Word Sense Disambiguation for All Words Without Hard Labor

Zhi Zhong and **Hwee Tou Ng**
Department of Computer Science
National University of Singapore
13 Computing Drive, Singapore 117417
{zhongzhi, nght}@comp.nus.edu.sg

Abstract

While the most accurate word sense disambiguation systems are built using supervised learning from sense-tagged data, scaling them up to all words of a language has proved elusive, since preparing a sense-tagged corpus for all words of a language is time-consuming and human labor intensive.

In this paper, we propose and implement a completely automatic approach to scale up word sense disambiguation to all words of English. Our approach relies on English-Chinese parallel corpora, English-Chinese bilingual dictionaries, and automatic methods of finding synonyms of Chinese words. No additional human sense annotations or word translations are needed.

We conducted a large-scale empirical evaluation on more than 29,000 noun tokens in English texts annotated in OntoNotes 2.0, based on its coarse-grained sense inventory. The evaluation results show that our approach is able to achieve high accuracy, outperforming the first-sense baseline and coming close to a prior reported approach that requires manual human efforts to provide Chinese translations of English senses.

1 Introduction

Word sense disambiguation (WSD) is the task of identifying the correct meaning of a polysemous word in context. As a fundamental problem in natural language processing (NLP), WSD is important for applications such as machine translation and information retrieval. Previous SensEval competitions [Palmer *et al.*, 2001; Snyder and Palmer, 2004; Pradhan *et al.*, 2007] show that the supervised learning approach is the most successful approach to WSD. Current state-of-the-art WSD systems are based on supervised learning, and they require a lot of sense-annotated training examples to achieve good performance. However, sense-annotation is expensive and labor intensive. Among the existing sense-annotated corpora, the SEMCOR corpus [Miller *et al.*, 1994] is the most widely used. Content words in SEMCOR were manually annotated with WordNet senses. However, each word type has just 10 instances on average, so SEMCOR is too small to train a supervised WSD system for all words of English. The lack of sense-annotated data becomes the bottleneck of the supervised learning approach to WSD.

In recent years, researchers have tried to automate the sense annotation process. Many efforts have been involved in exploiting training data for WSD from multilingual resources, like parallel corpora [Resnik and Yarowsky, 1997; Diab and Resnik, 2002; Ng *et al.*, 2003; Chan and Ng, 2005; Wang and Carroll, 2005]. For example, different senses of an English word typically have distinct Chinese translations, so it is possible to identify the sense of an English word in context if we know its Chinese translation. In our previous work [Chan *et al.*, 2007], we take advantage of this observation to extract training examples from English-Chinese parallel corpora. Evaluation results show that when evaluated on the English all-words tasks of SemEval 2007, this approach is able to achieve state-of-the-art WSD accuracy higher than the WordNet first-sense baseline. However, the implemented approach still requires manual human efforts to select suitable Chinese translations for every sense of every English word, which is still a time-consuming process.

In this paper, we tackle this problem by making use of bilingual dictionaries and statistical information. The selection of Chinese translations is done without any additional manual human efforts. As such, the entire process of extracting training data for WSD from parallel corpora is *fully automatic* and *unsupervised*. We conducted a large-scale empirical evaluation on more than 29,000 noun tokens in English texts annotated in OntoNotes 2.0, based on its coarse-grained sense inventory. The evaluation results show that our approach is able to achieve high accuracy, outperforming the first-sense baseline and coming close to our prior reported approach that requires manual human efforts to provide Chinese translations of English senses.

The remainder of this paper is organized as follows. In Section 2, we give a brief description of the method of [Chan and Ng, 2005]. Section 3 describes the details of our approach to automatically select Chinese target translations. Section 4 briefly describes our WSD system. In Section 5, we evaluate our approach on all noun tokens in OntoNotes 2.0 English texts, and compare the results with those of manual translation assignment. Finally, we conclude in Section 6.

2 Training Data from Parallel Texts

In this section, we briefly describe the process of gathering training data from parallel texts proposed by [Chan and Ng, 2005].

