Dual Distributional Verb Sense Disambiguation with Small Corpora and Machine Readable Dictionaries*

Jeong-Mi Cho and **Jungyun Seo**
Dept. of Computer Science,
Sogang University
Sinsu-dong, Mapo-gu, Seoul, 121-742, Korea
jmcho@nlprep.sogang.ac.kr
seojy@ccs.sogang.ac.kr

Gil Chang Kim
Dept. of Computer Science,
Korea Advanced Institute of
Science and Technology
Taejon, 305-701, Korea
gckim@csking.kaist.ac.kr

Abstract

This paper presents a system for unsupervised verb sense disambiguation using small corpus and a machine-readable dictionary (MRD) in Korean. The system learns a set of typical usages listed in the MRD usage examples for each of the senses of a polysemous verb in the MRD definitions using verb-object co-occurrences acquired from the corpus. This paper concentrates on the problem of data sparseness in two ways. First, extending word similarity measures from direct co-occurrences to co-occurrences of co-occurred words, we compute the word similarities using not co-occurred words but co-occurred clusters. Second, we acquire IS-A relations of nouns from the MRD definitions. It is possible to cluster the nouns roughly by the identification of the IS-A relationship. By these methods, two words may be considered similar even if they do not share any words. Experiments show that this method can learn from very small training corpus, achieving over 86% correct disambiguation performance without a restriction of word's senses.

1 Introduction

Much recent research in the field of natural language processing has focused on an empirical, corpus-based approach, and the high accuracy achieved by a corpus-based approach to part-of-speech tagging and parsing has inspired similar approaches to word sense disambiguation. For the most successful approaches to such problems, correctly annotated materials are crucial for training learning-based algorithms. Regardless of whether or not learning is involved, the prevailing evaluation methodology requires correct test sets in order to rigorously assess the quality of algorithms and compare their performance. This seems to require manual tagging of the training corpus with appropriate sense for each occurrence of an ambiguous word. However, in marked contrast to annotated training material for part-of-speech tagging, (a) there is no coarse-level set of sense distinctions widely agreed upon (whereas

headword : *open*[2]	
sense	usage examples
open	*Open the window a bit, please.*
	He opened the door for me to come in.
	Open the box.
start	*Our chairman opened the conference by*
	welcoming new delegates/
	Open a public meeting.

Table 1: The entry of *open*(vt.) in OALD

part-of-speech tag sets tend to differ in the detail); (b) sense annotation has a comparatively high error rate (Miller, personal communication, reports an upper bound for human annotators of around 90% for ambiguous cases, using a non-blind evaluation method that may make even this estimate overly optimistic(Resnik, 1997)); (c) in conclusion, a sense-tagged corpus large enough to achieve broad coverage and high accuracy word sense disambiguation is not available at present. This paper describes an unsupervised sense disambiguation system using a POS-tagged corpus and a machine-readable dictionary (MRD). The system we propose circumvents the need for the sense-tagged corpus by using MRD's usage examples as the sense-tagged examples. Because these usage examples show the natural examples for headword's each sense, we can acquire useful sense disambiguation context from them. For example, *open* has several senses and usage examples for its each sense listed in a dictionary as shown in Table 1. The words within usage examples *window*, *door*, *box*, *conference*, and *meeting* are useful context for sense disambiguation of *open*.

Another problem that is common for much corpus-based work is data sparseness, and the problem especially severe for work in WSD. First, enormous amounts of text are required to ensure that all senses of a polysemous word are represented, given the vast disparity in frequency among senses. In addition, the many possible co-occurrences for a given polysemous word are unlikely to be found in even a very large corpus, or they occur too infrequently to be significant. In this paper, we propose two methods

that attack the problem of data sparseness in W!
using small corpus and dictionary. First, extendi
word similarity measures from direct co-occurren
to co-occurrences of co-occurred words, we compi
the word similarities using not co-occurred wo
but co-occurred clusters. Second, we acquire IS
relations of nouns from the MRD definitions. D
tionary definitions of nouns are normally written
such a way that one can identify for each headw
(the word being defined), a "genus term" (a w
more general that the headword), and these are
lated via an IS-A relation(Amsler, 1979). It is po
ble to cluster the nouns roughly by the identificati
of the IS-A relationship.

