

# WoNeF, an improved, expanded and evaluated automatic French translation of WordNet

Quentin Pradet, Gaël de Chalendar and Jeanne Baguenier Desormeaux

CEA, LIST, Laboratoire Vision et Ingénierie des Contenus

Gif-sur-Yvette, F-91191, France

quentin.pradet|gael.de-chalendar@cea.fr

## Abstract

Automatic translations of WordNet have been tried to many different target languages. JAWS is such a translation for French nouns using bilingual dictionaries and a syntactic language model. We improve its precision and coverage, complete it with translations of other parts of speech and enhance its evaluation method. The result is named WoNeF. We produce three final translations balanced between precision (up to 93%) and coverage (up to 109 447 (literal, synset) pairs).

## 1 Introduction

Reproducing the lexicographic work of WordNet (Fellbaum, 1998) for other languages is costly and difficult to maintain. Even with some theoretical problems, (Fellbaum and Vossen, 2007; de Melo and Weikum, 2008) show that translating Princeton WordNet literals while keeping its structure and its synsets leads to useful linguistic resources.

WordNet automatic translations use the *expand approach*: its structure is preserved and only literals are translated. Three main techniques represent this approach in the literature. The simplest one seeds WordNet using bilingual dictionaries (Rigau and Agirre, 1995), which can be filtered manually by lexicographers (Vossen, 1998; Tufiş et al., 2004). A second translation method uses parallel corpora, which avoids the use of dictionaries that may cause lexical bias. Back-translations between Norwegian and English were first explored (Dyvik, 2002), while (Sagot and Fišer, 2008) combine a multilingual lexicon and the different BalkaNet wordnets to help disambiguation. Finally, the bilingual dictionaries extracted from the Wiktionary and the Wikipedia interlanguage links allow to create new wordnets (de Melo and Weikum, 2009; Navigli and Ponzetto, 2010) or improve existing ones (Hanoka and Sagot, 2012).

Three French WordNets exist. The French EuroWordNet (Vossen, 1998) has a limited coverage and requires significant improvements to be used (Jacquin et al., 2007). It is also neither free nor freely accessible, which prevented the community from using and improving it. WOLF is a second French translation originally built using parallel corpora (Sagot and Fišer, 2008) and since then expanded using various techniques (Apidianaki and Sagot, 2012). WOLF is distributed under a free LGPL-compatible license. Finally, JAWS (Mouton and de Chalendar, 2010) is a translation of WordNet nouns developed using bilingual dictionaries and a syntactic language model.

Our work expands and improves the techniques used in JAWS and evaluates it based on the adjudication of two annotators work. The result is called WoNeF<sup>1</sup> and is distributed under the LGPL-LR licence. To our knowledge, all current WordNet machine translations only exist in one version where the authors decide what metric to optimize. We provide such a version, but add two resources that can serve different needs and have been obtained using different means. The main WoNeF has an F-score of 70.9%. Another version has a precision of 93.3%, and the last one contains 109 447 (literal, synset) pairs. The main contributions of this paper are the improvement and completion of JAWS with all parts of speech (section 3) and its evaluation (sections 4 and 5). The evaluation is done through an adjudication itself validated by measuring the inter-annotator agreement, which validates the expand approach to translate WordNet.

## 2 JAWS

### 2.1 Translation process

JAWS was built with a weakly supervised algorithm that does not require any manually anno-

<sup>1</sup>This work was partially funded by the ANR ASFALDA ANR-12-CORD-0023 project.

tated data, only the links between the French and English Wiktionaries and a target syntactic language model. The language model was trained on a large corpus extracted from the Web (Grefenstette, 2007). The corpus was analyzed by LIMA (Besançon et al., 2010), a rule-based parser producing fine-grained syntactic dependencies. For a given relation  $r$  and a word  $x$ , the language model indicates what are the first 100 words co-occurring most frequently with  $x$  through the relation  $r$ . Thanks to the dictionary, JAWS does not need to select each synset literals from the entire vocabulary but only among a small number of candidates (9 on average). The translation process is done in three steps. First, an empty wordnet is created, preserving WordNet structure, but with no literal associated to synsets. Then, the easiest translations among dictionaries candidates are selected to start filling JAWS. Finally, JAWS is extended incrementally using the language model, relations between synsets and the existing JAWS.

