An Unsupervised Method for Word Sense Tagging using Parallel Corpora

Mona Diab and Philip Resnik
Department of Linguistics and Institute for Advanced Computer Studies
University of Maryland, College Park, MD 20742
{mdiab,resnik}@umiacs.umd.edu

Abstract

We present an unsupervised method for word sense disambiguation that exploits translation correspondences in parallel corpora. The technique takes advantage of the fact that cross-language lexicalizations of the same concept tend to be consistent, preserving some core element of its semantics, and yet also variable, reflecting differing translator preferences and the influence of context. Working with parallel corpora introduces an extra complication for evaluation, since it is difficult to find a corpus that is both sense tagged and parallel with another language; therefore we use pseudo-translations, created by machine translation systems, in order to make possible the evaluation of the approach against a standard test set. The results demonstrate that word-level translation correspondences are a valuable source of information for sense disambiguation.

1 Introduction

Word sense disambiguation (WSD) has long been a central question in computational linguistics, and in recent years the literature has seen a large number of advances as a result of three main factors: an increased attention to machine learning techniques, widespread dissemination of sense inventories (especially WordNet, Fellbaum 1998), and availability of large corpora and the means to do broad-coverage identification of relevant linguistic features in them.

On average, supervised methods (Bruce and Wiebe, 1994; Lin, 1999; Yarowsky, 1993), which learn from correctly sense tagged corpora, yield better performance results (Kilgarriff and Rosenzweig, 2000); however, they are highly tuned to the training corpus, and they need large quantities of high quality annotated data to produce reliable results. Unfortunately, large sense annotated corpora are expensive and labor intensive to create, and the data acquisition bottleneck is particularly severe when moving to less studied languages and genres. Unsupervised methods (e.g., (Agirre et al., 2000; Litkowski, 2000; Lin, 2000; Resnik, 1997; Yarowsky, 1992; Yarowsky, 1995)) have the advantage of making fewer assumptions about availability of data, but they generally tend to perform less well.

Parallel corpora present a new opportunity for combining the advantages of the two approaches, as well as an opportunity for exploiting translation correspondences in the text. In this paper, we present an unsupervised approach that utilizes parallel corpora for word sense tagging. We investigate the feasibility of automatically sense annotating (tagging) large amounts of data in parallel corpora using an unsupervised algorithm, making use of two languages simultaneously, only one of which has an available sense inventory. The method aims at achieving two main goals: first, producing large quantities of reasonably (if not perfectly) sense-annotated data for the language with the sense inventory, in order to bootstrap supervised learning techniques without the need for manual annotation;

second, achieving sense tagging using that same sense inventory for the second language, thus creating a sense-tagged corpus and automatically making a connection to the first language's sense inventory. In this paper we focus primarily on the first goal.

The crux of this research is the observation that translations can serve as a source of sense distinctions (Brown et al., 1991; Dagan, 1991; Dagan and Itai, 1994; Dyvik, 1998; Ide, 2000; Resnik and Yarowsky, 1999). A word that has multiple senses in one language is often translated as distinct words in another language, with the particular choice depending on the translator and the contextualized meaning; thus the corresponding translation can be thought of as a sense indicator for the instance of the word in its context. Looking at parallel translations, it becomes evident that two factors are at play. On the one hand, instances of a word/sense combination are translated with some consistency into a relatively small handful of words in the second language. On the other hand, that handful of words is rarely a singleton set even for a single word/sense, because the preferences of different translators and the demands of context produce semantically similar words that differ in their nuances.

For example, in a French-English parallel corpus, the French word *catastrophe* could be found in correspondence to English *disaster* in one instance, and to *tragedy* in another. Each of those English words is itself ambiguous — e.g., *tragedy* can refer to a kind of play (as opposed to *comedy*) — but we can take advantage of the fact that both English word instances appeared in correspondence with *catastrophe* to infer that they share some common element of meaning, and we can use that inference in deciding which of the English senses was intended. Having done so, we can go further: we can project the English word sense chosen for this instance of tragedy to the French word *catastrophe* in this context, thus tagging the two languages in tandem with a single sense inventory.

