Computers and the Humanities **00** : 00-00, 1999.

Cross-lingual Sense Determination: Can It Work?

NANCY IDE
Department of Computer Science, Vassar College, 124 Raymond Avenue, Poughkeepsie, NY 12604-0520 USA (email: ide@cs.vassar.edu)

Key words: sense disambiguation, parallel corpora, translation

Abstract. This article reports the results of a preliminary analysis of translation equivalents in four languages from different language families, extracted from an on-line parallel corpus of George Orwell's *Nineteen Eighty-Four* The goal of the study is to determine the degree to which translation equivalents for different meanings of a polysemous word in English are lexicalized differently across a variety of languages, and to determine whether this information can be used to structure or create a set of sense distinctions useful in natural language processing applications. A *coherence index* is computed that measures the tendency for different senses of the same English word to be lexicalized differently, and from this data a clustering algorithm is used to create sense hierarchies.

1. Introduction

It is well known that the most nagging issue for word sense disambiguation (WSD) is the definition of just what a word sense is. At its base, the problem is a philosophical and linguistic one that is far from being resolved. However, work in automated language processing has led to efforts to find practical means to distinguish word senses, at least to the degree that they are useful for natural language processing tasks such as summarization, document retrieval, and machine translation. Several criteria have been suggested and exploited to automatically determine the sense of a word in context (see Ide and Véronis, 1998), including syntactic behavior, semantic and pragmatic knowledge, and especially in more recent empirical studies, word co-occurrence within syntactic relations (e.g., Hearst, 1991; Yarowsky, 1993), words co-occurring in global context (e.g., Gale *et al.*, 1993; Yarowsky, 1992; Schütze, 1992, 1993), etc. No clear criteria have emerged, however, and the problem continues to loom large for WSD work.

The notion that cross-lingual comparison can be useful for sense disambiguation has served as a basis for some recent work on WSD. For example, Brown *et al.* (1991) and Gale *et al.* (1992a, 1993) used the parallel, aligned *Hansard Corpus* of Canadian Parliamentary debates for WSD, and Dagan *et al.* (1991) and Dagan and Itai (1994) used monolingual corpora of Hebrew and German and a bilingual dictionary. These studies rely on the assumption that the mapping between words and word senses varies significantly among languages. For example, the word *duty* in English translates into French as *devoir* in its obligation sense, and *impôt* in its tax sense. By determining the translation equivalent of *duty* in a parallel French text, the correct sense of the English word is identified. These studies exploit this information in order to gather co-occurrence data for the different senses, which is then used to disambiguate new texts. In related work, Dyvik (1998) used patterns of translational relations in an English-Norwegian parallel corpus (ENPC, Oslo University) to define semantic properties such as synonymy, ambiguity, vagueness, and semantic fields and suggested a derivation of semantic representations for signs (e.g., lexemes), capturing semantic relationships such as hyponymy etc., from such translational relations.

Recently, Resnik and Yarowsky (1997) suggested that for the purposes of WSD, the different senses of a word could be determined by considering only sense distinctions that are lexicalized cross-linguistically. In particular, they propose that some set of target languages be identified, and that the sense distinctions to be considered for language processing applications and evaluation be restricted to those that are realized lexically in some minimum subset of those languages. This idea would seem to provide an answer, at least in part, to the problem of determining different senses of a word: intuitively, one assumes that if another language lexicalizes a word in two or more ways, there must be a conceptual motivation. If we look at enough languages, we would be likely to find the significant lexical differences that delimit different senses of a word.

However, this suggestion raises several questions. For instance, it is well known that many ambiguities are preserved across languages (for example, the French *intérêt* and the English *interest*), especially languages that are relatively closely related. Assuming this problem can be overcome, should differences found in closely related languages be given lesser (or greater) weight than those found in more distantly related languages? More generally, which languages should be considered for this exercise? All languages? Closely related languages? Languages from different language families? A mixture of the two? How many languages, and of which types, would be "enough" to provide adequate information for this purpose?