**Initial selectors** Four algorithms called initial selectors choose correct translations among those proposed by the dictionary. First, words appearing in only one synset are not ambiguous: all their translations are added to the French wordnet. This is the monosemy selector. For example, all translations of *grumpy* are selected in the only synset where it appears. Second, the uniqueness selector identifies words with only one translation and selects this translation in all synsets where the words appear. The five synsets containing *pill* in English are thus completed with *pilule*. These two first selectors were previously used in (Atserias et al., 1997) and (Benítez et al., 1998). A third selector translates words that are not in the dictionary using the English word itself: the direct translation selector. A fourth selector uses the Levenshtein edit distance: despite some false friends, if the distance between an English word and its translation is short, it can be considered that they have the same sense. Two examples are *portion* and *university*).

**JAWS expansion** JAWS being partially filled, a new expansion phase leverages the relationships between WordNet synsets to propose new translations. For example, if a synset S1 is a meronym of a synset S2 in WordNet and there is a context where a selected literal in S1 is a meronym of a candidate literal C in S2, then the literal C is con-

sidered correct. The translation task is thus reduced to the task of comparing on the one hand the lexical relations between WordNet synsets and on the other hand the lexical relations between French lexemes.

Let's take as an example the literal *quill* which can be translated to *piquant* or *plume* (Figure 1). In WordNet, *quill* is a meronym of *porcupine* which has already been translated by *porcupine* by an initial selector. In the language model, *piquant* is a noun modifier of *porcupine* but this is not the case of *plume*. Here, the noun-complement relation implies meronymy. It is thus *piquant* that must be chosen as the correct translation of *quill*. The language model allowed to choose between the two possible translations.

A potential problem with this approach could be that the noun modifier relationship is not limited to meronymy. For example, *mémoire* in the language model comes from a book entitled *Mémoires d'un porc-épic* ("Memoirs of a porcupine"). Fortunately, *mémoire* is not in the *quill* translation candidates and thus cannot be chosen. Paradoxically, the language model cannot choose between two very different words, but is able to choose the correct translation of a polysemous word. While automatically translating WordNet only with a dictionary or a syntactic language model is impossible, combining the two resources can solve the problem.

Each such syntactic selector follows the same principle as the meronymy selector and translates new synsets by identifying relationships between lexemes through the syntactic language model. The match between the noun modifier relation and the meronymy relation is direct, but this is not the case for all relations: there is for example no syntactic relationship that directly expresses the synonymy between two literals. For these relations, JAWS uses second order syntactic relations (Lenci and Benotto, 2012). See (Mouton and de Chalendar, 2010) for more details and other selectors.

## 2.2 JAWS limits

JAWS suffers from two main limitations. Above all, it only contains nouns, which prevents its use in many applications. Also, its evaluation procedure makes it difficult to judge its quality. Indeed, JAWS was evaluated by comparing it to the French EuroWordNet and WOLF 0.1.4 (released in 2008). These two French wordnets are not gold standards:

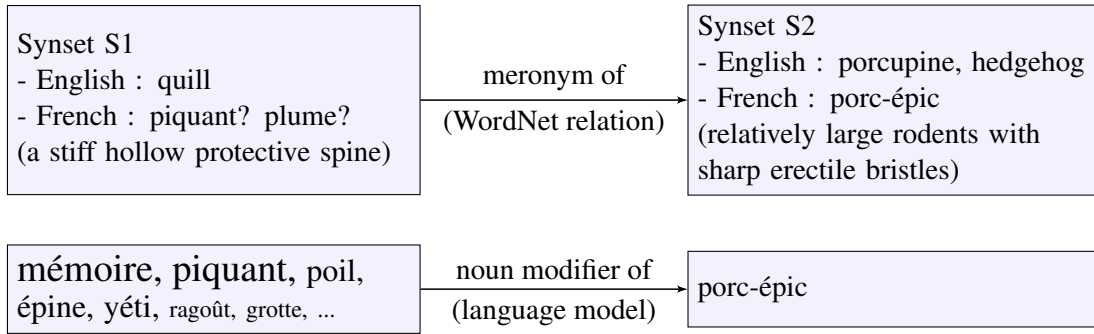


Figure 1: Translation through the part-of meronym relation.

they suffer from either limited coverage or limited accuracy. The authors decided to supplement this limited evaluation by a manual evaluation of literals that do not exist in WOLF, but it has been done on 120 (literal, synset) pairs only by a single annotator. The accuracy of JAWS is evaluated to 67.1%, which is lower than WOLF 0.1.4 and significantly lower than the accuracy of WOLF 1.0b. Furthermore this score should be taken with caution because of the size of the test sample: the confidence interval is approximately 25%.

### 3 WoNeF: JAWS improved and extended to other parts of speech

This section presents three key enhancements that have been made to JAWS and its extension to cover verbs, adjectives and adverbs. A change that is not detailed here is the one that led to a dramatically higher execution speed: JAWS built in several hours *versus* less than a minute for WoNeF, which helped to run many more experiments.

#### 3.1 Initial selectors

JAWS initial selectors are not optimal. While we keep the monosemy and uniqueness selectors, we changed the other ones. The direct translation selector is deleted as its precision was very low, even for nouns. A new selector considers candidate translations coming from several different English words in a given synset: the multiple sources selector, a variant the variant criterion of (Atserias et al., 1997). For example, in the synset *line, railway line, rail line*, the French literals *ligne de chemin de fer* and *voie* are translations of both *line* and *railway line* and are therefore chosen as translations.

Finally, the Levenshtein distance selector has been improved. 28% of English vocabulary is of French origin (Finkenstaedt and Wolff, 1973)

and anglicization produced predictable changes. It is possible to apply the same changes to the French candidate literal before computing the Levenshtein distance, bringing related words closer. We remove diacritics before applying several operations to word tails (Table 1). For example, reversing the "r" and "e" letter takes into account (*ordre/order*) and (*tigre/tiger*).<sup>2</sup> As before, false friends are not taken into account.

#### 3.2 Learning thresholds

In JAWS, each English literal can only correspond to the highest scoring French translation, regardless of the scores of lower-rated translations. This rejects valid candidates and accepts wrong ones. For example, JAWS does not include *particulier* in the *human being* synset because *personne* is already included with a higher score.

In WoNeF, we learned a threshold for each part of speech and selector. We first generated scores for all (literal, synset) candidate pairs, then sorted these pairs by score. The 12 399 pairs present in the WOLF 1.0b manual evaluation (our training set) were considered to be correct, while the pairs

<sup>2</sup>The Damerau-Levenshtein distance which takes into account transpositions anywhere in a word (Damerau, 1964) led to poorer results.

-que	-k	banque	→	bank
-aire	-ary	tertiaire	→	tertiary
eur	er	chercheur	→	researcher
ie	y	cajolerie	→	cajolery
-té	-ty	extrémité	→	extremity
-re	-er	tigre	→	tiger
ais	ese	libanais	→	lebanese
-ant	-ing	changeant	→	changing

Table 1: Changes to French word tails before applying the Levenshtein distance.

outside this set were not. We then calculated the thresholds maximizing precision and F-score.

Once these thresholds are defined, the selectors choose all candidates above the new threshold. This has two positive effects: valid candidates are not rejected when only the best candidate is already selected (improving both recall and coverage) and invalid candidates which were previously accepted are now rejected thanks to a stricter threshold (increasing precision).