2 Dual Distributional Similarity

We attempt to have the system learn to disa
biguate the appearances of a polysemous verb w
its senses defined in a dictionary using the
occurrences of syntactically related words in a PC
tagged corpus. We consider two major word class
V and N, for the verbs and nouns and a single re
tion between them, in our experiments the relati
between a transitive main verb and the head no
of its direct object. Thus, a noun is represented
a vector of verbs that takes the noun as its object,
and a verbs by a vector of nouns that appears as the
verb's object. Commonly used corpus-based models
depend on co-occurrence patterns of words to deter-
mine similarity. If word w_1's co-occurrence patterns
is similar to word w_2's patterns, then w_1 is similar
to w_2 contextually. Note that contextually similar
words do not have to be synonym, or to belong to
the same semantic category. We define a word being
computed the similarity as **a target word** and a
word occurring in the co-occurrence pattern of the
target word as **a co-occurred word**. The overlap
of words between co-occurrence patterns of two tar-
get words determines the similarity of them. How-
ever, in case of small training corpus, it is difficult
to confide in the similarity depending on statistics of
co-occurrences. The reason is that when two words
have no overlap of co-occurrence patterns, we can
not discriminate whether two words are not similar
or it fails to find the similarity due to sparse data To
distinguish two cases, we expand the co-occurrences
of the target word to the co-occurrences of the co-
occurred words with the target word. According
to the co-occurrence patterns of the co-occurred
words, it is possible to cluster the co-occurred words
roughly. And we can overcome the problem of data
sparseness by applied not co-occurred words but co-
occurred clusters to the similarity of target words.
A dual distributional similarity is an extension
to word similarity measure reflecting the distribu-
tions of the co-occurred words with the target word
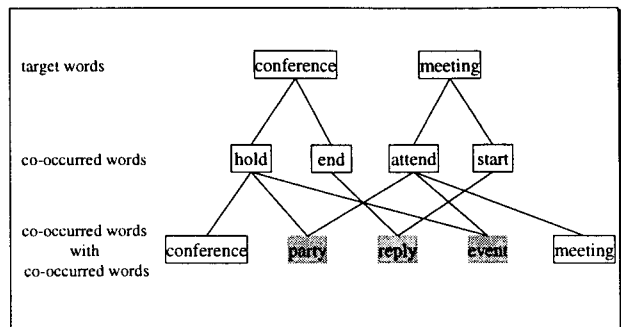as well as the distribution of the target word.

Figure 1: The example of dual distributional simi-
larity

Figure 1 demonstrates the advantage of the dual
distributional similarity. in comparison with the
unitary distributional similarity. The simple com-
parison with co-occurrence patterns of *conference*
and *meeting* fails to find the similarity between the
two nouns because there in no overlap in the co-
occurrence patterns. However, dual distributional
similarity measure can be find that the two nouns are
similar even if the co-occurrence patterns of the two
nouns do not overlap. First, since the co-occurred
verbs *attend, end, hold*, and *start* with *conference*
and *meeting* share several objects such as *event, re-
ply*, and *party*, we can find that the co-occurred verbs
are similar. And since *conference* and *meeting* share
similar verbs, they are similar even if they do not
share any verbs.

3 The WSD System Using a Corpus and a MRD

The architecture of the WSD system using a cor-
pus and a MRD is given in Figure 2. Our system
consists of two parts, which are the knowledge ac-
quisition system and the sense disambiguation sys-
tem. The knowledge acquisition system also consists
of two parts, one of the acquisition of selectional re-
striction examples from a POS-tagged corpus and
another of the acquisition of each verb's sense indi-
cators and noun clustering cues from a MRD. The
sense disambiguation system assigns an appropriate
sense to an ambiguous verb by computation of sim-
ilarity between its object in a sentence and its sense
indicators. The overall process for verb sense disam-
biguation is as follows:

- Extract all selectional restriction examples from
  a POS-tagged corpus.