2.1 Parallel Text Alignment

Parallel corpora	Size of English texts (million words (MB))	Size of Chinese texts (million chars (MB))
Hong Kong Hansards	39.9 (223.2)	35.4 (146.8)
Hong Kong News	16.8 (96.4)	15.3 (67.6)
Hong Kong Laws	9.9 (53.7)	9.2 (37.5)
Sinorama	3.8 (20.5)	3.3 (13.5)
Xinhua News	2.1 (11.9)	2.1 (8.9)
English translation of Chinese Treebank	0.1 (0.7)	0.1 (0.4)
Total	72.6 (406.4)	65.4 (274.7)

Table 1: Size of English-Chinese parallel corpora

Table 1 lists the 6 English-Chinese parallel corpora used in the experiment of [Chan and Ng, 2005]. These corpora were already aligned at sentence level. After tokenizing the English texts and performing word segmentation on the Chinese texts, the GIZA++ software [Och and Ney, 2000] was used to perform word alignment on the parallel texts.

2.2 Selection of Chinese Translations

In this step, Chinese translations C were manually selected for each sense s of an English word e. With WordNet [Miller, 1990] as the sense inventory, Chan and Ng [2005] manually assigned Chinese translations to the top 60% most frequently occurring noun types in the Brown corpus. From the word alignment output of GIZA++, the occurrences of an English word e which were aligned to one of the manually assigned Chinese translations c were selected. Since we know the sense s associated with a Chinese translation c, occurrences of the word e in the English side of the parallel corpora that are aligned to c will be assigned the sense s. These occurrences of e and their 3-sentence surrounding contexts were extracted as sense-annotated training data.

In this paper, the manually selected Chinese translations are those used in [Chan and Ng, 2005]. We also adopt the approach of [Chan and Ng, 2005] to extract training examples from the 6 parallel corpora listed above.

3 Automatic Selection of Chinese Translations

Compared to sense-annotating training examples directly, the human effort needed in the approach of [Chan and Ng, 2005] is relatively reduced. However, in WSD, different sense-annotated data are needed for different word types. Considering the huge number of word types in a language, manually assigning translations to the senses of words still needs a large amount of human effort. If we can find a completely automatic way to collect such translations in a second language for senses of a word, the whole process of extracting training examples from parallel texts for WSD will be completely unsupervised.

In this section, we propose several methods to find Chinese translations for English WordNet senses without any additional human effort, by making use of bilingual dictionaries and bilingual corpora.

3.1 Sinica Bilingual Ontological WordNet

WordNet is widely used as a sense inventory of English. Synsets are the foundations of WordNet. A WordNet synset is constructed with a set of synonyms and semantic pointers which describe its relationships with other synsets. Each sense of an English word has a unique corresponding synset.

Sinica Bilingual Ontological WordNet (BOW) [Huang *et al.*, 2004] integrates WordNet and two other resources, Suggested Upper Merged Ontology (SUMO) and the English-Chinese Translation Equivalents Database (ECTED). WordNet was manually mapped to SUMO and ECTED in BOW. With the integration of these three resources, BOW functions as an English-Chinese bilingual WordNet. That is, each WordNet synset has a set of corresponding Chinese translations in BOW. After carrying out some preprocessing, we extract 94,874 Chinese translations from BOW for all of the 66,025 WordNet noun synsets.

3.2 Extracting Chinese Translations from a Common English-Chinese Bilingual Dictionary

BOW provides Chinese translations for all WordNet synsets, but each noun synset has only 1.4 Chinese translations on average. As reported in our evaluation results, these Chinese translations available in BOW are not adequate for us to extract sufficient training examples from parallel texts. As such, we propose a method to extract more Chinese translations for WordNet synsets from a common English-Chinese bilingual dictionary, Kingsoft PowerWord 2003.[1]

PowerWord 2003 contains Chinese translations of English sense entries in the American Heritage Dictionary. For an English word sense, PowerWord lists both Chinese translations and English glosses. Because the sense definitions of PowerWord and WordNet are quite different and it is hard to map the English word senses in PowerWord to WordNet senses, the Chinese translations in PowerWord cannot be directly mapped to WordNet senses. Here we propose two ways to make use of the Chinese translations provided by PowerWord.

