# Multilingual Word Sense Disambiguation Using Wikipedia

**Bharath Dandala**
Dept. of Computer Science
University of North Texas
Denton, TX
BharathDandala@my.unt.edu

**Rada Mihalcea**
Dept. of Computer Science
University of North Texas
Denton, TX
rada@cs.unt.edu

**Razvan Bunescu**
School of EECS
Ohio University
Athens, OH
bunescu@ohio.edu

## Abstract

We present three approaches to word sense disambiguation that use Wikipedia as a source of sense annotations. Starting from a basic monolingual approach, we develop two multilingual systems: one that uses a machine translation system to create multilingual features, and one where multilingual features are extracted primarily through the interlingual links available in Wikipedia. Experiments on four languages confirm that the Wikipedia sense annotations are reliable and can be used to construct accurate monolingual sense classifiers. The experiments also show that the multilingual systems obtain on average a substantial relative error reduction when compared to the monolingual systems.

## 1 Introduction and Motivation

Ambiguity is inherent to human language. In particular, word sense ambiguity is prevalent in all natural languages, with a large number of the words in any given language carrying more than one meaning. For instance, the English noun *plant* can mean *green plant* or *factory*; similarly the French word *feuille* can mean *leaf* or *paper*. The correct sense of an ambiguous word can be selected based on the context where it occurs, and correspondingly the problem of *word sense disambiguation* is defined as the task of automatically assigning the most appropriate meaning to a polysemous word within a given context.

Two well studied categories of approaches to word sense disambiguation (WSD) are represented by knowledge-based (Lesk, 1986; Galley and McKeown, 2003; Navigli and Velardi, 2005) and data-driven (Yarowsky, 1995; Ng and Lee, 1996; Pedersen, 2001) methods. Knowledge-based methods rely on information drawn from wide-coverage lexical resources such as WordNet (Miller, 1995). Their performance has been generally constrained by the limited amount of lexical and semantic information present in these resources.

Among the various data-driven WSD methods proposed to date, supervised systems have been observed to lead to highest performance in the Senseval evaluations [1]. In these systems, the sense disambiguation problem is formulated as a supervised learning task, where each sense-tagged occurrence of a particular word is transformed into a feature vector which is then used in an automatic learning process. Despite their high performance, the supervised systems have an important drawback: their applicability is limited to those few words for which sense tagged data is available, and their accuracy is strongly connected to the amount of available labeled data.

In this paper, we address the sense-tagged data bottleneck problem by using Wikipedia as a source of sense annotations. Starting with the hyperlinks available in Wikipedia, we first generate sense annotated corpora that can be used for training accurate and robust monolingual sense classifiers (WIKIMONOSENSE, in Section 2). Next, the sense tagged corpus extracted for the *reference* language is translated into a number of *supporting* languages. The word alignments between the reference sentences and the supporting translations computed by Google Translate are used to generate complementary features in our first approach to multilingual WSD (WIKITRANSSENSE, in Section 3). The reliance on machine translation (MT) is significantly reduced during the training phase of our second approach to multilingual WSD, in which sense tagged corpora in the supporting languages are created through the interlingual links available in Wikipedia. Separate classifiers are

---

[1]http://www.senseval.org

trained for the reference and the supporting languages and their probabilistic outputs are integrated at test time into a joint disambiguation decision for the reference language (WIKIMUSENSE, in Section 4).

Experimental results on four languages demonstrate that the Wikipedia annotations are reliable, as the accuracy of the WIKIMONOSENSE systems trained on the Wikipedia dataset exceeds by a large margin the accuracy of an informed baseline that selects the most frequent word sense by default. We also show that the multilingual sense classifiers WIKITRANSSENSE and WIKIMUSENSE significantly outperform the WIKIMONOSENSE systems (Section 5).

## 2 The WikiMonoSense System

In an effort to alleviate the sense-tagged data bottleneck problem that affects supervised learning approaches to WSD, the WIKIMONOSENSE system uses Wikipedia both as a repository of word senses and as a rich source of sense annotations. Wikipedia is a free online encyclopedia, representing the outcome of a continuous collaborative effort of a large number of volunteer contributors. Virtually any Internet user can create or edit a Wikipedia webpage, and this "freedom of contribution" has a positive impact on both the quantity (fast-growing number of articles) and the quality (potential mistakes are quickly corrected within the collaborative environment) of this online resource. Wikipedia editions are available for more than 280 languages, with a number of entries varying from a few pages to three millions articles or more per language.