The remainder of this paper is organized as follows. Section 2 describes the approach. Section 3 lays out evaluation experiments, using SENSEVAL-2 data, showing the results of several different variations of the approach and comparing performance with other SENSEVAL-2 systems. Section 4 contains discussion and we conclude in Section 5.

2 Approach

For the sake of exposition, let us assume that we are working with an English-French parallel corpus and that we are using an English sense inventory.[1] Although there is no necessary assumption of directionality in translation, we will sometimes refer to the English language corpus as the target corpus and the French corpus as the source corpus, which corresponds to the characterization, in the previous section, of the French word (*catastrophe*) being translated into two different words (*disaster* and *tragedy*) in two diferent contexts. The process we described can be viewed more abstractly as follows:

1. Identify words in the target (English) corpus and their corresponding translations in the source (French) corpus.
2. Group the words of the target language — forming target sets — that were translated into the same orthographic form in the source corpus.
3. Within each of these target sets, consider all the possible sense tags for each word and select sense tags informed by semantic similarity with the other words in the group.
4. Project the sense tags from the target side to the source side of the parallel corpus.

The first step of the process assumes a sentence- or segment-aligned parallel corpus; suitable data are now available for many languages via organizations such LDC and ELRA and the Web is a promising source of data in new language pairs and in new genres (Nie et al., 1999; Resnik, 1999a). After identifying and tokenizing sentences, we obtain word-level alignments for the parallel corpus using the GIZA++

[1]The method has little dependence on language; in our evaluation section we report on work using English-French and English-Spanish.

implementation of the IBM statistical MT models (Och and Ney, 2000). For each French word instance f, we collect the word instance e with which it is aligned. Positions of the word instances are recorded so that in later stages we can project the eventual semantic annotation on e to f. For example, the alignment of *The accident was a tragedy* with *L'accident était une catastrophe* might associate these two instances of *catastrophe* and *tragedy*.

In the second step, we collect for each word type F the set of *all* English word types with which it is aligned anywhere in the corpus, which we call the *target set* for F. For example, the target set for French *catastrophe* might contain English word types *disaster*, *tragedy*, and *situation*, the last of these arising because some translator chose to render *la catastrophe* in English as *the awful situation*. In extracting correspondences we take advantage of WordNet to identify English nominal compounds in order to help reduce the number of ambiguous terms in the target set.[2] For example, without nominal compound identification on the English side, the target set for French *abeille* will contain *bee*, which is ambiguous (SPELLING-BEE vs. INSECT). With compound identification, the target set for *abeille* still contains *bee*, but it is also rich in unambiguous terms like *alkali_bee*, *honey_bee*, and *queen_bee*. In the semantic similarity computation, the presence of these monosemous words provides strong reinforcement for the INSECT sense of *bee*. Moreover, it enables us to tag instances of *bee* with their more specific compound-noun senses when they appear within a compound that is known to the sense inventory.

In the third step, the target set is treated as a problem of monolingual sense disambiguation with respect to the target-language sense inventory. Consider the target set {*disaster*, *tragedy*, *situation*}: to the human reader, the juxtaposition of these words within a single set automatically brings certain senses to the foreground. The same intuitive idea is exploited by Resnik's (1999b) algorithm for disambiguating groups of related nouns, which we apply here. For a target set $\{e_1, \ldots, e_n\}$, the algorithm considers each pair of words $\langle e_i, e_j \rangle (j \neq i)$, and identifies which senses of the two words are most similar semantically. Those senses are then reinforced by an amount corresponding to that degree of similarity.[3] After comparison across all pairs, each word sense $s_{i,k}$ of word e_i ends up having associated with it a confidence $c(s_{i,k}) \in [0, 1]$ that reflects how much reinforcement sense $s_{i,k}$ received based on the other words in the set. In our example, the KIND-OF-DRAMA sense of *tragedy* would have received little support from the senses of the other two words in the set; on the other hand, the CALAMITY sense would have been reinforced and therefore would receive higher confidence.

