Data-driven Synset Induction and Disambiguation for Wordnet Development

Marianna Apidianaki, Benoît Sagot

Noname manuscript No.
(will be inserted by the editor)

Data-driven Synset Induction and Disambiguation for Wordnet Development

Marianna Apidianaki · **Benoît Sagot**

Received: date / Accepted: date

Abstract Automatic methods for wordnet development in languages other than English generally exploit information found in Princeton WordNet (PWN) and translations extracted from parallel corpora. A common approach consists in preserving the structure of PWN and transferring its content in new languages using alignments, possibly combined with information extracted from multilingual semantic resources. Even if the role of PWN remains central in this process, these automatic methods offer an alternative to the manual elaboration of new wordnets. However, their limited coverage has a strong impact on that of the resulting resources. Following this line of research, we apply a cross-lingual word sense disambiguation method to wordnet development. Our approach exploits the output of a data-driven sense induction method that generates sense clusters in new languages, similar to wordnet synsets, by identifying word senses and relations in parallel corpora. We apply our cross-lingual word sense disambiguation method to the task of enriching a French wordnet resource, the WOLF, and show how it can be efficiently used for increasing its coverage. Although our experiments involve the English-French language pair, the proposed methodology is general enough to be applied to the development of wordnet resources in other languages for which parallel corpora are available. Finally, we show how the disambiguation output can serve to reduce the granularity of new wordnets and the degree of polysemy present in PWN.

Keywords cross-lingual word sense disambiguation, word sense induction, sense clustering, parallel corpora, WordNet

Marianna Apidianaki
LIMSI-CNRS
Rue John von Neumann
91403 Orsay, France
E-mail: marianna.apidianaki@limsi.fr

Benoît Sagot
Alpage, INRIA Paris-Rocquencourt & Université Paris-Diderot
Bâtiment Olympe de Gouges, Rue Albert Einstein
75013 Paris, France
E-mail: benoit.sagot@inria.fr

1 Introduction

The growing need for lexical and semantic knowledge in Natural Language Processing (NLP) applications has steered several initiatives for resource development in recent years. A common trend has been to develop multilingual resources based on Princeton WordNet (PWN) [15]: the structure of PWN is generally preserved and its contents are translated in new languages [40,31,39]. The advantages of this approach, which explain its wide adoption, are that it avoids the time-consuming manual elaboration of the semantic hierarchy in new languages and allows the alignment of the resulting wordnets, a feature particularly useful for multilingual NLP. Its main limitation is the strong bias imposed by PWN on the content and structure of the newly built wordnets. The structure of PWN is preserved in the target language and its content is transferred based on an assumption of language independence of concepts and semantic relations. However, concepts present in PWN might not be present in the target language, in which case the corresponding synsets in the new wordnet cannot be filled. This issue becomes more important in case of fine-grained sense distinctions in PWN, where finding target language counterparts becomes difficult or impossible. Combined to limitations due to the quantity of information available in the bilingual dictionaries used for manually translating PWN synsets into new languages, these factors have a strong impact on the coverage of the resulting wordnets which is generally much smaller than that of PWN.

In an attempt to address these weaknesses, several automatic wordnet development methods have been proposed that exploit information found in parallel corpora. These methods permit to acquire semantic information from texts and to circumvent, in this way, the need for pre-defined resources. Moreover, by exploiting alignment information, these methods offer an alternative to the manual filling of wordnets: translations are extracted from parallel corpora and are automatically integrated in the wordnet hierarchy. Nevertheless, the success and coverage of these methods highly depends on the nature of the parallel corpora and on the way the extracted information is used for wordnet filling. Coverage becomes an issue especially when automatically acquired translations are used to fill PWN-based resources, as rare senses (present in PWN) might not be found in the parallel corpora. Additionally, automatic methods often fail to retrieve semantic information as fine-grained as the one found in PWN, leaving numerous synsets empty. As a consequence, the resulting wordnet resources are rather sparse and methods for increasing their coverage are needed.

Following this line of research, we propose a novel automatic approach to wordnet development. We demonstrate how a cross-lingual word sense disambiguation (WSD) method can be applied to the wordnet development task for creating new resources or for enriching existing ones. WSD is the task of automatically identifying the meaning of words in context [25] while its cross-lingual variant predicts semantically correct translations [2,21]. In this work, we apply a cross-lingual WSD method [2] to the enrichment of an automatically built wordnet for French, the WOLF [34]. The disambiguation method exploits the output of a cross-lingual word sense induction (WSI) method which identifies word senses and their relations in parallel corpora. The WSI method generates clusters of semantically

related words in the new language (French) similar to wordnet synsets, which are integrated in the WOLF hierarchy by the cross-lingual WSD method.

The paper is organized as follows. Section 2 presents related work on wordnet development in languages other than English. We first describe the classical approach to cross-lingual transfer of WordNet information based on pre-defined lexical and semantic resources. Then we move to a number of automatic methods which combine existing resources with information extracted from corpora and describe the method that initially served to build WOLF, the French resource that we aim to enrich. Finally, we present a number of purely data-driven semantic analysis methods and explain our choice to apply a WSD method to the wordnet development task. Section 3 presents our data-driven synset induction method as well as the disambiguation method that serves to integrate the newly acquired synsets in the French resource. In Section 4, we present the results of a manual evaluation intended to estimate the quality of the clustering and the correctness of the new WOLF entries. The main findings of this study are summed up in the last section where we also present some avenues worth pursuing in future work, before concluding.

2 Cross-lingual approaches to wordnet development

2.1 Transfer of WordNet information to new languages

Multilingual wordnet development has always strongly relied on Princeton WordNet (PWN) [15]. Large-scale projects aiming the creation of wordnets in languages other than English, such as EuroWordNet, BalkaNet and MultiWordNet [40,31,39], have adopted a translation-driven approach: the structure of PWN was preserved while its contents were imported in the newly built resources by applying various translation-based methods. The main advantage of this approach, also called the *expand model*, is that it permits to avoid the time-consuming and expensive manual elaboration of the semantic hierarchy in new languages. An additional advantage is that the newly built wordnet is automatically aligned to PWN and to other wordnets built following the same principle. The resulting resources are thus interesting for contrastive semantic analysis and can be particularly useful in multilingual NLP tasks, such as Multilingual Information Retrieval.

Despite its strengths, the translation approach to wordnet development also presents a number of drawbacks. One of them is the strong bias imposed by PWN on the content and structure of the new wordnets. The structure of PWN is generally preserved and its content is transferred in the new languages based on the assumption that concepts and semantic relations between them are – at least to a large extent – language independent. This assumption is not theoretically valid and has important practical implications during the compilation of new wordnets. Several senses present in PWN have no target language counterpart, a problem that becomes more apparent in the case of fine-grained WordNet senses. As a consequence, a varying number of target language synsets may be left unfilled, depending on the language, and this sparseness limits the usefulness of the newly built resource in NLP applications.