### 3.3 Vote

After applying all selectors, our WordNet is large but contains some noisy synsets. In WoNeF, noise comes from several factors: selectors try to infer semantic information from a syntactic analysis without taking into account the full complexity of the syntax-semantics interface; the parser itself produces some noisy results; the syntactic language model is generated from a noisy corpus extracted from the Web (poorly written text, non-text content, non French sentences); and selected translations in one step are considered valid in the following steps while this is not always the case.

For the high-precision resource, we only keep literals for which the selectors were more confident. Since multiple selectors can now choose a given translation (section 3.2), our solution is simple and effective: translations proposed by multiple selectors are kept while the others are deleted. This voting principle is inspired from ensemble learning in machine learning. It is also similar to the combination method used in (Atserias et al., 1997) but we can avoid their manual inspection of samples of each method thanks to the development of our gold standard.

This cleaning operation retains only 18% of translations (from 87 757 (literal, synset) pairs to 15 625) but the accuracy increases from 68.4% to 93.3%. This high precision resource can be used as training data for other French WordNets. A typical voting methods problem is to choose only easier and poorly interesting examples, but the resource obtained here is well balanced between synsets containing only monosemic words and other synsets containing polysemous and more difficult to disambiguate words (section 5.2).

### 3.4 Extension to verbs, adjectives and adverbs

The work on JAWS began with nouns because they represent 70% of the synsets in WordNet. We

continued this work on all other parts of speech: verbs, adjectives and adverbs. Here, generic selectors have been modified, but in the future, we will develop selectors taking into account the different parts of speech characteristics in WordNet.

**Verbs** Selectors chosen for verbs are the uniqueness and monosemy selectors. Indeed, the Levenshtein distance gave poor results for verbs: only 25% of the verbs chosen by this selector were correct translations. For syntactic selectors, only the selector by synonymy gave good results, while the selector by hyponymy had the performance of a random classifier.

**Adjectives** For adjectives, all initial selectors are chosen, and the selected syntactic selector is the selector by synonymy.

**Adverbs** The configuration is the same than for adjectives. We have no gold standard for adverbs, which explains why they are not included in our evaluation. However, comparison with WOLF (section 5.4) shows that adverbs are better than other parts of speech.

## 4 WoNeF: an evaluated JAWS

### 4.1 Gold standard development

Evaluation of JAWS suffers from a number of limitations (section 2.2). We produced a gold standard for rigorous evaluation to evaluate WoNeF. For nouns, verbs and adjectives, 300 synsets have been annotated by two authors of this paper, both computational linguists, both native French speakers and respectively with a background in computer science and linguistics. For each candidate provided by our translation dictionaries, they had to decide whether or not it belonged to the synset. They used WordNet synsets to examine their neighbors, the Merriam-Webster dictionary, the French electronic dictionary TLFi and search engines to demonstrate the use of different senses of the words in question. Because dictionaries do not provide candidates for all synsets and some synsets have no suitable candidate, the actual number of non-empty synsets is less than 300 (section 4.2).

During manual annotation, we encountered difficulties arising from the attempt to translate the Princeton WordNet to French. Most problems come from verbs and adjectives appearing in a collocation. In WordNet, they can be grouped in a

way that makes sense in English, but that is not reflected directly in another language. For example, the adjective *pointed* is the only element of a synset defined as *Direct and obvious in meaning or reference; often unpleasant, “a pointed critique”, “a pointed allusion to what was going on”, “another pointed look in their direction”*. These three examples would result in three different translations in French: *une critique dure*, *une allusion claire* and *un regard appuyé*. There is no satisfactory solution in translating such a synset: the resulting synset contains either too many or too few translations. We view this issue as a mainly linguistic one in the way WordNet has grouped those three usages of *pointed*. We marked the concerned synsets and will handle them in a future work, either manually or with other approaches. These granularity problems concern 3% of nominal synsets, 8% of verbal synsets and 6% of adjectival synsets.