1. If two or more English synonyms in a WordNet synset syn share the same Chinese translation c in PowerWord, we assign c as a Chinese translation for synset syn.

   For example, in WordNet 1.6, synset "10969750.n", which means "a time interval during which there is a temporary cessation of something", has 5 synonyms: pause, intermission, break, interruption, and suspension. In PowerWord, *pause* and *suspension* have the same Chinese translation "中止"; *break*, *pause*, and *suspension* share the same Chinese translation "暂停". As such, "中止" and "暂停" are assigned as Chinese translations to synset "10969750.n".

[1] http://www.iciba.com/

2. Suppose an English word e is monosemous. Let syn be the WordNet synset corresponding to the only sense of e. Then all Chinese translations of e from PowerWord are assigned as the Chinese translations for synset syn.

   For example, in WordNet 1.6, synset "10382904.n", which means "a desirable state", has two synonyms: blessing and boon. Because the noun *boon* is monosemous in WordNet, all Chinese translations of *boon* "恩惠", "实惠", and "福利" in PowerWord are assigned to synset "10382904.n".

Via the above two ways, 52,599 Chinese translations are extracted from PowerWord for 29,066 out of 66,025 noun synsets. On average, each English synset has 1.8 Chinese translations.

So far, Chinese translations are gathered from both BOW and PowerWord for WordNet synsets. For each English word *e*, we can find the Chinese translations for its senses by referring to their corresponding synsets. Because WordNet senses are ordered such that a more frequent sense appears before a less frequent one, if several senses of *e* share an identical Chinese translation *c*, only the least numbered sense among these senses will have *c* assigned as a translation. In this way, a Chinese translation *c* is only assigned to one sense of a word *e*.

3.3 Shortening Chinese Translations

For an English word, some of its Chinese translations from dictionaries may have no occurrences in parallel texts aligned to this English word. In this case, no training examples can be extracted from parallel texts with such Chinese translations. For instance, the Chinese translation "尤指国家的税收" (especially referring to federal tax) extracted from dictionary for the second WordNet sense of *revenue* is not aligned to the English word *revenue* in parallel texts. As a result, no training examples for *revenue* will be extracted with this Chinese translation. But as a good Chinese definition for sense 2 of *revenue*, "尤指国家的税收" is supposed to contain some useful information related to *revenue*. In this subsection, we propose a method to make use of these Chinese translations by shortening them.

Suppose sense s of an English word e has a Chinese translation c from dictionary, and there are no occurrences of c aligned to e in parallel texts. For every such Chinese translation c, we first generate its longest prefix pre and longest suffix suf which happen to align to e in parallel texts. pre and suf, if found, are the possible shortened candidate translations of c that may be selected as translations of s. Among these shortened translation candidates, we further discard a candidate if it is a substring of any Chinese translations from dictionary for a different sense s' of e. The remaining translation candidates are then selected for use. Each chosen prefix or suffix of c is a Chinese translation of the sense s associated with c.

Using this method, we generate a shortened Chinese translation "税收" (tax) for "尤指国家的税收". Similarly, we also generate two shortened Chinese translations "价值观" (value concept) and "观念" (concept) for the Chinese translation "价值观念" (value concept), for sense 6 of the English noun *value*.

3.4 Adding More Chinese Translations Using Word Similarity Measure

Let *selected(e)* be the set of Chinese translations selected for an English word *e* (associated with any of its senses). With the previous methods, *selected(e)* contains Chinese translations from the dictionaries BOW and PowerWord, and their prefixes and suffixes. The occurrences of a Chinese translation *c* in parallel texts which are aligned to *e* will be extracted as training examples for *e* if and only if *c*∈*selected(e)*. Accordingly, if a Chinese translation *c*∉*selected(e)*, its occurrences in parallel texts that are aligned to e will be wasted.

So, in this subsection, we propose a method to assign Chinese translations which are not in *selected(e)*, but have occurrences aligned to e in parallel texts, to appropriate senses by measuring their similarities with Chinese translations in *selected(e)*. The assumption of this method is that two Chinese words are synonymous if they have the same translation and their distributional similarity is high.