Other issues posed by the translation approach are its heavy reliance on external lexico-semantic resources and the manual work needed for transfer. In EuroWordNet, BalkaNet and MultiWordNet, PWN literals were mainly translated by human lexicographers using external resources such as dictionaries, thesaurus and taxonomies.[1] Apart from limiting the approach to specific language pairs, the reliance on pre-defined resources introduces a new bias as their coverage has a strong impact on the one of the newly built wordnets.

2.2 Automatic cross-lingual information transfer

2.2.1 Combining lexicographic resources and parallel corpora

In spite of the theoretical and practical drawbacks inherent to the translation approach, new wordnets are still heavily based on Princeton WordNet. The methods used for transferring information into new languages have however evolved towards becoming more or less automatic, limiting the cost of the manual methods employed before. Moreover, these methods often exploit lexico-semantic information extracted from monolingual or multilingual corpora, instead of solely relying on pre-defined semantic resources. For instance, the French hierarchy WOLF [34], that we intend to enrich in this work, was automatically built by combining information from several multilingual resources (the EUROVOC thesaurus and Wikipedia-related resources) with information extracted from a multilingual parallel corpus. Another PWN-based resource for French, the JAWS network, was compiled by combining a bilingual dictionary and syntactic information acquired from corpora for disambiguating polysemous nouns and correctly integrating them in the hierarchy [24].

The multilingual semantic network BabelNet goes a step further by jointly exploiting PWN, Wikipedia and the output of statistical Machine Translation systems [26]. In BabelNet, PWN and Wikipedia are combined by automatically mapping WordNet senses and Wikipages and complementing their respective concept inventories. Multilingual lexicalizations of the concepts are acquired from the human-generated translations provided in Wikipedia and by using a statistical Machine Translation system to translate occurrences of the concepts within sense-tagged corpora. The resulting multilingual network has a wide coverage, as it contains 9 million entries (concepts and named entities) in 50 languages.[2]

In a different setting, aiming the semantic annotation of new languages, Diab and Resnik [9] combine translation information from a parallel corpus with semantic information in PWN. The possible semantic tags provided in PWN for the English translations of a foreign word are found, and the one characterizing the whole set of translations is selected and used as the foreign word's sense tag.

All these approaches successfuly combine information in PWN and other lexical and semantic resources with information learned from corpora. In the next section,

[1] In these projects, the *expand* model was occasionally combined with the *merge* model which is based on monolingual resources and permits to include language-specific properties in the wordnets of different languages.

[2] The BabelNet resource is available here: `http://babelnet.org`

we provide more information on the WOLF resource that we intend to enrich, the way it was compiled, its content and its coverage.

2.2.2 The WOLF resource

WOLF [34] is a freely available wordnet for French. Its first version (WOLF 0.1.4) was created on the basis of PWN (version 2.0) by following the *expand* model for wordnet development and relying on both existing resources and parallel corpora. In this section, we briefly sketch the approach that was used for building WOLF version 0.1.4. A more detailed discussion on the construction of the resource and its evaluation can be found in earlier publications [34].

To fill the WOLF, monosemous literals in the PWN were automatically translated using a bilingual French-English lexicon built from various multilingual resources. In particular, data was extracted from Wikipedia using inter-wiki links, from the English and French Wiktionary, the Wikispecies encyclopedia of living beings and the EUROVOC thesaurus (`http://europa.eu/eurovoc`).

Polysemous PWN literals were handled by an *alignment* approach (cf. Section 2.3) based on the multilingual parallel corpus SEE-ERA.NET [38], which is composed of the English, French, Romanian, Czech and Bulgarian parts of the JRC-Acquis corpus. The corpus was lemmatized, part of speech (POS) tagged and word aligned, and bilingual lexicons were automatically built including the translations of English words in different languages. These lexicons were then combined into various multilingual lexicons (3-lingual to 5-lingual) and a synset id was assigned to each lexicon entry by gathering all possible ids for this entry in all languages from the corresponding BalkaNet wordnets, which share the same inventory of synset ids. The underlying assumption being that it is unlikely that the same polysemy occurs in different languages, the intersection of the possible senses was expected to output only the correct synset. In this way, the ids shared by all non-French lexicon entries were assigned to their French translation. For example, among the various French-Czech-Bulgarian-English alignments involving the French word *droit* and the English word *law*, the alignment *droit-právo-право-law* was found 56 times. The only synset that contained *právo* in the Czech wordnet, *право* in the Bulgarian wordnet and *law* in the PWN is the synset ENG20-05791721-n. Therefore, *droit* could be added in the WOLF in this synset.

The synsets obtained for monosemous and polysemous literals by these two approaches were merged. The resulting network preserves the hierarchy and structure of PWN 2.0 and contains the definitions and usage examples provided in PWN for each synset. As information was not found for all PWN synsets by the employed automatic methods, the version 0.1.6 of WOLF which is used in our experiments is rather sparse.[3] In total, it contains 32,351 non-empty synsets including 37,991 unique literals (vs. 115,424 synsets with 145,627 literals in PWN 2.0). These synsets are filled with 34,827 unique French noun literals, 1,521 adjectives, 979 verbs and 664 adverbs.

The work presented in this paper is aimed at enriching this resource and increasing its coverage. However, in spite of the focus on this particular resource, the

[3] Compared to the initial version of WOLF (0.1.4), version 0.1.6 has an extended coverage on adverbs as a result of the work by Sagot et al. [35].

proposed methodology is general enough to be applied to the development of new wordnet resources in other languages. Before presenting our method in more detail, we will refer to a number of purely data-driven semantic analysis works developed in a multilingual setting and will explain the reasons for choosing a cross-lingual WSD method for this task.

2.3 Data-driven approaches to wordnet development

2.3.1 Corpus-based semantic analysis

The methods presented in this section do not use PWN but rely only on information coming from parallel corpora for building semantic resources. The basic assumption underlying these methods is that the translations of words in real texts offer insights into their semantics [33]. This strong assumption, which has been widely exploited in works on cross-lingual semantic analysis [27,1,4] and the integration of semantics in Machine Translation [5,6], has also been shown to be useful for creating semantic resources in new languages.

The first purely automatic method for semantic resource creation based on parallel corpora was the Semantic Mirrors method [11,12], which discovers word senses by treating each language in a bilingual parallel corpus as the mirror of the other. Concepts and semantic relations are discovered by going back and forth between the two sides of the parallel corpus on the basis of alignment links. The extracted relations permit to organize the concepts retained in the new language in a complex lexico-semantic network similar to PWN. However, the structure of the obtained network is different than the one of PWN and the resource is not aligned to other wordnets. Translation information is also used by Ide et al. [17] for inducing word senses from a multilingual parallel corpus. The translations of source (English) words in six languages found in the corpus serve as features for building translation vectors. The vectors are then clustered according to their similarity and the obtained clusters describe the English words' senses. In the same vein, van der Plas and Tiedemann [32] propose an alignment approach to synonym extraction. Translation vectors are built from the alignments of source (Dutch) words in ten languages found in a multilingual parallel corpus and the similarity of the vectors of different source words reveals their semantic proximity.