The other main difficulty stems from translations in our bilingual dictionaries. Rare meanings of a word are sometimes missing. For example, there is a WordNet synset containing the *egg* verb for its *coat with beaten egg* sense. Our dictionaries only consider *egg* as a noun: neither our gold standard nor JAWS can translate this synset. This case appeared rarely in practice, and none of these senses are in the most polysemous synsets (BCS synsets as defined in section 5.2), confirming that it doesn't affect the quality of our gold standard for the most important synsets. Yet WoNeF could be improved by using specific dictionaries for species (as in (Sagot and Fišer, 2008) with WikiSpecies), medical terms, etc. Unwanted translations are another issue. Our dictionaries translate *unkindly* to *sans aménité* (*without amenity*) which is a compositional phrase. While such a translation is expected in a bilingual dictionary, it should not be integrated in a lexical resource. The last difficulty lied in judgment adjectives: for example, there is no good translation of *weird* in French. Although most dictionaries provide *bizarre* as a translation, this one does not provide the *stupid* aspect of *weird*. There is no translation that would fit in all contexts: the synset meaning is not fully preserved after translation.

## 4.2 Inter-annotators agreement

Table 2 shows the inter-annotator agreement measured through Fleiss kappa for the three annotated

	Nouns	Verbs	Adj.
Fleiss Kappa	0.715	0.711	0.663
Synsets	270	222	267
Candidates	6.22	14.50	7.27

Table 2: Gold standard inter-annotator agreement

parts of speech. Even if it is a discussed metric (Powers, 2012), all existing evaluation tables consider these scores as high enough to describe the inter-annotator agreement as "good" (Gwet, 2001), which allows us to say that our gold standard is good. The expand approach for the translation of WordNets is also validated : it is possible to produce useful resource in spite of the difficulties mentioned in section 4.1.

## 5 Results

We present in this section the results of WoNeF. We first describe the initial selectors and proceed with the full resource. Our gold standard is divided into two parts: 10% of the literals form the validation set used to choose the selectors that apply to different versions of WoNeF, while the remaining 90% form the evaluation set. No training was performed on our gold standard. Precision and recall are based on the intersection of synsets present in WoNeF and our gold standard. Precision is the fraction of correct (literal, synset) pairs in the intersection while recall is the fraction of correctly retrieved pairs.

### 5.1 Initial selectors

For nouns, verbs and adjectives, we calculated the efficiency of each initial selector on our development set, and used this data to determine which ones should be included in the high precision version, the high F-score version and the large coverage one. Scores are reported on the test set.

Table 3 shows the results of this operation. Coverage gives an idea of the size of the resource. Depending on the objectives of each resource, the selected initial selectors are different. Since different selectors can choose the same translation, the sum of coverages is greater than the coverage of the high coverage resource.

### 5.2 Global results

We now focus on the overall results which include the application of initial selectors and syntactic selectors (Table 4). The high-precision method also

	P	R	F1	C
monosemy	71.5	76.6	74.0	54 499
unicity	91.7	63.0	75.3	9 533
mult. sources	64.5	45.0	53.0	27 316
Levenshtein	61.9	29.0	39.3	20 034
high precision	<b>93.8</b>	50.1	65.3	13 867
high F-score	71.1	<b>72.7</b>	<b>71.9</b>	82 730
high coverage	69.0	69.8	69.4	<b>90 248</b>

Table 3: Top part: Precision, Recall and F1-measure of initial selectors on all translations (nouns, verbs and adjectives). Bottom part: scores for various combinations of them. Coverage C is the total number of pairs (literal, synset).

applies a vote (section 3.3). As in the previous table, the coverage C is the number of (literal, synset) pairs. Without using structure-based nor conceptual distance-based selectors as in (Farreres et al., 2010), we obtain a coverage at 93% precision for our French wordnet (15 625) equal to their Spanish one (11 770) and larger than their Thai one (2 013).

