Class Based Sense Definition Model for Word Sense Tagging and Disambiguation

Tracy Lin
Department of Communication Engineering
National Chiao Tung University,
1001, Ta Hsueh Road, Hsinchu, 300, Taiwan, ROC
tracylin@cm.nctu.edu.tw

Jason S. Chang
Department of Computer Science
National Tsing Hua University
101, Kuangfu Road, Hsinchu, 300, Taiwan, ROC
jschang@cs.nthu.edu.tw

Abstract

We present an unsupervised learning strategy for word sense disambiguation (WSD) that exploits multiple linguistic resources including a parallel corpus, a bilingual machine readable dictionary, and a thesaurus. The approach is based on Class Based Sense Definition Model (CBSDM) that generates the glosses and translations for a class of word senses. The model can be applied to resolve sense ambiguity for words in a parallel corpus. That sense tagging procedure, in effect, produces a semantic bilingual concordance, which can be used to train WSD systems for the two languages involved. Experimental results show that CBSDM trained on Longman Dictionary of Contemporary English, English-Chinese Edition (LDOCE E-C) and Longman Lexicon of Contemporary English (LLOCE) is very effectively in turning a Chinese-English parallel corpus into sense tagged data for development of WSD systems.

1. Introduction

Word sense disambiguation has been an important research area for over 50 years. WSD is crucial for many applications, including machine translation, information retrieval, part of speech tagging, etc. Ide and Veronis (1998) pointed out the two major problems of WSD: sense tagging and data sparseness. On one hand, tagged data are very difficult to come by, since sense tagging is considerably more difficult than other forms of linguistic annotation. On the other hand, although the data sparseness is a common problem, it is especially severe for WSD. The problems were attacked in various ways. Yarowsky (1992) showed a class-based approach under which a very large untagged corpus and thesaurus can be used effectively for unsupervised training for noun homograph disambiguation. However, the method does not offer a method that explicitly produces sense tagged data for any given sense inventory. Li and Huang (1999) described a similar unsupervised approach for Chinese text based on a Chinese thesaurus. As noted in Merialdo (1994), even minimal hand tagging improved on the results of unsupervised methods. Yarowsky (1995) showed that the learning strategy of bootstrapping from small tagged data led to results rivaling supervised training methods. Li and Li (2002) extended the approach by using corpora in two languages to bootstrap the learning process. They showed *bilingual bootstrapping* is even more effective. The bootstrapping approach is limited by lack of a systematic procedure of preparing seed data for any word in a given sense inventory. The approach also suffers from errors propagating from one iteration into the next. Li and Huang

Another alternative involves using a parallel corpus as a surrogate for tagged data. Gale, Church and Yarowsky (1992) exploited the so-called *one sense per translation constraint* for WSD. They reported high precision rates of a WSD system for two-way disambiguation of six English nouns based on their translations in an English-French Parallel corpus. However, when working with a particular sense inventory, there is no obvious way to know whether the one sense per translation constraint holds or how to determine the relevant translations automatically.

Diab and Resnik (2002) extended the translation-based learning strategy with a weakened constraint that many instances of a word in a parallel corpus often correspond to lexically varied but semantically consistent translations. They proposed to group those translations into a *target set*, which can be automatically tagged with correct senses

based on the hypernym hierarchy of WordNet. Diab and Resnik's work represents a departure from previous unsupervised approaches in that no seed data is needed and explicit tagged data are produced for a given sense inventory (WordNet in their case). The system trained on the tagged data was shown to be on a par with the best "supervised training" systems in SENSEVAL-2 competition. However, Diab and Resnik's method is only applicable to nominal WordNet senses. Moreover, the method is seriously hampered by noise and semantic inconsistency in a target set. Worse still, it is not always possible to rely on the hypernym hierarchy for tagging a target set. For instance, the relevant senses of the target set of {**serve**, **tee off**} for the Chinese counterpart "發球" [**faqiu**] do not have a common hypernym:

Sense 15
serve – (put the **ball** into play; as in **games** like tennis)
⇨ move – (have a turn; make one's move in a game)
Sense 1
Tee off – (strike a golf **ball** from a tee at the start of a **game**)
⇨ play – (participating in game or sports)
⇨ compete – (compete for something)

