Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Computer Speech and Language 18 (2004) 1–21
COMPUTER SPEECH AND LANGUAGE
www.elsevier.com/locate/csl

Unsupervised word sense disambiguation using WordNet relatives

Hee-Cheol Seo [a,*], Hoojung Chung [a], Hae-Chang Rim [a,*], Sung Hyon Myaeng [b], Soo-Hong Kim [c]

[a] *Department of Computer Science and Engineering, Korea University 1, 5-ka, Anam-dong, Seongbuk-Gu, Seoul 136-701, Republic of Korea*
[b] *Information and Communications University, 58-4 Hwaam-dong, Yuseong-gu, Daejeon 305-732, Republic of Korea*
[c] *Department of Computer Software Engineering, College of Engineering, Sangmyung University, San 98-20, Anso-Dong, Chonan, Chungnam, Republic of Korea*

Received 1 October 2003; received in revised form 11 May 2004; accepted 20 May 2004

Abstract

This paper describes a sense disambiguation method for a polysemous target noun using the context words surrounding the target noun and its WordNet relatives, such as synonyms, hypernyms and hyponyms. The result of sense disambiguation is a relative that can substitute for that target noun in a context. The selection is made based on co-occurrence frequency between candidate relatives and each word in the context. Since the co-occurrence frequency is obtainable from a raw corpus, the method is considered to be an unsupervised learning algorithm and therefore does not require a sense-tagged corpus. In a series of experiments using SemCor and the corpus of SENSEVAL-2 lexical sample task, all in English, and using some Korean data, the proposed method was shown to be very promising. In particular, its performance was superior to that of the other approaches evaluated on the same test corpora.

* Corresponding authors. Tel.: +82-2-3290-3195; fax: +82-2-929-7914.
E-mail addresses: hcseo@nlp.korea.ac.kr (H.-C. Seo), hjchung@nlp.korea.ac.kr (H. Chung), rim@nlp.korea.ac.kr (H.-C. Rim), myaeng@icu.ac.kr (S.H. Myaeng), soohkim@smuc.ac.kr (S.-H. Kim).

0885-2308/$ - see front matter
doi:10.1016/j.csl.2004.05.004

1. Introduction

Word sense disambiguation (WSD) is the task of selecting the correct sense of a word in a specific context. Many applications of natural language processing (NLP), such as machine translation, information extraction, and question answering, require a semantic analysis, where WSD plays a crucial role. With its importance, WSD has been known as a very important field of NLP and studied steadily since the advent of NLP in the 1950s.

While there have been various studies to identify the sense of a word in a certain context, few WSD systems are known to be used for practical NLP applications, unlike part-of-speech (POS) taggers and syntactic parsers. This is because most WSD studies have focused only on a small number of polysemous words based on supervised learning approaches that require a sense tagged corpus. Since the construction of a sense tagged corpus is quite labor-intensive, only a small number of polysemous words were sense tagged and used for training WSD systems.

More specifically, the following difficulties may be encountered in constructing a sense tagged corpus:

1. The total number of sense tags used in a lexical database, such as a dictionary or WordNet, is very large. For example, there are about 60,000 distinct senses for nouns in WordNet alone, while the number of POS tags or tree tags is less than 1000.
2. The distinction between different sense tags for a word is sometimes unclear even for human judges. It is known that inter-agreement among human taggers is far from perfect in fine-grained sense distinctions (Ng and Lee, 1996). Consequently, it is not easy to find a corpus where all the words are properly tagged.

Unlike supervised learning approaches that require hand-labeled data, unsupervised approaches use a raw corpus.[1] or a lexical database without sense tags. Based on the types of resources used, unsupervised approaches are classified into the following approaches: raw corpus based, dictionary based, and WordNet based. Each approach are described in detail below.

Schütze (1998) presented a typical unsupervised approach based on a raw corpus. He clustered example sentences of a polysemous word based on the word similarity and regarded each cluster as a sense of the word. A new context with the polysemous word was assigned to the nearest cluster, and the sense of that word was determined by the sense related to the cluster. This approach has several deficiencies. Since clusters do not exactly correspond to the meanings of the words in a lexical database, it is not easy to identify the meaning of a word in the context by using only the cluster information. Also, it is difficult to apply the approach to the task of identifying the senses of all the words in a corpus because it requires a significant amount of time and space for clustering and storing the example sentences.

Another approach uses definitions of words in a dictionary. The words used in the definition of a sense of a word are distinct from those used in the definitions of the other senses of the same word. As a result, the definitions of the words can help disambiguate the senses. Given a context containing a polysemous word, WSD is reduced to a selection of a definition of the word that is

[1] In WSD field, the raw corpus refers to the corpus that is not sense tagged: thus it can be a POS-tagged corpus or a tree-tagged corpus.

most similar to the context words. A drawback of this approach is that definitions consisting of one or two short sentences are sometimes insufficient for WSD.