A large number of the concepts mentioned in Wikipedia are explicitly linked to their corresponding article through the use of links or piped links. Interestingly, these links can be regarded as *sense annotations* for the corresponding concepts, which is a property particularly valuable for words that are ambiguous. In fact, it is precisely this observation that we rely on in order to generate sense tagged corpora starting with the Wikipedia annotations (Mihalcea, 2007; Dandala et al., 2012).

### 2.1 A Monolingual Dataset through Wikipedia Links

Ambiguous words such as e.g. *plant*, *bar*, or *argument* are linked in Wikipedia to different articles, depending on their meaning in the context

where they occur. Note that the links are *manually* created by the Wikipedia users, which means that they are most of the time accurate and referencing the correct article. The following represent four example sentences for the ambiguous word *bar*, with their corresponding Wikipedia annotations (links):

1. In 1834, Sumner was admitted to the [[bar (law)|bar]] at the age of twenty-three, and entered private practice in Boston.

2. It is danced in 3/4 time (like most waltzes), with the couple turning approx. 180 degrees every [[bar (music)|bar]].

3. Jenga is a popular beer in the [[bar (establishment)|bar]]s of Thailand.

4. This is a disturbance on the water surface of a river or estuary, often cause by the presence of a [[bar (landform)|bar]] or dune on the riverbed.

To derive sense annotations for a given ambiguous word, we use the links extracted for all the hyperlinked Wikipedia occurrences of the given word, and map these annotations to word senses, as described in (Dandala et al., 2012). For instance, for the *bar* example above, we extract five possible annotations: *bar (establishment), bar (landform), bar (law)*, and *bar (music)*.

In our experiments, the WSD dataset was built for a subset of the ambiguous words used during the SENSEVAL-2, SENSEVAL-3 evaluations and a subset of ambiguous words in four languages: English, Spanish, Italian and German. Since the Wikipedia annotations are focused on nouns (associated with the entities typically defined by Wikipedia), the sense annotations we generate and the WSD experiments are also focused on nouns. We also avoided those words that have only one Wikipedia label. This resulted in a set of 105 words in four different languages: 30 for English, 25 for Italian, 25 for Spanish, and 25 for German. Table 1 provides relevant statistics for the corresponding monlingual dataset.

### 2.2 The WikiMonoSense Learning Framework

Provided a set of sense-annotated examples for a given ambiguous word, the task of a supervised WSD system is to automatically learn a disambiguation model that can predict the correct sense

| Language | #words | #senses | #examples |
|----------|--------|---------|-----------|
| English | 30 | 5.3 | 632 |
| German | 25 | 4.5 | 550 |
| Italian | 25 | 5.4 | 815 |
| Spanish | 25 | 4.6 | 484 |

Table 1: #words = number of ambiguous words, #senses = average number of senses, #examples = average number of examples.

for a new, previously unseen occurrence of the word. Assuming that such a system can be reliably constructed, the implications are two-fold. First, accurate disambiguation models suggest that the data is reliable and consists of correct sense annotations. Second, and perhaps more importantly, the ability to correctly predict the sense of a word can have important implications for applications that require such information, including machine translation and automatic reasoning.

The WIKIMONOSENSE system integrates local and topical features within a machine learning framework, similar to several of the top-performing supervised WSD systems participating in the SENSEVAL-2 and SENSEVAL-3 evaluations. The disambiguation algorithm starts with a preprocessing step, where the text is tokenized, stemmed and annotated with part-of-speech tags. Collocations are identified using a sliding window approach, where a collocation is defined as a sequence of words that forms a compound concept defined in Wikipedia. Next, local and topical features are extracted from the context of the ambiguous word. Specifically, we use the current word and its part-of-speech, a local context of three words to the left and right of the ambiguous word, the parts-of-speech of the surrounding words, the verb and noun before and after the ambiguous words, and a global context implemented through sense-specific keywords determined as a list of words occurring at least three times in the contexts defining a certain word sense. We used TreeTagger for part-of-speech tagging[2] and Snowball stemmer[3] for stemming as they both have publicly available implementations for multiple languages. The features are integrated in a Naive Bayes classifier, which was selected for its state-of-the-art performance in previous WSD systems.