This paper describes a new WSD approach to simultaneously attack the problems of tagging and data sparseness. The approach assumes the availability of a parallel corpus of text written in *E* (the first language, L1[+]) and *C* (the second language, L2), an L1 to L2 bilingual machine readable dictionary *M*, and a L1 thesaurus *T*. A so-called *Mutually Assured Resolution of Sense* Algorithm (*MARS*) and Class Based Sense Definition Model (CBSDM) are proposed to identify the word senses in *I* for each word in a semantic class of words *L* in *T*. Unlike Diab and Resnik, we do not apply the MARS algorithm directly to *target sets* to avoid the noisy words therein. The derived classes senses and their relevant glosses in L1 and L2 make it possible to build Class Based Sense Definition and Translation Models (CBSDM and CBSTM), which subsequently can be applied to assign sense tags to words in a parallel corpus.

The main idea is to exploit the defining L1 and L2 words in the glosses to resolve the sense ambiguity. For instance, for the class containing "*serve*" and "*tee off*," the approach exploits common defining words, including "*ball*" and "*game*" in two relevant **serve**-15 and **tee off**-1 to assign the correct senses to "*serve*" and "*tee off*." The character bigram "發 球" [**faqiu**] in an English-Chinese MRD:

> **serve** v **10** [I∅; T1] to begin play by striking (the ball) to the opponent **發球** (LDOCE E-C p. 1300),

would make it possible to align and sense tag "*serve*" or "*tee off*" in a parallel corpus such as the bilingual citations in Example 1:

(1C) 發球前先喝一小碗藥酒。
(1E) drink a capful before **teeing off** at each hole. (Source: *Sinorama, 1999, Nov. Issue, p.15, Who Played the First Stroke?*).

That effectively attaches semantic information to bilingual citations and turns a parallel corpus into a Bilingual Semantic Concordance (BSC). The BSC enables us to simultaneously attack two critical WSD problems of sense tagging difficulties and data sparseness, thus provides an effective approach to WSD. BSC also embodies a projection of the sense inventory from L1 onto L2, thus creates a new sense inventory and semantic concordance for L2. If *I* is based on WordNet for English, it is then possible to obtain an L2 WordNet. There are many additional applications of BSC, including bilingual lexicography, cross language information retrieval, and computer assisted language learning.

The remainder of the paper is organized as follows: Sections 2 and 3 lay out the approach and describe the MARS and SWAT algorithms. Section 4 describes experiments and evaluation. Section 5 contains discussion and we conclude in Section 6.

2. Class Based Sense Definition Model

We will first illustrate our approach with an example. A formal treatment of the approach will follow in Section 2.2.

2.1 An example

[+] This has nothing to do with the direction of translation and is not to be confused with the native and second language distinction made in the literature of Teaching English As a Second Language (TESL) and Computer Assisted Language Learning.

To make full use of existing machine readable dictionaries and thesauri, some kind of linkage and integration is necessary (Knight and Luk, 1994). Therefore, we are interested in linking thesaurus classes and MRD senses: Given a thesaurus class *S*, it is important that the relevant senses for each word *w* in *S* is determined in a MRD-based sense inventory *I*. We will show such linkage is useful for WSD and is feasible, based solely on the words of the glosses in *I*. For instance, given the following set of word (N060) in Longman Lexicon of Contemporary English (McArthur 1992):

L = **{difficult**, **hard**, **stiff**, **tough**, **arduous**, **awkward}.**

Although those words are highly ambiguous, the juxtaposition immediately brings to mind the relevant senses. Specifically for the sense inventory of LDOCE E-C, the relevant senses for *L* are as follows:

- **difficult** *adj* **1**. not easy; hard to do, make, understand, etc. 不容易 [bu rongyi] ; 難 [nan]
- **hard** *adj* **2**. difficult (to do or understand) 難的 [nan de]; 困難的 [kunnan de]
- **stiff** *adj* **6**. difficult to do 難做的 [nanzuo de] ; 棘手的 [jishou de]
- **tough** *adj* **4**. difficult to do; not easy; demanding effort 難做的 [nanzuo de] ; 費力的 [feili de]
- **arduous** *adj* **1**. needing much effort; difficult 費力的 [feilide] ; 艱難的 [jian-nan de]
- **awkward** *adj* **2**. not well made for use; difficult to use; causing difficulty 難於使用的 [nan yu shiyong de] ; 不便的 [buban de]

Therefore, we have the intended senses, *S*

S = {**difficult-1**, **hard-2**, **stiff-6**, **tough-4**, **arduous-1**, **awkward-2}**.