<b>All synsets</b>	P	R	F1	C
high precision	<b>93.3</b>	51.5	66.4	15 625
high F-score	68.9	73.0	<b>70.9</b>	88 736
high coverage	60.5	<b>74.3</b>	66.7	<b>109 447</b>
<b>BCS synsets</b>	P	R	F1	C
high precision	<b>90.4</b>	36.5	52.0	1 877
high F-score	56.5	62.8	<b>59.1</b>	14 405
high coverage	44.5	<b>66.9</b>	53.5	<b>23 166</b>

Table 4: Global results for all synsets and BCS synsets only.

In WordNet, most words are monosemous, but a small minority of polysemous words are the most represented in texts. It is precisely on this minority that we wish to create a quality resource. To evaluate this, we use the list of **BCS** (Basic Concept Set) synsets provided by the BalkaNet project (Tufig et al., 2004). This list contains 8 516 synsets lexicalized in six different translations of WordNet. They should represent the most frequent synsets and those with the most polysemous words. While the high F-score and the high coverage resources lose precision for BCS synsets, this is not the case for the high precision resource. In fact, the voting mechanism makes the high-precision resource very robust, even for the BCS synsets.

### 5.3 Results by part of speech

Table 5 shows the detailed results for each part of speech. Concerning nouns, the high precision mode uses two selectors, both based on the noun modifier syntactic relation: the meronymy selector described in section 2.1 and the hyponymy selector. The high precision resource for nouns is our best resource. The high F-score version has an F-score of 72.4%, which ensures that present (literal, synset) pairs have good quality and that it does not miss too many translations. The nominal version is better than JAWS by 2.8% points of F-score.

		P	R	F1	C
PR	nouns	<b>96.8</b>	56.6	71.4	11 294
	verbs	<b>68.4</b>	41.9	52.0	1 110
	adj.	<b>90.0</b>	36.7	52.2	3 221
F1R	nouns	71.7	73.2	<b>72.4</b>	59 213
	<b>JAWS</b>	70.7	68.5	69.6	55 416
	verbs	48.9	<b>76.6</b>	<b>59.6</b>	9 138
	adj.	69.8	71.0	70.4	20 385
CR	nouns	61.8	<b>78.4</b>	69.1	<b>70 218</b>
	verbs	45.4	61.5	52.2	<b>18 844</b>
	adj.	69.8	<b>71.9</b>	<b>70.8</b>	<b>20 385</b>

Table 5: Results by part of speech. Horizontal parts give scores for the high-precision resource (PR), the high-F1-measure one (F1R) and the high coverage one (CR). JAWS containing only nouns, it is compared with the high F-score nominal WoNeF resource.

Results for verbs are lower. The main reason is that verbs are on average more polysemous in WordNet and our dictionaries than other parts of speech: verbal synsets have twice as many candidates as nouns and adjectives synsets (Table 2). This shows the importance of the dictionary to limit the number of literals from which algorithms must choose. The synonymy selector is the only syntactic selector applied to verbs: it uses second-order syntactic relations for three types of verbal syntactic dependencies: if two verbs share the same objects, they are likely to be synonyms or near-synonyms. This is the case for *dévoré* and *manger* which both accept the object *pain*. Other syntactic selectors have not been used for verbs because of their poor results. Indeed, while the detection of hyponymy using only the inclusion of contexts was effective on the nouns, it has the performance of a random classifier for verbs. This

highlights the complexity of verbal polysemy.

For adjectives and verbs, only the synonymy selector was applied. For high F-score and high coverage resources, the same selectors (initial and syntactic) are applied, which is why the results are the same. While the inter-annotator agreement was lower on adjectives than on verbs, results are much better for adjectives. This is mainly due to the number of candidates from which to select: there are twice as less candidates for adjectives. This highlights the importance of dictionaries.