The cross-lingual method of Apidianaki [1] combines translation and distributional information found in a bilingual parallel corpus for word sense induction (WSI). The translations of a source word are represented by weighted feature vectors built from the corresponding source contexts. The distributional vectors serve to cluster the translations according to their similarity and the obtained translation clusters describe the senses of source words in the corpus. The cross-lingual WSD method that we use in this work exploits the translation clusters built by applying this word sense induction method to an English-French parallel corpus. The automatically acquired French sense clusters constitute the new synsets to be integrated into WOLF by the WSD method. The vectors that serve for translation clustering are exploited during disambiguation for finding the most adequate anchor points for the new clusters in

the hierarchy. The main advantages of the proposed method are that it is fully automatic, it requires no manual translation and the semantic information is directly derived from a publicly available corpus (Europarl) which was not used for the initial construction of WOLF. The method is still dependent on the structure of PWN, however it offers an alternative way for automatically creating new wordnets or adding synsets to already existing ones while preserving the alignment of the resources.

In the next section we explain this methodological choice, as well as the advantages of using a cross-lingual rather than a monolingual method for filling wordnets in new languages.

2.3.2 Adapting a cross-lingual WSD method to wordnet development

Filling empty synsets in a wordnet can be divided into two subtasks: (a) creating new clusters of synonyms (synsets), and (b) defining the place where the synsets should be located in the hierarchy. New synonym clusters can be acquired automatically by a word sense induction method. For the second subtask, a word sense disambiguation method is needed.

For enriching WOLF, one option would be to acquire new synsets from monolingual French corpora and integrate them in the hierarchy. Monolingual word sense induction methods generally discover senses by clustering word usages on the basis of distributional information which can subsequently be used for disambiguation [22]. If sense induction was performed in French, then the disambiguation method would need to exploit information in WOLF to identify where the new synsets should be placed in the hierarchy. However, as WOLF contains a high number of empty synsets, its sparsity would have a negative impact on disambiguation.

Given that WOLF has the same structure as PWN (version 2.0), an alternative to using a monolingual disambiguation method is to exploit information in the English WordNet for disambiguating the new French synsets. In this case, the new French synsets would be included in the hierarchy by means of a cross-lingual WSD classifier, based on information found in the English WordNet. The cross-lingual WSD method proposed by Apidianaki [2] is well adapted to the task at hand for several reasons. First, it exploits the results of a WSI method that generates synset-like clusters of the translations of words in a parallel corpus [1]. The translations are grouped together according to their semantic similarity, calculated on the basis of source language distributional information. More precisely, the translations are characterized by source language feature vectors whose similarity serves to group the translations into clusters. When applied to the English-French language pair, the method clusters the French translations of English words by comparing the corresponding English context feature vectors. The obtained clusters of translations describe the senses of the English words in the corpus and contain semantically close words in French, similar to wordnet synsets. These automatically built French clusters constitute the synsets to be included in the resource.

The second reason that makes this cross-lingual WSD method well suited for this task is that the proposed WSD classifier selects French clusters for filling the empty synsets based on source language (English) information. This is due to the nature of

the output of the WSI method: the generated translation clusters are characterized by weighted English feature vectors that can be used for assessing the similarity between a cluster and a synset. During disambiguation, the comparison of the vectors to information extracted from WordNet would serve to identify the most adequate synset for each French cluster. The word sense induction and disambiguation methods employed in this study are presented in detail in the next section.

3 Data-driven synset induction and disambiguation

3.1 Synset acquisition through word sense induction

3.1.1 Training

Our WSI method is trained on the sentence aligned English-French (EN–FR) part of the Europarl corpus (release v6) [20]. Prior to word alignment, we apply standard pre-processing steps: the corpus is tokenized and lowercased, and imbalanced sentence pairs, which are hard to word align, are omitted.[4] Both sides of the corpus are then lemmatized and part-of-speech (POS) tagged using the TreeTagger [36], and the corpus is aligned at the level of word types using GIZA++ [29] in both directions. Two bilingual lexicons are extracted from the alignment results, one for each translation direction (EN–FR/FR–EN). To discard noisy alignments, the translations are filtered on the basis of their alignment score (threshold: 0.01) and according to their POS, keeping for each word only high confidence translations pertaining to the same grammatical category (i.e. we retain the noun translations of nouns, the verb translations of verbs, etc.). This eliminates erroneous alignment correspondences due to translation divergences. Finally, an intersection filter discards any translation correspondences not found in both lexicons. The translations used for clustering are the ones that translate a source word (w) more than 10 times in the training corpus. This threshold, which was experimentally shown to perform well, leaves out some translations of the source words but has a double merit: it reduces data sparseness issues that pose problems during clustering and eliminates erroneous translations which may be present in the lexicons because of spurious alignments.

3.1.2 Semantic similarity calculation

For each translation of a source word w, we extract the content words that co-occur with w in the corresponding source sentences of the parallel corpus (i.e. words that occur in the same sentence as w whenever it is translated by that translation). The retained source language words constitute the features of the vector built for each translation. For instance, four vectors are built for the translations retained from the training corpus for the English noun *stage*: *stade*, *phase*, *étape* and *scène*. The features of each vector are the content words that cooccur with *stage* in the source

[4] Sentence pairs with a great difference in length, where one sentence is more than three times longer than the corresponding sentence in the other language.

- *stade*: {procedure, point, process, agreement, priority, negotiation, proposal, support, ...}
- *étape*: {work, development, system, policy, procedure, proposal, area, support, council, ...}
- *phase*: {process, development, treaty, proposal, policy, report, negotiation, monetary, ...}
- *scène*: {political, economic, spectator, player, global, citizen, people, partner, action, ...}

Fig. 1 Source language features retained for the translations of the English noun *stage*.

side of the aligned sentences where it is translated by each French translation, as shown in Figure 1.

A similarity score is computed for each pair of translations using the weighted Jaccard measure [16]. The input of the similarity calculation consists of the co-occurrence counts of the source language features retained for the translations of the source word. The score assigned to a pair of translations indicates their degree of similarity and is computed as follows.

Let F be the total number of features retained from the contexts of w and let $N \in F$ be the number of features retained for each translation T_i. Each feature F_j ($1 \leq j \leq N$) receives a total weight with the translation ($w(F_j, T_i)$), defined as the product of the feature's global weight ($gw(F_j)$) and its local weight with that translation ($lw(F_j, T_i)$). The global weight of a feature F_j is a function of the number n of translations to which F_j is related and of the probability (p_{ij}) that F_j co-occurs with instances of w translated by each of the translations:

$$gw(F_j) = 1 - \frac{\sum_{i=1}^{n} p_{ij} \log(p_{ij})}{n} \tag{1}$$

Each p_{ij} is computed as the ratio of the co-occurrence counts of F_j with w when translated as T_i to the total number of features (N) seen with this translation.