It is reasonable to assume each sense in *I* is accompanied by a sense definition written in the same language (L1). We use D(*S*) to denote the glosses of *S*. Therefore we have

D(*S*) = "not easy; hard to do, make, understand, etc.; difficult to do or understand; difficult to do; difficult to do; not easy; demanding effort; needing much effort; difficult; not well made for use; difficult to use; causing difficulty;"

The intuition of bringing out the intended senses of semantically related words can be formalized by Class Based Sense Definition Model (CBSDM), which is a *micro* language model generating D(*S*), the glosses of *S* in *I*. For simplicity, we assume an unigram language model P(*d*) that generates the content words *d* in the glosses of *S*. Therefore, we have

D(*S*) = "easy hard do make understand difficult do understand difficult do difficult do easy demanding effort needing much effort difficult well made use difficult use causing difficulty"

If we have the relevant senses, it is a simple matter of counting to estimate P(*d*). Conversely, with P(*d*) available to us, we can pick the relevant sense of *S* in *I* which is most likely generated by P(*d*). The problem of learning the model P(*d*) lend itself nicely to an iterative relaxation method such as the Expectation and Maximization Algorithm (Dempster, Laird, Rubin, 1977).

Initially, we assume all senses of *S* word in *I* is equally likely and use all the defining words therein to estimate P(*d*) regardless of whether they are relevant. For LDOCE senses, initial estimate of the relevant glosses is as follows:

D(*S*) = "easy hard do make understand people unfriendly quarrelling pleased ... firm stiff broken pressed bent difficult do understand forceful needing using force body mind ...bent painful moving moved ... strong weakened suffer uncomfortable conditions cut worn broken ...needing effort difficult lacking skill moving body parts body CLUMSY made use difficult use causing difficulty"

Table 1. The initial CBSDM for n-word list {difficult, hard, stiff, tough, arduous, awkward} based on the relevant and irrelevant LDOCE senses, $n = 6$.

Defining word *d*	Count, *k*	P(*d*) = *k*/*n*
Difficult	5	0.83
Effort	3	0.50
Understand	2	0.33
Bad	2	0.33
Bent	2	0.33
Body	2	0.33
Broken	2	0.33
Difficulty	2	0.33
Easy	2	0.33
Firm	2	0.33
Hard	2	0.33
Moving	2	0.33
Needing	2	0.33
Water	2	0.33

As evident from Table 1, the initial estimates of P(*d*) are quite close to the true probability distribution (based on the relevant senses only). The three top ranking defining words "*difficult*," "*effort*," and "*understand*" appear in glosses of relevant senses,

and not in irrelevant senses. Admittedly, there are still some noisy, irrelevant words such as "*bent*" and "*broken.*" But they do not figure prominently in the model from the start and will fade out graduately with successive iterations of re-estimation. We estimate the probability of a particular sense *s* being in *S* by P(D(*s*)), the probability of its gloss under P(*d*). For intance, we have

P(**hard**-1) = P(D(**hard**-1)) = P("firm and stiff; which ..."),

P(**hard**-2) = P(D(**hard**-2)) = P("difficult to do or understand").

On the other hand, we re-estimate the probability P(*d*) of a defining word *d* under CBSDM by how often *d* appears in a sense *s* and P(*s*). P(*d*) is positively prepositional to the frequency of *d* in D(*s*) and to the value of P(*s*). Under that re-estimation scheme, the defining words in relevant senses will figure more prominently in CBSDM, leading to more accurate estimation for probability of *s* being in *S*. For instance, in the first round, "*difficult*" in the gloss of **hard**-2 will weigh twice more than "*firm*" in the gloss of irrelevant **hard**-1, leading to relatively higher unigram probability for "*difficult*." That in turn makes **hard**-2 even more probable than **hard**-1. See Table 2.

Table 2. First round estimates for P(*s*), the probability of sense *s* in *S*.

Sense*	Definition	P(*s*)
hard-1	**firm** and stiff; which cannot easily be broken	0.2857
hard-2	**difficult** to do or understand	0.7143
stiff-1	not easily bent	0.2857
stiff-6	**difficult** to do	0.7143

* in LDOCE.

** Assuming $P(s) \approx \max_{d \in D(s)} P(d)$

Often the senses in *I* are accompanied with glosses written in a second language (L2); exclusively (as in a simple bilingual word list) or additionally (as in LDOCE E-C). Either way, the words in L2 glosses can be incorporated into D(*s*) and P(*d*). For instance, the character unigrams and/or overlapping bigrams in the Mandarin glosses of *S* in LDOCE E-C and their appearance counts and probability are shown in Table 3.

