conference*, Szeged, Hungary.
- Sagot, Benoît and Darja Fišer. 2008. Building a free French wordnet from multilingual resources. In *Proceedings of OntoLex*.
- Vossen, Piek. 1998. *EuroWordNet: a multilingual database with lexical semantic networks*. Springer.