Both Luk (1999) and Karov and Edelman (1998) proposed a dictionary based approach. Luk (1999) employed concepts instead of words to supplement insufficient information regarding definitions. He defined *1792 defining concepts* from the definitions in the Longman Dictionary of Contemporary English (LDOCE) and acquired co-occurrence frequencies between the defining concepts from a raw corpus. Then, he calculated the similarity between a context and a definition by using the co-occurrence frequency between definition concepts. Karov and Edelman (1998) iteratively measured word similarity, sentence similarity and similarity between a definition and a sentence, and then made clusters by assigning sentences in a corpus to the most similar definition. Finally, they determined a sense of the word according to the similarity between the context and each cluster. However, the reliability of the additional information is not guaranteed because the size of the initial data in a definition is too small.

WordNet based approaches can be classified into the following three categories: WordNet gloss based, conceptual density based, and relative based. As gloss in WordNet is a definition of a synonym set, the WordNet gloss based approach is similar to the dictionary-based approach. However, the WordNet gloss based approach can utilize more disambiguation information than the dictionary based approach because the gloss of relatives of the word as well as the gloss of that word is available in WordNet. Both Fernandez-Amoros et al. (2001b) and Haynes (2001) augmented the definition of a sense with the definitions of relatives in WordNet in their WSD work. Nonetheless, they did not take into account the fact that words in a higher position of the WordNet hierarchy are less semantically related to each other than those in a lower position. Therefore, it is not appropriate to use glosses of relatives for a word in the higher position. Moreover, the definitions still do not contain sufficient information for WSD.

The conceptual density based approach identifies a sense by using the conceptual distance among the senses of a word in a context. It selects the sense with the shortest conceptual distance from other words in the context. A conceptual distance is usually defined as the number of links between two concepts in a hierarchical lexical database such as WordNet or a thesaurus. The more links between concepts, the longer the conceptual distance. Both Agirre and Rigau (1996) and Fernandez-Amoros et al. (2001a) utilized various relations among concepts in WordNet to calculate a conceptual distance.

WordNet specifies relationships among the meanings of words. Relatives of a word are defined as words that have a relation with it, e.g. they are synonyms, antonyms, superordinates (hypernyms), or subordinates (hyponyms). Relatives, especially those in a synonym class, usually have related meanings and tend to share similar contexts. Hence, relative-based approaches extract relatives of each sense of a polysemous word from WordNet, collect example sentences, and learn the senses from the example sentences for WSD. Yarowsky (1992) proposed this approach using International Roget's Thesaurus as a hierarchical lexical database instead of WordNet. However, the approach seems to suffer from examples irrelevant to the senses of a polysemous word since many of the relatives are polysemous. Leacock et al. (1998) attempted to exclude irrelevant or spurious examples by using only monosemous relatives in WordNet. However, some senses do not have short distance monosemous relatives through a relation such as synonym, child, and parent. A possible alternative of using only monosemous relatives in the long distance, however, is problematic because the longer the distance of two synsets in WordNet, the weaker the relationship between them. In other words, the monosemous relatives in the long distance may provide irrelevant examples for WSD.

Our approach to WSD also uses relatives in a lexical database similar to that of Yarowsky (1992) and Leacock et al. (1998). It is similar to other relative based approaches in that it acquires relatives from WordNet and extracts co-occurrence frequencies of the relatives from a raw corpus. However, it differs from the others in that it uses polysemous as well as monosemous relatives. To avoid the negative effect of weakly related relatives and polysemous relatives on co-occurrence frequency calculation, the proposed approach handles the example sentences of each relative separately instead of putting the example sentences of all relatives together into a pool.

The remaining sections of this paper are organized as follows: Section 2 explains the organization and characteristics of WordNet; Section 3 describes the proposed approach based on the relatives in WordNet; Section 4 presents experimental results on English data (SemCor and the corpus of SENSEVAL-2 lexical sample task) and Korean data; and Section 5 summarizes the characteristics of the proposed approach to WSD, provides some future research directions, and concludes the paper.

2. WordNet

2.1. Organization

WordNet [2] ongoing in Princeton University since the 1980s, is an on-line lexical database with a hierarchical structure, where a node is a synset (a set of synonyms) and a link is a relationship between two synsets. As a synset represents a meaning in WordNet, a polysemous word is present in more than one synset. A synset is associated with a gloss, where a definition and some example sentences of words in the synset are provided. Fig. 1 shows four synsets involving the word *chair*. Therefore *chair* is a polysemous word with four senses. Each numbered item represents a frequency of each sense (*35*, *2*, *0*, *0*), a synset and a gloss, in that order. Senses are sorted by the frequency that is extracted from the semantic concordance (SemCor).

WordNet consists of four parts: nominal, verbal, adverbial, and adjectival. Each part is organized differently in WordNet. In this paper, only the nominal part is described. [3] The relationships used for the nominal part are synonymy, antonymy, hypernymy/hyponymy, and meronymy/holonymy. The relationships are classified into two types: lexical and semantic. Synonymy and antonymy belong to the former, defined as being between word forms, while hypernymy/hyponymy and meronymy/holonymy belong to the latter, defined as being between word meanings (i.e. synsets).

Synonymy relates semantically similar words while antonymy relates semantically opposite words (e.g. *victory* has a synonymous relation with *triumph* [4] and an antonymous relation with

[2] We use WordNet 1.7.1 version.

[3] See Fellbaum (1998) for other parts.