Table 3. Classes Based Sense Translation Model for **{**difficult-1, hard-2, stiff-6, tough-4, arduous-1, awkward-2**}** in LDOCE*.

Translation	Count	Prob	Translation	Count	Prob
難	6	1.00	棘	1	0.17
力	2	0.33	艱	1	0.17
不	2	0.33	難做	3	0.50
做	2	0.33	費力	2	0.33
費	2	0.33	不容	2	0.33
手	1	0.17	不便	1	0.17
用	1	0.17	困難	1	0.17
困	1	0.17	使用	1	0.17
使	1	0.17	於使	1	0.17
於	1	0.17	容易	1	0.17
易	1	0.17	棘手	1	0.17
便	1	0.17	艱難	1	0.17
容	1	0.17			

* After removing the stop word 的 [**de**]

We call the part of CBSDM that are involved with words written in L2, *Class Based Sense Translation Model*. CBSTM trained on a thesaurus and a bilingual MRD can be exploited to align words and translation counter part as well as to assign word sense in a parallel corpus. For instance, given a pair of aligned sentences in a parallel corpus:

(2C) 一位接近李約瑟的學者分析他所以能成就如此巨構，實有旁人**難**及的條件。

(2E) A scholar close to Needham analyses the reasons that he was able to achieve this huge work as being due to a combination of factors that would be **hard** to find in any other person. (*Source: 1990, Dec Issue Page 24, Giving Justice Back to China --Dr. Joseph Needham and the History of Science and Civilisation in China)*

It is possible to apply CBSTM to obtain the following pair of translation equivalent, (難 [**nan**], "hard") and, at the same time, determine the intended sense. For instance, we can label the citation with **hard**-2_{LDOCE}, leading to the following quadruple:

(3) (**hard**, 難 [nan], **hard**-2_{LDOCE}, (2C, 2E))

After we have done this for all pairs of word and translation counterpart, we would in effect establish a Bilingual Semantic Concordance (BSC).

2.2 The Model

We assume that there is a Class Based Sense Definition Model, which can be viewed as a language model that generates the glosses for a class of senses *S*. Assume that we are given *L*, the words of *S* but not explicitly the intended senses *S*. In addition, we are given a sense inventory *I* in the form of an MRD with the regular glosses, which are written in L1 and/or L2. We are concerned with two problems: (1) Unsupervised training of *M*, CBSDM for *S;* (2) Determining *S* by identifying a relevant sense in *I*, if existing, for each word in *L*. Those two problems can be solved based on Maximum Likelihood Principle: Finding *M* and *S* such that *M generates* the glosses of *S* with maximum probability. For that, we utilize the Expectation and Maximization Algorithm to derive *M* and *S* through *Mutually Assured Resolution of Sense Algorithm* (MARS) given below:

Mutual Assured Resolution of Sense Algorithm

Determine the intended sense for each of a set of semantic related words.

Input: (1) Class of words $L = \{w_1\ w_2 \ldots w_n\}$;
(2) Sense inventory *I*.

Output: (1) Senses *S* from *I* for words in *L;*
(2) CBSTM *M* from L1 to L2.

1. Initially, we assume that each of the senses $w_{i,j}$, j = 1, m_i in *I* is equally probable to be in *S* with probability $P(w_{i,j} \mid i,L) = \frac{1}{m_i}$, j = 1, m_i; where m_i is the number of senses in *I* for the word w_i.
2. Estimate CBSDM P($d \mid L$) for *L* ,
$$P(d \mid L) = \frac{\sum_i \max_{j,k} P(w_{i,j} \mid i,L) EQ(d, d_{i,j,k})}{n},$$
where *d* is a unigram or overlapping bigram in L1 or L2, $d_{i,j,k}$ = the *k*th word in D($w_{i,j}$), and *EQ*(*x*, *y*) = 1, if $x = y$ and 0 otherwise;
3. Re-estimate P($w_{i,j} \mid i,L$) according to $d_{i,j,k}$, $k = 1, n_{i,j}$:
$$P_1(w_{i,j} \mid i,L) = 0.5 \max_k P(d_{i,j,k} \mid L) + 0.5 \sum_k \frac{1}{n_{i,j}} P(d_{i,j,k} \mid L),$$
$$P(w_{i,j} \mid i,L) = \frac{P_1(w_{i,j} \mid i,L)}{\sum_{j=1,\ m_i} P_1(w_{i,j} \mid i,L)};$$
4. Repeat Steps 2 and 3 until the values of P($d \mid L$) and P($w_{i,j} \mid i, L$) converge;
5. For each *i*, find the most probable sense w_{i,j^*} , j^*=argmax$_j$ P($w_{i,j} \mid i, L$) ;
6. Output $S = \{ w_{i,j^*} \mid j^* = \text{argmax}_j P(w_{i,j} \mid i, L)\}$;
7. Estimate and output CBSTM for *L*,
$$P(c \mid L) = \frac{\sum_{i=1,n} I(c \in t_{i,j^*})}{n},$$
where *c* is a unigram or overlapping bigram in L2 and $t_{i,j}$ is the L2 gloss of $w_{i,j}$.