At the end of the third step, we highlight the significance of variability in translation: since the method relies on semantic similarities between multiple items in a target set, the target set must contain at least two members. If throughout the parallel corpus the translator *always* chose to translate the French word *catastrophe* to *tragedy*, the target set for *catastrophe* will contain only a single element. Our algorithm will have no basis for assigning reinforcement differently to different senses, and as a result, none of these instances of *tragedy* — the ones corresponding to *catastrophe* — will be tagged.

At this point we take advantage of the bookkeeping information recorded earlier. We know which instances of *tragedy* are associated with the target set {*disaster*, *tragedy*, *situation*}, and so those instances can be labeled with the most confident sense (CALAMITY) — or, for that matter, with the confidence distribution over all possible senses as determined by the noun-group disambiguation algorithm.

In the fourth and final step, we take advantage of the English-side tagging and the word-level alignment to project the sense tags on

[2] We used a small set of compound-matching rules considering a window of two tokens to the right and left, and also used the "satellite" annotations in SENSEVAL data as part of our preprocessing.

[3] Since we use WordNet as our sense inventory, we also adopt the information-theoretic measure of semantic similarity based on that taxonomy.

English to the corresponding words in French. For example, the tagging *The accident was a tragedy*/CALAMITY would yield *L'accident était une catastrophe*/CALAMITY. As a result, a large number of French words will receive tags from the English sense inventory.

Corpus	Tokens	Lines
BC-SV1	2498405	101841
SV2-AW	5815	242
SV2-LS	1760522	74552
WSJ	1290297	49679
Total	5555039	226314

Table 1: Sizes of corpora used in experiments

3 Evaluation

In order to provide a useful formal evaluation of this approach for English sense disambiguation, there were three requirements. We needed:

- a parallel corpus with English on one side, large enough to train stochastic translation models,
- gold-standard sense tags on the English side for some subset of the corpus,
- performance figures for other systems on the same subset, in order to compare results.

Meeting all three requirements simultaneously presented something of a challenge. There are a few human-tagged English corpora available for word sense disambiguation, but most are relatively small by model-training standards and none have associated translations in other languages. Conversely, there are some parallel corpora large enough for training alignment models, but to our knowledge none of these have been even partially sense tagged.

3.1 Corpora and Sense Inventory

To solve this problem, we adopted a "pseudo-translation" approach (Diab, 2000). A suitably large English corpus is constructed, containing as a subset an English corpus for which we have an existing set of associated gold-standard sense tags. The entire corpus, including the subset, is translated using commercial MT technology, producing an artificial parallel corpus. This corpus is then used as described in Section 2, and the quality of sense tagging on the English gold-standard subset is assessed using community-wide evaluation standards, with results suitable for inter-system comparison with other algorithms that have been tested on the same data.

The pseudo-translation approach has advantages and disadvantages. On the one hand, using commercial MT systems does not necessarily result in performance figures representing what could be obtained with better quality human translations. On the other hand, a pseudo-translated corpus is far easier to produce, and this approach to evaluation allows for controlled experimentation using English paired with multiple languages.

We used the the English "all words" portion of the SENSEVAL-2 test data (henceforth SV2-AW) as our gold-standard English subset. The corpus comprises three documents from the Wall Street Journal, totaling 242 lines with 5826 tokens in all. To fill out this English-side corpus, we added the raw unannotated texts of the Brown Corpus (BC) (Francis and Kučera, 1982), the SENSEVAL-1 corpus (SV1), the SENSEVAL-2 English Lexical Sample test, trial and training corpora (SV2-LS), and Wall Street Journal (WSJ) sections 18-24 from the Penn Treebank. We will refer to this unwieldy merged corpus with the unwieldy but informative label BCSV1SV2WSJ. Table 1 shows the sizes of the component corpora.