[4] In fact, synonymy in WordNet is not expressed by a link, but by a synset. In other words, *victory* and *triumph* are in the same synset {*victory*, *triumph*}.

1. (35) {chair} -- (a seat for one person, with a support for the back; " *he put his coat over the back of the chair and sat down* ")
2. (2) {professorship, chair} -- (the position of professor; " *he was awarded an endowed chair in economics*")
3. (0) {president, chairman, chairwoman, chair, chairperson} -- (the officer who presides at the meetings of an organization; " *address your remarks to the chairperson*")
4. (0) {electric chair, chair, death chair, hot seat} -- (an instrument of execution by electrocution; resembles a chair; " *the murderer was sentenced to die in the chair*")

Fig. 1. Example: synsets including a word *chair*.

defeat). Hypernymy/hyponymy in WordNet links a synset to other synsets with more general/specific meanings (e.g. the synset {*seat*} is a hypernym of the synset {*chair*} and {*chair*} is a hyponym of {*seat*}). Meronymy/holonymy in WordNet is a part/whole relation between synsets (e.g. {*back*, *backrest*} is a meronym of {*chair*}, and {*chair*} is a holonym of {*back*, *backrest*}).

In particular, synonymy and hypernymy/hyponymy have central roles in WordNet: synonymy forms synsets that are basic units of WordNet, and hypernymy/hyponymy organize a hierarchical structure with synsets in WordNet.

2.2. Characteristics

The following characteristics of WordNet are pertinent to the WSD method we propose in this paper:

(1) Relatives of a word corresponding to a sense do not necessarily have a strong relationship among each other, although each relative is strongly related with the word. Especially, the relatives in a higher position in the WordNet hierarchy are less semantically related among each other than those in a lower position. For example, Fig. 2 shows the children of the word *object* at depth 2 [5] and the children of the word *chair* at depth 8, respectively. In the figure, we can observe that the children of *chair* are semantically closely related to each other, but those of *object* are not.

(2) Many senses of polysemous words do not have monosemous synonyms, children, and parents since they are usually located in a relatively high position in WordNet while monosemous words are located in relatively low positions. Fig. 3 represents the ratio between the number of senses of polysemous words and that of monosemous words according to WordNet depths. This figure shows that more polysemous words exist at depths 1 through 4 but more monosemous words exist at deeper levels. Thus, senses in higher position have fewer monosemous relatives than those in lower positions. Fig. 4 shows a distribution of the senses of polysemous words which do not have monosemous relatives with respect to WordNet depths.

(3) There are many polysemous relatives in WordNet. The number of senses of polysemous words is 40,002 in the nominal part of WordNet, and the number of polysemous relatives (syn-

[5] The depth of the root synset is 1.

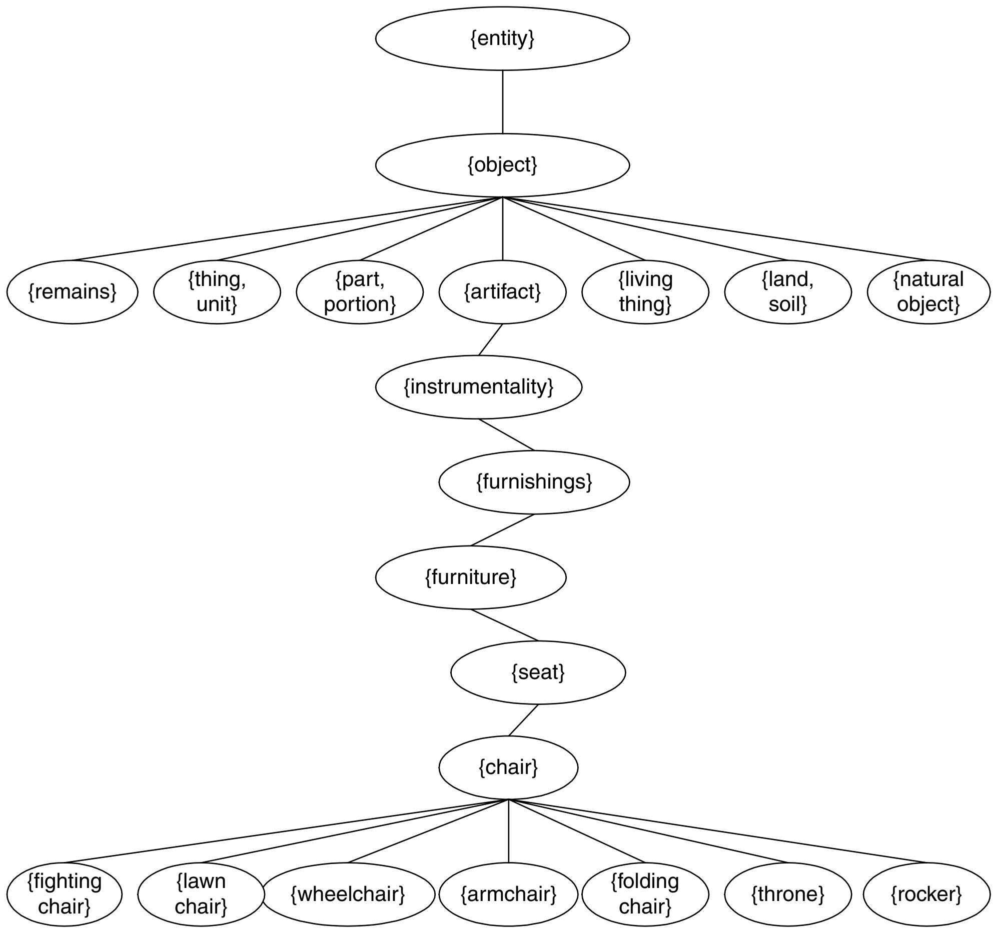

Fig. 2. Example: children of a word *object* at depth 2 and a word *chair* at depth 8.