$$p_{ij} = \frac{\text{cooc\_count}(F_j, T_i)}{N} \tag{2}$$

On the other hand, the local weight between feature F_j and translation T_i ($lw(F_j, T_i)$) directly depends on the number of times they occur together:

$$lw(F_j, T_i) = \log(\text{cooc\_count}(F_j, T_i)) \tag{3}$$

The intuition underlying this weighting scheme is that if an interesting semantic relation exists between a feature F_j and a translation T_i of w, then we expect the probability (p_{ij}) of the feature F_j occuring in the contexts where w is translated by this translation to be higher than if they were independent. In other words, a feature gets a high total weight with a translation when it appears frequently in the corresponding source contexts and rarely in the contexts of the other translations of w. Recall now that the total weight of a feature with a translation is defined as follows:

$$w(F_j, T_i) = gw(F_j) \cdot lw(F_j, T_i) \tag{4}$$

The Weighted Jaccard (WJ) similarity of two translations T_m and T_n is then calculated using the total weight of the features that occur with each translation:

$$WJ(T_m, T_n) = \frac{\sum_{r=1}^{|F|} \min(w(F_r, T_m), w(F_r, T_n))}{\sum_{r=1}^{|F|} \max(w(F_r, T_m), w(F_r, T_n))} \quad (5)$$

Translation pairs with a score above a threshold are considered as semantically related. The threshold is defined locally for each source word using the dynamic thresholding procedure proposed by Apidianaki and He [3]. The threshold for a word w is initially set to the mean of the scores (above 0) of its translation pairs. The set of translation pairs of w is then divided into two sets ($G1$ and $G2$) according to whether they exceed or are inferior to the threshold. The average of scores of the translation pairs in each set is computed ($m1$ and $m2$) and constitutes the new threshold which serves to re-partition the translation pairs into two sets. The procedure is repeated until convergence.

The similarity calculation output and the similarity threshold are exploited by the clustering algorithm which groups closely related translations into *sense clusters*, describing the senses of the source language words. The clusters generated for the noun *stage*, for example, describe its two senses in the training corpus: {*stade*, *phase*, *étape*} and {*scène*} (i.e., the "phase" sense and the "platform" sense). The clustering procedure is detailed in the next section.

3.1.3 Semantic clustering

The semantic clustering algorithm used in our experiments groups the translations into clusters by exploiting the results of the similarity calculation described in the previous section [1,3]. The input of the algorithm for a source word w consists in: (a) the list of w's translations; (b) their similarity scores, and (c) the similarity threshold. The clustering is performed in two steps. First, each pair of translations with a similarity score above the threshold is considered as semantically close and forms an *initial* cluster (C). These two-element clusters are derived directly from the similarity table. During the second step, they may be enriched by additional translations by a recursive function which takes as input the cluster C and the list of translations of w, and outputs C eventually enriched by other translations. A new translation is included in a cluster if it is strongly related to all the other elements in the cluster (i.e. their similarity score exceeds the threshold). The clustering stops when all the translations of w are included in some cluster and all their relations have been checked. In graph theory terms, the final clusters are characterized by *global connectivity* given that all their elements are linked between them by strong relations. The translations having no strong relations to any other translation of w are included in separate one-element clusters.

Using this cross-lingual WSI method, two sense cluster inventories are created from our training data: an EN–FR inventory, where the senses of English words are described by clusters of their French translations, and a FR–EN inventory, where the senses of French words are described by clusters of their English translations. The sense clusters group semantically similar words in the target language and could be compared to wordnet synsets. In Table 1, we present some examples of English and

Table 1 Entries from the sense cluster inventories.

Language	POS	Source word	Sense clusters
EN–FR	Nouns	omission	{carence}{lacune, oubli ,négligence} {lacune, omission}
		assessment	{analyse, appréciation, bilan, estimation, étude} {évaluation} {jugement, estimation}
	Verbs	accommodate	{adapter, répondre} {satisfaire, répondre} {accueillir}
		combine	{conjuguer, combiner, associer} {fusionner} {ajouter} {réunir, unir, conjuguer, concilier} {conjuguer, concilier, réunir, associer} {regrouper, rassembler, réunir}
	Adjs	dubious	{suspect}{douteux, discutable} {discutable, contestable}
		outstanding	{excellent, suspens, remarquable} {exceptionnel, extraordinaire} {remarquable, exceptionnel, excellent}
FR–EN	Nouns	diffusion	{broadcasting, dissemination, distribution} {circulation} {distribution, diffusion} {broadcasting, distribution, broadcast}
		peine	{sentence, penalty, punishment} {trouble, bother}
	Verbs	menacer	{threaten} {endanger, risk, jeopardise}
		lier	{link, connect, relate} {attach} {combine}
	Adjs	lisible	{comprehensible, legible} {legible, readable}
		malheureux	{sad, unhappy, wretched} {unfortunate}

French entries of different POS and degrees of polysemy. The EN verb *accommodate*, for instance, has four translations (*adapter, répondre, satisfaire, accueillir*) which are grouped in three sense clusters: {*adapter, répondre*} ("adapt" sense), {*satisfaire, répondre*} ("satisfy") and {*accueillir*} ("put up" sense). The first two clusters overlap (they both contain the French verb *répondre*), which means that the described senses are probably related. The cluster overlaps could actually serve as clues to their merge, if coarser-grained sense descriptions were needed. However, as a translation might be found in the intersection of two clusters because of being ambiguous between the two senses, a merge would be more reliable if the intersection contained more than one element.

Here, the sense clusters are used for filling French synsets corresponding to PWN synsets, which are characterized by fine granularity. As wordnet synsets might in general contain the same literals, the cluster overlaps pose no problem in this context. Consequently, clusters are not merged but used as proposed by the WSI method.

3.2 Sense clusters integration into WOLF

The automatically built EN–FR inventory contains entries for English words of different parts of speech. In this first experiment, we focus on word meanings that correspond to empty synsets in WOLF. In future work, we intend to further enrich

non empty synsets (i.e. synsets that already contain one or more French literals) with additional information found in the sense clusters.

We use the cross-lingual WSD method proposed by Apidianaki [2] which exploits the output of the WSI method presented in the previous section. In a monolingual data-driven WSD task, the clusters obtained for a word by clustering its instances in a monolingual corpus would constitute its candidate senses from which the most adequate one would have to be selected for new instances of the word in context. This selection would be performed by comparing the cluster vectors to information in the new context.

In the current setting, the goal of the WSD method is to assign French clusters to empty synsets in WOLF using English feature vectors. So, the information exploited for WSD consists in the words found in the corresponding English synsets (in PWN) and their related synsets, their definitions and usage examples. Given that information in the vectors built from the training corpus is lemmatized, the information retained from PWN is lemmatized as well [36] and gathered in a bag of words. The adequacy of a cluster (C) for filling a given synset (S) is estimated by comparing the vectors of the clustered translations to the information retained from PWN for the synset. If common features (CFs) are found with just one cluster, this cluster is selected. Otherwise, each 'cluster-synset' association is assigned a score corresponding to the mean of the weights of the CFs with the clustered translations (weights assigned to each feature during WSI (cf. Section 3.1)). In formula 6, $(CF_j)_{j=1}^{|CF|}$ is the set of common features between the cluster and the synset, and N_{CF} is the number of translations T_i in the cluster characterized by a feature CF (i.e. translations having the CF in their vector). The cluster that receives the highest score is selected and assigned to the empty synset.

$$assoc\_score(C,S) = \frac{\sum_{i=1}^{N_{CF}} \sum_{j=1}^{|CF|} w(T_i, CF_j)}{N_{CF} \cdot |CF|} \tag{6}$$

For instance, the empty synset 'odd#a#2' (definition: "not easily explained"; usage: "it is odd that his name is never mentioned"), is correctly filled by the French cluster {*curieux*, *bizarre*}. The other clusters available for *odd*, which do not fit this synset and get lower scores, are: {*contradictoire*, *singulier*, *bizarre*} and {*curieux*, *étrange*}. More examples of synsets filled by the WSD method are shown in Table 2. We provide the PWN id of the empty synsets in WOLF, the English headword, the literals in the corresponding PWN synsets, as well as their definitions and usage examples.[5] The French literals in the sense cluster most strongly associated with a PWN synset, which are used to fill the corresponding synset in WOLF, are given in the last column of Table 2.