[2] www.cis.uni-muenchen.de/~schmid/tools/TreeTagger
[3] snowball.tartarus.org

# 3 The WikiTransSense System

Consider the examples centered around the ambiguous noun "chair", as shown in Figure 1, where English is the reference language and German is a supporting language. The figure shows only 2 out of the 5 possible meanings from the Wikipedia sense inventory. The two examples illustrate two important ways in which the translation can help disambiguation. First, two different senses of the target ambiguous word may be translated into a different word in the supporting language. Therefore, assuming access to word alignments, knowledge of the target word translation can help in disambiguation. Second, features extracted from the translated sentence can be used to enrich the feature space. Even though the target word translation is a strong feature in general, there may be cases where different senses of the target word are translated into the same word in the supporting language. For example, the two senses "bar (unit)" and "bar (establishment)" of the English word "bar" translate to the same German word "bar". In cases like this, words in the context of the German translation may help in identifying the correct English meaning.

## 3.1 A Multilingual Dataset through Machine Translation

In order to generate a multilingual representation for the monolingual dataset, we used Google Translate to translate the data from English into several other languages. The use of Google Translate is motivated by the fact that Google's statistical machine translation system is available for many languages. Furthermore, through the University Research Program, Google Translate also provides the word alignments. Given a target word in an English sentence, we used the word alignments to identify the position of the target word translation in the translated sentence. Each of the four languages is used as a reference language, with the remaining three used as supporting languages. Additionally, French was added as a supporting language in all the multilingual systems, which means that each reference sentence was translated in four supporting languages.

## 3.2 The WikiTransSense Learning Framework

Similar to the WIKIMONOSENSE approach described in Section 2.2, we extract the same types

| | An airline seat is a <u>chair</u> on an airliner in which passengers are accommodated for the duration of the journey. |
| | Ein Flugzeugsitz ist ein <u>Stuhl</u> auf einem Flugzeug, in dem Passagiere fr die Dauer der Reise untergebracht sind. |
| | For a year after graduation, Stanley served as <u>chair</u> of belles-lettres at Christian College in Hustonville. |
| | Seit einem Jahr nach dem Abschluss, diente Stanley als <u>Vorsitzender</u> Belletristik bei Christian College in Hustonville. |

Figure 1: English to German translations from Google Translate, with the target words aligned.

| Language | WikiTransSense | WikiMuSense |
|----------|----------------|-------------|
| English  | 75,832         | 13,151      |
| German   | 54,984         | 8,901       |
| Italian  | 81,468         | 4,697       |
| Spanish  | 48,384         | 6,560       |

Table 2: Total number of sentence translations per language, in the two multilingual approaches.

of features from the reference sentence, as well as from the translations in each of the supporting languages. Correspondingly, the feature vector will contain a section with the reference language features, followed by a multilingual section containing features extracted from the translations in the supporting languages. The resulting multilingual feature vectors are then used with a Naive Bayes classifier.

## 4 The WikiMuSense System

The number of sentence translations required to train the WIKITRANSSENSE approach is shown in the second column of Table 2. If one were to train a WSD system for all ambiguous nouns, the large number of translations required may be prohibitive. In order to reduce the dependency on the machine translation system, we developed a second multilingual approach to WSD, WIKIMUSENSE, that exploits the interlingual links available in Wikipedia.

### 4.1 A Multilingual Dataset through Interlingual Wikipedia Links

Wikipedia articles on the same topic in different languages are often connected through interlingual links. These are the small navigation links that show up in the "Languages" sidebar in most Wikipedia articles. For example, the English Wikipedia sense "Bar (music)" is connected through an interlingual link to the German Wikipedia sense "Takt (Musik)". Given a sense inventory for a word in the reference language, we automatically build the sense repository for a

supporting language by following the interlingual links connecting equivalent senses in the two languages. Thus, given the English sense repository for the word "bar" *EN* = {*bar (establishment), bar (landform), bar (law), bar (music)*}, the corresponding German sense repository will be *DE* = {*Bar (Lokal), noteank, NIL, Takt (Musik)*}[4]. The resulting sense repositories can then be used in conjunction with Wikipedia links to build sense tagged corpora in the supporting languages, using the approach described in Section 2.1. However, this approach poses the following two problems:

1. There may be reference language senses that do not have interlingual links to the supporting language. In the "bar" example above, the English sense *bar (law)* does not have an interlingual link to German.