onyms, children, parents) is 162,067. Therefore, there are 4.05 polysemous relatives per sense on average.

(4) There are many polysemous terms in the nominal part of WordNet: the number of polysemous words and phrases is 12,794 and 1687, respectively.

We argue that a WSD approach using the relatives in WordNet should take advantage of the above WordNet characteristics. In particular, we should avoid using weakly related relatives in sense disambiguation. Although it is not possible to distinguish weakly related words from strongly related ones without a sense-tagged corpus, it should be possible to measure how strongly a relative is related to the words forming the context around the target word by using co-occurrence frequency between the relative and the context words. More specifically, given a context for a target word, the relative most strongly related to the context is used to determine the sense of the word. Polysemous relatives can be handled in a similar manner. If there is a polysemous relative which is strongly related to a context, it is very helpful for disambiguating the sense of the target word in the context.

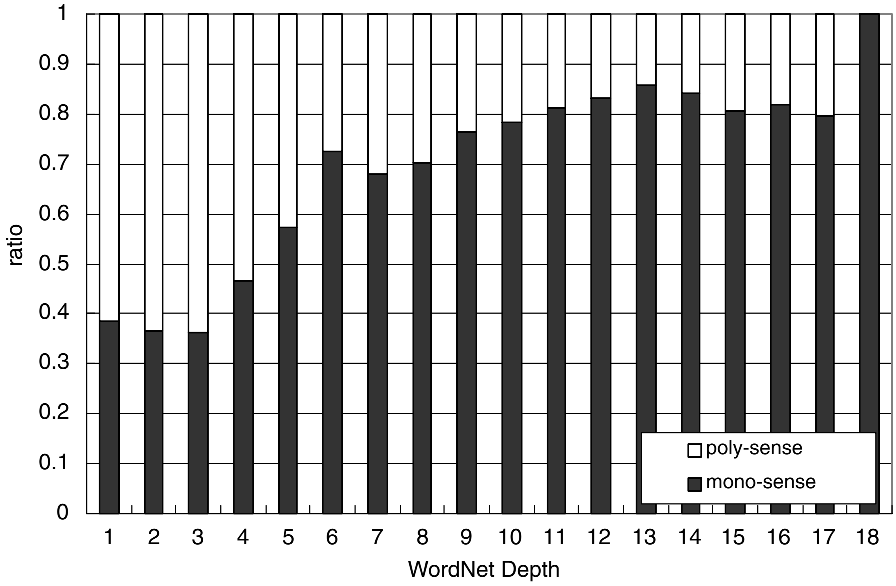

Fig. 3. Ratio between number of senses of polysemous words and monosemous words.

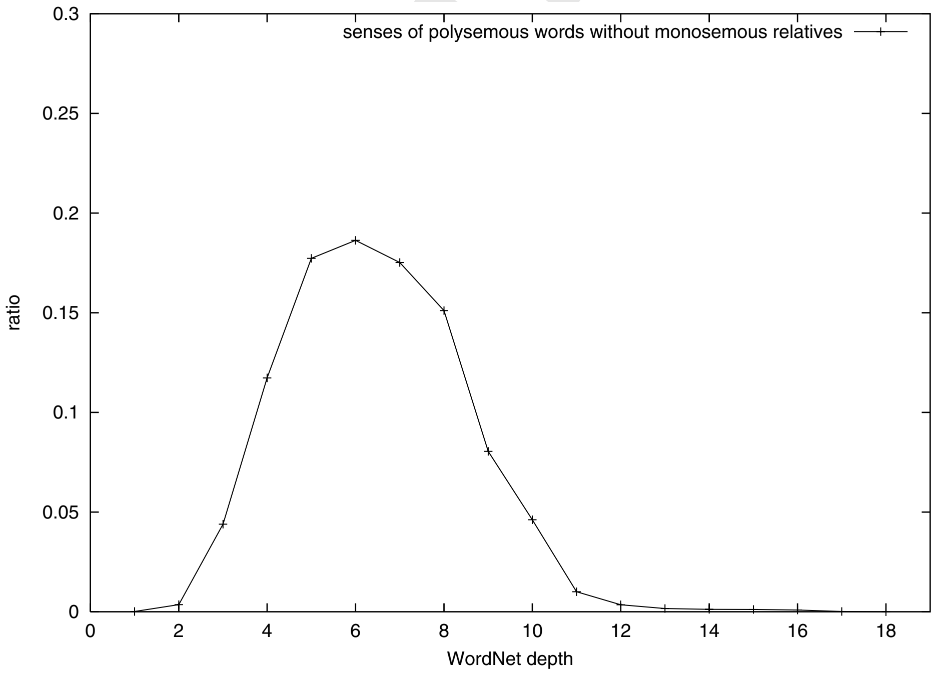

Fig. 4. Distribution of senses of polysemous words which do not have monosemous relatives (synonyms, children, parents).