Note that the purpose of Step 2 is to estimate how likely a word will appear in the definition of S based on the defininig word for the senses, $w_{i,j}$ and relevant probability P($w_{i,j} \mid i,L$). This likelihood of the word *d* being used to define senses in questions is subsequently used to re-estimate P($w_{i,j} \mid i,L$), the likelihood of the *j*th sense, $w_{i,j}$ of w_i being in the intended senses of *L*.

3. Application to Word Sense Tagging

Armed with the Class Based Sense Translation Model, we can attack the word alignment and sense tagging problems simultaneously. Each word in a pair of aligned sentences in a parallel corpus will be considered and assigned a counterpart translation and intended sense in the given context through the proposed algorithm below:

Simutaneous Word Alignment and Tagging Algorithm (SWAT)

Align and sense tag words in a give sentence and translation.

Input: (1) Pair of sentences (*E*, *C*);
(2) Word *w*, POS *p* in question;
(3) Sense Inventory *I;*
(4) CBSTM, P($c|L$).

Output: (1) Translation *c* of *w* in *C;*
(2) Intended sense *s* for *w.*

1. Perform part of speech tagging on *E;*
2. Proceed if *w* with part of speech *p* is found in the results of tagging *E;*
3. For all classes *L* to which (*w*, *p*) belongs and all words *c* in *C*:
$$L^* = \arg\max_L \left(\max_{LINK(w,c)} P(c \mid L) \right),$$
$$c^* = \arg\max_c \left(P(c \mid L^*) \right),$$
where LINK(*x*, *y*) means *x* and *y* are two word aligned based on Competitive Linking Alignment
4. Output *c** as the translation;
5. Output the sense of *w* in *L** as the intended sense.

To make sense tagging more precise, it is advisable to place constraint on the translation counterpart *c* of *w*. SWAT considers only those translations *c* that has been linked with *w* based the Competitive

Linking Algorithm (Melamed 1997) and logarithmic likelihood ratio (Dunning 1993).

Table 4. The experimental results of assigning LDOCE senses to classes of LLOCE.

Word	Class	pos	Sense**	Ex
Star	K082	N	**8** a famous or very skilful performer. 明星；主角；名角	Y
Star	L002	N	**1** a brightly-burning heavenly body of great size, … 星	Y
Interest	F028	N	**2** a readiness to give attention. … 趣味；引起注意之性質	Y
Interest	F228	N	**1** a readiness to give attention 興趣	Y
Interest	J112	N	**6** money paid for the use of money 利息	Y
Interest	K006	N	**3** an activity, subject, etc., which one gives time and attention to 所愛好之事物；嗜好	Y
Issue	A020	N	**7** children 孩子；子嗣	Y
Issue	G180	N	**4** something printed, brought out again … 發行之印刷物	Y
Issue	G243	N	**2b** something which comes or is given out 流來或給出之事物	N
Issue	N153	N	**6** the result 結果	Y
Serve	C263	V	**9** to spend (a period of time) in prison 服刑	Y
Serve	D108	V	**5** to be good enough or satisfying for… 適合	N
Serve	E015	V	**7** to offer (food, a meal, etc.) for eating 送（食物、餐食等）	Y
Serve	I028	V	**1** to work (faithfully) for; do a useful job for 服務；為效力	N
Hard	N060	A	**2** difficult (to do or understand) 難的；困難的	Y
Hard	N264	A	**1** firm and stiff; … 堅硬	Y

4. Experiments and evaluation

In order to assess the feasibility of the proposed approach, we carried out experiments and evaluation on an implementation of MARS and SWAT based on LDOCE E-C, LLOCE, and Sinorama.

