[TMI 99: *Proceedings of 8th International Conference on Theoretical and Methodological Issues in Machine Translation*, August 1999, Chester, UK]

Lexical Selection with a Target Language Monolingual Corpus and an MRD

Hyun Ah Lee, Jong C. Park and Gil Chang Kim
Dept. of Computer Science
Korean Advanced Institute of Science and Technology
373-1, Kusong-Dong, Yusong-District
Taejon, 305-701 Republic of Korea
halee@csone.kaist.ac.kr, {park,gckim}@cs.kaist.ac.kr

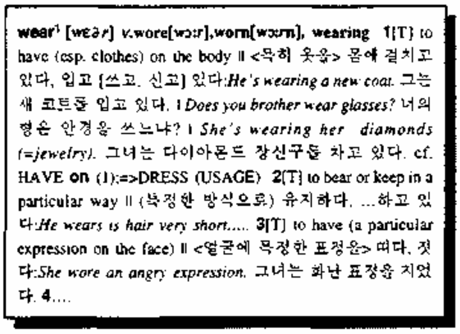
Abstract

In this paper, we propose a lexical selection method with three steps: sense disambiguation of source words, sense-to-word mapping, and selection of the most appropriate target language lexical item. The knowledge for each step is extracted from a machine readable dictionary and a target language monolingual corpus. By splitting the process of lexical selection into three steps and extracting the essential knowledge for each step from existing resources, our system can select appropriates word for translation with high extensibility and robustness.

1 Introduction

Lexical selection in a transfer-based machine translation system is a process that selects an appropriate target word corresponding to a source word. Like other problems in MT, knowledge acquisition is crucial for lexical selection, so many researches have attempted to extract knowledge from existing resources. Among them, corpus based approaches select a target word using statistic information that is automatically extracted from corpora. Some of the corpus-based approaches use a bilingual corpus as a knowledge source to extract statistic information (Brown et al. 1991). Although a bilingual corpus is a good resource for retrieving translation knowledge, these techniques are not preferred in general since such resources are difficult to obtain. In other corpus-based approaches that use a target language monolingual corpus as a knowledge source (Dagan & Itai 1994), a selected lexical item could have an inappropriate sense, since they ignore the word sense in a source language sentence and use only the target language information. Some researches have used a machine readable dictionary (MRD) as a knowledge source. Copestake et al (1994) attempted to acquire relations of lexical translation from MRDs. However, their method is applicable/useful only when a large knowledge base (LKB) already exists. Klavans & Tzoukermann (1996) used both MRDs and corpora to build a bilingual lexicon. But it is an open-question that their proposal is applicable for extracting the knowledge for lexical selection, since they used a bilingual corpus and are mostly concerned with the motion verbs.

A word can have many senses and each of those senses can be mapped into many target language words. On the basis of this observation, we propose a three-step lexical selection method and an automatic knowledge extraction method. The first step of lexical selection disambiguates word senses in source language sentences using rules for



wear[1] [wεər] v.wore[wɔːr],worn[wɔːrn], wearing **1**[T] to have (esp. clothes) on the body ‖ <특히 옷을> 몸에 걸치고 있다, 입고 [쓰고, 신고] 있다:*He's wearing a new coat.* 그는 새 코트를 입고 있다. | *Does you brother wear glasses?* 너의 형은 안경을 쓰느냐? | *She's wearing her diamonds (=jewelry).* 그녀는 다이아몬드 장신구를 차고 있다. cf. HAVE **on** (1);=>DRESS (USAGE) **2**[T] to bear or keep in a particular way ‖ (특정한 방식으로) 유지하다, ...하고 있다:*He wears is hair very short.....* **3**[T] to have (a particular expression on the face) ‖ <얼굴에 특정한 표정을> 띠다, 짓다:*She wore an angry expression.* 그녀는 화난 표정을 지었다. **4**....