2. The distribution of examples per sense in the automatically created sense tagged corpus for the supporting language may be different from the corresponding distribution for the reference language. Previous work (Agirre et al., 2000; Agirre and Martinez, 2004) has shown that the WSD performance is sensitive to differences in the two distributions.

We address the first problem using a very simple approach: whenever there is a sense gap, we randomly sample a number of examples for that sense in the reference language and use Google Translate to create examples in the supporting language. The third column in Table 2 shows the total number of sentence translations required by the WIKIMUSENSE system. As expected, due to the use of interlingual links, it is substantially smaller than the number of translations required in the WIKITRANSSENSE system.

To address the second problem, we use the distribution of reference language as the true distribution and calculate the number of examples to

---
[4]NIL stands for a missing corresponding sense in German.

be considered per sense from the supporting languages using the statistical method proposed in (Agirre and Martinez, 2004).

## 4.2 The WikiMuSense Learning Framework

Once the datasets in the supporting languages are created using the method above, we train a Naive Bayes classifier for each language (reference or supporting). Note that the classifiers built for the supporting languages will use the same senses/classes as the reference classifier, since the aim of using supporting language data is to disambiguate a word in the reference language. Thus, for the word "bar" in the example above, if English is reference and German is supporting, the Naive Bayes classifier for German will compute probabilities for the four English senses, even though it is trained and tested on German sentences.

For each classifier, the features are extracted using the same approach as in the WIKI-MONOSENSE system.

At test time, the reference sentence is translated into all four supporting languages using Google Translate. The five probabilistic outputs – one from the reference ($P_R$) and four from the supporting classifiers ($P_S$) – are combined into an overall disambiguation score using Equation 1 below. Finally, disambiguation is done by selecting the sense that obtains the maximum score.

$$P = P_R + \sum_S P_S * min(1, |D_S|/|D_R|) \quad (1)$$

In Equation 1, $D_R$ is the set of training examples in the reference language $R$, whereas $D_S$ is the set of training examples in a source language $S$. When the number of training examples in a supporting language is smaller than the number of examples in the reference language, the probabilistic output from the corresponding supporting classifier will have a weight smaller than 1 in the disambiguation score, and thus a smaller influence on the disambiguation output. In general, the influence of the supporting classifier will always be less than or equal with the influence of the reference classifier.

## 5 Experimental Evaluation

We ran 10-fold cross-validation experiments on the Wikipedia dataset [5], with all three systems: WIKIMONOSENSE (WMS), WIKITRANSSENSE

---

[5]The dataset is available from http://lit.csci.unt.edu.

| Language | MFS | WMS | WTS | WMuS |
|----------|-----|-----|-----|------|
| English | 62.2 | 78.9 | **81.9** | 81.3 |
| German | 69.5 | 81.2 | 84.6 | **85.6** |
| Italian | 66.0 | 81.8 | 84.0 | **84.7** |
| Spanish | 66.8 | 76.0 | 78.7 | **79.7** |

Table 3: WSD macro accuracies.

| Language | MFS | WMS | WTS | WMuS |
|----------|-----|-----|-----|------|
| English | 59.2 | 79.3 | 80.6 | **80.3** |
| German | 75.6 | 83.9 | 86.5 | **87.0** |
| Italian | 74.3 | 84.6 | 86.3 | **87.5** |
| Spanish | 72.6 | 79.8 | 81.1 | **82.7** |

Table 4: WSD micro accuracies.

(WTS), and WIKIMUSENSE (WMUS). For the WIKIMUSENSE system, since the gaps in the supporting language datasets are addressed using reference language translations, we enforced the constraint that a translation of the test example does not appear in the training data of the supporting language.

We used two different accuracy metrics to report the performance:

1. *macro accuracy*: an accuracy number was calculated separately for each ambiguous word. Macro accuracy was then computed as the average of these accuracy numbers.

2. *micro accuracy*: the system outputs for all ambiguous words were pooled together and the micro accuracy was computed as the percentage of instances that were disambiguated correctly.