We use the distributional similarity measure based on syntactic relations as described in [Lin, 1998] as our word similarity measure. Suppose (w, r, m) is a dependency triple extracted from a corpus parsed by a dependency parser, where r is the dependency relation, w is the head word, and m is the modifier together with its part-of-speech. $||w, r, m||$ denotes the frequency count of the dependency triple (w, r, m) in a parsed corpus. If w, r, or m is '*', the value will be the sum of frequency counts of all the dependency triples that match the rest of the expression. Define $I(w, r, m)$ as the amount of information contained in (w, r, m), whose value is

$$I(w, r, m) = \log \frac{||w, r, m|| \times ||*, r, *||}{||w, r, *|| \times ||*, r, m||}.$$

Let $T(w)$ be the set of pairs (r, m) such that $I(w, r, m)$ is positive. The similarity $sim(w_1, w_2)$ between two words w_1 and w_2 is calculated as

$$\frac{\sum_{(r,m)\in T(w_1)\cap T(w_2)} (I(w_1, r, m) + I(w_2, r, m))}{\sum_{(r,m)\in T(w_1)} I(w_1, r, m) + \sum_{(r,m)\in T(w_2)} I(w_2, r, m)} \quad (1)$$

We first train the Stanford parser [de Marneffe *et al.*, 2006] on Chinese Treebank 5.1 (LDC2005T01U01), and then parse the Chinese side of the 6 parallel corpora with the trained parser to output dependency parses.[2] We only consider the triples of subject relation, direct object relation, and modifying relation. Dependency triples whose head word's frequency is less than 10 are removed. From the parsed corpus, we extract a total of 13.5 million dependency triples. The similarity between two Chinese words is calculated using the above similarity measure on the set of 13.5 million dependency triples.

Suppose e is an English word, and c is a Chinese translation of e. Define $sense(c)$ as the sense of e that c is assigned to, and $count(c)$ as the number of occurrences of c aligned to e in the parallel corpora. The function avg calculates the average value of a set of values, and the function σ calculates the standard deviation of a set of values.

[2] Due to computational consideration, all sentences that are longer than 50 words are not included.

$\Phi \leftarrow$ the set of Chinese translations that are aligned to e in parallel texts but not in *selected(e)*
$count_{avg} \leftarrow avg(\{count(c) : c \in \Phi\})$
for each $c \in \Phi$
    if $count(c) < count_{avg}$
        $\Phi \leftarrow \Phi - \{c\}$
        continue
    end if
    $S[c] \leftarrow \max_{c' \in selected(e)} sim(c, c')$
    $C[c] \leftarrow argmax_{c' \in selected(e)} sim(c, c')$
end for
$threshold \leftarrow \min(avg(S) + \sigma(S), \theta)$
for each $c \in \Phi$
    if $S[c] \geq threshold$
        set c as a Chinese translation for $sense(C[c])$
    end if
end for

Figure 1: Assigning Chinese translations to English senses using word similarity measure.

Figure 1 shows the process in which we assign the set of Chinese translations Φ that are aligned to e in parallel texts but not selected as Chinese translation for e in our previous methods. Because most of the Chinese translations aligned to e with low frequency are erroneous in the word alignment output of GIZA++, in the first step, we eliminate the Chinese translations in Φ whose occurrence counts are below average. For each Chinese translation c remaining in Φ, we calculate its similarity scores with the Chinese translations in *selected(e)*. Suppose c_{max} is the Chinese translation in *selected(e)* which c is most similar to. We consider c as a candidate Chinese translation for the sense associated with c_{max}. To ensure that c is a Chinese synonym of c_{max}, we require that the similarity score between c and c_{max} should be high enough. A threshold $arg(S) + \sigma(S)$ is set to filter those candidates with low scores, where $arg(S) + \sigma(S)$ is the mean plus standard deviation of the scores of all candidates. To ensure that $arg(S) + \sigma(S)$ is not too high such that most of the candidates are filtered out, we set an upper bound θ for the threshold. In our experiment, θ is set to be 0.1. Finally, each candidate whose score is higher than or equal to the threshold will be assigned to the sense of its most similar Chinese translation.