There is also the question of the criteria that would be used to establish that a sense distinction is "lexicalized cross-linguistically". How consistent must the distinction be? Does it mean that two concepts are expressed by *mutually non-interchangeable* lexical items in some significant number of other languages, or need it only be the case that the *option* of a different lexicalization exists in a certain percentage of cases?

Another consideration is where the cross-lingual information to answer these questions would come from. Using bilingual dictionaries would be extremely tedious and error-prone, given the substantial divergence among dictionaries in terms of the kinds and degree of sense distinctions they make. Resnik and Yarowsky (1997) suggest EuroWordNet (Vossen, 1998) as a possible source of information, but, given that EuroWordNet is primarily a lexicon and not a corpus, it is subject to many of the same objections as for bi-lingual dictionaries.

An alternative would be to gather the information from parallel, aligned corpora. Unlike bilingual and multi-lingual dictionaries, translation equivalents in parallel texts are determined by experienced translators, who evaluate each instance of a word's use in context rather than as a part of the meta-linguistic activity of classifying senses for inclusion in a dictionary. However, at present very few parallel aligned corpora exist. The vast majority of these are bi-texts, involving only two languages, one of which is very often English. Ideally, a serious evaluation of Resnik and Yarowsky's proposal would include parallel texts in languages from several different language families, and, to maximally ensure that the word in question is used in the exact same sense across languages, it would be preferable that the same text were used over all languages in the study. The only currently available parallel corpora for more than two languages are Orwell's *Nineteen Eighty-Four* (Erjavec and Ide, 1998), Plato's *Republic* (Erjavec, *et al.*, 1998), the MULTEXT *Journal of the Commission* corpus (Ide and Véronis, 1994), and the Bible (Resnik, *et al.*, in press). It is likely that these corpora do not provide enough appropriate data to reliably determine sense distinctions, Also, it is not clear how the lexicalization of sense distinctions across languages is affected by genre, domain, style, etc.

This paper attempts to provide some preliminary answers to the questions outlined above, in order to eventually determine the degree to which the use of parallel data is viable to determine sense distinctions, and if so, the ways in which this information might be used. Given the lack of large parallel texts across multiple languages, the study is necessarily limited; however, close examination of a small sample of parallel data can, as a first step, provide the basis and direction for more extensive studies.

2. Methodology

I have conducted a small study using parallel, aligned versions of George Orwell's *Nineteen Eighty-Four* (Erjavec and Ide, 1998) in five languages: English, Slovene, Estonian, Romanian, and Czech.[1] The study therefore involves languages from four language families (Germanic, Slavic, Finno-Ugrec, and Romance), two languages from the same family (Czech and Slovene), as well as one non-Indo-European language (Estonian).

Nineteen Eighty-Four is a text of about 100,000 words, translated directly from the original English in each of the other languages. The parallel versions of the text are sentence-aligned to the English and tagged for part of speech. Although *Nineteen Eighty-Four* is a work of fiction, Orwell's prose is not highly stylized and, as such, it provides a reasonable sample of modern, ordinary language that is not tied to a given topic or sub-domain (such as newspapers, technical reports, etc.). Furthermore, the translations of the text seem to be relatively faithful to the original: for instance, over 95% of the sentence alignments in the full parallel corpus of seven languages are one-to-one (Priest-Dorman, *et al.,* 1997).

Four ambiguous English words were considered in this study: *hard, line, country* and *head*. *Line* and *hard* were chosen because they have served in various WSD studies to date (e.g., Leacock *et al.,* 1993) and a corpus of occurrences of these words from the *Wall Street Journal* corpus was generously made available for comparison.[2] *Serve*, another word frequently used in these studies, did not appear frequently enough in the Orwell text to be considered, nor did any other suitable ambiguous verb.[3] *Country* and *head* were chosen as substitutes because they appeared frequently enough for consideration.