Tables 3 and 4 show the micro and macro accuracies for the three systems. The tables also show the accuracy of a simple WSD baseline that selects the Most Frequent Sense (MFS).

Overall, the Wikipedia-based sense annotations were found reliable, leading to accurate sense classifiers for the WIKIMONOSENSE system with an average relative error reduction of 44%, 38%, 44%, and 28% compared to the most frequent sense baseline in terms of macro accuracy. WIKIMONOSENSE performed better for 76 out of the 105 words in the four languages compared to the MFS baseline, which further indicates that Wikipedia data can be useful for creating accurate and robust WSD systems.

Compared to the monolingual WIKIMONOSENSE system, the multilingual WIKITRANSSENSE system obtained an average relative error reduction of 13.7%, thus confirming the utility of using translated contexts. Relative to the MFS baseline, WIKITRANSSENSE performed better on 83 of the 105 words. Finally, WIKIMUSENSE had an even higher average error reduction of 16.5% with respect to WIKIMONOSENSE, demonstrating that the multilingual data available in Wikipedia can successfully replace the machine translation component during training. Relative to the MFS baseline, the multilingual WIKIMUSENSE system performed better on 89 out of the 105 words.

Since WIKIMUSENSE is still using machine translation when interlingual links are missing, we ran an additional experiment in which MT was completely removed during training to demonstrate the advantage of sense-annotated corpora available in supporting language Wikipedias. Thus, for the 105 ambiguous words, we eliminated all senses that required machine translation to fill the sense gaps. After filtering, 52 words from the four languages had 2 or more sense in Wikipedia for which all interlingual links were available. The results averaged over the 52 words are shown in Table 5 and demonstrate that WIKIMUSENSE still outperforms WIKIMONOSENSE substantially.

| Accuracy | WikiMonoSense | WikiMuSense |
| --- | --- | --- |
| Macro | 83.9 | **87.2** |
| Micro | 87.5 | **89.8** |

Table 5: WSD performance with no sense gaps.

We have also evaluated the proposed WSD systems in a coarse-grained setting on the same dataset. Two annotators were provided with the automatically extracted sense inventory from Wikipedia along with the corresponding Wikipedia articles and requested to discuss and create clusters of senses for the 105 words in the four languages. The results on this coarse-grained sense inventory are shown in Tables 6 and 7 indicate that our multilingual systems outperform the monolingual system.

## 5.1 Learning Curves

One aspect that is particularly relevant for any supervised system is the learning rate with respect to the amount of available data. To determine the

| | MFS | WMS | WTS | WMuS |
| --- | --- | --- | --- | --- |
| English | 72.9 | 87.3 | 88.9 | **89.9** |
| German | 72.8 | 84.1 | 87.8 | **87.9** |
| Italian | 71.7 | 87.6 | 89.4 | **90.0** |
| Spanish | 73.6 | 83.2 | 86.1 | **86.9** |

Table 6: Coarse grained macro accuracies.

| | MFS | WMS | WTS | WMuS |
| --- | --- | --- | --- | --- |
| English | 69.7 | 88.9 | 89.7 | **91.0** |
| German | 78.5 | 86.7 | **89.6** | 89.3 |
| Italian | 78.4 | 88.7 | 90.3 | **90.9** |
| Spanish | 79.8 | 87 | 88.7 | **90.0** |

Table 7: Coarse grained micro accuracies.

learning curve, we measured the disambiguation accuracy under the assumption that only a fraction of the data were available. We ran 10-fold cross-validation experiments using 10%, 20%, ..., 100% of the data, and averaged the results over all the words in the data set. The learning curves for the four languages are plotted in Figure 2. Overall, the curves indicates a continuously growing accuracy with increasingly larger amounts of data. Although the learning pace slows down after a certain number of examples (about 50% of the data currently available), the general trend of the curve seems to indicate that more data is likely to lead to increased accuracy. Given that Wikipedia is growing at a fast pace, the curve suggests that the accuracy of the word sense classifiers built on this data is likely to increase for future versions of Wikipedia.