18

object verb
↓
object
↑
similarity?
↓
sense indicator¹
sense indicator²
sense indicator³
↓
object verb²

Sense disambiguation system

CORPUS
↓
CORPUS Analyzer
↓
word co-occurrences
within syntactic relation

sense indicators
clustering cues
↓
MRD Analyzer
↓
MRD
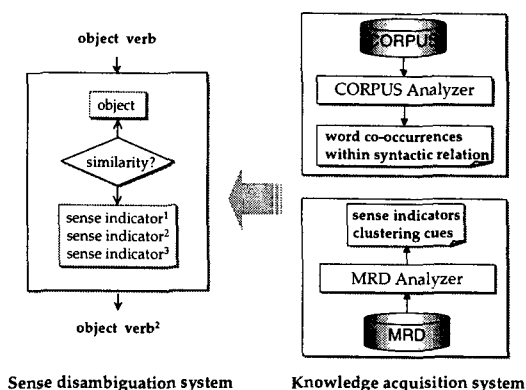
Knowledge acquisition system

Figure 2: The WSD system using a corpus and a MRD

- Extract each polysemous verb's sense indicators from a MRD.

- For a target verb, compute similarities between its object and its sense indicators using the selectional restriction examples acquired from the corpus and clustering cues from the MRD.

- Determine the sense of the most similar sense indicator as the verb's disambiguated sense.

3.1 Context for verb sense disambiguation

Presumably verbs differ in their selectional restrictions because the different actions they denote are normally performed with different objects. Thus we can distinguish verb senses by distinguishing selectional restrictions. (Yarowsky, 1993) determined various disambiguating behaviors based on syntactic category; for example, that verbs derive more disambiguating information from their objects than from their subjects, and adjectives derive almost all disambiguating information from nouns they modify. We use verb-object relation for verb sense disambiguation. For example, consider the sentences *Susan opened the meeting* and *Susan opened the door*. In deciding which *open*'s senses in Table 1 are tagged in the two sentences, the fact that *meeting* and *door* appear as the direct object of *open* respectively gives some strong evidence.

3.2 Lexical knowledge acquisition

3.2.1 Machine-readable dictionaries

In previous works using MRDs for word sense disambiguation, the words in definition texts are used as sense indicators. However, the MRD definitions alone do not contain enough information to allow reliable disambiguation. To overcome this problem,

we use the MRD usage examples as the sense-tagged examples as well as definitions for acquiring sense indicators. We acquire all objects in the MRD definitions and usage examples of a polysemous verb as its sense indicators. We use objects as sense indicators by same reason of using verb-object selection relation for verb sense disambiguation. These sense indicators is very useful to verb sense disambiguation because the objects in usage examples are typical and very often used with the sense of the verb.

The entries of *wear* in OALD and *ipta* (wear) and *ssuta* (write) in Korean dictionary and the sense indicator sets acquired from them are shown in Table 2.

We acquire another information from the dictionary definition. Dictionary definitions of nouns are normally written in such a way that one can identify for each headword (the word being defined), a "genus term" (a word more general that the headword), and these are related via an IS-A relation(Bruce,1992; Klavans, 1990; Richardson,1997). We use the IS-A relation as noun clustering cues. For example, consider the following definitions in OALD.

hat covering for the head with a brim, worn out of doors.
cap¹ soft covering for the head without a brim. bonnet.
shoe¹ covering for the foot, esp. one that does not reach above the ankle.

Here *covering* is common genus term of the headwords, *hat*, *cap¹*, and *shoe²*. That is, we can say that *"hat* IS-A *covering"*, *"cap¹* IS-A *covering"*, and *"shoe²* IS-A *covering"*, and determine these three nouns as same cluster *covering*. In *cap¹*'s definition, *bonnet* is a synonym of *cap¹*. We also use the synonyms of a headword as another clustering cues.

Our mechanism for finding the genus terms is based on the observation that in Korean dictionary, the genus term is typically the tail noun of the defining phrase as follows:

ilki   nalmata kyekkun il, sayngkakul cekun *kilok*(record).
(diary)   (daily *record* of events, thoughts, etc.)

Because these clustering cues are not complete and consistent, we use parent and sibling clusters without multi-step inference for acquired IS-A relations.

3.2.2 Corpora

We acquire word co-occurrences within syntactic relations for learning word similarity from a POS-tagged corpus in Korean. To acquire word co-occurrences within syntactic relations, we have to get the required parsing information. Postpositions in Korean are used to mark the syntactic relations of the preceding head components in a sentence. For example, the postpositions *ka* and *i* usually mark

headword	sense definition	usage examples	sense indicators
English Dictionary(OALD)			
wear[2]	have one the body, carry on one's person or on some part of it;ace (of looks) have on the face	*He was wearing a hat/spectacles/ a beard/heavy shoes/a ring on his finger/a troubled look.*	{ one, hat, spectacles, beard, shoes, ring, look }
Korean Dictionary(Grand Korean Dictionary)			
ipta[1]	mom-ey os-ul kelchikena tuluta (wear clothes)	*hanbok-ul ipta/chima-lul ipta* (wear Korean clothes)/(wear a skirt)	{os(clothes), hanbok(Korean clothes), chima(skirt)}
ssuta[2]	kul-ul cista (write an article)	*sosel-lul ssuta/phyenci-ul ssuta* (write a novel/write a letter)	{kul(article), sosel(novel), chima(skirt)}