First experiment was involved with the trainability of CBSDM and CBSTM via MARS. The second experiment was involved with the effectiveness of using SWAT and CBSTM to annotate a parallel corpus with sense information. Evaluation was done on a set of 14 nouns, verbs, adjectives, and adverbs studies in previous work. The set includes the nouns "*bass*," "*bow*," "*cone*," "*duty*," "*gallery*," "*mole*," "*sentence*," "*slug*," "*taste*," "*star*," "*interest*," "*issue*," the adjective "*hard*," and the verb "*serve*."

Table 5. Evaluation of the MARS Algorithm based on 12 nouns, 1 verb, 1 adjective in LDOCE.

Word	Pos	#Senses	#Done	#Correct	Prec (LB*)	Prec.
Bass	N	4	1	1	0.25	1.00
Bow	N	5	2	2	0.25	1.00
Cone	N	3	3	2	0.33	0.67
Duty	N	2	2	2	0.13	1.00
Galley	N	3	3	2	0.33	0.67
Mole	N	3	2	2	0.33	1.00
Sentence	N	2	2	2	1.00	1.00
Slug	N	2	2	2	0.20	1.00
Taste	N	6	1	1	0.17	1.00
Star	N	8	2	2	0.13	1.00
Interest	N	6	4	4	0.17	1.00
Issue	N	7	4	3	0.14	0.75
Serve	V	13	4	2	0.08	0.50
Hard	A	12	2	2	0.08	1.00
Avg.		4.14	1.36	1.29	0.26	0.90

* The lower bound of precision of picking one sense in random.

Table 6. Experimental results of sense tagging the Sinorama parallel Corpus.

Word	Instance	#done	#correct	Precision
Star	173	86	82	0.95
Hard	325	37	33	0.89

4.1 Experiment 1: Training CBSDM

We applied MARS to assign LDOCE senses to word classes in LLOCE. Some results related to the test set are shown in Tables 4. The evaluation in Tables indicates that MARS assigns LDOCE senses to an LLOCE class with a high average precision rate of 90%.

4.2 Experiment 2: Sense Tagging

We applied SWAT to sense tag English words in some 50,000 reliably aligned sentence pairs in Sinorama parallel Corpus based on LDOCE sense inventory. The results are shown in Tables 6. Evaluation indicates an average precision rate of around 90%.

5. Discussion

The proposed approach offers a new method for automatic learning for the task of word sense disambiguation. The class based approach attacks the problem of tagging and data sparseness in a way similar to the Yarowsky approach (1992) based on

thesaurus categories. We differ from the Yarowsky's approach, in the following ways:

i. The WSD problem is solved for two languages instead of one within a single sense inventory. Furthermore, an explicit sense tagged corpus is produced in the process.
ii. It is possible to work with any number of sense inventories.
iii. The method is applicable not only to nouns but also to adjectives and verbs, since it does not rely on topical context, which is effective only for nouns as pointed out by Towell and Voorhees (1998).

The approach is very general and modular and can work in conjunction with a number of learning strategies for word sense disambiguation (Yarowsky, 1995; Li and Li, 2002).

The approach is limited by the comprehensiveness of L2 glosses in MRD to cover in-context translations. For instance, the translation equivalent ("hard", "吃力") in (3) is not covered by CBSTM for N060 words trained on LDOCE E-C.

(3C) 不少雄中學生念起現在的課本都覺得很吃力
(3E) At Kaohsiung High, … quite a few of the students find the textbooks hard to follow. (*Source: Sinorama, 2000, May Issue Page 4, Star Wars--The Controversy over Elite High Schools)*

That limitation can be partially alleviated by a smoothing technique of backing off from the probability of rare translation to that of its more frequent synonym in an L2 thesaurus. For instance, P("吃力" | N060) = 0 under maximum likelihood estimation, but using the back-off scheme, we have

P("吃力"|N060) = *c* P("困難"|N060),
for some constant *c*.

By backing off to the probability of "困難," the most frequent synonym of "吃力" (Mei, et al. 1984), we are able to correctly align and tag "*hard*" In Example 3.