Another relevant aspect is the dependency between the amount of data available in supporting languages and the performance of the WIKIMUSENSE system. To measure this, we ran 10-fold cross-validation experiments using all the data from the reference language and varying the amount of supporting language data from 10% to 100%, in all supporting languages. The accuracy results were averaged over all the words. Figure 3 shows the learning curves for the 4 languages. When using 0% fraction of supporting data, the results correspond to the monolingual WIKIMONOSENSE system. When using 100% fraction of the supporting data, the results correspond to the final multilingual WIKIMUSENSE system. We can see that WIKIMUSENSE starts to perform better than WIKIMONOSENSE when
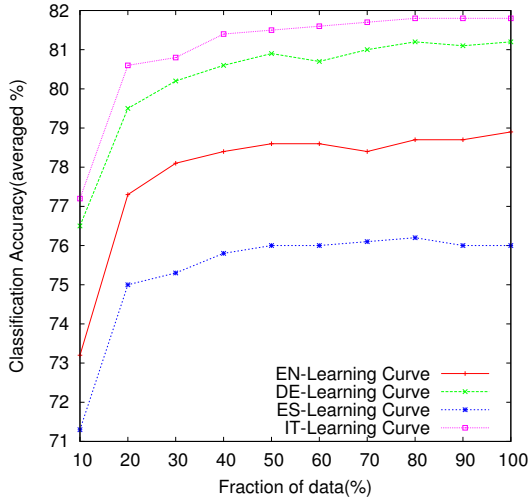
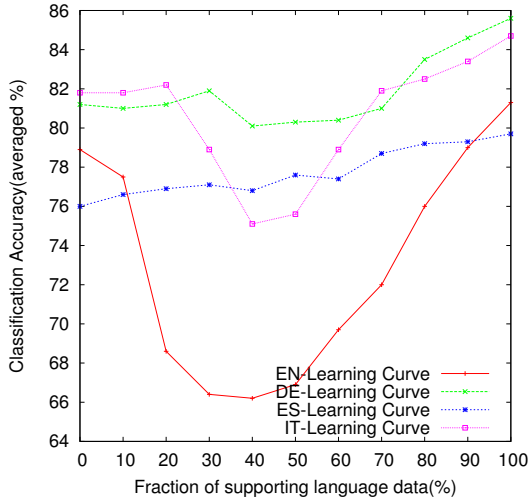Figure 2: Learning curves for WIKIMONOSENSE.



Figure 3: Learning curves for WIKIMUSENSE.

at least 70-80% of the available supporting data is used, and continues to increase its performance with increasing amounts of supporting data.

Finally, we also evaluated the impact that the number of supporting languages has on the performance of the two multilingual WSD systems. Both WIKITRANSSENSE and WIKIMUSENSE are evaluated using all possible combinations of 1, 2, 3, and 4 supporting languages. The resulting macro accuracy numbers are then averaged for each number of supporting languages. Figure 4 indicates that the accuracies continue to improve as more languages are added for both systems.

## 6    Related Work

Despite the large number of WSD methods that have been proposed so far, there are only a few methods that try to explore more than one lan-
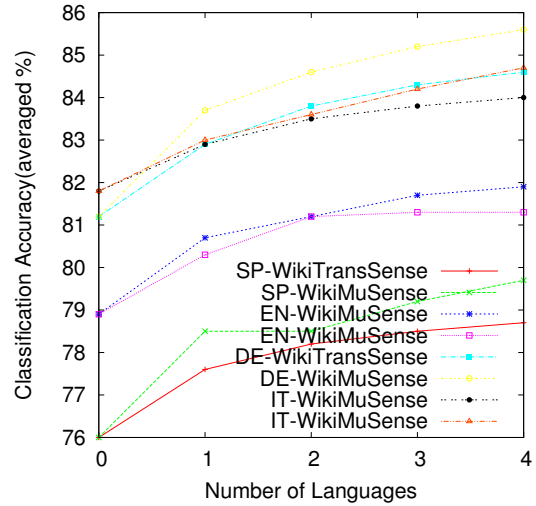


Figure 4: Impact of the number of supporting languages on the two multilingual WSD systems.

guage at a time.

Brown et al. (1991) made the observation that mappings between word-forms and senses may differ across languages and proposed a statistical machine learning technique that exploits these mappings for WSD. Subsequently, several works (Gale et al., 1992; Resnik and Yarowsky, 1999; Diab and Resnik, 2002; Diab, 2004; Ng et al., 2003; Chan and Ng, 2005; Chan et al., 2007) explored the use of parallel translations for WSD.