Two different commercially available MT systems were used for the pseudo-translations: Globalink Pro 6.4 (GL) and Systran Professional Premium (SYS). The motivation behind using two MT systems stems from a desire to more closely approximate the variability of human translation in a very large corpus, where one translator would be unlikely to have performed the entire task, and to help offset the possible tendency of any single MT system to be unnaturally consistent in its lexical selection.

The English BCSV1SV2WSJ was translated into French and Spanish, resulting in four parallel corpora: BCSV1SV2WSJ paired with the French GL translation (yielding parallel corpus FRGL), with French SYS translation (FRSYS), with Spanish GL (SPGL), and with Spanish SYS (SPSYS).[4]

Each of the four parallel corpora just described (FRGL, FRSYS, SPGL, SPSYS) represents a separate experimental variant. Consistent with Diab (2000), we added one more variant for each language in order to more closely approach the variability associated with multiple translations: in Step 2 we combined the target sets from the two MT systems. For example, if the word types *shore*, *bank* are in the target set of *orilla* in SPGL, and *coast*, *bank*, and *shore* are in the target set for *orilla* in SPSYS, the union of the target sets is taken and the result is a merged target set for *orilla* containing {*bank*, *coast*, *shore*}. These last two variations are labeled MFRGLSYS and MSPGLSYS.

We restricted our experiments to disambiguation of nouns, for which there were 1071 instances in SV2-AW not marked "unassignable" by SENSEVAL's human annotators. Nouns were identified on the basis of human-assigned part-of-speech tags where available (BC, WSJ and SV2-AW) and using the Brill tagger elsewhere (Brill, 1993). The choice of SV2-AW as our gold standard corpus determined our choice of sense inventory: SENSEVAL-2 produced a gold standard for the English "all words" task using a pre-release of WordNet 1.7 (Fellbaum, 1998), and we restricted our attention to the noun taxonomy.

3.2 Sense Selection Criterion

Because the algorithm for disambiguating noun groupings returns a confidence value for every sense of a word, some threshold or other criterion is needed to decide which sense or senses to actually assign. We simply assign the sense tag that scored the maximum confidence level, or all such tags, equally weighted, if there is a tie. (The SENSEVAL evaluation measures allow for partial credit.)

Variant	Precision	Recall
FRGL	58.1	50.9
FRSYS	58.0	49.0
MFRGLSYS	59.4	54.5
SPGL	57.9	48.6
SPSYS	60.0	51.5
MSPGLSYS	59.4	53.3

Table 2: Results on SENSEVAL-2 nouns (%)

This criterion is fairly sensitive to noise in target sets; for example, in a real corpus the French *catastrophe* is aligned with English {*catastrophe*, *disaster*, *shocker*, *tragedy*}. *Shocker* is an outlier in this set and its presence affects the overall confidence score assignment for all the words in the set. We observed that this is similar to what happens when the French word underlying the target set is homonymous; such cases are part of our discussion in Section 4.

3.3 Results

We evaluated the algorithm's performance using the standard SENSEVAL-2 evaluation software, obtaining figures for precision and recall for sense tagging the nouns in our gold standard. In this evaluation, partial credit is given in cases where a system assigns multiple sense tags.[5] We report results using the "fine-grained" scoring variant; this is the strictest variant, which sometimes requires systems to discern among WordNet senses that even linguists have a difficult time distinguishing.

Table 2 summarizes the results, and Figure 1 shows our algorithm's results (triangles) compared to the performance of the 21 SENSEVAL-2 English All Words participants, when the evaluation is restricted to the same set of noun test instances.[6] Hollow circles represent supervised

[4]The choice of languages was partly a question of available software for reasonably high quality translation, and partly motivated by the longer-term aim of performing evaluation of sense tags propagated back into the source languages via comparison with EuroWordNet.