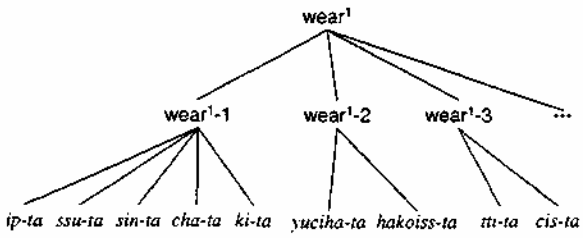


Figure 1: Part of LDEEK about a verb, 'wear'

word sense disambiguation (WSD). Source language WSD rules are extracted from an MRD. In the second step, the selected sense of a source language word is mapped into a set of target language words. The mapping information from a source language word sense to a set of target language words can also be extracted from an MRD. A target language word that is properly situated in the target language sentence is selected in the third step. To select a target lexical item, we use "collocating lexical items" (henceforth collocations) in target language sentences with syntactic relations and syntactic-relation weights. Target language collocations and their syntactic relations are extracted from a tree annotated target language monolingual corpus.

The rest of the paper is organized as follows. We introduce our lexical selection method in Section 2, and explain how to extract knowledge from a corpus and an MRD in Section 3. We present our machine translation system in Section 4, show experimental results in Section 5, and conclude in Section 6. In the paper, we use examples in English-to-Korean translation, and use the Yale romanization system to transcribe Korean.

2 3-step Lexical Selection

The figure on the left in Figure 1 shows part of the English-English-Korean Dictionary (LDEEK 1993).[1,2] Notice that a word can have many different senses and each of those senses can be mapped to many target language words. On the basis of this observation, we propose a new method for lexical selection, which consists of three steps: sense disambiguation of source words, sense-to-word mapping, and selection of a target language lexical item.

2.1 Source Word Sense Disambiguation and Word Mapping

In the first step, our method disambiguates source language word senses using the knowledge in a lexicon.

[1] The English-English-Korean Dictionary that is used here is an English-English dictionary, where English definitions and example sentences are paired with Korean translations.

[2] In this paper, we use the convention that the j-th meaning of the i-th dictionary entry of X is denoted by X^i-j.

Table 1: Part of WSD knowledge for 'wear' and 'old'

	relation	collocation	word sense	mapped target words
wear	**object**	coat, glasses	wear^1-1 (put on)	*ip-ta, ssu-ta, sin-ta, cha-ta*
		expression	wear^1-3 (express)	*tti-ta, cis-ta*
old	modifiee	man, book	old^1-2 (not young)	*nai tun, nonyen-uy, nulk-un*
		shoe, car	old^1-3 (not new)	*nalk-un, hen*

Based on "one sense per one collocation" (Yarowsky 1993), we use collocations of a source word to disambiguate its sense. And, for more precise disambiguation, we additionally use syntactic relation information between a source word and its collocation. Collocations and their syntactic relation are described in a MT lexicon.

(1) 'He wears an angry expression.'

그는 화난 표정을 짓는다.

ku-nun hwana-n phyoceng-ul cis-nun-ta

he-TOP angry-MOD expression-ACC wear-PRESENT-DECL

(2) 'She wears old shoes.'

그녀는 낡은 신발을 신는다.

kunye-nun nalk-un sinpal-ul sin-nun-ta.

she-TOP old-MOD shoes-ACC wear-PRESENT-DECL

In (1), the meaning of 'wear' is "to have (a particular expression on the face)", but in (2), the meaning of 'wear' is "to have on the body". The meaning of 'old' in (2) is "not new", but that in the phrase 'the oldest student' is "not young". Table 1 shows a piece of WSD knowledge in our lexicon for the verb 'wear' and the adjective 'old'. The object of the verb 'wear' in (2) is 'shoes'. By calculating the similarity between 'shoes' and its collocations in Table 1, we can determine that the 'wear' in (2) has the sense wear^1-1. We use WordNet for calculating word similarity. WSD knowledge in a lexicon is automatically extracted from an MRD. We will explain our knowledge extracting method in the next section.

In the second step, the selected source language word sense is mapped to a set of target words that are synonymous but have different usages or nuances. The last column of Table 1 shows mapping information from source word senses to their target words. This mapping information is also extracted from an MRD. A word in the set, which is considered the correct target word of the source word, is selected in the third step.

2.2 Word Selection using Statistic Information of Target Language

The third step selects, among the target language words, one that properly embodies a target language sentence. A target word that frequently occurs in a similar target language context should be an appropriate target word of the source word sense. The target language context can be modeled by collocation words, among others. We use statistic information of target language collocations to select an appropriate target word.