The process of selecting the French synset that best suits a cluster on the basis of English contextual information is illustrated with the example given in Table 3. It details the case of the English adjective *peaceful* which belongs to three synsets in PWN, all empty in WOLF. The WSD method has to fill one of these synsets with the

[4] The weights of the features are omitted for the sake of readability.

[5] The table does not include information on all the neighboring PWN synsets, which was used during WSD. This information can however be easily recovered from PWN.

Table 2 WOLF synsets filled by our WSD method.

POS	EN entry	PWN synset	FR sense cluster
Nouns	**presentation**	ENG20-06725607-n: {presentation#n#4} the act of presenting a proposal	{présentation, exposé}
	scam	ENG20-00709982-n: {scam#n#1, cozenage#n#1} a fraudulent business scheme	{arnaque, escroquerie}
	loyalty	ENG20-04639012-n: {loyalty#n#1} the quality of being loyal	{fidélité, loyauté}
Verbs	**discourage**	ENG20-00841635-v: {warn#v#2, discourage#v#3, admonish#v#1, monish#v#2} admonish or counsel in terms of someone's behavior; "I warned him not to go too far"; "I warn you against false assumptions"; "She warned him to be quiet"	{décourager, dissuader}
	distance	ENG20-02602279-v: {distance#v#1} keep at a distance; "we have to distance ourselves from these events in order to continue living"	{éloigner, distancier}
	divide	ENG20-02543903-v: {separate#v#1, divide#v#3} act as a barrier between; stand between; "The mountain range divides the two countries"	{partager, séparer, répartir}
Adjectives	**horrific**	ENG20-01575285-a: {hideous#a#1, horrid#a#2, horrific#a#1, outrageous#a#1} grossly offensive to decency or morality; causing horror; "subjected to outrageous cruelty"; "a hideous pattern of injustice"; "horrific conditions in the mining industry"	{atroce, terrible, épouvantable}
	sure	ENG20-00331475: {indisputable#a#2, sure#a#9} impossible to doubt or dispute; "indisputable (or sure) proof"	{sûr, certain}
	clever	ENG20-00413048-a: {cagey#a#1, cagy#a#1, canny#a#1, clever#a#2} showing self-interest and shrewdness in dealing with others; "a cagey lawyer"; "too clever to be sound"	{habile, astucieux}
Adverbs	**brutally**	ENG20-00204148-b: {viciously#r#1, brutally#r#1, savagely#r#1} in a vicious manner; "he was viciously attacked"	{brutalement, sauvagement}
	early	ENG20-00101775-b: {early_on#r#1, early#r#1} during an early stage; "early on in her career"	{rapidement, tôt}
	exactly	ENG20-00372187-b: {precisely#r#2, incisively#r#2, exactly#r#3} in a precise manner; "she always expressed herself precisely"	{exactement, précisément}

Table 3 Comparison of vector and PWN information during WSD. Synset ids correspond to Princeton WordNet version 2.0.

PWN entry *peaceful* (adj)

French cluster	**The English vector of the cluster represented as a bag of words**
{*paisible*, *pacifique*}	***absence acceptance*** *achieve action* ***activity aggressive*** *agreement* ***atmosphere attitude*** *authority* ***be become*** *believe bring* ***call calm*** *can* ***citizen*** *clear coexistence commission community conflict continue cooperation council* ***country*** *crisis democracy democratic demonstration demonstrator development dialogue* ***dispute*** *do east economic effort election* ***emotional*** *energy ensure…*

Corresponding PWN synsets	**Synset-related information represented as a bag of words**
01615936-a {*law-abiding*, *peaceful*} Def.: *(of groups) not violent or disorderly* Usage: *the right of peaceful assembly* Neighboring synsets: 01615787-a {*orderly*}	*an assembly confront crowd devoid disorderly disruption group law-abiding not of or* ***orderly*** *peaceful* ***president right*** *the* ***violence*** *violent*
01686906-a {*peaceful*} Def.: *not disturbed by strife or turmoil or war* Usage: *a peaceful nation; peaceful times; a far from peaceful Christmas; peaceful sleep* Neighboring synsets: 01687344-a {*calm*, *serene*, *tranquil*} 00302191-a {*calm*} 01202829-a {*amicable*} …	*a* ***absence*** *abstain* ***acceptance activity aggressive*** *agitation almost amicable an and antagonist assertiveness at* ***atmosphere attitude be become*** *by* ***call calm*** *characterize* ***citizen*** *conducive* ***country*** *directly dispose* ***dispute*** *disturb disturbance dovish* ***emotional*** …
02425529-a {*passive*, *peaceful*} Def.: *peacefully resistant in response to injustice* Usage: *passive resistance* Neighboring synsets: 02425368-a {*nonviolent*}	*abstain from in injustice nonviolent of on* ***passive*** *peaceful peacefully principle resistance resistant response the to use violence*

Cluster-to-synset mapping: the bag of words representing the synset ENG20-01686906-a is the closest to that of the vector of the French cluster {*paisible*, *pacifique*} and it also gets the highest score during WSD.

Outcome: *paisible* and *pacifique* are added to synset ENG20-01686906-a in WOLF

French cluster {*paisible*, *pacifique*}. Each of the PWN synsets for *peaceful* is shown in Table 3 (including literals, definition and usage examples) together with some of the related synsets that are used to build the corresponding bags of words. The features that characterize, at the same time, the cluster vector and one of the synsets are shown in boldface. The bag of words representing the synset ENG20-01686906-a is the closest to that of the vector of the French cluster, and the synset also gets the highest score during WSD. Therefore, *paisible* and *pacifique* are added to synset ENG20-01686906-a in the WOLF.

4 Evaluation results

Our approch fills 3,904 previously empty synsets in WOLF: 2,333 nominal, 576 verbal, 709 adjectival and 286 adverbial synsets. Given that no gold standard is available for this task – which would permit to perform an automatic evaluation – we have manually examined 10% of the synsets filled for each POS to evaluate the quality of the proposed clusters and the correctness of their assignment to some synset in WOLF according to the following criteria:

- a cluster is considered as a good quality one if it groups words that share the same meaning,
- the assignment of a cluster to a synset is considered as correct if its contents correctly describe the sense in the corresponding PWN synset.