Li and Li (2004) introduced a bilingual bootstrapping approach, in which starting with in-domain corpora in two different languages, English and Chinese, word translations are automatically disambiguated using information iteratively drawn from the bilingual corpora. Khapra et al. (2009; 2010) proposed another bilingual bootstrapping approach, in which they used an aligned multilingual dictionary and bilingual corpora to show how resource deprived languages can benefit from a resource rich language. They introduced a technique called *parameter projections*, in which parameters learned using both aligned multilingual Wordnet and bilingual corpora are projected from one language to another language to improve on existing WSD methods.

In recent years, the exponential growth of the Web led to an increased interest in multilinguality. Lefever and Hoste (Lefever and Hoste, 2010) introduced a SemEval task on cross-lingual WSD in SemEval-2010 that received 16 submissions. The corresponding dataset contains a collection of sense annotated English sentences for a few words

with their contextually appropriate translations in Dutch, German, Italian, Spanish and French.

Recently, Banea and Mihalcea (2011) explored the utility of features drawn from multiple languages for WSD. In their approach, a multilingual parallel corpus in four languages (English, German, Spanish, and French) is generated using Google Translate. For each example sentence in the training and test set, features are drawn from multiple languages in order to generate more robust and more effective representations known as *multilingual vector-space representations*. Finally, training a multinomial Naive Bayes learner showed that a classifier based on multilingual vector representations obtains an error reduction ranging from 10.58% to 25.96% as compared to the monolingual classifiers. Lefever (2012) proposed a similar strategy for multilingual WSD using a different feature set and machine learning algorithms. Along similar lines, (Fernandez-Ordonez et al., 2012) used the Lesk algorithm for unsupervised WSD applied on definitions translated in four languages, and obtained significant improvements as compared to a monolingual application of the same algorithm. Although these three methodologies are closely related to our WIK-ITRANSSENSE system, our approach exploits a sense inventory and tagged sense data extracted automatically from Wikipedia.

Navigli and Ponzetto (2012) proposed a different approach to multilingual WSD based on *BabelNet* (2010), a large multilingual encylopedic dictionary built from WordNet and Wikipedia. Their approach exploits the graph structure of *BabelNet* to identify complementary sense evidence from translations in different languages.

## 7 Conclusion

In this paper, we described three approaches for WSD that exploit Wikipedia as a source of sense annotations. We built monolingual sense tagged corpora for four languages, using Wikipedia hyperlinks as sense annotations. Monolingual WSD systems were trained on these corpora and were shown to obtain relative error reductions between 28% and 44% with respect to the most frequent sense baseline, confirming that the Wikipedia sense annotations are reliable and can be used to construct accurate monolingual sense classifiers.

Next, we explored the cumulative impact of features originating from multiple supporting languages on the WSD performance of the reference language, via two multilingual approaches: WIK-ITRANSSENSE and WIKIMUSENSE. Through the WIKITRANSSENSE system, we showed how to effectively use a machine translation system to leverage two relevant multilingual aspects of the semantics of text. First, the various senses of a target word may be translated into different words, which constitute unique, yet highly salient signals that effectively expand the target words feature space. Second, the translated context words themselves embed co-occurrence information that a translation engine gathers from very large parallel corpora. When integrated in the WIKITRANSSENSE system, the two types of features led to an average error reduction of 13.7% compared to the monolingual system.

In order to reduce the reliance on the machine translation system during training, we explored the possibility of using the multilingual knowledge available in Wikipedia through its interlingual links. The resulting WIKIMUSENSE system obtained an average relative error reduction of 16.5% compared to the monolingual system, while requiring significantly fewer translations than the alternative WIKITRANSSENSE system.

## Acknowledgments

## References

E. Agirre and D. Martinez. 2004. Unsupervised word sense disambiguation based on automatically retrieved examples: The importance of bias. In *Proceedings of EMNLP 2004*, Barcelona, Spain, July.

E. Agirre, G. Rigau, L. Padro, and J. Asterias. 2000. Supervised and unsupervised lexical knowledge methods for word sense disambiguation. *Computers and the Humanities*, 34:103–108.

