

COMBINING MACHINE READABLE LEXICAL RESOURCES AND BILINGUAL CORPORA FOR BROAD WORD SENSE DISAMBIGUATION

Jason J.S. Chang, Jen-Nan Chen, Huei-Hong Sheng, and Sur-Jin Ker
Department of Computer Science
National Tsing Hua University
Hsinchu 30043, Taiwan, ROC

Tel: +886 35 731-069 Fax: +886 35 723-694
e-mail: jschang@cs.nthu.edu.tw

Abstract

This paper describes a new approach to word sense disambiguation (WSD) based on automatically acquired "word sense division. The semantically related sense entries in a bilingual dictionary are arranged in clusters using a heuristic labeling algorithm to provide a more complete and appropriate sense division for WSD. Multiple translations of senses serve as outside information for automatic tagging of bilingual corpora and acquisition of WSD rules. We describe and implement a WSD method using the English-Chinese bilingual version (LecDOCE) of the Longman Dictionary of Contemporary English (LDOCE). For this purpose, we draw on information about topics and topical sets in the Longman Lexicon of Contemporary English (LLOCE) to represent and disambiguate LecDOCE senses. Example sentences and their translations from LecDOCE are employed as training materials for WSD, while further examples from the Brown corpus are used for testing. Quantitative results of disambiguating 12 words are also presented.

1. Introduction

Word sense disambiguation (WSD) has been found useful in many NLP applications, including information retrieval (McRoy 1992; Krovetz and Croft 1992) and machine translation (Brown et al. 1991; Dagan et al. 1991; Dagan and Itai 1994). Recent work on computational linguistics have paid increasing attention to WSD (Lesk 1986; Schütze 1992; Gale et al. 1992; Yarowsky 1992 and 1995; Luk 1995). Given a polysemous word in running text, the task of WSD involves examining contextual information to determine the intended sense from a predefined set. If the set of senses is chosen to be possible translations of the word of interest, the WSD becomes the problem of lexical selection (Dagan et al. 1991; Dagan and Itai 1994) in machine translation (MT) process. WSD methods can be characterized by how word senses are divided and by how WSD knowledge is represented and acquired.

Word sense is an abstract concept frequently based on subjective and subtle distinctions in topic, register, dialect, collocation, part-of-speech and valency (McRoy 1992). Researchers have experimented with various knowledge sources for WSD system, including (1) the defining words in every-day dictionaries (Lesk 1986; Cowie et al. 1992; Luk 1995), (2) indicative words in the context of words listed under a thesaurus category acquired from corpus (Yarowsky 1992; Chen and Chang 1994), (3) bilingual corpora or monolingual corpora in the target language (Gale, Church and Yarowsky 1992; Dagan et al. 1991; Dagan

and Itai 1994), (4) automatic induced clusters with sublexical representation (Schütze 1992), and (5) hand-crafted lexicon containing knowledge from multiple sources (McRoy 1992).

This study is motivated by several observations. First, word-based approaches trained on dictionaries or corpora offer limited coverage for unrestricted text. Lesk (1986) described a word-sense disambiguation technique that is based on the number of overlap between words in a dictionary definition and words in the local context of the word to be disambiguated. The author reported that WSD performance ranged from 50 to 70%. On the other hand, the performance of corpus-trained word-based models (Yarowsky 1992 and 1995; Gale, Church and Yarowsky 1992) were shown to be very effective. However, Luk (1995) pointed out that these methods were tested on technical writing or text in a limited domain, therefore it remains doubtful whether these models can perform as well for text with all kinds of genre such as the Brown corpus.