A cluster can be correctly assigned to a synset only if it is of good quality according to the first evaluation criterion. Consequently, all WSD assignments involving noisy clusters are considered as wrong assignments. We consider as noisy the clusters that contain one or more translations that are not semantically close to the others, even if the rest of the translations in the cluster are synonymous.

Both aspects have been evaluated by two annotators. The inter-annotator agreement was measured at $\kappa = 0.67$ for cluster quality and 0.59 for the WSD results, which is conventionally interpreted as "good" agreement [7]. The lower agreement obtained for disambiguation is unsurprising, given the closeness of the considered synsets which correspond to fine-grained sense distinctions in PWN. As has been shown by Erk and McCarthy [14], multiple WordNet senses might apply to an occurrence of a polysemous word in context, an argument in favor of graded sense assignments [19]. Here, we are looking for the best-fitting WordNet sense for a cluster of translations (not for words in context) but the same effect of varying sense applicability occurs in this setting.

The evaluation results are presented in Table 4. According to the results obtained for all POS, the clusters group semantically similar words in 75.5% of the cases. Significant variations are however observed for different POS. The first row of the table contains the percentage of good quality clusters in the test set. The second row of the table shows the percentage of the clusters that were correctly assigned to WOLF synsets by the WSD method. Given that according to our evaluation criteria only good clusters can be correctly integrated into WOLF, we calculate the (conditional) accuracy of the WSD method by reference to the number of good clusters. This

Table 4 Results on empty WOLF synsets (%)

	Nouns	Verbs	Adjs	Advs	All POS
good quality clusters	72.1	62.9	81.0	86.2	75.5
correct WSD	64.6	53.0	75.1	73.7	66.6
overall: good quality clusters with correct WSD	46.6	33.3	60.8	63.5	51.0

accuracy for WSD insertions on all POS is 67%, which is very encouraging.[6] Table 4 shows that for WSD, as is the case for WSI, the results vary from one POS to another. We also provide overall accuracy results for the whole task, i.e. the proportion of automatically obtained clusters that are both good clusters (grouping words that share the same meaning) and assigned to the correct synset. This overall accuracy is computed directly as the WSI accuracy times the WSD conditional accuracy (recall that the latter was computed only on good clusters). In a setting where a manually compiled bilingual dictionary would be available and the candidate senses would not contain any noise, the focus would be put only on the accuracy of the WSD method.

The divergences observed between different parts of speech are due to the restrictive cluster quality criterion according to which one incorrect word in an otherwise correct cluster turns the whole cluster into an incorrect one. This strict criterion unfairly penalizes and rejects interesting although noisy clusters. We notice that this constraint has a strong impact during evaluation especially on clusters with many translations, like the verb clusters. We plan to proceed to a more detailed and flexible evaluation in order to more accurately estimate the actual merit of the clustering method, which will also imply devising methods for cleaning noisy clusters. In a semi-automatic setting, the manual cleaning of the noisy clusters by a lexicographer would considerably improve their integration in the French resource.

We should highlight the difficulty of the disambiguation task as the WSD method is asked to fill synsets that were left empty by the methods initially employed for creating WOLF. These empty synsets often correspond to rare senses in PWN that may not exist in the training corpus, or to senses for which little information is available in PWN. The role of the training corpus is particularly important in data-driven word sense induction and disambiguation. Given that the training corpus used in our experiments contains parliamentary proceedings, the derived senses are not always adequate for filling a general language resource like WOLF, as senses represented by empty synsets might not be present in the corpus.

In order to more fairly estimate the performance of the WSD method in this setting, we also tested it on the whole resource. In this case, the method was asked to select the most appropriate synset for each cluster from *all* synsets in WOLF (not only the empty ones). In this setting, the WSD method reaches a much higher performance of 80.13%, as shown in Table 5, which shows that it is particularly well adapted to the

[6] All accuracy scores reported for our system in Table 4 have been computed with respect to the judgments of the two annotators. More precisely, we first computed an accuracy score separately for each annotator and then retained the average of the two scores.

Table 5 Results on all WOLF synsets (%)

	Nouns	**Verbs**	**Adjs**	**Advs**	**All POS**
good quality clusters	72.1	62.9	81.0	86.2	75.5
correct WSD	69.7	72.7	85.1	93.0	80.1
overall: good quality clusters with correct WSD	50.3	45.7	68.9	80.2	61.2

wordnet development task. Table 5 contains detailed results for words of different parts of speech.

5 Discussion and perspectives

5.1 Analysis of the errors in the clustering output

From a close examination of the clustering results, two main sources of errors were identified. In some cases, the noise found in the clusters is due to alignment errors that were not detected and eliminated by the filters that served to clean the lexicons (cf. Section 3.1). In other cases, the noise is introduced during clustering. The error analysis indicates some cases of problematic clustering that fall into the second category:

(a) cases where multiword units were not considered during word alignment. This is observed in the cluster {*considération*, *compte*} corresponding to the English noun *consideration*, which should ideally be {*prise en compte*, *considération*}. This issue could be addressed if multiword expressions were identified prior to word alignment.
(b) clustering of topically related but not synonymous words, as in the cluster {*raisin*, *moût*} corresponding to the noun *grape*.
(c) clustering of antonymous but distributionally similar words, as in the case of {*sain*, *malsain*} (cluster of *unhealthy*). Antonymous words can be found in the alignment results when the negation is expressed paraphrastically in one of the languages and is not captured by the alignment, as is here the case with the translation *sain* retained for *unhealthy*. Then, as antonymous words often appear in similar contexts, it happens that they end up in the same cluster.

5.2 Generating coarser wordnets using cross-lingual WSD

A common criticism of PWN is the high granularity of the proposed semantic descriptions which, combined to the great number and similarity of the senses, might hamper the efficient use of this resource for WSD [13,28]. Fine sense distinctions increase the processing complexity and the risk of information loss when a forced choice among closely related senses has to be made without considering their relations [10]. As pointed out by Ide and Wilks [18], this fine granularity is not

even necessary for efficient WSD in NLP applications where disambiguation, when needed, mostly involves homonym-level distinctions. In the rare cases where finer-grained distinctions are needed, they should be handled by more robust types of processing. As discussed earlier in this paper, an additional problem posed by the high granularity of PWN during the building of new wordnets is that it is difficult to find correspondences for fine-grained senses in the new languages. This results in resources much sparser than PWN, containing numerous empty synsets. Furthermore, the difficulty to establish correspondences between fine-grained senses and their translations makes difficult the exploitation of the resources in multilingual applications [37].