[5]The *scorer2* program, disseminated by Rada Mihalcea in conjunction with the SENSEVAL-2 exercise, implements a version of Melamed and Resnik's (2000) framework for tagger evaluation given hierarchical tag sets. For discussion see Kilgarriff and Rosenzweig (2000).

[6]We computed results for other systems on our only-nouns subset of the task by subsetting those systems'

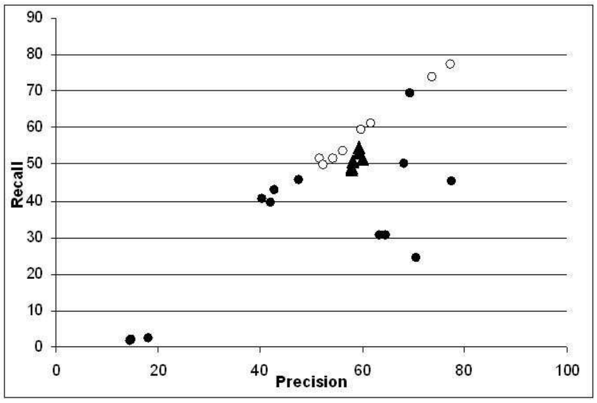

Figure 1: Comparison with SENSEVAL-2 systems

systems and filled circles represent unsupervised systems.[7] Of the systems that are unsupervised, and can therefore be included in a fair comparison, only one is clearly better on both precision and recall.

4 Discussion

The results show that the performance of our approach is comparable or superior to most other unsupervised systems, even though it is based on cross-language lexical correspondences, a radically different source of evidence, and even though those correspondences were derived from machine translations rather than clean human translations. Here we briefly consider issues that bear on recall and precision, respectively.

4.1 Considerations Affecting Recall

Some of the sentences in the test corpus could not be automatically aligned because our aligner discards sentence pairs that are longer than a pre-defined limit. For these sentences, therefore, no attempt could be made at disambiguation. Future experiments will attempt to increase the acceptable sentence length, as limited by real memory, and to break longer sentence pairs into logical sub-parts for alignment.

A second issue that affects recall is the lack of variability in pseudo-translations. Of the English nouns that are aligned with source-language words, approximately 35% are always aligned with the *same* word, rendering them untaggable using an approach based on semantic similarity within target sets. Some cases may reflect preserved ambiguity in the language pair — e.g. French *intérêt* and English *interest* are ambiguous in similar ways — and others may simply reflect the fact that commercial MT systems are just not very creative or context sensitive in their lexical choices. It should be possible to increase variability by extending the corpus to include human-translated parallel text, or by combining evidence from multiple or more distantly related source languages in the spirit of Resnik and Yarowsky (1999).

4.2 Considerations Affecting Precision

On inspecting the target sets qualitatively, we find that they contain many outliers, largely owing to noisy alignment. The problem worsens when the outliers are monosemous, since a monosemous word with a misleading sense will erroneously bias the sense tag assignment for the other target set words. For example, the word types *adolescence*, *idol*, *teen*, and *teenager* form a target set for the French source word *adolescence*, and the presence of *idol* has a negative impact on the sense assignment for the other members of the set. In addition, semantically distant words can align with the same source word; e.g., *amorce* in French may align with *initiation*, *bait*, and *cap*, which are all correct translations in suitable contexts but provide no suitable basis for semantic reinforcement.

These problems reflect the algorithm's implicit assumption that the source words are monosemous, reflected in its attempt to have every word in a target set influence the semantics of every other word. Inspecting the data pro-

answers from the full all-words task, which are available at http://www.cis.upenn.edu/~cotton/senseval/answers+misc.tgz. A scatterplot of the results for the 21 systems including all parts of speech appears at http://www.sle.sharp.co.uk/senseval2/Results/all_graphs.htm.