With a set of programs described in this paper, we try to find out whether it is possible to exploit the existing machine readable lexical resources to put to use in automatic acquisition of knowledge bases of word sense division and disambiguation. Furthermore, we also experimented with the concept-based approach to see if it can provide broader coverage, while maintaining comparable high precision. We hope to achieve a degree of generality, so it is possible to resolve word sense ambiguity, even for a word found in a particular unfamiliar context in terms of word overlapping. For instance, the sense of "bank" in the context of "vole" (see Table 1) is arguably difficult to disambiguate, even with a very large training corpus. Consider the sentences that contains the word "bank" among some 25,000 example sentences in LecDOCE (Proctor 1988). The intended sense of "bank" in these sentences is predominately FINANCE. Table 1 provides further details for other sentences which are related to the GEOGRAPHY sense of "bank," taking note of the indicative context and the Chinese translation of "bank."

Table 1. Example sentences in LecDOCE containing a GEOGRAPHY sense of "bank"

Example Sentences	Translation of "bank"
a bank vole.	堤 (ti)
a ribbon of mist along the river bank.	岸 (an)
a small excavation in the river bank.	岸 (an)
He sat down and rested on a mossy bank in the woods.	堤坡 (bienpo)
The air became moister as we descended the hill towards the river bank.	堤 (bien)
The Dogger Bank in the North Sea can be dangerous for ships.	沙洲 (shazhou)
The hunter recognized the footprints of a deer near the river bank.	岸 (an)
the left bank of the stream.	岸 (an)
The men built banks of earth to hold back the rising flood waters.	堤防 (tifang)
The Romans founded a great city on the banks of this river.	岸 (an)

First, observe that this sample of sentential context for the GEOGRAPHY sense of "bank" is so small that it does not include a whole lot of re-occurring words, except for the word "river" Therefore, it is very difficult to guarantee broad coverage, when disambiguation is dependent on word overlaps. However, it is evident that even with a sample as small as this, there are many re-occurring topics or concepts such as GEOGRAPHY ("stream," "sea," "lake," "earth," "flood," "water," "woods," "hill," "river," "to flow," and "to overflow"), DIRECTION ("north," "south," "east," and "west"), POSITION ("left," "right," and "side"), and ANIMALS ("vole" and "deer"). The data seems to indicate that a class-based approach will be effective for WSD.

Second, the translations are quite diversified; six distinct translations for ten instances of "bank"

reflecting subtle sense difference. The data show that using a single hand-picked translation (Gale, Church and Yarowsky 1992) to label bilingual training data will result in high precision but low applicability. For instance, using the translation, “岸” (*an*), as condition for labeling “bank” as GEOGRAPHY results in 100% precision and 50% coverage for the bilingual sentences shown in Table 1.

Table 2. Example sentences in LDOCE containing a Geography sense of “bank”

Sense Definition	Translation
land along the side of a river, lake, etc.	岸 (<i>an</i>) ; 堤 (<i>ti</i>)
earth which is heaped up in a field or garden, often making a border or division	田埂 (<i>tigeng</i>)
a slope made at bends in a road or race-track, so that they are safer for cars to go round	邊坡 (<i>bienpo</i>)
= SANDBANK	沙洲 (<i>shazhou</i>)

A comprehensive dictionary such as LDOCE, usually provides quite sufficient information about such subtle sense distinction. For instance, the LDOCE lists twelve senses for the word “bank”, four of which are related to the GEOGRAPHY sense of “bank.” Table 2 provides further details. If we can identify and merge such closely related senses, their translation can then be collected to be used in broad and precise labeling of bilingual training corpora for WSD. For instance, if all four senses in Table 2 can be labeled as GEOGRAPHY, then we can collect Chinese characters in their translation as a tuple, [“岸,” “堤,” “田,” “埂,” “邊,” “坡,” “沙,” “洲”]. With this tuple, the word “bank” in all ten example sentences in Table 1 will be labeled correctly as GEOGRAPHY.