The granularity of wordnet-like resources can be reduced by identifying the similarity of the proposed senses.[7] Several attempts have been made for reducing the polysemy of words in PWN and the granularity of their senses. Peters et al. [30] perform automatic sense clustering of nouns and verbs using the relations defined in WordNet (sisters, autohyponyms, twins and cousins). Mihalcea and Moldovan [23] apply three principles from the lexical semantics litterature [8] to measuring the ambiguity level between PWN synsets of all parts of speech. The proposed rules take into account the overlaps between the elements in different synsets and their relations to other synsets in the hierarchy (hypernyms, antonyms and pertainyms). Synsets whose similarity is high enough according to these rules are collapsed together into one. Furthermore, the polysemy of WordNet is reduced based on the frequency of senses and the probability of a synset occurring in texts (as measured on SemCor, a corpus sense-tagged with PWN synsets). Synsets with very low probability of occurrence are dropped and the number of senses of polysemous words is reduced. So, the semantic principles result in collapsed synsets, while the probabilistic principles determine which synsets can be discarded. The application of these two types of rules results in a reduction in the number of synsets and, consequently, in the number of word senses.

The output of the cross-lingual WSD method presented in this paper could also serve to reduce the granularity of the new wordnet resource and, consequently, the degree of polysemy in PWN. This can be done by collapsing highly similar synsets together in the new wordnets and identifying the relations between the corresponding PWN synsets. As explained in the previous sections, the enrichment of WOLF synsets with new literals was performed by finding the most adequate synset *for each translation cluster*. The contents of the cluster were then used to fill the selected synset. It would however also be possible to proceed the other way around, i.e. to seek the most adequate cluster *for each synset*. In this way, the same cluster could be associated to different PWN synsets and this would serve as a clue for measuring the similarity of the synsets and merging them. The examples presented in Table 6 illustrate cases of one-to-many associations established between clusters and PWN/WOLF synsets, where the clusters given in the last column are associated to several empty synsets of the source word. For instance, the third cluster of noun *waste*: {*perte, gaspillage*} (other clusters for the word: {*ordure*}, {*déchet, gaspillage*}) is

[7] This information can also be highly useful for the evaluation of WSD systems as it would permit to penalize differently WSD errors involving close and distant senses [33].

source word	synset id	definition	score	literals
waste	00699179-n	useless or profitless activity; using or expending or consuming thoughtlessly or carelessly	2.279	perte gaspillage
	04652558-n	the trait of wasting resources	1.760	
	01182120-n	the collection of rules imposed by authority	1.582	
disturbance	13588082-n	an unhappy and worried mental state	1.741	perturbation trouble
	13182390-n	a disorderly outburst or tumult	1.858	
	00318377-n	the act of disturbing something or someone; setting something in motion	1.417	
	01110261-n	the act of fighting; any contest or struggle	0.852	
birthday	14390624-n	the date on which a person was born	3.2	anniversaire
	14388733-n	an anniversary of the day on which a person was born (or the celebration of it)	1.654	

Table 6 One-to-many associations between clusters and synsets

associated by the WSD method to three empty synsets of *waste* in WOLF, whose glosses are given in the third column of Table 6. Each cluster-synset association is weighted during disambiguation (cf. Section 3.2) and the scores in column 4 show the strength of the association; the higher the score, the stronger the association between the cluster and the synset. In the same way, the cluster {*perturbation, trouble*} is assigned to four empty synsets of the word *disturbance*, while both empty synsets of *birthday* are assigned the cluster {*anniversaire*}.

An alternative way to establish one-to-many correspondences would be to retain *for each cluster* not just the synset that gets the best association score during WSD but more than one high scored synsets. In this case, a careful study of the proposed associations could help to define a threshold that would reflect good assignments.

The establishment of one-to-many correspondences can serve to fill multiple WOLF synsets at once or to merge them into one, before assigning the contents of the cluster. Furthermore, these 'cluster-synset' associations can serve to merge the corresponding PWN synsets, to reduce the granularity of the resource and facilitate the establishment of correspondences with words in the target language. This information seems thus to be particularly relevant for creating coarser-grained resources. Nevertheless, although successful in several cases, the output of this process should be treated with caution as noisy associations might lead to erroneous merges. In future work, we intend to explore ways for identifying strong correspondences and ruling out erroneous ones but until then, the method would be better suited to a semi-supervised setting where the proposed associations could be manually validated by a lexicographer.

6 Conclusion

We have proposed a novel approach to developing and enriching wordnet resources in languages other than English. We have shown how a cross-lingual word sense disambiguation method can be used to assign content to otherwise empty synsets in newly built wordnet resources. Our WSD method exploits the results of a data-driven sense induction method which discovers the senses of English words by grouping

their French translations into sense clusters. The obtained clusters are integrated by the WSD method into the French wordnet resource WOLF based on information found in Princeton WordNet, to which the WOLF is aligned. The results indicate that the proposed methods are particularly useful for building wordnets in new languages. Moreover, given that wordnet resources in languages other than English are often aligned to PWN, the cross-lingual WSD method can be used to enrich these resources and increase their coverage.

Our work shows that word sense induction and disambiguation methods can efficiently collaborate for enriching lexico-semantic resources by mapping senses automatically extracted from parallel data to a manually developed sense inventory like Princeton WordNet. It is therefore a valuable and complementary alternative to more common approaches that leverage multilingual lexicons extracted from dictionaries. Still, if such lexicons are available they could replace the sense induction output and be directly exploited by the disambiguation method for enriching wordnet resources or creating new ones.

Based on these encouraging results, we aim at extending the use of the proposed cross-lingual WSD method for enriching non-empty WOLF synsets with additional literals. This is because some synsets are incomplete and lack some literals which could be retrieved from corpora by the methods introduced in this paper. As explained above, the quality of the automatically acquired synsets and the performance of the disambiguation method strongly depend on the parallel corpora used for training. The present study was carried out using information extracted from the Europarl corpus. We would like to include more diverse training corpora in order to obtain richer semantic representations and further extend WOLF's coverage. Last but not least, we intend to explore methods for discarding synsets corresponding to rare senses in PWN or to senses that have no counterpart in French, and for merging semantically close synsets in order to reduce the granularity and the sparseness of the French resource. The output of the disambiguation method would be particularly useful to this aim given that the method can establish one-to-many correspondences between clusters and synsets, highlighting in this way their semantic similarity that can serve to their grouping.

References

1. Apidianaki, M.: Translation-oriented word sense induction based on parallel corpora. In: Language Resources and Evaluation Conference (LREC), pp. 3269–3275. Marrakech, Morocco (2008)
2. Apidianaki, M.: Data-driven semantic analysis for multilingual WSD and lexical selection in translation. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL), pp. 77–85. Association for Computational Linguistics, Athens, Greece (2009)
3. Apidianaki, M., He, Y.: An algorithm for cross-lingual sense clustering tested in a MT evaluation setting. In: Proceedings of the 7th International Workshop on Spoken Language Translation (IWSLT), pp. 219–226. Paris, France (2010)
4. Bannard, C., Callison-Burch, C.: Paraphrasing with Bilingual Parallel Corpora. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), pp. 597–604. Association for Computational Linguistics, Ann Arbor, Michigan (2005)
5. Carpuat, M., Wu, D.: Improving Statistical Machine Translation using Word Sense Disambiguation. In: Proceedings of the Joint EMNLP-CoNLL Conference, pp. 61–72. Prague, Czech Republic (2007)