Table 2: The entries of *wear* (vt.) in OALD and *ipta* and *ssuta* in Korean dictionary

the subjective relation and *ul* and *lul* the objective relation.[1] Given the sentence[2] *kunye-ka(she) phyenci-lul(letter) ssu-ta(write)*, we can know that *kunye* (she) is the subject head and *phyenci* (letter) is the direct object head according to the postpositions *ka* and *lul*. We call guessing the syntactic relation by postpositions as Postposition for Syntactic Relation heuristic. When there are multiple verbs in a sentence, we should determine one verb in relation to the object component. In such attachment ambiguity, we apply the Left Association heuristic, corresponding to the Right Association in English. This heuristic states that the object component prefers to be attached to the leftmost verb in Korean. With the two heuristics, we can accurately acquire word co-occurrences within syntactic relations from the POS-tagged corpus without parsing(Cho, 1997).

3.3 Dual distributional sense disambiguation

In our system, verb sense disambiguation is the clustering of an ambiguous verb's objects using its sense indicators as seeds. As noted above, a noun is represented by a vector of verbs that takes the noun as its object, and a verbs by a vector of nouns that appears as the verb's object. We call the former **a noun distribution** and the latter **a verb distribution**. The noun distribution is probabilities of how often each verb had the noun as object, given the noun as object, that the verb is $v_1, v_2, ... v_{|V|}$. That is,

$$d(n) =< p(v_1|n), p(v_2|n), ..., p(v_{|V|}|n) > \quad (1)$$

$$p(v_i|n) = \frac{freq(v_i, n)}{\sum_{j=1}^{|V|} freq(v_j, n)} \quad (2)$$

where $|V|$ is the number of verbs used as transitive verb in training corpus, and $freq(v, n)$ is the frequency of verb v that takes noun n as direct object.

A verb distribution is a vector of nouns that appears as the verb's direct object. We define the verb

distribution as containing binary value, "1" if each noun occurring as its direct object and "0" otherwise.

$$d(v) =< b(n_1, v), b(n_2, v), ..., b(n_{|N|}, v) > \quad (3)$$

$$b(n_i, v) = 1 \; if \; n_i \; appeared \; as \; v's \; directed \; object \quad (4)$$

$$0 \; otherwise \quad (5)$$

where $|N|$ is the number of nouns appeared as transitive verb's direct object.

The process of object clustering is as follows:

1. Cluster the objects according to clustering cues acquired from the MRD.

2. Cluster the objects excepted from Step 1 using the dual distribution.

3. Cluster the objects excepted from Step 2 to the MRD's first sense of the polysemous verb.

3.3.1 Clustering using IS-A relations implicit in MRD definition

We define cluster $cluster(w)$ and synonym set $synonym(w)$ of a word w using IS-A relations implicit in the MRD definition. The criteria of clustering word w_1 and word w_2 as same cluster are as follows:

- $w_1 \in cluster(w_2)$

- $w_1 \in synonym(w_2)$

- $w_i \in cluster(w_2)$ where $w_i \in cluster(w_1)$

- $w_i \in synonym(w_2)$ where $w_i \in cluster(w_1)$

- $w_i \in synonym(w_2)$ where $w_i \in synonym(w_1)$

3.3.2 Measuring similarities between nouns

To compute the similarities between nouns we use the *relative entropy* or *Kullback-Leibler(KL) distance* as metric to compare two noun distributions. The relative entropy is an information-theoretic measure of how two probability distributions differ. Given two probability distributions p and q, their relative entropy is defined as