The rest of the paper is organized as follows. We begin by giving the details of material used, including the characteristics of definition sentences in LecDOCE and the organization of words in LLOCE. Next, a set of four algorithms for (1) labeling LecDOCE senses, (2) tagging bilingual corpora, (3) acquisition of WSD rules, and (4) rule-based WSD, are described. Examples demonstrating the effectiveness of the algorithms are given for illustration. After describing the algorithm, the experimental results for a twelve-word test set are presented. Moreover, the proposed method is compared with other approaches in computational linguistics literature. Finally, concluding remarks are made.

2. Labeling dictionary senses

The labeling of dictionary definition sentences with a coarse sense distinction like the set labels in LLOCE is a special form of the WSD problem. No simple method can solve the general problem of WSD in unrestricted text. We will show that this labeling task is made simpler for several reasons. For instance, consider the definition sentences for the first few senses of “bank” in LecDOCE as shown in Table 1.

First of all, only simple words are used in the definitions. Furthermore, the text generation schemes are rather regular. The scheme that lexicographers used in generating above definitions is similar to the DEFINITION scheme described in McKeown (1985). A DEFINITION scheme begins with a *genus term* (i.e., conceptual parent or ancestor of the sense), followed by the so-called *differentia* which consists of words semantically related to the sense to provide specifics about the sense. Those relations have been shown to be very effective knowledge sources for WSD (McRoy 1992). For the most part, those relations exist conveniently among words under the same topic or across topics of cross reference in LLOCE. For instance, most of the above mentioned words are listed under the same topic *Ld* (Geography) of intended label *Ld099*, or its cross reference *Me* (Places). Therefore, these definitions can be disambiguated very effectively on the base of similarity between the defining keywords and the words lists in LLOCE.

2.1. Organizing information in LLOCE

In this work, the labels used for tagging dictionary definitions are taken from the LLOCE (McArthur 1992). The category of LLOCE is organized mainly according to subject matter. Nearly 2,500 sets of related words in LLOCE are organized according to 14 subjects and 129 topics (TOP). Cross references (REF) between sets, topics, and subjects are also given to show various inter-sense relations not captured within the same topic. The cross references in LLOCE are primarily between topics.

The sets under which the word is listed in LLOCE are considered as initial candidates for labeling. For instance, the candidates for labeling senses of "bank" are the following 4 sets:

Je104 (banks, exchange, etc.), *Je106* (banking and saving),
Ld099 (river banks), and *Nj295* (bending and leaning).

The set label *Je104* (as well as *Je106*) is listed under the topic *Je* (Banking, Wealth, and Investment), while *Ld099* and *Nj295* are listed under *Ld* (Geography) and *Nj* (Action and Position) respectively. For instance, there is a REF link (in Fig. 1) from topic *Je* to topic *De* (Belonging and Owning, Getting and Giving) (McArthur 1992). To facilitate measurement of similarity between a definition sentence and a topic, we use TOP_S to denote the list of words under a LLOCE topic S , while REF_S denotes that under cross references of S . For instance, the label *Je104* (as well as *Je106*) is associated with a list of words from its topic (TOP_{Je104}) and cross reference ($REF_{Je104} = TOP_{De}$):

$$TOP_{Je104} = TOP_{Je} = \{affluent, budget, cut\ down, deficit, economize, fortune, giro, \\ income, keep, luxury, maintenance, needy, pay, windfall, amenity, \dots\}$$