All sentences containing an occurrence or occurrences (including morphological variants) of each of the three words were extracted from the English text, together with the parallel sentences in which they occur in the texts of the four comparison languages (Czech, Estonian, Romanian, Slovene). The English occurrences were first separated according to part of speech, retaining the noun senses of *line, country,* and *head,* and the adjective and adverb senses of *hard.* As Wilks and Stevenson (1998) have pointed out, part-of-speech tagging accomplishes a good portion of the work of semantic disambiguation; therefore only occurrences with the same part of speech have been considered.[4] The selected English occurrences were then grouped using the sense distinctions in WordNet, (version 1.6) [Miller *et al.*, 1990; Fellbaum, 1998]). The sense categorization was performed by the author and two student assistants; results from the three were compared and a final, mutually agreeable grouping was established. The occurrence data for each sense of each of the four words is given in Table 1.[5]

For each of the four comparison languages, the corpus of sense-grouped parallel sentences for English and that language was sent to a linguist and native speaker of the comparison language. The linguists were asked to provide the lexical item in each parallel sentence that corresponds to the ambiguous English word; if inflected, they were asked to provide both the inflected form and the root form. In addition, the linguists were asked to indicate the type of translation, according to the distinctions given in Table 2. Additional information about possible synonyms, etc. was also asked for.

For over 85% of the English word occurrences (corresponding to types 1 and 2 in Table 2), a specific lexical item or items could be identified as the translation equivalent for the corresponding English word. Translations of type 5, involving phrases whose meaning encompassed a larger phrase in the English, were considered to be translation equivalents on a case-by-case basis. For example, the Czech translation of "grow[n] hard" is translated in a single verb (closer in meaning to the English "harden") and as such was judged not to be an equivalent for "hard", whereas the translation of "stretch of country" in all four comparison languages by a single lexical word was considered to be equivalent, since the translation does not combine two (necessarily) separable concepts.[6] Each translation equivalent was represented by its lemma (or the lemma of the root form in the case of derivatives), for comparison purposes, and associated with the WordNet sense to which it corresponds.[7]

Table 1 : Corpus statistics for parallel data from Orwell's *Nineteen Eighty-Four*

WORD	SENSE DESCRIPTION (WordNet)	WordNet SENSE #	# OF OCC	TOTAL OCC
hard	difficult	1.1	4	13
	metaphorically hard	1.2	2	
	not yielding to pressure ; vs. “soft”	1.3	3	
	very strong or vigorous, arduous	1.4	1	
	with force or vigor (adv.)	2.1	2	
	earnestly, intently (adv.)	2.3	1	
line	direction, course	1.10	3	28
	acting in conformity	1.16	1	
	a linear string of words	1.5	8	
	contour, outline	1.4	3	
	formation of people/things beside one another	1.1	1	
	wrinkle, furrow, crease	1.12	3	
	logical argument	1.8	1	
	something long, thin, flexible	1.18	4	
	fortified position	1.7	1	
	spatial location	1.11	2	
	formation of people/things behind one another	1.3	1	
country	a politically organized body of people	1.2	16	19
	area outside cities and towns	1.5	3	
head	part of the body	1.1	50	65
	intellect	1.3	12	
	ruler, chief	1.4	2	
	front, front part	1.7	1	
TOTAL NUMBER OF OCCURRENCES OF ALL WORDS				125
TOTAL NUMBER OF SAMPLES (TOTAL OCC x 4 LANGUAGES)				500

Table 2 : Translation types and their frequencies

TYPE	MEANING	# OCC	%OCC
1	A single lexical item is used to translate the English equivalent (possibly a different part of speech)	**395**	86%
2	The English word is translated by a phrase of two or more words or a compound, which has the same meaning as the single English word	**5**	1%
3	The English word is not lexicalized in the translation	**29**	6%
4	A pronoun is substituted for the English word in the translation	**3**	.6%
5	An English phrase containing the ambiguous word is translated by a single word in the comparison language which has a broader or more specific meaning, or by a phrase in which the specific concept corresponding to the English word is not explicitly lexicalized	**28**	6%

In order to determine the degree to which the assigned sense distinctions correspond to translation equivalents, a *coherence index (CI)* was computed that measures the degree to which each pair of senses is translated using the same word as well as the consistency with which a given sense is translated with the same word.[8] Note that the CIs do not determine whether or not a sense distinction *can be* lexicalized in the target language, but only the degree to which they *are* lexicalized differently in the translated text. However, it can be assumed that the CIs provide a measure of the *tendency* to lexicalize different WordNet senses differently, which can in turn be seen as an indication of the degree to which the distinction is valid.