4 The WSD System

We use the WSD system built with the supervised learning approach described in [Lee and Ng, 2002]. Individual classifiers are trained for all word types using the knowledge sources of local collocations, parts-of-speech (POS), and surrounding words.

We use 11 local collocations features: $C_{-1,-1}$, $C_{1,1}$, $C_{-2,-2}$, $C_{2,2}$, $C_{-2,-1}$, $C_{-1,1}$, $C_{1,2}$, $C_{-3,-1}$, $C_{-2,1}$, $C_{-1,2}$, and $C_{1,3}$, where $C_{i,j}$ refers to the ordered sequence of tokens in the local context of an ambiguous word *w*. Offsets *i* and *j* denote the starting and ending position (relative to *w*) of the sequence, where a negative (positive) offset refers to a token to its left (right). For parts-of-speech, 7 features are used: P_{-3}, P_{-2}, P_{-1}, P_0, P_1, P_2, P_3, where P_0 is the POS of *w*, and P_{-i} (P_i) is the POS of the *i*th token to the left (right) of *w*. We use all unigrams (single words) in the surrounding context of *w* as surrounding word features. Surrounding words can be in a different sentence from *w*. In this paper, SVM is used as our learning algorithm, which was shown to achieve good WSD performance in [Lee and Ng, 2002; Chan *et al.*, 2007].

5 Evaluation on OntoNotes

In this section, we evaluate some combinations of the above translation selection methods on all noun types in OntoNotes 2.0 data.

5.1 OntoNotes

The OntoNotes project [Hovy *et al.*, 2006] annotates coreference information, word senses, and some other semantic information on the Wall Street Journal (WSJ) portion of the Penn Treebank [Marcus *et al.*, 1993] and some other corpora, such as *ABC*, *CNN*, *VOA*, etc. In its second release (LDC2008T04) through the Linguistic Data Consortium (LDC), the project manually sense-annotated nearly 83,500 examples belonging to hundreds of noun and verb types, with an interannotator agreement rate of at least 90%, based on a coarse-grained sense inventory.

Noun Set	No. of noun types	Average no. of senses	No. of noun tokens
T60Set	257	4.3	22,353
All nouns	605	3.5	29,510

Table 2: Statistics of sense-annotated nouns in OntoNotes 2.0

As shown in Table 2, there are 605 noun types with 29,510 noun tokens in OntoNotes 2.0.[3] These nouns have 3.5 senses on average. Among the top 60% most frequent nouns with manually annotated Chinese translations from [Chan and Ng, 2005], 257 of them have sense-annotated examples in our test data set. We refer to this set of 257 nouns as *T60Set*. The nouns in this set have a higher average number of senses (4.3).

5.2 Quality of the Automatically Selected Chinese Translations

In this part, we manually check the quality of the Chinese translations generated by the methods described above.

In Section 3.2, Chinese translations are extracted from PowerWord for WordNet synsets in two ways. We randomly evaluate 100 synsets which get extended Chinese translations with the first way. 134 out of 158 (84.8%) extended Chinese translations in these 100 synsets are found to be good translations. Similarly, 100 synsets, which get extended Chinese translations from PowerWord with the second way, are

[3] We remove erroneous examples which are simply tagged with 'XXX' as sense-tag, or tagged with senses that are not found in the sense inventory provided. Also, since we map our training data from WordNet senses to OntoNotes senses, we remove examples tagged with OntoNotes senses which are not mapped to WordNet senses. On the whole, about 7.6% of the original OntoNotes noun examples are removed as a result.

randomly selected for evaluation. 214 out of 261 (82.0%) extended Chinese translations in these synsets are good.

Chinese translations from dictionaries are shortened with the method described in Section 3.3. We randomly evaluate 50 such Chinese translations, and find that 70% (35/50) of these shortened Chinese translations are appropriate.

In Section 3.4, we extend the Chinese translations of each English word by finding Chinese synonyms. 329 Chinese synonyms of 100 randomly selected English words which get Chinese translations in this method are manually evaluated. About 77.8% (256/329) of them are found to be good Chinese translations.