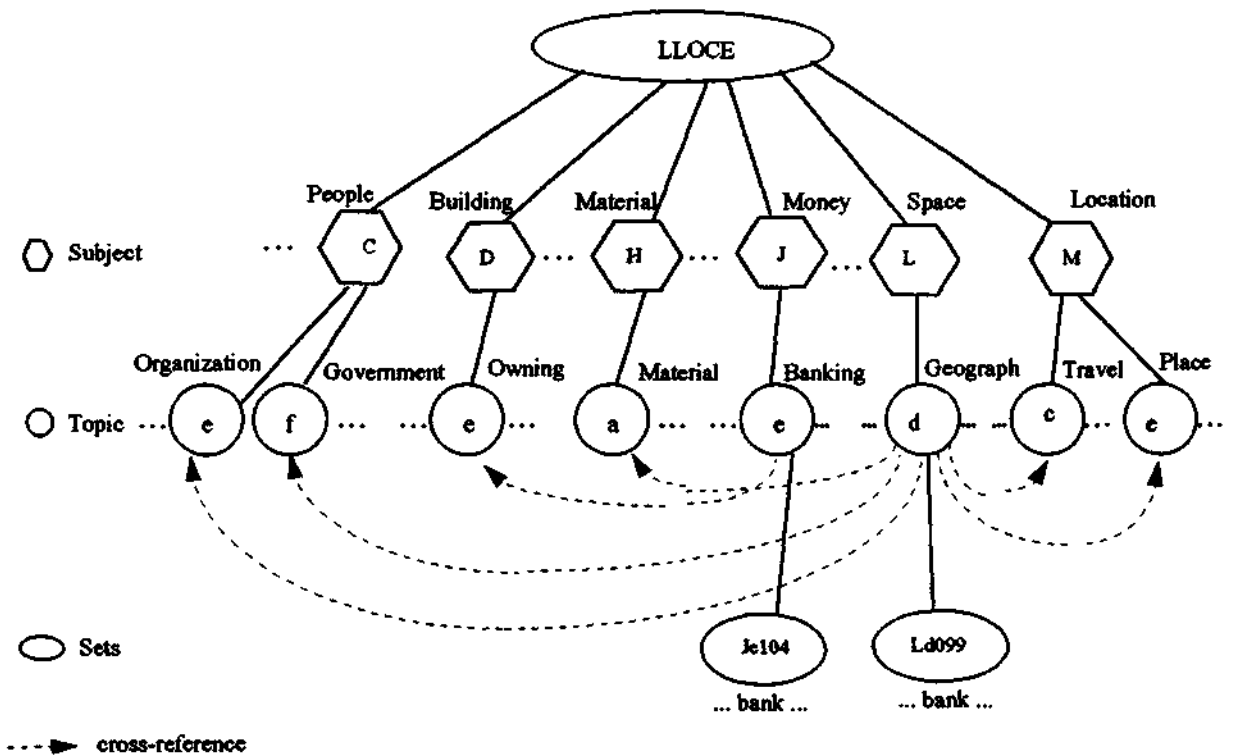
$$REF_{Je104} = TOP_{De} = \{bring\ back, contribution, doff, equip, facility, keep, yield, \dots\}.$$


Figure 1. Subjects, topics, sets, and cross reference between topics in LLOCE.

3. The algorithms

We sum up the above descriptions and outline the procedure for the four algorithms in the following subsections.

3.1. Labeling dictionary entries

Algorithm 3.1 Labeling sense entries in MRD.

Step 1: Given a head word h , read its definition, DEF_h , from LecDOCE.

Step 2: For each definition D of DEF_h , tag each word in D with POS information.

Step 3: Remove all stop words in D to obtain a list of keyword-POS pair, KEY_D .

Step 4: Lookup LLOCE for headword h to obtain a list of sets SET_h that contain h . For each S in SET_h , compile a set of words TOP_S that listed under the topic of S and REF_S the set of words listed under its cross references.

Step 5: Compute similarity $Sim(D, S)$ based on Dice Coefficient for all definitions $D \in DEF_h$ and labels $S \in SET_h$.

$$Sim(D, S) = \sum_{k \in KEY_D} \frac{2 \times w_k \times (In(k, TOP_S) + \gamma In(k, REF_S))}{|TOP_S \cup REF_S| + |KEY_D|}$$

where KEY_D = the set of POS-keyword pairs in definition D ,

γ = the overall relevancy of cross references to a topic,

w_k = 1/the degree of ambiguity of the keyword k ,

$In(a, B) = 1$, if $a \in B$,

$In(a, B) = 0$, if $a \notin B$.