Table 3 : Number of words used to translate the test words

WORD	#Senses	RO	ES	SL	CS
hard	6	8	7	5	6
country	3	2	4	3	4
line	11	9	14	12	11
head	4	9	6	9	4

For each ambiguous word, the CI is computed for each pair of senses, as follows:

$$CI(s_q s_r) = \frac{\sum_{i=1}^{n} s_{<q,r>}^{(i)}}{m_{s_q} m_{s_r} n}$$

where:

- n is the number of comparison languages under consideration;
- m_{sq} and m_{sr} are the number of occurrences of sense s_q and sense s_r in the English corpus, respectively, including occurrences which have no identifiable translation;
- $s_{<q,r>}^{(i)}$ is the number of times that senses q and r are translated by the same lexical item in language i, i.e.,

$$\sum_{x \in trans(q), y \in trans(r)} x = y$$

The CI is a value between 0 and 1, computed by examining clusters of occurrences translated by the same word in the other languages. If sense i and sense j are consistently translated with the same word in each comparison language, then $CI(s_i, s_j) = 1$; if they are translated with a different word in every occurrence, $CI(s_i, s_j) = 0$. In general,, the CI for pairs of different senses provides an index of their relatedness, i.e., the greater the value of $CI(s_i, s_j)$, the more frequently occurrences of sense i and sense j are translated with the same lexical item. When $i = j$, we obtain a measure of the coherence of a given sense.

The CIs were computed over four sets of comparison languages, in order to determine the effects of language-relatedness on the results:

- Estonian (Finno-Ugric), Romanian (Romance), and Czech and Slovene (Slavic);
- Estonian, Romanian, and Slovene (three different language families);
- Czech and Slovene (same language family);

- Romanian, Czech, and Slovene (Indo-European) for comparison with Estonian (non-Indo-European).

CIs were also computed for each language individually. To better visualize the relationship between senses, a hierarchical clustering algorithm was applied to the CI data to generate trees reflecting sense proximity.[9] Finally, in order to determine the degree to which the linguistic relation between languages may affect coherence, a correlation was run among CIs for all pairs of the four target languages.

3. Results

Although the data sample is small, it gives some insight into ways in which a larger sample might contribute to sense discrimination.

Table 4 : CIs for *hard* and *head*

HARD							HEAD				
WordNet Sense No	2.1	2.3	1.4	1.3	1.1	1.2		1.1	1.3	1.4	1.7
2.1	0.50						1.1	0.69			
2.3	0.13	1.00					1.3	0.53	0.45		
1.4	0.00	0.25	1.00				1.4	0.12	0.07	0.50	
1.3	0.04	0.50	0.17	0.56			1.7	0.40	0.00	0.00	1.00
1.1	0.19	0.00	0.00	0.00	0.63						
1.2	0.00	0.00	0.25	0.21	0.00	0.50					

The CI data for *hard* and *head* are given in Table 4. CIs measuring the affinity of a sense with itself—that is, the tendency for all occurrences of that sense to be translated with the same word-- show that all of the six senses of *hard* show greater internal consistency than affinity with other senses, with senses 1.1 ("difficult" - CI = .56) and 1.3 ("not soft" – CI = .63) registering the highest internal consistency.[10] The same holds true for three of the four senses of *head,* while the CI for senses 1.3 (“intellect”) and 1.1 (“part of the body”) is higher than the CI for 1.3/1.3.

Figure 1 shows the sense clusters for *hard* generated from the CI data.[11] The senses fall into two main clusters, with the two most internally consistent senses (1.1 and 1.3) at the deepest level of each of the respective groups. The two adverbial forms[12] are placed in separate groups, reflecting their semantic proximity to the different adjectival meanings of *hard.* The clusters for *head* (Figure 2) similarly show two distinct groupings, each anchored in the two senses with the highest internal consistency and the lowest mutual CI (“part of the body” (1.1) and “ruler, chief” (1.4)).