$$D(p \parallel q) = - \sum_x p(x) log \frac{p(x)}{q(x)} \quad (6)$$

where we define $0 log \frac{0}{q} = 0$ and otherwise $plog\frac{p}{0} = \infty$. This quantity is always non-negative, and $D(p||q) = 0$ iff $p = q$. Note that relative entropy is not a metric (in the sense in which the term is used in mathematics): it is not symmetric in p and q, and it does not satisfy a triangle equality. Nevertheless, informally, the relative entropy is used as the "distance" between two probability distribution in many previous works(Pereira, 1993; Resnik, 1997). The relative entropy can be applied straightforwardly to the probabilistic treatment of selectional restriction. As noted above, the noun distribution $d(n)$ is verb v_i's condition probability given by noun n. Given two noun distributions $d(n_1)$ and $d(n_2)$, the similarity between them is quantified as:

$$D_n(d(n_1) \| d(n_2)) = -\sum_{v_i \in V} p(v_i|n_1)log\frac{p(v_i|n_1)}{p(v_i|n_2)} \quad (7)$$

3.3.3 Measuring similarities between verbs

The noun distributions p and q is easy to have zero probabilities by the problem of sparse data with small training corpus[3]. In such case, the similarity of the distributions is not reliable because of $0 log \frac{0}{q} = 0$ and $plog\frac{p}{0} = \infty$. This can be known from the results of sense disambiguation experiments using only noun distributions (see Section 4.2). The verb distributions play complementary roles when the noun distributions have zero probabilities. For all verbs where $p(v_i|n_2) = 0$ and $p(v_i|n_1) > 0$ or the reverse case:

1. execute OR operation with all distributions for the verbs v_i where $p(v_i|n_2) = 0$ and $p(v_i|n_1) > 0$ in the noun distribution $d(n_1)$ and make new distribution, dv_1.

$$dv_1 = \bigvee d(v_i), for \, p(v_i|n_1) > 0 \, and \, p(v_i|n_2) = 0$$

2. execute OR operation with all distributions for the verbs v_i where $p(v_i|n_2) > 0$ and $p(v_i|n_1) = 0$ in the noun distribution $d(n_2)$ and make new distribution, dv_2.

$$dv_2 = \bigvee d(v_i), for \, p(v_i|n_2) > 0 \, and \, p(v_i|n_1) = 0$$

3. execute inner product with new distributions, dv_1 and dv_2

$$D_v(d(v_1), d(v_2)) = dv_1 \cdot dv_2$$

We use a stop verb list to discard from Steps 1 and 2 verbs taken too many nouns as objects, such as *hata* (do), which do not contribute to the disambiguation process. The verb distribution has the binary values, 1 or 0 according to its object distributions in the training corpus. Thus, the inner product $D_{verb}(d(v_1), d(v_2))$ with dv_1 and dv_2 means the number of common objects to two distributions. We can compute the similarities of the co-occurred verbs in the two noun distributions with the number of common objects. Although the two noun distribution do not share any verbs, if they have similar verbs in common, they are similar.

Combining similarities of noun distributions and verb distributions, we compute total similarity between the noun distributions.

$$D_t = \alpha D_n + \beta D_v \quad (8)$$

The α, β are the experimental constants(0.71 for α and 0.29 for β).

4 Experimental Evaluation

We used the KAIST corpus, which contains 573,193 eojeols[4] and is considered a small corpus for the present task. As the dictionary, we used the *Grand Korean Dictionary*, which contains 144,532 entries.

The system was tested on a total of 948 examples of 10 polysemous verbs extracted from the corpus: *kamta, kelta, tayta, tulta, ttaluta, ssuta, chita, thata, phwulta,* and *phiwuta* (although we confined the test to transitive verbs, the system is applicable to intransitive verbs or adjectives). For this set of verbs, the average number of senses per verb is 6.7. We selected the test verbs considering the frequencies in the corpus, the number of senses in the dictionary, and the usage rates of each sense.

We tested the systems on two test sets from KAIST corpus. The first set, named C23, consists of 229,782 eojeols and the second set, named C57, consists of 573,193 eojeols. The experimental results obtained are tabulated in Table 3. As a baseline against which to compare results we computed the percentage of words which are correctly disambiguated if we chose the most frequently occurring sense in the training corpus for each verb, which resulted in 42.4% correct disambiguation. Columns 3-5 illustrate the effect of adding the dual distribution and the MRD information. When the dual distribution is used, we can see significant improvements of about 22% for recall and about 12% for the precision. Specially, in smaller corpus (C23), the improvement of recall is remarkable as 25%. This represents that the dual distribution is effective to overcome the problem of sparse data, especially for small corpus. Moreover, by using both the dual distribution and the MRD information, our system achieved

---

[3] As many of the possible co-occurrences are not observed even in a large corpus(Church, 1993), actually the noun distributions have not many common verbs.