Unlike the rough sketch of the logic behind our proposed method, the previously discussed approaches using relatives in WordNet did not appropriately utilize the characteristics of WordNet. Since they used example sentences of all relatives of a sense, the weakly related relatives may have a negative effect on the disambiguation by adding irrelevant or spurious examples to the training corpus. For example, the relatives *artifact* and *remains* of the word *object* in Fig. 2 are not strongly associated with each other, and thus the example sentences of the two words cannot provide disambiguation information to determine the correct sense of *object* in those contexts. The approach also suffers from polysemous relatives, because their example sentences includes different senses irrelevant to the target word that often make the training corpus inappropriate. For instance, the polysemous word *knight* has two senses in the nominal part of WordNet, and the second sense of *knight* (i.e. "*a chessman in the shape of a horse's head*") has the word *horse* as a synonym. The word *horse* with six senses in the nominal part of WordNet stands for a meaning *animal* in most contexts, but other senses (including the sense related to *knight*) of the word scarcely occurs in the corpus. Hence, the example sentences of *horse* are not helpful to identify the correct sense of *knight* in contexts. Besides, the approach cannot be practically applied to sense disambiguation of many words because collecting example sentences of relatives of many words requires too much space and time.

2.3. Korean WordNet

We have constructed Korean WordNet by manually mapping Korean words into synsets of English WordNet. [6] Since some senses of Korean words are not in English WordNet because of linguistic and cultural gaps, English WordNet has been expanded with synsets that are unique to Korean. The structure of Korean WordNet is the same as that of English WordNet, and is organized with all the relationships in English WordNet. At present, Korean WordNet consists of two parts: nominal and verbal. The nominal part contains 26,825 Korean nouns and 19,787 synsets, and the verbal part 1405 Korean verbs and 2058 synsets. Since Korean WordNet were constructed with the most frequent words in some Korean corpora, it covers over 90% of words in the corpus. Unlike English WordNet, Korean WordNet does not contain frequency information, and thus sense numbers of a word are not ordered by frequencies.

3. Word sense disambiguation using WordNet relatives

This section describes WordNet relatives, the proposed WSD method using WordNet relatives, and co-occurrence frequency matrix, which is used to efficiently disambiguate all polysemous nouns in WordNet.

[6] Hereafter, Princeton's WordNet is referred to as English WordNet as distinguished from Korean WordNet.

3.1. WordNet relatives

WordNet relatives of a word are defined to be the words in WordNet that are associated with the target word in terms of relationships such as synonyms, hypernyms, and meronyms. Relatives have two important characteristics for WSD. First, the relatives of a word sense are usually different from those of other senses of the same word. For example, *slope*, *incline*, and *riverbank* are relatives of the word *bank* when it has the meaning of *sloping land*, not the meaning of *financial institution*. As a result, it is possible to determine a sense of a word if appropriate relatives of the word can be selected with the help of the context in which the word occurs.

Second, the relatives of a sense tend to share common context words. As synonyms are interchangeable in most contexts, a hypernym or a hyponym of a word can also substitute for the word having a particular context even if the meaning of the context becomes more general or specific. For example, *chair* in a context "*address your remarks to the chair*" can be substituted with its synonyms, *president* or *chairman*. The hypernym *presiding officer* and the hyponym *vice chairman* can also be replaced for *chair*, without changing the original context drastically. Similarly, a meronym or a holonym can also replace the word with "*word with meronym*" or "*word of holonym*" without altering the overall meaning of the context. For example, the word *wheel* in the context "*I held the wheel*" can be expanded with its holonym *car* such as "*I held the wheel of the car*". In this example, the phrase "*the wheel of the car*" rather than the holonym *car* substitutes for the word *wheel*. Therefore, holonyms/meronyms of words can be regarded as possible substituents.

3.2. Word sense disambiguation

We disambiguate senses of a noun in a context [7] by selecting a substituent word from the relatives of the noun. Fig. 5 represents a flowchart of the proposed approach. Given a target word and its context, a set of relatives of the target word is created by searches in WordNet. Next, the most appropriate relative that can be substituted for the word in the context is chosen. In this step, co-occurrence frequency is used. Finally, the sense of the target word that is related to the selected relative is determined. If the selected relative is related to several senses of the target word, then the several senses are deemed to be proper senses. [8] For example, the word *slope* is a relative of the second and the ninth sense [9] of the word *bank* in WordNet. When the word *slope* is selected as a substituent word for the word *bank* in a context, both the second and the ninth senses are determined to be proper senses.

The example in Fig. 6 illustrates how the proposed approach disambiguates senses of the target word *chair* given the context. The set of relatives {*president*, *professorship*,...} of *chair* is built by WordNet searches, and the probability, "$Pr(professorship|Context)$", that a relative can be

[7] In this paper, a context indicates a sentence including a target word.