Figure 1 : Cluster tree and distance measures for the six senses of *hard*

```
  ____________________________|--------------------> hard2.1
 |                            |--------------------> hard1.1
-|                             ______________________|---------> hard2.3
 |----------------------------|                      |---------> hard1.3
                              |______________________|---------------> hard1.4
                                                     |---------------> hard1.2
       minimum distance = 0.249399     ( 1.3 )      ( 2.3 )
       minimum distance = 0.434856     ( 1.2 )      ( 1.4 )
       minimum distance = 0.555158     ( 1.1 )      ( 2.1 )
       minimum distance = 0.602972     ( 1.4 1.2 )      ( 2.3 1.3 )
       minimum distance = 0.761327     ( 2.3 1.3 1.4 1.2 )      ( 2.1 1.1 )
```

Figure 2 : Cluster tree and distance measures for the four senses of *head*

```
 |-------------------------> 1.4
-|                          _______________________|----------------> 1.1
 |-------------------------|                       |----------------> 1.3
                           |----------------------> 1.7

      minimum distance = 0.441022     ( 1.3 )    ( 1.1 )
      minimum distance = 0.619052     ( 1.7 )    ( 1.1 1.3 )
      minimum distance = 0.723157     ( 1.1 1.3 1.7 )  ( 1.4 )
```

The hierarchies apparent in the cluster graphs make intuitive sense. Structured like dictionary entries, the clusters for *hard* and *head* might appear as in Figure 3. This is not dissimilar to actual dictionary entries for *hard* and *head;* for example, the entries for *hard* in four differently constructed dictionaries (*Collins English (CED), Longman's (LDOCE), Oxford Advanced Learner's (OALD),* and *COBUILD*) all list the "difficult" and "not soft" senses first and second, which, since most dictionaries list the most common or frequently used senses first, reflects the gross division apparent in the clusters. Beyond this, it is difficult to assess the correspondence between the senses in the dictionary entries and the clusters. The remaining WordNet senses are scattered at various places within the entries or, in some cases, split across various senses. The hierarchical relations apparent in the clusters are not reflected in the dictionary entries, since the senses are for the most part presented in flat, linear lists. However, it is interesting to note that the first five senses of *hard* in the *COBUILD* dictionary, which was constructed on the basis of corpus examples and presents senses in order of frequency, correspond to five of the six WordNet senses in this study; WordNet's "metaphorically hard" is spread over multiple senses in the *COBUILD*, as it is in the other dictionaries.

Figure 3 : Clusters for *hard* and *head* structured as dictionary entries

```
HARD  I   1 difficult                HEAD  I  1  a. part of the body
          2 vigorously                           b. intellect
     II   1 a. not soft                       2  front, front part
            b. strong                     II  ruler, chief
          2 a. earnestly
            b. metaphorically hard
```

The results for different language groupings show that the tendency to lexicalize senses differently is not affected by language distance (Table 5). The mean CI for Estonian, the only non-Indo-European language in the study, is lower than that for any other group, indicating that WordNet sense distinctions are slightly less likely to be clearly distinguished in Estonian. However, the difference (z = -1.43) is not statistically significant. Correlations of CIs for each language pair (Table 6) also show no relationship between the degree to which sense distinctions are lexicalized differently and language distance. This is contrary to results obtained by Resnik and Yarowsky (submitted), who found that that non-Indo-European languages tended to lexicalize English sense distinctions, especially at finer-grained levels, more than Indo-European languages. However, their translation data was generated by native speakers presented with isolated sentences in English who were asked to provide the translation for a given word in the sentence. It is not clear how this data compares to translations generated by trained translators working with full context.