Step 6: Assign to D the label S such that $Sim(D, S)$ is maximized and greater than a certain threshold.

Step 7: If all initial candidates to be dissimilar in Step 6, a second sweep is executed with the number of candidates expanded to include all topical set in LLOCE.

3.2 Labeling training data

Algorithm 3.2 Labeling training a bilingual corpus.

Step 1: For each $L \in LAB_h$, compute TT_L as follows where $TT_L = \{C \mid L_D = L, D \in DEF_h, \text{ and } C \in T_D\}$.

Step 2: For each E of the example sentences containing h , and its translation in Chinese C_E and a label $L \in LAB_h$, compute similarity $Sim(C_E, L)$ based on Dice Coefficient.

$$Sim(C_E, L) = \sum_{c \in C_E} \frac{2 \times In(c, C_L)}{|C_E| + |C_L|}$$

Step 3: Label h in E with the label L such that $Sim(C_E, L)$ is maximized and greater than a certain threshold.

3.3 Acquisition of WSD rules

Algorithm 3.3 Acquisition of decision rules for WSD

- Step 1. Initialize the quadruple table, *COL* to an empty list.
 Step 2. For each example sentence *E* containing *h* with a POS of *p*, let *KEY_E* be the list of keywords in *E* and each *K* in *KEY_E* is labeled with a topical set label *L_K* using Algorithm 3.2.
 Step 3. For each *K* ∈ *KEY_E* and *K* ≠ *h*, insert the following quadruples into *COL*
 (*h*, *p*, *K*, *L_h*)
 (*h*, *p*, *L_K*, *L_h*)
 (*h*, *p*, *topic(L_K)*, *L_h*), where *topic(L_K)* is the topic of *L_K*
 Step 4. For each tuple (*h*, *p*, *c*, *L*) in *COL*, compute logarithmic likelihood ratio as follows:

$$R_L(h, p, c, L) = \log \frac{\text{Count}(h, p, c, L)}{\text{Count}(h, p, c, \bar{L})}$$
 where \bar{L} is a label other than *L*
 Count(*h*, *p*, *c*, *L*) = the number of instances of (*h*, *p*, *c*, *L*) in *COL*
 Count(*h*, *p*, *c*, \bar{L}) = the number of instances of (*h*, *p*, *c*, \bar{L}) in *COL*
 Step 5. For all tuples in *COL*, remove the tuple with *R_L* smaller than a certain threshold.
 Step 6. Sort the list of tuples in *COL* by head word and by log-likelihood ratio in descending order.

3.4 Word Sense Disambiguation

Algorithm 3.4 Word Sense Disambiguation

- Step 1. Initialize the quadruple table, *QUAD* to an empty list.
 Step 2: For an input sentence *INP* containing *h*, tag each word in *INP* with POS information.
 Step 3: Remove all stop words in *INP* to obtain a list of keyword-POS pair, *KEY_{INP}*.
 Step 4: Let *LAB_{hp}* be the topical set labels acquired to represent the word sense division of *h* with POS *p*, using Algorithm 3.1. For each *h*, *K* ∈ *KEY_{INP}* (*K* ≠ *h*) and *L* ∈ *LAB_{hp}*, insert (*h*, *p*, *K*, *L*), (*h*, *p*, *K'*, *L*), and (*h*, *p*, *K''*, *L*) into the *QUAD*, where *p* is the POS of *h* in *INP* and *K'* ∈ *SET(K)* and *K''* = *TOP(K')*.
 Step 5: For a polysemous word *h* in *INP*, find the quadruple (*h*, *p*, *c*, *L*) ∈ *QUAD* that maximizes the following score:

$$\text{Score}(h, p, c, L) = \text{Count}(h, p, c, L) \times R_L(h, p, c, L)$$
 where *Count*(*h*, *p*, *c*, *L*) = the number of instances of (*h*, *p*, *c*, *L*) in *QUAD*.
 Step 6: If *Score*(*h*, *p*, *c*, *L*) is greater than a preset threshold, then report that the sense of *h* is *L*. Otherwise, report the most frequent sense label for the word *h* with POS *p*.