[8] In Section 4, we evaluated our approach on English SemCor, SENSEVAL-2 data, and Korean data. Among these data, only SENSEVAL-2 data allows multiple senses for an instance to be suggested. For SemCor and Korean data, we regard multiple senses of our system for an instance as an incorrect answer.

[9] The second sense of the word *bank* means "*sloping land* (*especially the slope beside a body of water*)", and the ninth sense means "*a slope in the turn of a road or track*; *the outside is higher than the inside in order to reduce the effects of centrifugal force*".

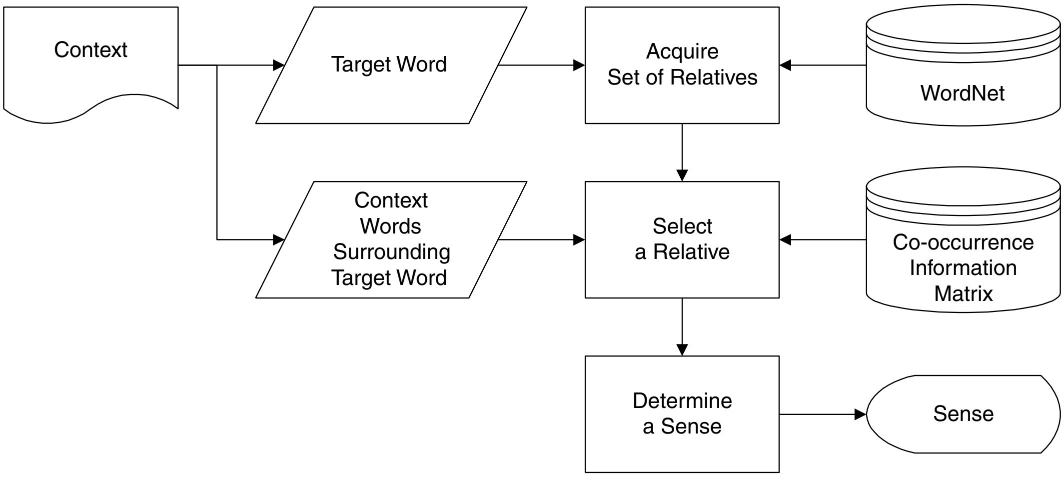

Fig. 5. Flowchart of the proposed approach.

substituted for the target word in the given context is estimated by the co-occurrence frequency between the relative and each of the context words. In this example, the relative, *seat*, is selected with the highest probability and the proper sense, "*a seat for one person*, *with a support for the back*", is chosen.

Thus, the second step of the proposed approach (i.e. selecting a relative) has to be carefully implemented to select the proper relative that can substitute for the target word in the context, while the first step (i.e. acquiring the set of relatives) and the third step (i.e. determining a sense) are done simply through searches in WordNet.

The substituent word of the ith target word tw_i in context C is defined to be the relative of tw_i which has the largest co-occurrence probability with the words in the context

$$\mathrm{SW}(tw_i, C) \stackrel{\text{def}}{=} \arg\max_{r_{ij}} P(r_{ij}^{\alpha}|C), \tag{1}$$

where SW is the substituent word, r_{ij} is the jth relative of tw_i, and r_{ij}^{α} is the αth sense related to tw_i. [10] If α is 2, the second sense of r_{ij} is related to tw_i. The right-hand side of Eq. (1) is calculated logarithmically under the assumption that words in C occur independently:

$$\begin{aligned}
\arg\max_{r_{ij}} P(r_{ij}^{\alpha}|C) &= \arg\max_{r_{ij}} \frac{P(C|r_{ij}^{\alpha})P(r_{ij}^{\alpha})}{P(C)} && (2)\\
&= \arg\max_{r_{ij}} P(C|r_{ij}^{\alpha})P(r_{ij}^{\alpha}) && (3)\\
&= \arg\max_{r_{ij}} \log P(C|r_{ij}^{\alpha}) + \log P(r_{ij}^{\alpha}) && (4)\\
&\approx \arg\max_{r_{ij}} \sum_{k=1}^{n} \log P(w_k|r_{ij}^{\alpha}) + \log P(r_{ij}^{\alpha}), && (5)
\end{aligned}$$

[10] α is a function with two parameters tw_i and r_{ij}, but it can be written briefly without parameters.

where w_k is the kth word in C and n is the number of words in C. In Eq. (5), we assume independence among the words in C.