C. Banea and R. Mihalcea. 2011. Word sense disambiguation with multilingual features. In *International Conference on Semantic Computing*, Oxford, UK.

P. F Brown, S. A. Pietra, V. J. Pietra, and R. Mercer. 1991. Word-sense disambiguation using statistical methods. In *Proceedings of the 29th annual meeting on Association for Computational Linguistics*, pages 264–270. Association for Computational Linguistics.

Y.S. Chan and H.T. Ng. 2005. Scaling up word sense disambiguation via parallel texts. In *Proceedings of the 20th national conference on Artificial intelligence - Volume 3*, AAAI'05, pages 1037–1042.

Y.S. Chan, H.T. Ng, and D. Chiang. 2007. Word sense disambiguation improves statistical machine translation. In *Proceedings of the Association for Computational Linguistics*, Prague, Czech Republic.

B. Dandala, R. Mihalcea, and R. Bunescu. 2012. Word sense disambiguation using wikipedia. *The People's Web Meets NLP: Collaboratively Constructed Language Resources*.

M. Diab and P. Resnik. 2002. An unsupervised method for word sense tagging using parallel corpora. In *Proceedings of the 40st Annual Meeting of the Association for Computational Linguistics (ACL 2002)*, Philadelphia, PA, July.

M. Diab. 2004. Relieving the data acquisition bottleneck in word sense disambiguation. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL 2004)*, Barcelona, Spain, July.

E. Fernandez-Ordonez, R. Mihalcea, and S. Hassan. 2012. Unsupervised word sense disambiguation with multilingual representations. In *Proceedings of the Conference on Language Resources and Evaluations (LREC 2012)*, Istanbul, Turkey.

W. Gale, K. Church, and D. Yarowsky. 1992. A method for disambiguating word senses in a large corpus. *Computers and the humanities*, 26(5-6):415–439.

M. Galley and K. McKeown. 2003. Improving word sense disambiguation in lexical chaining. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI 2003)*, Acapulco, Mexico, August.

M. Khapra, S. Shah, P. Kedia, and P. Bhattacharyya. 2009. Projecting parameters for multilingual word sense disambiguation. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 459–467.

M. Khapra, S. Sohoney, A. Kulkarni, and P. Bhattacharyya. 2010. Value for money: balancing annotation effort, lexicon building and accuracy for multilingual wsd. In *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING '10, pages 555–563.

E. Lefever and V. Hoste. 2010. Semeval-2010 task 3: Cross-lingual word sense disambiguation. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 15–20. Association for Computational Linguistics.

E. Lefever. 2012. *ParaSense: parallel corpora for word sense disambiguation*. Ph.D. thesis, Ghent University.

M.E. Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the SIGDOC Conference 1986*, Toronto, June.

H. Li and C. Li. 2004. Word translation disambiguation using bilingual bootstrapping. *Computational Linguistics*, 30(1):1–22.

R. Mihalcea. 2007. Using Wikipedia for automatic word sense disambiguation. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*, Rochester, New York, April.

G. Miller. 1995. Wordnet: A lexical database. *Communication of the ACM*, 38(11).

R. Navigli and S. Ponzetto. 2010. BabelNet: Building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden.

R. Navigli and S. P. Ponzetto. 2012. Joining forces pays off: Multilingual joint word sense disambiguation. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1399–1410, Jeju Island, Korea, July.

R. Navigli and P. Velardi. 2005. Structural semantic interconnections: A knowledge-based approach to word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27.

H.T. Ng and H.B. Lee. 1996. Integrating multiple knowledge sources to disambiguate word sense: An examplar-based approach. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL 1996)*, Santa Cruz.

H.T. Ng, B. Wang, and Y.S. Chan. 2003. Exploiting parallel texts for word sense disambiguation: An empirical study. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003)*, Sapporo, Japan, July.

T. Pedersen. 2001. A decision tree of bigrams is an accurate predictor of word sense. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL 2001)*, pages 79–86, Pittsburgh, June.

P. Resnik and D. Yarowsky. 1999. Distinguishing systems and distinguishing senses: New evaluation methods for word sense disambiguation. *Natural Language Engineering*, 5(2):113–134.

D. Yarowsky. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL 1995)*, Cambridge, MA, June.