[7]We classified systems based on their descriptions at http://www.cogs.susx.ac.uk/lab/nlp/mccarthy/SEVALsystems.html. Per the SENSEVAL reporting guidelines, we do not identify individual systems.

duces many counterexamples, e.g. French *canon* (*cannon*, *cannonball*, *canon*, *theologian*) *bandes* (*band*, *gang*, *mob*, *strip*, *streak*, *tape*), and *baie* (*bay*, *berry*, *cove*).

A sensible alternative would be apply automatic clustering techniques to the target sets (e.g. (Diab and Finch, 2000; Schütze, 1992)), providing target sub-clusters of words that should be treated as related, with no cross-cluster reinforcement. For example, the target set for French *canon* would have two coherent sub-clusters containing {*cannon*, *cannonball*} and {*canon*, *theologian*)}, respectively. Manual inspection of target sets in our experiments suggests that when target sets are semantically coherent — e.g. *adversaires* (*antagonists*, *opponents*, *contestants*), *accident*: (*accident*, *crash*, *wreck*) — sense assignment is generally highly accurate.

5 Conclusions

This paper presents an unsupervised approach to word sense disambiguation that exploits translations as a proxy for semantic annotation across languages. The observation behind the approach, that words having the same translation often share some dimension of meaning, leads to an algorithm in which the correct sense of a word is reinforced by the semantic similarity of other words with which it shares those dimensions of meaning.

Performance using this algorithm has been rigorously evaluated and is comparable with other unsupervised WSD systems, based on fair comparison using community-wide test data. Because it achieves this performance using cross-language data alone, it is likely that improved results can be obtained by also taking advantage of monolingual contextual evidence. Although in the end all unsupervised systems are likely to produce precision results inferior to the best supervised algorithms, they are often more practical to apply in a broad-vocabulary setting. Moreover, noisy annotations can serve as seeds both for monolingual supervised methods and for bootstrapping cross-linguistic sense disambiguation and sense inventories, complementing other research on the complex problem of mapping sense tags cross linguistically (e.g. (Alonge et al., 1998; Rodriguez et al., 1998; Vossen et al., 1999)).

Acknowledgments

This work has been supported, in part, by ONR MURI Contract FCPO.810548265, NSA RD-02-5700, and DARPA/ITO Cooperative Agreement N660010028910. The authors would like to thank the anonymous reviewers for their comments, Rebecca Hwa and Okan Kolak for helpful assistance and discussion, Franz Josef Och for his help with GIZA++, Adwait Ratnaparkhi for the use of MXTERMINATOR, and our collaborators at Johns Hopkins for the use of their computing facilities in parts of this work.

References

R. Agirre, L. Padro, and J. Atserias. 2000. Combining supervised and unsupervised lexical knowledge methods for word sense disambiguation. *Computers and the Humanities: Special issue on SENSEVAL*, 34:103–108.

A. Alonge, N. Calzolari, P. Vossen, L. Loksma, I. Casrellon, M. A. Marti, and W. Peters. 1998. The linguistic design of the eurowordnet database. *Computers and the Humanities: Special issue on EuroWordNet*, 32(2-3).

Eric Brill. 1993. *A Corpus-Based Approach to Language Learning*. Ph.D. thesis, Computer and Information Science, University of Pennsylvania, June.

P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, and R. L. Mercer. 1991. A statistical approach to sense disambiguation in machine translation. In *Proc. of the Speech and Natural Language Workshop*, pages 146–151, Pacific Grove, CA.

Rebecca Bruce and Janyce Wiebe. 1994. A new approach to sense identification. In *ARPA Workshop on human Language Technology*, Plainsboro, NJ, March.

Ido Dagan and Alon Itai. 1994. Word sense disambiguation using a second language monolingual corpus.

Ido Dagan. 1991. Lexical disambiguation: sources of information and their statistical realization. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, June. Berkeley, California.