#### 5.4 Evaluation against WOLF

Using our gold standard to compare WOLF and WoNeF would unfairly penalize WOLF for all correct words not present in our dictionaries. Conversely, we cannot consider WOLF as a direct reference as WOLF itself is not fully validated. The last publication giving overall WOLF figures (Sagot and Fišer, 2012) indicates a number of pairs around 77 000 with 86% precision<sup>3</sup>. We thus compare the intersections between the high-precision WoNeF (93.3% precision) and WOLF 0.1.4 and 1.0b (Table 6). It shows that although WoNeF is still smaller than WOLF, it is a complementary resource. The comparison of the differences between WOLF 0.1.4 and WOLF 1.0b is instructive as it highlights WOLF improvements.

<b>WOLF 0.1.4</b>	$\subset$	$\supset$	$\oplus$
Nouns	18.7	3.0	10 526
Verbs	6.5	0.8	1 743
Adjectives	26.9	5.8	3 710
Adverbs	23.8	5.6	757
<b>WOLF 1.0b</b>	$\subset$	$\supset$	$\oplus$
Nouns	49.7	8.6	6 503
Verbs	26.5	2.6	1 338
Adjectives	36.4	13.3	2 530
Adverbs	41.2	12.6	543

Table 6: Intersections between the high precision WoNeF and WOLF 0.1.4 and 1.0b.  $\subset$  is the percentage of WoNeF pairs included in WOLF and  $\supset$  is the percentage of WOLF pairs included in WoNeF.  $\oplus$  is the number of new elements contributed by WoNeF.

The  $\oplus$  column gives the number of translations that are present in WoNeF but not in WOLF.

<sup>3</sup>The detailed results for WOLF 1.0b are not currently available.

For nouns, verbs and adjectives, it means that we contribute 10 914 new high precision (literal, synset) pairs by merging WoNeF and WOLF 1.0, in other words 94% of the high precision WoNeF pairs which shows how much the two approaches are complementary: different literals are selected. This produces a French wordnet 10% larger than WOLF with an improved accuracy. A merging with the high F-score resource would be slightly less precise, but it would provide 81 052 new (literal, synset) pairs comparing to WOLF 1.0b, resulting in a merge containing 73 712 non-empty synsets and 188,657 (literal, synset) pairs, increasing WOLF coverage by 75% and the WoNeF one by 63%.

#### Conclusion

In this work, we have shown that the use of a syntactic language model to identify lexical relations between lexemes is possible in a constrained environment and leads to results with a state of the art precision for the task of translating WordNet. We offer three different resources, each with a different purpose. Finally, we provide a validated high quality gold standard that has enabled us to demonstrate both the validity of the approach of translating WordNet by extension and the validity of our specific approach. This gold standard can also be used to evaluate and develop other French WordNet translations. WoNeF is freely available on <http://wonef.fr/> under the LGPL-LR licence. A web interface based on sloWTool (Fi[Pleaseinsertintopreamble]er and Novak, 2011) (initially developed for sloWNet, the Slovenian WordNet) allows to browse the resulting WordNet online. The current distribution formats are the DEBVisDic XML and WordNet-LMF formats. This allows to integrate WoNeF into the Global WordNet Grid and facilitates access and conversions into any lexical resource format.

Future work on WoNeF will focus on verbs, adjectives and adverbs, for which dedicated new selectors may be considered to improve coverage. For example, the synonymy selector can be extended to the WordNet adjectival quasi-synonymy relationship because distributional semantic techniques tend to identify quasi-synonyms rather than synonyms.

Another important source of improvement will be to enrich our syntactic language model by taking into account reflexive verbs and multi-word

expressions. We would also like to move towards a continuous language model (Le et al., 2012). This will be coupled with the collection of a more recent and larger Web corpus analyzed with a recent version of our linguistic analyzer. This will allow us to measure the impact of the language model quality on the WordNet translation.

The WOLF French wordnet was built using several techniques. Merging WoNeF and WOLF will soon improve again the status of the French translation of WordNet: we are working with WOLF authors to merge WOLF and WoNeF.