We also manually evaluate 500 randomly selected sense-tagged instances from parallel texts for 50 word types (10 instances for each word type). The accuracy of these sample instances is 80.4% (402/500).

5.3 Evaluation

In the experiment, training examples with WordNet senses are mapped to OntoNotes senses. One of our baselines is strategy "WNs1". It always assigns the OntoNotes sense which is mapped to the first sense in WordNet as the answer to each noun token. As mentioned previously, SEMCOR is the most widely used sense-annotated corpus. We use the strategy "SC", which uses only the SEMCOR examples as training data, as a baseline of supervised systems. In the following strategies, the SEMCOR examples are merged with a maximum of 1,000 examples gathered from parallel texts for each noun type:

- strategy "SC+BOW" uses Chinese translations from BOW to extract examples from parallel texts for all noun types;
- strategy "SC+Dict" uses the Chinese translations from both BOW and PowerWord;
- strategy "SC+Dict+Sht" applies the method described in Section 3.3 to extend the Chinese translations in strategy "SC+Dict";
- strategy "SC+Dict+Sht+Sim" extends the Chinese translations in strategy "SC+Dict+Sht" using the method described in Section 3.4;
- strategy "SC+Manu" only extracts training examples from parallel texts for the noun types in *T60Set* with their manually annotated Chinese translations.

For each noun type, the examples from the parallel corpora are randomly chosen according to the sense distribution of that noun in SEMCOR corpus. When we use the Chinese translations automatically selected to gather training examples from parallel texts, we prefer the examples related to the Chinese translations from dictionary BOW and PowerWord. If a word type has no training data, a random OntoNotes sense will be selected as the answer.

Table 3 shows the WSD accuracies of different strategies on *T60Set* and all of the nouns in OntoNotes 2.0. Comparing to WNs1 baseline, all the strategies using training examples from parallel texts achieve higher or comparable accuracies on both *T60Set* and all noun types. In Table 4, we list the error reduction rate of the supervised learning strategies comparing to the supervised baseline strategy "SC".

Strategy	Evaluation Set	
	T60Set	All nouns
SC+Manu	80.3%	77.0%
SC+Dict+Sht+Sim	77.7%	75.4%
SC+Dict+Sht	77.1%	74.9%
SC+Dict	76.7%	74.3%
SC+BOW	76.2%	73.7%
SC	73.9%	72.2%
WNs1	76.2%	73.5%

Table 3: WSD accuracy on OntoNotes 2.0

Strategy	Evaluation Set	
	T60Set	All nouns
SC+Manu	24.5%	17.3%
SC+Dict+Sht+Sim	14.6%	11.5%
SC+Dict+Sht	12.3%	9.7%
SC+Dict	10.7%	7.6%
SC+BOW	8.8%	5.4%

Table 4: Error reduction comparing to *SC* baseline

Comparing to the supervised baseline "SC", our approach "SC+Dict+Sht+Sim" achieves an improvement in accuracy of 3.8% for *T60Set* and 3.2% for *All nouns*. That is, our *completely automatic* approach is able to obtain more than half (59%) of the improvement obtained using the manual translation assignment approach of "SC+Manu" for *T60Set*, and 67% of the improvement for *All nouns*.

5.4 Significance Test

We conducted one-tailed paired t-test with a significance level $p = 0.01$ to see whether one strategy is statistically significantly better than another. The t statistic of the difference between each test example pair is computed.

The significance test results on all noun types in OntoNotes 2.0 are as follow:

SC+Manu $>$ SC+Dict+Sht+Sim
$>$ SC+Dict+Sht
$>$ SC+Dict
$>$ SC+BOW $\sim$ WNs1
$>$ SC

The significance tests on the *T60Set* have similar results. So we will discuss the significance test results without differentiating these two sets of noun types. In each step where we extend the automatic Chinese translation selection, a significant improvement is achieved in the WSD accuracy.

The "WNs1" baseline is only significantly better than strategy "SC". It is comparable to strategy "SC+BOW" but significantly worse than the other strategies. Strategy "SC+Manu" is significantly better than all other strategies.