Mona Diab and Steven Finch. 2000. A statistical word level translation model for comparable corpora. In *Proceedings of Conference on Content based multimedia information Access RIAO'00*, Paris, France. Content Based Multimedia Information Access.

Mona Diab. 2000. An unsupervised method for multilingual word sense tagging using parallel corpora: A preliminary investigation. In *SIGLEX2000: Word Senses and Multi-linguality*, Hong Kong, October.

Helge Dyvik. 1998. Translations as semantic mirrors. In *Proceedings of Workshop W13: Multilinguality in the lexicon II*, pages 24–44, Brighton, UK. The 13th biennial European Conference on Artificial Intelligence ECAI 98.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

W. Francis and H. Kučera. 1982. *Frequency Analysis of English Usage*. Houghton Mifflin Co.: New York.

Nancy Ide. 2000. Cross-lingual sense determination: Can it work? *Computers and the Humanities: Special issue on SENSEVAL*, 34:223–234.

Adam Kilgarriff and Joseph Rosenzweig. 2000. Framework and results for english SENSEVAL. *Computers and the Humanities: Special issue on SENSEVAL*, 34:15–48.

Dekang Lin. 1999. A case-base algorithm for word sense disambiguation. In *Proceedings of Conference Pacific Association for Computational Linguistics*, Waterloo, Canada. Pacific Association for Computational Linguistics.

Dekang Lin. 2000. Word sense disambiguation with a similarity based smoothed library. *Computers and the Humanities: Special issue on SENSEVAL*, 34:147–152.

K. Litkowski. 2000. SENSEVAL: The cl-research experience. *Computers and the Humanities: Special issue on SENSEVAL*, 34:153–158.

I. Dan Melamed and Philip Resnik. 2000. Evaluation of sense disambiguation given hierarchical tag sets. *Computers and the Humanities*, 34(1–2).

J. Nie, P. Isabelle, M. Simard, and R. Durand. 1999. Cross-language information retrieval based on parallel texts and automatic mining of parallel texts from the web. In *Proceedings of ACM-SIGIR conference*, pages 74–81, Berkeley, CA.

Franz Josef Och and Hermann Ney. 2000. Improved statistical alignment models. In *Proceedings of 37th Annual Meeting of the Association for Computational Linguistics (ACL'00)*, Hong Kong, October.

Philip Resnik and David Yarowsky. 1999. Distinguishing systems and distinguishing senses: New evaluation methods for word sense disambiguation. *Natural Language Engineering*, 5(2):113–133.

Philip Resnik. 1997. Selectional preference and sense disambiguation. In *ANLP Workshop on Tagging Text with Lexical Semantics*, Washington, D.C., April.

Philip Resnik. 1999a. Mining the Web for bilingual text. In *37th Annual Meeting of the Association for Computational Linguistics (ACL'99)*, College Park, Maryland, June.

Philip Resnik. 1999b. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research (JAIR)*, 11:95–130.

H. Rodriguez, S. Climent, P. Vossen, L. Loksma, W. Peters, A. Alonge, F. Bertagna, and A. Roventini. 1998. The top-down strategy for building eurowordnet: Vocabulary coverage, base concepts and top ontology. *Computers and the Humanities: Special issue on EuroWordNet*, 32(2-3).

Hinrich Schütze. 1992. Dimensions of meaning. In *Proceedings of Supercomputing '92*.

Piek Vossen, Wim Peters, and Julio Gonzalo. 1999. Towards a universal index of meaning. In *Proceedings of the ACL SIGLEX workshop*, Maryland, MD, USA.

David Yarowsky. 1992. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, pages 454–460, Nantes, France, July.

David Yarowsky. 1993. One sense per collocation. ARPA Workshop on Human Language Technology, March. Princeton.

David Yarowsky. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pages 189–196, Cambridge, MA. Association for Computational Linguistics.