6. Conclusion

In this paper, we present the Mutual Assured Resolution of Sense (MARS) Algorithm for assigning relevant senses to word classes in a given sense inventory (i.e. LDOCE or WordNet). We also describe the SWAT Algorithm for automatic sense tagging of a parallel corpus.

We carried out experiments on an implementation of the MARS and SWAT Algorithms for all the senses in LDOCE and LLOCE. Evaluation on a set of 14 highly ambiguous words showed that very high precision CBSDM and CBSTM can be constructed. High applicability and precision rates were achieved, when applying CBSTM to sense tagging of a Chinese-English parallel corpus.

A number of interesting future directions present themselves. First, it would be interesting to see how effectively we can broaden the coverage of CBSTM via backing off smoothing. Second, a CBSTM trained directly on a parallel corpus would be more effective in word alignment and sense tagging. The approach of training CBSTM on the L2 glosses in a bilingual MRD may lead to occasional mismatch between MRD translations and in-context translations. Third, there is a lack of research for a more abstractive and modular representation of sense differences and commonality. There is potential of developing Sense Definition Model to identify and represent semantic and stylistic differentiation reflected in the MRD glosses pointed out in DiMarco, Hirst and Stede (1993). Last but not the least, it would be interesting to apply MARS to both LDOCE E-C and WordNet and project WordNet's sense inventory to a sencond language via CBSDM and a parallel corpus, thus creating a Chinese WordNet and semantic concordance.

Acknowledgement

We acknowledge the support for this study through grants from National Science Council and Ministry of Education, Taiwan (NSC 90-2411-H-007-033-MC and MOE EX-91-E-FA06-4-4).

References

Dagan, Ido; A. Itai, and U. Schwall (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.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38.

Diab, M. and P. Resnik, (2002). An Unsupervised Method for Word Sense Tagging using Parallel Corpora, *Proceedings of ACL*, 255-262.

DiMarco, C., G. Hirst, M. Stede, (1993). "The semantic and stylistic differentiation of synonyms and near-

synonyms." In: *Working notes of the AAAI Spring Symposium on Building Lexicons for Machine Translation.* Stanford University.

Dunning, T (1993) Accurate methods for the statistics of surprise and coincidence, Computational Linguistics 19:1, 61-75.

Gale, W., K. Church, and D. Yarowsky, (1992). Using Bilingual Materials to Develop Word Sense Disambiguation Methods. In *Proceedings, Fourth International Conference on Theoretical and Methodological Issues in Machine Translation.* Montreal, 101-112, 1992.

Ide, N. and J. Véronis (1998). Word sense disambiguation: The state of the art. *Computational Linguistics*, 24:1, 1-40.

Knight, K, and A. Luk, (1994). Building a Large-Scale Knowledge Base for Machine Translation, *Proc. of the National Conference on Artificial Intelligence (AAAI)*.

Knight, K., I. Chander, M. Haines, V. Hatzivassiloglou, E. Hovy, M. Iida, S. Luk, A. Okumura, R. Whitney, K. Yamada, (1994). "Integrating Knowledge Bases and Statistics in MT, *Proc. of the Conference of the Association for Machine Translation in the Americas (AMTA)*.

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

Li, C, and H. Li (2002). Word Translation Disambiguation Using Bilingual Bootstrapping, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, July 2002, 343-351.

Li, Juanzi and C. Huang (1999). A Model for Word Sense Disambiguation. In Computational Linguistics and Chinese Language Processing,4(2), August 1999, pp.1-20

McArthur, T. (1992) *Longman Lexicon of Contemporary English*, Longman Group (Far East) Ltd., Hong Kong.

Mei, J. J., et al. (1984) Tongyici Cilin, Shanghai, Commercial Press. (in Chinese)

Melamed, I.D. (1997). "A Word-to-Word Model of Translational Equivalence". In Procs. of the ACL97. pp 490-497. Madrid Spain.

Merialdo, B, (1994). Tagging English Text with a Probabilistic Model, *Computational Linguistics*, 20(2):155-171.

Miller, G., A, R.. Beckwith, C. Fellbaum, D. Gross and K.J. Miller. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography,* 3(4), 235- 244.

Proctor, P. (1988) Longman English-Chinese Dictionary of Contemporary English, Longman Group (Far East) Ltd., Hong Kong.

Towell, G. and E. Voorhees. (1998) Disambiguating Highly Ambiguous Words. *Computational Linguistics*, vol. 24, no. 1, 125-146.

Yarowsky, D. (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, D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, 189-196