The first probability in Eq. (5) is calculated as follows:

$$
\begin{aligned}
P(w_k|r_{ij}^{\alpha}) &= \frac{P(r_{ij}^{\alpha}|w_k)P(w_k)}{P(r_{ij}^{\alpha})} && (6)\\
&= \frac{P(r_{ij}^{\alpha}, r_{ij}|w_k)P(w_k)}{P(r_{ij}^{\alpha})} && (7)\\
&= \frac{P(r_{ij}|w_k)P(r_{ij}^{\alpha}|w_k, r_{ij})P(w_k)}{P(r_{ij}^{\alpha})} && (8)\\
&\approx \frac{P(r_{ij}|w_k)P(r_{ij}^{\alpha}|r_{ij})P(w_k)}{P(r_{ij}^{\alpha})} && (9)\\
&= \frac{P(r_{ij}|w_k)P(r_{ij}^{\alpha}, r_{ij})P(w_k)}{P(r_{ij}^{\alpha})P(r_{ij})} && (10)\\
&= \frac{P(r_{ij}|w_k)P(r_{ij}^{\alpha})P(w_k)}{P(r_{ij}^{\alpha})P(r_{ij})} && (11)\\
&= \frac{P(r_{ij}|w_k)P(w_k)}{P(r_{ij})}. && (12)
\end{aligned}
$$

We assume that r_{ij}^{α} is independent of w_k in Eq. (9).

The second probability in Eq. (5) is computed as follows:

$$P(r_{ij}^{\alpha}) = \beta(r_{ij}^{\alpha})P(r_{ij}), \tag{13}$$

where $\beta(r_{ij}^{\alpha})$ is the ratio of the frequency of r_{ij}^{α} to that of r_{ij}: [11]

$$\beta(r_{ij}^{\alpha}) = \frac{WNf(r_{ij}^{\alpha}) + 0.5}{n * 0.5 + WNf(r_{ij})}, \tag{14}$$

where $WNf(r_{ij}^{\alpha})$ is the frequency of r_{ij}^{α} in WordNet, $WNf(r_{ij})$ is the frequency of r_{ij} in WordNet, 0.5 is a smoothing factor, and n is the number of senses of r_{ij}.

Applying Eqs. (12) and (13) to Eq. (5), we have the following equation for acquiring the relative with the largest co-occurrence probability:

$$
\begin{aligned}
\arg\max_{r_{ij}} P(r_{ij}^{\alpha}|C) &\approx \arg\max_{r_{ij}} \sum_{k=1}^{n} \log\frac{P(r_{ij}|w_k)P(w_k)}{P(r_{ij})} + \log\beta(r_{ij}^{\alpha})P(r_{ij}) && (15)\\
&= \arg\max_{r_{ij}} \sum_{k=1}^{n} \log\frac{P(r_{ij}|w_k)}{P(r_{ij})} + \log\beta(r_{ij}^{\alpha})P(r_{ij}). && (16)
\end{aligned}
$$

In the case that several relatives have equally large co-occurrence probabilities, all senses related to the relatives are determined to be proper senses.

[11] As Korean WordNet does not contain sense frequency, β is defined as $1/n$ in Korean.

Context :
He should sit in the chair beside the desk.

Target Word :
'chair'

Set of Relatives :
{professorship, president, chairman, chairwoman, chairperson,
electronic chair, death chair, hot seat, seat, position, post, berth,
office, presiding officer, instrument of execution, ...}

Probability of Relative given the Context :
P(professorship | *He should sit in the beside the desk.*)
P(president | *He should sit in the beside the desk.*)
...
P(seat | *He should sit in the beside the desk.*)
...

Selected Relative :
'seat' - it is the most likely word occurred from the above context among the relatives of 'chair'

Determined Sense :
chair%1:06:00 - "a seat for one person, with a support for the back."
'seat' is the hypernym of the sense chair%1:06:00 of 'chair'.

Fig. 6. Example of sense disambiguation procedure for *chair*.

3.3. Co-occurrence frequency matrix

In order to select a substituent word for a target word in a given context, we must calculate the probabilities of finding relatives, given the context. These probabilities can be estimated based on the co-occurrence frequency between a relative and individual context words as follows:

$$P(r_{ij}) = \frac{freq(r_{ij})}{\mathrm{CS}}, \tag{17}$$

$$P(r_{ij}|w_k) = \frac{P(r_{ij}, w_k)}{P(w_k)} = \frac{freq(r_{ij}, w_k)}{freq(w_k)}, \tag{18}$$

where $freq(r_{ij})$ is the frequency of r_{ij}, CS is the corpus size, $P(r_{ij}, w_k)$ is the probability that r_{ij} and w_k co-occur, and $freq(r_{ij}, w_k)$ is the frequency that r_{ij} and w_k co-occur.

In order to calculate these probabilities, frequencies of words and word pairs are required. For this, we build a co-occurrence frequency matrix that contains co-occurrence frequencies of words pairs. In this matrix, an element m_{ij} represents the frequency that the ith word and jth word in the vocabulary co-occur in a corpus. [12] The frequency of a word can be calculated by counting all

[12] The co-occurrence frequency matrix is a symmetric matrix, thus m_{ij} is the same as m_{ji}.

frequencies in the same row or column. The vocabulary is composed of all content words in the corpus. Now, Eqs. (17) and (18) can be calculated through the matrix.

The matrix is easily built by counting each word pair in a given corpus. It is not necessary to make an individual matrix for each polysemous word, since the matrix contains co-occurrence frequencies of all word pairs. Hence, it is possible to disambiguate all words with only one matrix. In other words, the proposed method disambiguates the senses of all nominal words efficiently with only one matrix.