Table 5 : Average CI values for language groupings

Language group	Average CI
ALL	0.27
RO/ES/SL	0.28
SL/CS	0.28
RO/SL/CS	0.27
ES	0.26

Table 6 : Correlation among CIs for the four target languages

Language Pair	**Correlation**
ES/CS	0.74
RO/SL	0.80
RO/CS	0.72
SL/CS	0.71
RO/ES	0.73
ES/SL	0.80

4. Summary

The small sample in this study suggests that cross-lingual lexicalization can be used to define and structure sense distinctions. The cluster graphs above provide information about relations among WordNet senses that could be used, for example, to determine the granularity of sense differences, which in turn could be used in tasks such as machine translation, information retrieval, etc. For example, it is likely that as sense distinctions become finer, the degree of error is less severe. Resnik and Yarowsky (1997) suggest that confusing finer-grained sense distinctions should be penalized less severely than confusing grosser distinctions when evaluating the performance of sense disambiguation systems. The clusters also provide insight into the lexicalization of sense distinctions related by various semantic relations (metonymy, meronymy, etc.) across languages; for instance, the “part of the body” and “intellect” senses of *head* are lexicalized with the same item a significant portion of the time across all languages, information that could be used in machine translation. In addition, cluster data such as that presented here could be used in lexicography, to determine a more detailed hierarchy of relations among senses in dictionary entries.

It is less clear how cross-lingual information could be used to *determine* sense distinctions independent of a pre-defined set, such as the WordNet senses used here. More work needs to be done on this topic utilizing substantially larger parallel corpora that include a variety of language types. We are currently experimenting with clustering occurrences rather than senses (similar to Schütze, 1992), as well as using WordNet synsets and “back translations” (i.e., additional translations in the original language of the translations in the target language) to create semantic groupings, which could provide additional information for determining sense distinctions.

Acknowledgements

The author would like to gratefully acknowledge the contribution of those who provided the translation information: Tomaz Erjavec (Slovene), Vladimir Petkevic (Czech), Dan Tufis (Romanian), and Kadri Muischnek (Estonian); as well as Dana Fleur and Daniel Kline, who helped to transcribe and evaluate the data. Special thanks to Dan Melamed and Hinrich Schütze for their helpful comments on earlier drafts of the paper.

Notes

[1] The Orwell parallel corpus also includes versions of *Nineteen-Eighty Four* in Hungarian, Bulgarian, Latvian, Lithuanian, Serbian, and Russian.

[2] Claudia Leacock provided samples of *hard* and *line* from the *Wall Street Journal* corpus.

[3] The verb sense of *line* does not occur in the English Orwell.

[4] Both the adjective and adverb senses of *hard* were retained because the distinction is not consistent across the translations used in the study.

[5] The sense inventories and parallel corpus extracts used in this analysis are available at http://www.cs.vassar.edu/~ide/wsd/.

[6] That all four languages use a single lexical item to express this concept itself provides some basis to regard "stretch of country" as a collocation expressing a single concept.

[7] The number of translation equivalents for each word in the analysis is given in Table 3

[8] Note that the CI is similar in concept to semantic entropy (Melamed, 1997). However, Melamed computes entropy for word types, rather than word senses.

[9] Developed by Andreas Stolcke.

[10] Senses 2.3 and 1.4 have CIs of 1 because each of these senses exists in a single occurrence in the corpus, and have therefore been discarded from consideration of CIs for individual senses. We are currently investigating the use of the Kappa statistic (Carletta, 1996) to normalize these sparse data.

[11] For the purposes of the cluster analysis, CIs of 1.00 resulting from a single occurrrence were normalized to .5.

[12] Because root forms were used in the analysis, no distinction in translation equivalents was made for part of speech.

References

Carletta, Jean (1996). Assessing Agreement on Classification Tasks: The Kappa Statistic. Computational Linguistics, 22(2), 249-254.

Dagan, Ido and Itai, Alon (1994). Word sense disambiguation using a second language monolingual corpus. *Computational Linguistics,* 20(4), 563-596.

Dagan, Ido; Itai, Alon; and Schwall, Ulrike (1991). Two languages are more informative than one. *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics,* 18-21 June 1991, Berkeley, California, 130-137.

Dyvik, Helge (1998). Translations as Semantic Mirrors. *Proceedings of Workshop W13: Multilinguality in the Lexicon II, The 13th Biennial European Conference on Artificial Intelligence (ECAI 98)*, Brighton, UK, 24-44.