6 Conclusion

The bottleneck of current supervised WSD systems is the lack of sense-annotated data. In this paper, we extend [Chan and Ng, 2005]'s method by automatically selecting Chinese translations for English senses. With our approach, the process of

extracting sense-annotated examples from parallel texts becomes completely unsupervised. Evaluation on a large number of noun types in OntoNotes 2.0 data shows that the training examples gathered with our approach are of high quality, and results in statistically significant improvement in WSD accuracy.

References

[Chan and Ng, 2005] Yee Seng Chan and Hwee Tou Ng. Scaling up word sense disambiguation via parallel texts. In *Proceedings of AAAI05*, pages 1037–1042, 2005.

[Chan *et al.*, 2007] Yee Seng Chan, Hwee Tou Ng, and Zhi Zhong. NUS-PT: Exploiting parallel texts for word sense disambiguation in the English all-words tasks. In *Proceedings of SemEval-2007*, pages 253–256, 2007.

[de Marneffe *et al.*, 2006] Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC06*, pages 449–454, 2006.

[Diab and Resnik, 2002] Mona Diab and Philip Resnik. An unsupervised method for word sense tagging using parallel corpora. In *Proceedings of ACL02*, pages 255–262, 2002.

[Hovy *et al.*, 2006] Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. OntoNotes: The 90% solution. In *Proceedings of HLT-NAACL06*, pages 57–60, 2006.

[Huang *et al.*, 2004] Chu-Ren Huang, Ru-Yng Chang, and Hsiang-Pin Lee. Sinica BOW (bilingual ontological wordnet): Integration of bilingual WordNet and SUMO. In *Proceedings of LREC04*, pages 1553–1556, 2004.

[Lee and Ng, 2002] Yoong Keok Lee and Hwee Tou Ng. An empirical evaluation of knowledge sources and learning algorithms for word sense disambiguation. In *Proceedings of EMNLP02*, pages 41–48, 2002.

[Lin, 1998] Dekang Lin. Automatic retrieval and clustering of similar words. In *Proceedings of ACL98*, pages 768–774, 1998.

[Marcus *et al.*, 1993] Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. Building a large annotated corpus of english: The Penn Treebank. *Computational Linguistics*, 19(2):313–330, 1993.

[Miller *et al.*, 1994] George A. Miller, Martin Chodorow, Shari Landes, Claudia Leacock, and Robert G. Thomas. Using a semantic concordance for sense identification. In *Proceedings of ARPA Human Language Technology Workshop*, pages 240–243, 1994.

[Miller, 1990] George A. Miller. WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–312, 1990.

[Ng *et al.*, 2003] Hwee Tou Ng, Bin Wang, and Yee Seng Chan. Exploiting parallel texts for word sense disambiguation: An empirical study. In *Proceedings of ACL03*, pages 455–462, 2003.

[Och and Ney, 2000] Franz Josef Och and Hermann Ney. Improved statistical alignment models. In *Proceedings of ACL00*, pages 440–447, 2000.

[Palmer *et al.*, 2001] Martha Palmer, Christiane Fellbaum, Scott Cotton, Lauren Delfs, and Hoa Trang Dang. English tasks: All-words and verb lexical sample. In *Proceedings of SENSEVAL-2*, pages 21–24, 2001.

[Pradhan *et al.*, 2007] Semeer S. Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. SemEval-2007 task-17: English lexical sample, SRL and all words. In *Proceedings of SemEval-2007*, pages 87–92, 2007.

[Resnik and Yarowsky, 1997] Philip Resnik and David Yarowsky. A perspective on word sense disambiguation methods and their evaluation. In *Proceedings of SIGLEX97*, pages 79–86, 1997.

[Snyder and Palmer, 2004] Benjamin Snyder and Martha Palmer. The English all-words task. In *Proceedings of SENSEVAL-3*, pages 41–43, 2004.

[Wang and Carroll, 2005] Xinglong Wang and John Carroll. Word sense disambiguation using sense examples automatically acquired from a second language. In *Proceedings of HLT-EMNLP05*, pages 547–554, 2005.