3.5 Illustrative examples

We illustrate how the algorithms function using the sense entries for the word “bank” and an example sentence containing “bank.” Table 3 shows the results of running Algorithm 3.1 on the sense entries of “bank” in LecDOCE. Table 4 shows the collective translations for entries receiving the same label. Algorithm 3.2 is illustrated by using an example containing “bank.” The results of preprocessing and tagging an LecDOCE example are shown in Table 5. This sentence gives rise to the quadruples shown in Table 6 after being processed using Algorithm 3.3. The resulting decision list for “bank” after Algorithm 3.3 is finished is shown in Table 7.

Table 3. Results of Algorithm 3.1

Sense D	POS	Label L_D	Chinese
bank.1.n.1	n	Ld099	岸;堤
bank.1.n.2	n	Ld099	田埂
bank.1.n.3	n	Ha026	一堆;一團
bank.1.n.4	n	Ld099	邊坡
bank.1.n.5	n	Ld099	沙洲
bank.2.v.2	v	Nj235	傾斜轉彎
bank.3.n.1	n	Ha026	一排
bank.4.n.1	n	Je104	銀行
bank.4.n.2	n	Je104	儲存所;庫
bank.4.n.3	n	Je104	莊家
bank.5.v.1	v	Je106	存於銀行
bank.5.v.2	v	Je106	存款

Table 4. Translation Tuples for "bank"

Label L	POS	Translation tuple TT
Ld099	n	岸堤田埂邊坡沙洲
Ha026	n	一堆團排
Nj235	v	傾斜轉彎
Je104	n	銀行儲存所庫莊家
Je104	v	存於銀行存款

Table 5. Results of Running Algorithm 3.2

Example	The interest on my bank account accrued over the years.
Translation	我銀行帳戶的利息逐年有所增加。
Labeled Keywords	interest/Je112 bank/Je104 account/Je105 accrue/Nd092

Table 6. Quadruples representing boolean functions

	Word	POS	Context Word	Sense Label
word level	bank	n	interest	Je104
	bank	n	account	Je104
	bank	n	accrue	Je104
set level	bank	n	Je112	Je104
	bank	n	Je105 Nd092	Je104 Je104
topic level	bank	n	Je	Je104
	bank	n	Nd	Je104

Table 7. The Decision List for "bank"

Collocation	POS	Sense	#Cor	#Err	R_L
Cl	n	Je104	21	0	4.44
Je	n	Je104	20	0	4.39
Jh	n	Je104	19	0	4.34
rob	n	Je104	11	0	3.81
money	n	Je104	10	0	3.71
account	n	Je104	9	0	3.61
river	n	Ld099	7	0	3.37
criminal	n	Je104	5	0	3.04
police	n	Je104	4	0	2.83
thief	n	Je104	4	0	2.83
interest	n	Je104	3	0	2.56
work	n	Je104	3	0	2.56

3.6 Experiments and Evaluation

Experiments were performed using two test sets of sentences containing 12 polysemous words used in recent WSD experiments (Yarowsky 1992; Luk 1995). Table 8 displays a word by word performance of the WSD algorithm. The results show that the algorithm can disambiguate 92% of the instances of these words with a precision rate of 94% in close test of examples drawn from LecDOCE. The open test consisting of sentences from the Brown corpus showed that 88% of the instances are disambiguated with a precision rate of 84%, compared to the precision rate of 77% for the Definition-Based Concept Co-occurrence (DBCC) approach by Luk (1995). Table 8 provides further details. These results show that with much less data and simpler training procedure, the method described above can offer higher precision in WSD for text containing many types of genres. However, there are still room for improvement in the area of coverage. Evidence have shown that by exploiting the constraint of so-called "one sense per discourse," (Yarowsky 1993) and the strategy of bootstrapping (Yarowsky 1995), it is possible to boost

coverage, while maintaining about the same level of precision. We are currently putting together an experiment to show that this observation holds out for our WSD method.