Table 2: Sense of the object of wear[1]-1 and its appropriate target word

	body-wear	head-wear	face-wear	hand-wear	foot-wear	arm-wear
wear[1]-1	*sin-ta*	*ssu-ta*	*kki-ta*	*kki-ta*	*sin-ta*	*cha-ta*

Table 3: Part of cooccurrence statistic information

	"clothes"		"hat"	"glasses"	"glove"	"shoes"			"watch"
	os	*uypok*	*moca*	*ankyeng*	*cangkap*	*kwutwu*	*sinpal*	*sin*	*sikyey*
sin-ta	36	1	-	-	-	-	-	-	-
ssu-ta	1	-	8	11	-	-	-	-	-
kki-ta	-	-	-	2	2	-	-	-	-
sin-ta	-	-	-	-	-	2	5	1	-
cha-ta	-	-	-	-	-	-	-	-	1

For example, one of the senses of 'wear', wear[1]-1, is mapped to many target words or phrases: *mom-ey kelchi-ta, ip-ta, ssu-ta, sin-ta, chata,* etc. In Korean, if the object of the sense wear[1]-1 is a foot-wear, such as *kwutwu* "shoes" or *yangmal* "socks", the sense wear[1]-1 should be mapped to *sin-ta.* If the object of wear[1]-1 is a head-wear such as *moca* "hat", the sense should be translated to *ssu-ta.* Table 2 shows target words of wear[1]-1 and some senses of words which frequently co-occur as objects of these target words. And Table 3 shows part of the frequencies of collocations of target words of wear[1]-1 and their object words. From these two tables, we can find out that a collocation in a certain syntactic relation is a good indicator for lexical selection.

The statistic information of target language collocations is the frequency of each syntactic tuple in a target language corpus. A syntactic tuple denotes a syntactic relation between two or more words (Dagan & Itai 1994). Dagan & Itai (1994) use a source language syntactic relation in making syntactic tuples of target language due to the difficulty of transfer. But we make a syntactic tuple with a syntactic relation of the target language, because the third step of lexical selection comes after structural transfer.[3] Two target language words in a syntactic tuple are elements of the set of target language words that are mapped in the second step. Table 4 shows syntactic tuples for Korean and their frequencies of target words of wear[1]-1 in the sentence (2).[4]

Using statistic information of collocations, we estimate the probability *p(a,i)* of selecting a target word a_i as a word for translation. Let us assume that the selected word sense is a. We want to select one word from $a_1, a_2, \ldots, a_k$, the set of target words of the sense a. Let us denote a set of nodes that are connected in a sentence structure of target language to the node of the sense a as $\Theta(a)$. If we denote an element of the set Θ (a) as the node (a, m, γ), this means that the sense of the node is α, which can be in turn translated to m target words, $\alpha_1, \alpha_2, \ldots, \alpha_m$, and the node is connected with the syntactic relation γ to the node with the sense a. Therefore, this node represents a collocation of the sense a, and (γ a_i a_l) is a syntactic tuple on the sense a_i. Suppose $f(a_i, a_l, \gamma)$ is the number of the syntactic tuple (γ a_i a_l) in a target language corpus.

[3] We will discuss the transfer module in Section 4.

[4] In the table, the character '*' indicates an arbitrary word or relation.

Table 4: Syntactic tuples and their frequencies of wear[1]-1 of sentence (2)

(verb-obj *ssu-ta sin*)	0	(verb-obj *sin-ta sin*)	1	(verb-subj *kki-ta kunye*)	1
(verb-obj *ssu-ta sinpal*)	0	(verb-obj *sin-ta sinpal*)	5	(verb-obj *kki-ta* *)	0
(verb-obj *ssu-ta kwutwu*)	0	(verb-obj *sin-ta kwutwu*)	2	(* *ip-ta* *)	0
(verb-subj *ssu-ta kunye*)	2	(verb-subj *sin-ta kunye*)	0	(* *cha-ta* *)	0