[4] Eojeol is the smallest meaningful unit consisting of content words (nouns, verbs, adjectives, etc.) and functional words (postpositions, auxiliaries, etc.)

measure	corpus	noun dis.	dual dis.	dual dis. + MRD
recall	C23	40.9%	66.0%	80.0%
	C57	47.8%	67.7%	86.3%
precision	C23	47.7%	55.8%	80.0%
	C57	48.3%	61.0%	86.3%

Table 3: Experimental results

the improvements of about 16% for recall and about 25% for the precision.

The average performance of our system is 86.3% and this is a little behind comparing with other previous work's performance in English. Most previous works have reported the results in "70%-92%" accuracies for particular words. However, our system is the unsupervised learning with small POS-tagged corpus,and we do not restrict the word's sense set within either binary senses(Yarowsky,1995; Karov, 1998) or dictionary's homograph level(Wilks, 1997). Thus, our system is appropriate for practical WSD system as well as bootstrapping WSD system starting with small corpus.

5 Related Work

Using MRDs for word sense disambiguation was popularized by (Lesk, 1986). Several researchers subsequently continued and improved this line of work(Guthrie, 1991; Krovetz, 1989; Veronis, 1990; Wilks, 1997). Unlike the information in a corpus, the information in the dictionary definitions is presorted into senses. However, the dictionary definitions alone do not contain enough information to allow reliable disambiguation. Recently, many works combined a MRD and a corpus for word sense disambiguation(Karov, 1998; Luk, 1995; Ng, 1996; Yarowsky,1995). In (Yarowsky,1995), the definition words were used as initial sense indicators, automatically tagging the target word examples containing them. These tagged examples were then used as seed examples in a bootstrapping process. In (Luk, 1995), using the dictionary definition, co-occurrence data of concepts, rather than words, is collected from a relatively small corpus to tackle the data sparseness problem. In (Karov, 1998), all the corpus examples of the dictionary definition words, instead of those word alone were used as sense indicators. In comparison, we suggest to combine the MRD definition words and usage examples as the sense indicators. Because the MRD's usage examples can be used as the sense-tagged instances, the sense indicators extracted from them are very useful for word sense disambiguation. And this yield much more sense-presorted training information.

The problem of data sparseness, which is common for much corpus-based work, is especially severe for work in WSD. Traditional attempts to tackle the problem of data sparseness include the class-based approaches and similarity-based approaches. The class-based approaches(Brown, 1992; Luk, 1995; Pereira, 1993; Resnik, 1992) attempt to obtain the best estimates by combining observations of classes of words considered to belong to a common category. These methods answer in part the problem of data sparseness and eliminate the need for pretagged data. However, there is some information loss with these methods because the hypothesis that all words in the same class behave in a similar fashion is too strong. In the similarity-based approaches(Dagan, 1997; Karov, 1998), rather than a class, each word is modeled by its own set of similar words derived from statistical data extracted from corpora. However, deriving these sets of similar words requires a substantial amount of statistical data and thus these approaches require relatively large corpora. (Karov, 1998) proposed an extension to similarity-based methods by means of an iterative process at the learning stage with small corpus. Our system is similar to (Karov, 1998) with respect to similarity measure, which allows it to extract high-order contextual relationship. However, we attempt to concern a polysemous word's all senses in the training corpus, rather than restricting the word's sense set within binary senses and this allows our system to be more practical.

6 Conclusions

We have described an unsupervised sense disambiguation system using a small corpus and a MRD. Our system combines the advantages of corpus-based approaches (large number of word patterns) with those of the MRD-based approaches (data presorted by senses), by acquiring sense indicators from the MRD's usage examples as well as definitions and acquiring word co-occurrences from the corpus. Because the MRD's usage examples can be used as the sense-tagged instances, the sense indicators acquired from them are very useful for word sense disambiguation. In our system, Two nouns are considered similar even if they do not share any verbs if they appear as objects to similar verbs because the similarities between verbs simultaneously compute with the similarities between nouns. Thus, we can overcome effectively the problem of sparse data due to unobserved co-occurrences of words in the training corpus. Our experiments show that the results using the dual distribution and the MRD information lead to better performance on very sparse data.