## References

- Marianna Apidianaki and Benoît Sagot. 2012. Applying cross-lingual WSD to wordnet development. In *LREC'12*, May.
- Jordi Atserias, Salvador Climent, Xavier Farreres, German Rigau, and Horacio Rodr Guez. 1997. Combining Multiple Methods for the Automatic Construction of Multilingual WordNets. In *RANLP'97*, September.
- Laura Benítez, Sergi Cervell, Gerard Escudero, Mònica López, German Rigau, and Mariona Taulé. 1998. Methods and Tools for Building the Catalan WordNet. In *ELRA Workshop on Language Resources for European Minority Languages*, May.
- Romarc Besançon, Gaël de Chalendar, Olivier Ferret, Faiza Gara, and Nasredine Semmar. 2010. LIMA: A Multilingual Framework for Linguistic Analysis and Linguistic Resources Development and Evaluation. In *LREC 2010*, May.
- Fred J. Damerau. 1964. A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7(3):171–176, March.
- Gerard de Melo and Gerhard Weikum. 2008. On the Utility of Automatically Generated Wordnets. In *GWC 2008*, January.
- Gerard de Melo and Gerhard Weikum. 2009. Towards a universal wordnet by learning from combined evidence. In *CIKM 2009*, November.
- Helge Dyvik. 2002. Translations as Semantic Mirrors: From Parallel Corpus to WordNet. In *ICAME 23*, May.
- Javier Farreres, Karina Gibert, Horacio Rodríguez, and Charnyote Pluempitiwiriyawej. 2010. Inference of lexical ontologies. The LeOnI methodology. *Artificial Intelligence*, 174(1):1–19, January.
- Christiane Fellbaum and Piek Vossen. 2007. Connecting the Universal to the Specific: Towards the Global Grid. In *IWIC 2007*, January.
- Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA, May.
- Thomas Finkenstaedt and Dieter Wolff. 1973. *Ordered profusion: Studies in dictionaries and the English lexicon*, volume 13 of *Annales Universitatis Saraviensis*. C. Winter.
- Darja Fišer and Jernej Novak. 2011. Visualizing slownet. In *eLex 2011*, November.
- Gregory Grefenstette. 2007. Conquering language: Using NLP on a massive scale to build high dimensional language models from the web. In *CICLing 2007*, February.
- Kilem L. Gwet. 2001. *Handbook of inter-rater reliability*. Advanced Analytics, LLC, September.
- Valérie Hanoka and Benoît Sagot. 2012. Wordnet extension made simple: A multilingual lexicon-based approach using wiki resources. In *LREC'12*, may.
- Christine Jacquin, Emmanuel Desmontils, and Laura Monceaux. 2007. French eurowordnet lexical database improvements. In *CICLing 2007*, February.
- Hai-Son Le, Alexandre Allauzen, and François Yvon. 2012. Continuous Space Translation Models with Neural Networks. In *NAACL-HLT 2012*, June.
- Alessandro Lenci and Giulia Benotto. 2012. Identifying hypernyms in distributional semantic spaces. In *\*SEM 2012*, June.
- Claire Mouton and Gaël de Chalendar. 2010. JAWS: Just Another WordNet Subset. In *TALN 2010*, June.
- Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a very large multilingual semantic network. In *ACL 2010*, July.
- David M W Powers. 2012. The Problem with Kappa. In *EACL 2012*, April.
- German Rigau and Eneko Agirre. 1995. Disambiguating bilingual nominal entries against WordNet. In *Workshop "The Computational Lexicon"*. *ESSLLI'95*, August.
- Benoît Sagot and Darja Fišer. 2008. Building a free French wordnet from multilingual resources. In *Ontolex 2008 Workshop*, May.
- Benoît Sagot and Darja Fišer. 2012. Automatic Extension of WOLF. In *GWC 2012*, January.
- Dan Tufiş, Dan Cristea, and Sofia Stamou. 2004. BalkaNet: Aims, methods, results and perspectives. a general overview. *Romanian Journal of Information Science and Technology*, 7(1-2):9–43.
- Piek Vossen. 1998. *EuroWordNet: a multilingual database with lexical semantic networks*. Kluwer Academic, October.