Table 5: Probabilities of target words of wear[1]-1 of sentence (2)

a_i	*ip-ta*	*ssu-ta*	*sin-ta*	*cha-ta*	*kki-ta*
$n(a_i)$	0.000	0.600	2.667	0.000	0.300
$\hat{p}(a_i)$	0.000	0.168	0.748	0.000	0.084

in the case of w(wear[1]-1, obj-verb) = 1.0, w(wear[1]-1, subj-verb) = 0.3

And let $w(a, \gamma)$ be the weight representing the importance of the syntactic relation γ in selecting the target word of the sense a. Then the probability $\hat{p}(a_i)$ of selecting a_i as the target word of sense a is computed as follows :

$$n(a_i) = \sum_{(\alpha,m,\gamma)\in\Theta(a)} \frac{\sum_{l=1}^{m} f(a_i, \alpha_l, \gamma)}{m} \cdot w(a, \gamma) \quad (1)$$

$$\hat{p}(a_i) = \frac{n(a_i)}{\sum_{j=1}^{k} n(a_j)} \quad (2)$$

$n(a_i)$ is the weighted frequency of the appearance of the target word a_i in corpora. Using the set $(\alpha, m, \gamma) \in \Theta(a)$, the equation (1) is used to sum up the frequencies of all the nodes that are connected to the node with the sense a. And, by calculating the weighted-sum of the frequencies using the syntactic-relation weight $w(a, \gamma)$ in (1), we can account for the degree of importance of syntactic relation γ in selecting the suitable target word of the sense a.

Table 5 shows the estimated probabilities. Using the frequencies in Table 4 and the equations, we get the weighted-frequency $n(sin\text{-}ta) = \frac{1+5+2}{3} * 1.0 + \frac{0}{1} * 0.3 = 2.667$, and the probability $\hat{p}(sin\text{-}ta) = 0.748$. Then we use the same dynamic threshold for $\hat{p_1}/\hat{p_2}$ of Dagan & Itai (1994) to select the best word for translation. Since $\hat{p}(sin\text{-}ta)/\hat{p}(ssu\text{-}ta)$ is bigger than the threshold value in the sentence (2), we can select the appropriate target word using our method.

The target language collocations are extracted from a tree annotated target language monolingual corpus. In the next section, we will show our method to extract the knowledge from an MRD and a corpus.

3 Knowledge extraction from an MRD and corpus

We extract three types of knowledge from an MRD and corpus: WSD knowledge, sense-to-word mapping information, and collocation information. In this section, we focus on

Table 6: Some examples of the WSD clue extraction

knowledge source	sense entry	definition/example sentence	extracted clue	syntactic relation
typical object in definition sentence	bite^1-2	*(of insects and snakes) to prick the skin (of) an draw blood*	insects and snakes	subject
	aim^1-1	*to point or direct (a weapon, shot, remark, etc.) towards some object, esp. with the intention of hiiting it*	a weapon, shot, remark, etc.	object
example sentence	make^1-1	*Congress makes laws*	congress	subject
			laws	object
	firm^1-3	*a firm belief*	belief	modifiee

describing methods to extract each type of knowledge. Evaluation and experimental result of knowledge extraction will be detailed in Section 5.

3.1 WSD knowledge extraction from an MRD

WSD knowledge used in the first step of the lexical selection has collocations and their syntactic relations to a source word as a selectional restriction for disambiguating the sense of a source word. We use an MRD to automatically extract this information.

An MRD contains two kinds of information for WSD : typical objects in definition sentences and collocation words in example sentences. We use some simple heuristics to extract WSD clues, e.g. "for verbs, extract from definition sentences a typical object that appears after a verb or preposition as a WSD clue with the object syntactic relation", and "for verbs and adjectives, extract their subjects in example sentences as a WSD clue with the subject syntactic relation", and so on. Table 6 shows some results of WSD knowledge extraction from an MRD.

We use Brill's tagger (Brill 1994) to determine parts of speech (POS) of a word in definition sentences and example sentences. Abney's partial parser (Abney 1996) is used to analyze example sentences, and lexical items with appropriate grammatical relations, such as subject, object and modifiee, are determined using a word location in the parsed sentence structure.