Table 8: performance of the WSD algorithm

Word	Test set #1: LecDOCE examples				Test set #2: Brown Corpus				
	Class-based WSD				Class-based WSD				DBCC (Luk)
	Total	#Cor	#Err	Prec	Total	#Cor	#Err	Prec	Precision Rate
bass	5	5	0	100%	15	14	1	93%	94%
bow	13	13	0	100%	21	17	4	81%	78%
cone	3	2	1	67%	12	12	0	100%	100%
duty	18	17	1	94%	77	74	3	96%	59%
galley	3	3	0	100%	8	6	2	75%	100%
interest	98	88	10	90%	79	53	26	68%	49%
issue	16	16	0	100%	67	47	20	69%	59%
mole	3	3	0	100%	2	1	1	50%	67%
sentence	41	41	0	100%	35	28	7	86%	84%
slug	5	4	1	80%	18	15	3	83%	80%
star	36	34	2	94%	40	37	3	93%	53%
taste	35	33	2	94%	57	57	0	100%	98%
Average	276	259	17	94%	431	361	70	84%	77%

4. Discussion

Work in sense disambiguation can be categorized by the way word senses are represented and how sense-indicative factors are acquired and recorded. Recently, researchers have turned to machine readable dictionaries (MRD) in order to save labor-intensive effort in representing word sense and WSD knowledge (McRoy 1992). The sense number in the on-line LDOCE dictionary has been used to represent word senses (Lesk 1986). This approach has the advantage of having definition and example sentences explicitly associated with a certain word sense, so proto-typical words in the indicative context of a certain word sense can be obtained directly. However, as noted by Dolan (1994), dictionary dichotomy of meaning is in general too fine to be adequate for word sense representation and disambiguation. Performance is understandably low (50-70%), given that word distinction is very fine-grained.

Gale, Church and Yarowsky (1992) reported a disambiguation method that uses bilingual corpus. Six words including "duty," "drug," "land," "language," "position," and "sentence" are studied. Their method relies on two translations in French for each word to tag training material for WSD. For instance, a sentence contains an instance of "duty" is tagged with the TAX sense, if the translation contains the French word "droit," and the OBLIGATION sense if the translation contains "devoir." The content words in the 100-word context of the ambiguous words are taken as contextual indicators of word sense. The average precision is 90% for two-way sense disambiguation. Dagan et al. (1994) also used translations in a bilingual lexicon to represent both word sense and contextual indicators. The authors show that the use of corpus in English is very effective in resolve sense ambiguity of sentences in Hebrew. Experimental results shows that 70% of the polysemous words are applicable for disambiguation. About 92% of these applicable polysemous words are disambiguated correctly.

Yarowsky (1992) used the major categories in Roget's thesaurus as its sense tags. The statistical

approach of the disambiguation method is the concordance set, the collection of the words in the 100-word context of a Roget category in Grolier's Encyclopedia. An average 92% accuracy rate was reported for 12 polysemous words for 3-way disambiguation on the average.

In most of the above-mentioned works, the sense division is either determined by hand or is taken directly from an existing source faulted with gaps. We have shown that the fine-grained senses in MRD can be used to fill gaps in the topical sets in LLOCE to arrive at a more complete and appropriate sense division for WSD. We experimented with the inclusion of conceptual relation for enhancement on coverage. Directly comparing methods is often difficult. Nevertheless, it is evident that in comparison our algorithms are simpler, take up less time and space overhead, and most importantly require no human intervention in all phases of WSD (sense division, tagging of training material, deduction of rules). This assumption seems to hold out, for experiments have shown that with a much smaller training corpus, the algorithms still provide broad-coverage sense disambiguation at comparable level of precision for text with many types of genres.