6. Chan, Y.S., Ng, H.T., Chiang, D.: Word Sense Disambiguation Improves Statistical Machine Translation. In: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL-07), pp. 33–40. Association for Computational Linguistics, Prague, Czech Republic (2007)
7. Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurement **20**(1), 37–46 (1960)
8. Cruse, D.: Lexical Semantics. Cambridge University Press, Cambridge, UK (1986)
9. Diab, M., Resnik, P.: An Unsupervised Method for Word Sense Tagging using Parallel Corpora. In: Proceedings of the the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02), pp. 255–262. Association for Computational Linguistics, Philadelphia (2002)
10. Dolan, W.B.: Word sense ambiguation: clustering related senses. In: Proceedings of the 15th conference on Computational linguistics - Volume 2, COLING '94, pp. 712–716. Association for Computational Linguistics, Stroudsburg, PA, USA (1994)
11. Dyvik, H.: Translations as Semantic Mirrors: from Parallel Corpus to Wordnet. In: Proceedings of the Workshop Multilinguality in the lexicon II at the 13th biennial European Conference on Artificial Intelligence (ECAI'98), pp. 24–44. Brighton, UK (1998)
12. Dyvik, H.: Translations as a Semantic Knowledge Source. In: Proceedings of the Second Baltic Conference on Human Language Technologies. Tallinn, Estonia (2005)
13. Edmonds, P., Kilgarriff, A.: Introduction to the special issue on evaluating word sense disambiguation systems. Natural Language Engineering **8**(4), 279–291 (2002)
14. Erk, K., McCarthy, D.: Graded Word Sense Assignment. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pp. 440–449. Association for Computational Linguistics, Singapore (2009)
15. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database. MIT Press, Cambridge, Massachusetts (1998)
16. Grefenstette, G.: Explorations in Automatic Thesaurus Discovery. Dordrecht : Kluwer Academic Publisher (1994)
17. Ide, N., Erjavec, T., Tufiş, D.: Sense Discrimination with Parallel Corpora. In: Proceedings of ACL'02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions, pp. 54–60. Association for Computational Linguistics, Philadelphia (2002)
18. Ide, N., Wilks, Y.: Making sense about sense. In: Word Sense Disambiguation: Algorithms and Applications, *Text, Speech and Language Technology*, vol. 33, pp. 47–74. Springer, Dordrecht, The Netherlands (2006)
19. Jurgens, D., Klapaftis, I.: SemEval-2013 Task 13: Word Sense Induction for Graded and Non-Graded Senses. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), pp. 290–299. Association for Computational Linguistics, Atlanta, Georgia, USA (2013)
20. Koehn, P.: Europarl: A Parallel Corpus for Statistical Machine Translation. In: Proceedings of MT Summit X, pp. 79–86. Phuket, Thailand (2005)
21. Lefever, E., Hoste, V.: SemEval-2010 Task 3: Cross-Lingual Word Sense Disambiguation. In: Proceedings of the 5th International Workshop on Semantic Evaluation, pp. 15–20. Association for Computational Linguistics, Uppsala, Sweden (2010)
22. Manandhar, S., Klapaftis, I., Dligach, D., Pradhan, S.: SemEval-2010 Task 14: Word Sense Induction & Disambiguation. In: Proceedings of the 5th International Workshop on Semantic Evaluation, pp. 63–68. Association for Computational Linguistics, Uppsala, Sweden (2010)
23. Mihalcea, R., Moldovan, D.I.: Automatic generation of a coarse grained WordNet. In: 14th Flairs Conference, pp. 454–458 (2002)
24. Mouton, C., de Chalendar, G.: JAWS: Just Another WordNet Subset. In: Proceedings of the 10th TALN Conference. Montreal, Canada (2010)
25. Navigli, R.: Word Sense Disambiguation: A Survey. ACM Computing Surveys **41**(2), 1–69 (2009)
26. Navigli, R., Ponzetto, S.P.: BabelNet: Building a Very Large Multilingual Semantic Network. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 216–225. Association for Computational Linguistics, Uppsala, Sweden (2010)
27. Ng, H.T., Chan, Y.S.: SemEval-2007 Task 11: English Lexical Sample Task via English-Chinese Parallel Text. In: Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), pp. 54–58. Association for Computational Linguistics, Prague, Czech Republic (2007)
28. Ng, H.T., Wang, B., Chan, Y.S.: Exploiting Parallel Texts for Word Sense Disambiguation: An Empirical Study. In: Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, pp. 455–462. Association for Computational Linguistics, Sapporo, Japan (2003)

29. Och, F.J., Ney, H.: A Systematic Comparison of Various Statistical Alignment Models. Computational Linguistics **29**(1), 19–51 (2003)
30. Peters, W., Peters, I., Vossen, P.: Automatic sense clustering in EuroWordNet. In: Proceedings of the First International Conference on Language Resources and Evaluation (LREC), vol. 1. Granada, Spain (1998)
31. Pianta, E., Bentivogli, L., Girardi, C.: MultiWordNet: developing an aligned multilingual database. In: First International Conference on Global WordNet, pp. 293–302. Mysore, India (2002)
32. van der Plas, L., Tiedemann, J.: Finding Synonyms Using Automatic Word Alignment and Measures of Distributional Similarity. In: Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions, pp. 866–873. Association for Computational Linguistics, Sydney, Australia (2006)
33. Resnik, P., Yarowsky, D.: Distinguishing Systems and Distinguishing Senses: New Evaluation Methods for Word Sense Disambiguation. Natural Language Engineering **5**(2), 113–133 (1999)
34. Sagot, B., Fišer, D.: Building a free French wordnet from multilingual resources. In: Ontolex 2008. Marrakech, Morocco (2008)
35. Sagot, B., Fort, K., Venant, F.: Extending the adverbial coverage of a French wordnet. In: Proceedings of the NODALIDA 2009 workshop on WordNets and other Lexical Semantic Resources. Odense, Danemark (2009)
36. Schmid, H.: Probabilistic Part-of-Speech Tagging Using Decision Trees. In: Proceedings of the International Conference on New Methods in Language Processing, pp. 44–49. Manchester, UK (1994)
37. Specia, L., Stevenson, M., Nunes, M.D.G., Castelo, G., Ribeiro, B.: Multilingual versus Monolingual WSD. In: Proceedings of the EACL Workshop Making Sense of Sense: Bringing Psycholinguistics and Computational Linguistics Together, April 3-7, pp. 33–40. Association for Computational Linguistics, Trento, Italy (2006)
38. Steinberger, R., Pouliquen, B., Widiger, A., Ignat, C., Erjavec, T., Tufiş, D., Varga, D.: The JRC Acquis: A multilingual aligned parallel corpus with 20+ languages. In: Proceedings of LREC. Genoa, Italy (2006)
39. Tufiş, D., Cristea, D., Stamou, S.: BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. In: Romanian Journal on Information Science and Technology. Special Issue on BalkaNet, vol. 7, pp. 9–34 (2004)
40. Vossen, P. (ed.): EuroWordNet: a multilingual database with lexical semantic networks for European Languages. Kluwer, Dordrecht (1998)