3.2 Extracting sense-to-word mapping information from an MRD

The mapping information from a source language word sense to a set of target language words used in the second step is extracted from a bilingual MRD. We extract the mapping information from two parts of the MRD: translation definition sentences and target language example sentences.

It is not difficult to extract target language words in a translation definition sentence: We divide a translation definition into words on delimiters, and then delete special characters and expand expressions in brackets. Table 7 shows some of the mapping information extracted from translation definition sentences.[5]

[5] We show hangul (Korean) characters in the table without transcription, as the point here is to show that patterns are matched.

Table 7: Extraction of sense-to-word mapping information from translation definitions

sense	translation definition	extracted target words
$drop^2$-5	불쑥[허물없이] 방문하다[들르다]	불쑥 방문하다, 불쑥 들르다, 허물없이 방문하다, 허물없이 들르다
$bite^1$-1	(...을) 물다, 깨물어 자르다[부수다]; <사람, 사물을> 물어뜯다	물다, 깨물어 자르다, 깨물어 부수다 물어뜯다

On the other hand, it is not easy to extract target words that correspond to a source word from example sentences. In this paper, we just use a simple heuristic: "if only one word in a sentence has the same POS as that of a source word, select it as a target word of the source word". For example, the first sense of bite, $bite^1$-1, has 6 example sentences in an MRD and (3) is one of them. In this sentence, our heuristic can extract *kkaymwul-ta* as the target word of the sense $bite^1$-1. We can see that this word is not extracted from the translation definition in Table 7.

We use a tagger developed for Korean (J.H.Kim 1996) to determine the POS of a word in Korean sentences.

(3) 'The boy bit into a piece of cake.'

그	꼬마는	케이크	한	조각을	깨물었다.
ku	*kkoma-nun*	*kheyikhu*	*han*	*cokak-ul*	*kkaymwul-ess-ta.*
the	boy-TOP	cake	one	piece-OBJ	bite-PAST-DECL

3.3 Extraction of collocation information from a corpus

The collocation information is extracted from a tree annotated target language corpus. In Korean, a syntactic relation can be easily detected by the POS of a case particle. Details on the structure of a tree annotated Korean corpus and the method to extract syntactic collocations are described in (K.J.Lee et al. 1997).[6]

4 English-to-Korean Transfer System

The proposed method is implemented in a transfer system for English-to-Korean translation. Figure 2 shows the structure of our transfer system.

The input of the transfer system is an phrase structure tree of an English sentence. First, our transfer system selects the sense of a word and maps it into a set of Korean words. This process corresponds to the first and second steps of our lexical selection method. After disambiguating senses of English words, the system transfers the structure of English sentence into that of Korean sentence. Finally, using the statistic information of target language collocation we select a word of target language as the word for translation. This process corresponds to the third step of our lexical selection method. The output of our transfer system is a Korean sentence structure, and a generating system translated it into a Korean sentence.

[6] This tree annotated Korean corpus has been developed at KAIST and available for a nominal fee.