Erjavec, Tomaz and Ide, Nancy (1998). The MULTEXT-EAST Corpus. *Proceedings of the First International Conference on Language Resources and Evaluation*, 27-30 May 1998, Granada, 971-74.

Erjavec, Tomaz, Lawson, Ann, and Romary, Laurent (1998). East meets West: Producing Multilingual Resources in a European Context. *Proceedings of the First International Conference on Language Resources and Evaluation,* 27-30 May 1998, Granada, 981-86.

Fellbaum, Christiane (ed.) (1998). *WordNet: An Electronic Lexical Database.* MIT Press, Cambridge, Massachusetts.

Gale, William A., Church, Kenneth W. and Yarowsky, David (1993). A method for disambiguating word senses in a large corpus. *Computers and the Humanities,* 26, 415-439.

Hearst, Marti A. (1991). Noun homograph disambiguation using local context in large corpora. *Proceedings of the 7th Annual Conference of the University of Waterloo Centre for the New OED and Text Research,* Oxford, United Kingdom, 1-19.

Ide, Nancy and Véronis, Jean (1998). Word sense disambiguation: The state of the art. *Computational Linguistic,s* 24:1, 1-40.

Leacock, Claudia; Towell, Geoffrey and Voorhees, Ellen (1993). Corpus-based statistical sense resolution. *Proceedings of the ARPA Human Language Technology Worskshop*, San Francisco, Morgan Kaufman.

Melamed, I. Dan. (1997). Measuring Semantic Entropy. *ACL-SIGLEX Workshop Tagging Text with Lexical Semantics: Why, What, and How?* April 4-5, 1997, Washington, D.C., 41-46.

Miller, George A.; Beckwith, Richard T. Fellbaum, Christiane D.; Gross, Derek and Miller, Katherine J. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography,* 3(4), 235-244.

Priest-Dorman, Greg; Erjavec, Tomaz; Ide, Nancy and Petkevic, Vladimír (1997). Corpus Markup. COP Project 106 MULTEXT-East Deliverable D2.3 F. Available at http://nl.ijs.si/ME/CD/docs/mte-d23f/mte-D23F.html.

Resnik, Philip; Broman Olsen, Mari and Diab, Mona (in press). Creating a Parallel Corpus from the Book of 2000 Tongues. *Computers and the Humanities.*

Resnik, Philip and Yarowsky, David (submitted). Distinguishing systems and distinguishing senses: New evaluation methods for word sense disambiguation. Submitted to *Natural Language Engineering.*

Resnik, Philip and Yarowsky, David (1997). A perspective on word sense disambiguation methods and their evaluation. *ACL-SIGLEX Workshop Tagging Text with Lexical Semantics: Why, What, and How?* April 4-5, 1997, Washington, D.C., 79-86.

Schütze, Hinrich (1992). Dimensions of meaning. *Proceedings of Supercomputing'92.* IEEE Computer Society Press, Los Alamitos, California, 787-796.

Schütze, Hinrich (1993). Word space. In Hanson, Stephen J.; Cowan, Jack D.; and Giles, C. Lee (Eds.) *Advances in Neural Information Processing Systems 5*, Morgan Kauffman, San Mateo, California, 5, 895-902.

Vossen, Piek (ed.) (1998). EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Press, Dordrecht. Reprinted from *Computers and the Humanities,* 32:2-3.

Wilks, Yorick and Stevenson, Mark (1998). Word Sense Disambiguation Using Optimized Combinations of Knowledge Sources. *Proceedings of COLING/ACL-98*, Montreal, August, 1998.

Yarowsky, David (1992). Word sense disambiguation using statistical models of Roget's categories trained on large corpora. *Proceedings of the 14th International Conference on Computational Linguistics,* COLING'92, 23-28 August, Nantes, France, 454-460.

Yarowsky, David (1993). One sense per collocation. *Proceedings of the ARPA Human Language Technology Workshop,* Princeton, New Jersey, 266-271.