5. Conclusions and Future Work

The proposed method herein takes advantages of a number of linguistic phenomena: (1) Division of senses is primarily along the line of subject and topic. (2) Rather rigid schemes of text generation and predictable semantic relations are used to define senses in MRDs such as LDOCE. (3) The implicit links between instances of many of these relations are available in a thesaurus such as LLOCE. (4) Word senses seems to form semantic clusters that are effective knowledge sources for WSD.

This work also underscores the effectiveness of concept-based approach to WSD. Relation between concepts can be acquired efficiently from tagged bilingual corpora for broad and precise WSD.

Acknowledgment

The authors would like to thank the National Science Council of the ROC for financial support of this research under Contract No. NSC 85-2213-E-007-042.

References

1. Brown, P., S.A. Pietra, V.J.D. Pietra, and R. Mercer (1991), "Word Sense Disambiguation Using Statistical Methods," In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pp 264-270.
2. Chen, J.N., and Jason J.S. Chang (1994). "Towards Generality and Modularity in statistical Word Sense Disambiguation." In *Proceedings of the Asian Conference on Language, Information and Computation*, pp 45-48.
3. Cowie, Jim, Joe Guthrie and Louise Guthrie (1992). "Lexical disambiguation using simulated annealing." In *Proceedings of the International Conference on Computational Linguistics*, pp 359-365.

4. Dagan, Ido, Alon Itai and Ulrike Schwall (1991). "Two Language Are More Informative than One," In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pp 130-137.
5. Dagan, Ido, Alon Itai (1994), "Word sense disambiguation using a second language monolingual corpus," *Computational Linguistics*: 20(4), pp 563-596.
6. Dolan, W. B. (1994). "Word sense disambiguation - clustering related senses." In *Proceedings of the International Conference on Computational Linguistics*, pp. 712-716.
7. Gale, W., K. Church, and D. Yarowsky (1992), "Using Bilingual Materials to Develop Word Sense Disambiguation Methods," In *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation*, pp 101-112.
8. Krovetz, R. and Croft, W. B. (1992). "Lexical Ambiguity and Information Retrieval." *ACM Transaction on Information Systems*, pp 115-141.
9. Lesk, Michael E. (1986). "Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from a Ice-cream Cone," In *Proceedings of the ACM SIGDOC Conference*, Toronto, Ontario, pp 24-26.
10. Luk, Alpha K (1995). "Statistical sense disambiguation with relatively small corpora using dictionary definitions." In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp 181-188.
11. McArthur, Tom (1992). "Longman Lexicon of Contemporary English." *Longman Group (Far East) Ltd.*, Hong Kong.
12. McKeown, Katherine R. (1985). "Using discourse and focus constraints to generate natural language text." *Cambridge University Press*, Cambridge, England.
13. McRoy, Lucy (1992). "Using multiple knowledge sources for word sense discrimination." *Computational Linguistics* 18(1), pp 1-30.
14. Proctor, Paul (1988). "Longman English-Chinese Dictionary of Contemporary English." *Longman Group (Far East) Ltd.*, Hong Kong.
15. Schütze, Hinrich (1992). "Word Sense Disambiguation with Sublexical Representations," In *Proceedings of the 1992 AAAI Workshop on Statistically-based Natural Language Programming Techniques*, pp 100-104.
16. Yarowsky, David (1992). "Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora." In *Proceedings of the International Conference on Computational Linguistics*, pp 454-460.
17. Yarowsky, David (1993). "One Sense Per Collocation." In *Proceedings of the ARPA Human Language Technology Workshop*, Princeton.
18. Yarowsky, David (1995). "Unsupervised Word Sense Disambiguation Rivaling Supervised methods." In *Proceedings of Annual Meeting of the Association of Computational Linguistics*, pp 189-196.