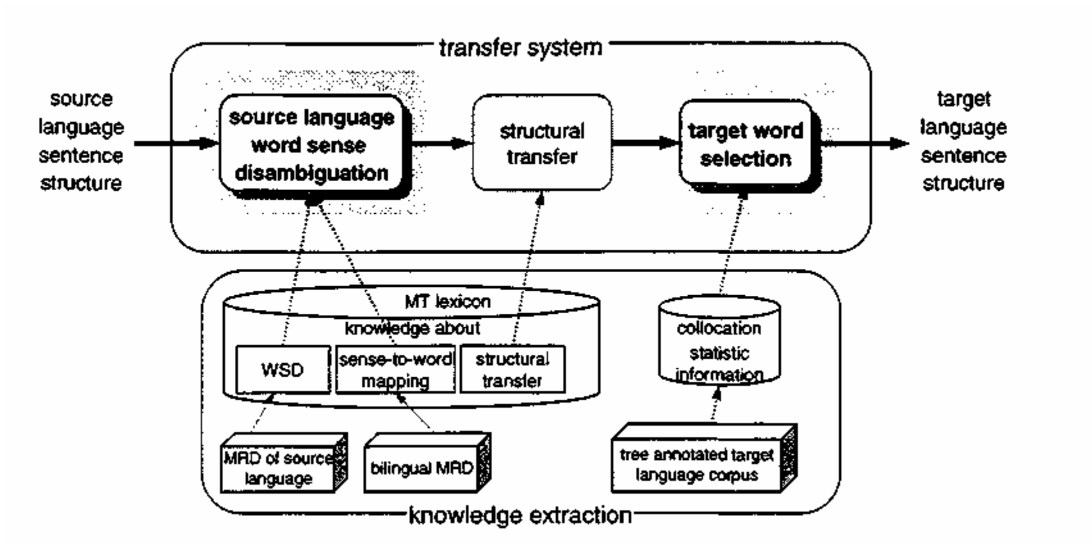


Figure 2: Transfer system using proposed method

Our system splits lexical selection into two subprocesses and transfers a structure between them, and then the first subprocess uses source language information of an MRD and the second subprocess uses target language information of a corpus. Therefore, we can use the syntactic information of target language and appropriate statistic information of target language in the second subprocess of lexical selection, i.e. the third step of our method.

In this paper, we use a hand-coded mono-bilingual MRD and a tree annotated Korean corpus as a source of the knowledge extraction. The MRD is constructed from Longman English-English-Korean Dictionary (LDEEK 1993). Using this mono-bilingual MRD, we can extract the WSD information and the sense-to-word mapping information from the same MRD without mapping between entries in different MRDs. The MRD contains 2114 words that are chosen from the subset of the LDOCE's defining vocabulary. The tree annotated Korean corpus consists of about 30000 sentences.

5 Evaluation

Table 8 shows syntactic tuples and their frequencies in Korean that are obtained from an English phrase "have a car".[7] In Korean, the word *cha* is polysemous and can translate into 'tea', 'car', 'difference', etc., so the syntactic tuple (verb-obj *masi-ta cha)* frequently occurs in a Korean corpus. From the table, the method that uses only target language information as in Dagan & Itai (1994) would select the inappropriate target word *masi-ta* "drink" as the word for translation. But our method selects the sense of source language word using WSD knowledge prior to using the statistic information of target language. So, *masi-ta* is filtered out in the WSD step, and we can avoid the use of improper statistic information.

[7] In the table, synonyms are clustered in a row: *kaci-ta soyuha-ta* "own", *et-ta, ipswuha-ta, patta-ta* "obtain", and *mek-ta* "eat" and *masi-ta* "drink".

Table 8: Frequencies of target words of 'have' according to the object *cha* "car"

(verb-obj *kaci-ta cha*) 2	(verb-obj *soyuha-ta cha*) 0	
(verb-obj *et-ta cha*) 0	(verb-obj *ipswuha-ta cha*) 0	(verb-obj *patta-ta car*) 0
(verb-obj *mek-ta car*) 0	(verb-obj *masi-ta car*) 9	

Table 9: Frequencies of collocations of the sentence (5)

(a) frequencies used in Dagan & Itai (1994)

verb-obj	*cilmwun*	*mwulum*	*mwuncey*	*uymwun*	etc
taytapha-ta	-	-	1	-	-
hwitapha-ta	-	-	-	-	-
phwul-ta	1	-	18	2	2
etc	-	-	-	-	-

(b) frequencies used in our model

verb-tar	*cilmwun*	*mwulum*
taytapha-ta	17	8
hwitapha-ta	-	-

(4) 'He has a new car.'

그는 새 차를 가지고 있다.

ku-nun say cha-lul kaciko iss-ta.

he-TOP new car-ACC have-PROGRESSIVE-DECL

The sentence (5) is another example. In this sentence, the relation between the words 'answer' and 'question' in an English sentence is verb-obj, but that in a Korean sentence is verb-tar.[8] In this case, previous methods that use the relation of a target language can not select the appropriate target word with (a) in Table 9. But since our method makes syntactic tuples with relations of target language with disambiguated word senses, we can select the appropriate target word with (b) in Table 9.

(5) 'You must answer the question.'