Our immediate plans are to test our system on various syntactic categories involving nouns as well as intransitive verbs and adjectives, and to suggest that different kinds of disambiguation procedures are needed depending on the syntactic category and other characteristics of the target word. Further-

more, we plan to build a large sense-tagged corpus, where the sense distinction is at the level of a dictionary in Korean. The sense-tagged corpus would be reused to achieve broad coverage, high accuracy word sense disambiguation.

References

Amsler, R. A. and J. White. 1979. Development of a computational methodology for deriving natural language semantic structures via analysis of machine-readable dictionaries. National Science Foundation, Technical Report. MCS77-01315.

Brown, Peter F., Vincent J. Dellar Peitra, Peter V. deSouza, Jenifer C. Lai, and Robert L. Mercer. 1992. Class-based n-gram models of natural language. *Computational Linguistics*, vol. 14, no. 4, pp.467-479.

Bruce, R. and L. Guthrie. 1992. Genus disambiguation: A study in weighted preference. In *Proceedings of COLING92*, pp.1187-1191.

Cho, Jeong-Mi, Young Hwan Cho, and Gil Chang Kim. 1997. Automatic acquisition of N(oun)-C(ase)-P(redicate) information from POS-tagged corpus. *Computer Processing of Oriental Languages*, vol. 11, no. 2, pp.191-204.

Church, Kenneth W. and Robert L. Mercer. 1993. Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics*, vol. 19, pp.1-24.

Dagan, Ido, Lillian Lee, and Fernando Pereira. 1997. Similarity-based methods for word sense disambiguation. In *Proceedings of the 35th Annual Meeting of the ACL*, pp.56-63.

Guthrie, Joe Al, Louise Guthrie, Yorick Wilks, and Homa Aidinejad. 1991. Subject-dependent cooccurrence and word sense disambiguation. In *Proceedings of the 29th Annual Meeting of the ACL*, pp.146-152.

Karov, Yael and Shimon Edelman. 1998. Similarity-based word sense disambiguation. *Computational Linguistics*, vol. 24, no. 1, pp.41-60.

Klavans, J., M. Chodorow, and N. Wacholder. 1990. From dictionary to knowledge base via taxonomy. In *Proceedings of the 4th Annual Conference of the University of Waterloo Center for the New Oxford English Dictionary: Electronic Text Research*, pp.110-132.

Krovetz, Robert and W. Bruce Croft. 1989. Word sense disambiguation using machine-readable dictionaries. In *Proceedings of ACM SIGIR'89*, pp.127-136.

Lesk, Michael. 1986. Automatic sense disambiguation: How to tell a pine cone from an ice cream cone. In *Proceedings of the 1986 ACM SIGDOC Conference*, pp.24-26.

Luk, Alpha K.. 1995. Statistical sense disambiguation with relatively small corpora using dictionary definitions. In *Proceedings of the 33rd Annual Meeting of the ACL*, pp.181-188.

Ng, Hwee Tou. 1996. Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach. In *Proceedings of the 34th Annual Meeting of ACL*, pp.40-47.

Pereira, Fernando, Naftali Tishby, and Lillian Lee. 1993. Distributional Clustering Of English Words. In *Proceedings of the 31st Meeting of the Association for Computational Linguistics*.

Resnik, Philip Stuart. 1992. WordNet and distributional analysis: a class-based approach to lexical discovery. In *Proceedings of AAAI Workshop on Statistically-Based NLP Techniques*.

Resnik, Philip Stuart. 1997. Selectional preference and sense disambiguation. In *Proceedings of ANLP Workshop, "Tagging Text with Lexical Semantics: Why, What, and How?"*.

Richardson, Stephen D.. Determining similarity and inferring relations in a lexical knowledge base. Ph.D. Dissertation. The City University of New York.

Veronis, Jean and Nancy Ide. 1990. Word sense disambiguation with very large neural networks extracted from machine-readable dictionaries. In *Proceedings of COLING-90*, pp.389-394.

Wilks, Yorick and Mark Stevenson. 1997. Combining independent knowledge sources for word sense disambiguation. In *Proceedings of RANLP*.

Yarowsky, David. 1993. One sense per collocation. In *Proceedings of the ARPA Human Language Technology Workshop*. pp.265-271.

Yarowsky, David. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting of the ACL*, pp.189-196.