너는 반드시 질문에 대답하여야 한다.

ne-nun pantusi cilmwun-ey taytapha-yeya han-ta.

you-TOP necessarily question-TARGET answer-MUST-DECL

Tables 10 and 11 show experimental results of our knowledge extraction method. For all 2114 words in the MRD, we use randomly selected 41 words for evaluation. In Table 10, we see that our heuristics gain high accuracy in extracting the sense-to-word mapping information from definition sentences, but that they do not from example sentences as predicted. Table 11 shows that we gain high precision in WSD knowledge extraction despite simple heuristics of some patterns.

6 Conclusion

In this paper, we proposed a three-step method for lexical selection that uses automatically extracted knowledge. Although we use automatically extracted knowledge, our

[8] In Korean, the relation verb-tar can be analyzed without any semantics, because a case marker *ey* denotes the target relation.

Table 10: Experimental results for extraction of sense-to-word mapping information

from definition sentences		from example sentences	
Precision	Recall	Precision	Recall
297/302 98.3%	297/307 96.7%	67/97 69.1%	67/170 39.4%

Table 11: Experimental results for knowledge extraction for WSD

	from typical objects		from example sentences	
	Precision	Recall	Precision	Recall
with a correct syntactic relation	20/25 80%	20/22 90.9%	127/133 95.5%	127/221 57.5%
with any syntacitic relation	21/25 83%	21/22 95.5%	131/133 98.5%	131/221 59.3%

method can select appropriate words unlike previous methods because of the following reasons: We disambiguate senses of source language words prior to using statistic information of target language; we use the syntactic relations of target language and the syntactic relation weights when using statistic information of target language ; and combine statistic information of all collocations of a word in sentences to estimate the probability. Other advantages of our method are robustness and extensibility. Because we use a target language corpus, our method can select better target words as the size of the corpus increases. Even though the size of the corpus is very small, we can avoid selecting or generating a target word the meaning of that is definitely improper in given sentences, since our system disambiguates senses of source language words using reliable knowledge in an MRD prior to selecting target lexical items.[9]

References

Steven Abney : 1996, 'Partial Parsing via Finite-State Cascades,' in *Proceedings of the ESSLLI '96 Robust Parsing Workshop.*

E. Brill : 1994, 'Some advances in rule-based part of speech tagging,' in *Proceedings of the Twelfth National Conference on Artificial Intelligence : AAAI-94,* Seattle, Wa., USA, pp 722-727.

Peter F. Brown, Stephen A. Della Pietra & Vincent J. Della Pietra: 1991, 'Word Sense Disamgituation Using Statistical Methods,' In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL),* pp 264-270.

Ann Copestake, Ted Briscoe, Piek Vossen, Alicia Ageno, Irene Casteloon, Grancesc Ribas, German Rigau, Horacio Rodriguez & Anna Samiotou : 1994, 'Acquisition of lexical translation relations from MRDs,' In *Machine Translation,* Vol.9, No.3-4, pp 183-219.

[9] We realize that there is no end-to-end evaluation, as one of the reviewers pointed out. A starting point would be to use target language fluency only, where sense divisions in the dictionary are ignored for a direct comparison with previous work. We are grateful to the reviewer for this suggestion, but we leave it for future work.

Ido Dagan & Alon Itai : 1994, 'Word Sense Disambiguation Using a Second Language Monoligual Corpus,' In *Computational Linguistics (CL),* Vol.20, No.4, pp 563-596.

Jae-Hoon Kim & Gil Chang Kim : 1996, 'Fuzzy Network Model for Part-of-Speech Tagging under Small Training Data,' In *Natural Language Engineering,* Vol. 2, No. 2.

Judith Klavans & Evelyne Tzoukermann : 1996, 'Combining corpus and machine-readable dictionary data for building bilingual lexicons' In *Machine Translation,* Vol.10, No.3, pp 1-34.

Kong Joo Lee, Jae-Hoon Kim. Young S. Han & Gil Chang Kim : 1997, 'Restricted Representation of Phrase Structure Grammar for Building a Tree Annotation Corpus of Korean,' In *Natural language Engineering,* Vol. 3, No. 2&3, pp 215-231.

Longman English-English-Korean Dictionary, KumSung Textbook Co., Ltd, 1993.

David Yarowsky : 1993. 'One sense per collocation,' In *Proceedings of the ARPA Human Language Technology Workshop,* 1993.