4. Experiment

Experiments were carried out on both English and Korean data. The English data consists of SemCor and the corpus of the SENSEVAL-2 lexical sample task. There are three directories in SemCor: *brown1*, *brown2* and *brownv*. In the files of the *brown1* and *brown2* directories, all content words (i.e. nouns, verbs, adjectives, adverbs) are annotated with the most appropriate WordNet senses and with POS tags, whereas in the files of the *brownv* directory, only verbs are tagged. Since our current research focuses on WSD of nouns, we only used the files of *brown1* and *brown2*.

For the SENSEVAL-2 lexical sample task, there is a corpus in which a small number of words are tagged with WordNet senses. Among the words with sense tags, we used only the noun component of the corpus for this experiment. The Korean data contains three Korean nouns [13] tagged with senses. Unlike the English data, which is tagged with fine-grained senses, the senses in the Korean data are coarse-grained. Detailed information of each corpus is described in Table 1. In this table, *num. of polysemous words* is the number of polysemous words, *num. of instances* is the number of instances of polysemous words in each corpus, *WordNet baseline* represents the recall when the first sense in WordNet are assigned to each word, and *most frequent sense baseline* represents the recall when each word is tagged with the most frequent sense in each sense tagged corpus. *Most frequent sense baseline* is usually used as the baseline for supervised approaches. In English WordNet, the order of senses is based on the frequency of senses on SemCor. Therefore, the first sense is the most frequent sense. Hence, the performance of *WordNet baseline* is similar to that of *most frequent sense baseline* on SemCor.

Co-occurrence frequency matrix for the English data is built based on the Wall Street Journal (WSJ) corpus in Penn Treebank II and some components of the LATIMES corpus in TREC. The WSJ and some parts of the LATIMES corpora contain about 3 million and 6 million words, respectively. The matrix for the Korean data is constructed based on the corpus containing ten million Korean words. Each matrix stores co-occurrence frequencies between words within a same sentence.

For the evaluation measure, we used the *recall* measure as defined for SENSEVAL, which is the percentage of right answers on all instances in the test set Edmonds and Cotton, 2001. [14] In

[13] Korean nouns are *bae*, *bam*, and *gogae*.
[14] As the proposed system disambiguates all instances, its coverage is 100% and its precision is the same as its recall.

Table 1
Experimental data

Corpus name	Number of polysemous words	Number of instance	WordNet baseline	Most frequent sense baseline
SemCor	5304	61,190	69.17%	70.62%
SENSEVAL-2	29	1754	42.47%	52.62%
Korean Data	3	9444	35.45%	82.21%

Table 2
Effects of relative types

Type of relatives	SemCor	SENSEVAL-2	Korean
Basic relative (BR)	48.39%	40.52%	69.16%
BR + antonym(a)	48.96%	40.24%	69.16%
BR + holonym(h)	50.30%	41.21%	72.24%
BR + meronym(m)	50.88%	45.42%	74.22%
BR + sister(s)	51.35%	40.64%	72.42%
BR + h + m	51.21%	45.37%	75.98%
BR + h + m + a	51.88%	**45.48**%	75.98%
BR + h + m + s	51.90%	42.97%	**76.60**%
BR + h + m + a + s	**52.34**%	43.03%	**76.60**%

SENSEVAL, three scoring schemes have been employed: fine-, coarse-, and mixed-grained. We adopted fine-grained scoring for the corpus of SENSEVAL-2, which scores the system with the match count between the system answers and the correct answers.

Two kinds of experiments were conducted in order to answer the following questions: how much each type of relatives in WordNet contributes to WSD, and how distant from a sense can be hypernyms/hyponyms to be considered for WSD.

4.1. Experiment 1: Contribution of relative types

In this experiment, we attempt to determine which type of relatives and which combination of types of relatives is useful for WSD. At first, we built the basic set of relatives (i.e. basic relatives) by using synonyms, hypernyms, and hyponyms, and then the basic relatives are extended with meronyms, holonyms, antonyms, and sisters. The experiments were conducted on the basic relatives, the extended relatives, and the various combinations of extended relatives. The results are presented in Table 2. From this table, we discover that the greater the number of types of relatives used, the better performance achieved on all data. Particularly, meronyms and holonyms are very valuable for WSD. Our approach turns out to be better than *WordNet baseline* on SENSEVAL-2 and the Korean corpus.

However, the combinations of types of relatives that achieve the highest performance differ according to the test data. For SemCor and Korean data, every type of relatives improves

Table 3
Effects of relative types on words in SENSEVAL-2 data

Word	WNB	BR	+a	+s	+h + m	+h + m + s
Art	44.90%	51.02%	51.02%	50.00%	51.02%	50.00%
Authority	40.22%	36.96%	36.96%	23.91%	36.96%	23.91%
Bar	41.06%	49.01%	49.01%	48.34%	49.01%	48.34%
Bum	2.22%	73.33%	73.33%	73.33%	73.33%	73.33%
Chair	79.71%	84.06%	84.06%	84.06%	84.06%	84.06%
Channel	13.70%	17.81%	17.81%	19.18%	17.81%	19.18%
Child	54.69%	45.31%	45.31%	37.50%	39.06%	40.63%
Church	56.25%	26.56%	26.56%	17.19%	50.00%	43.75%
Circuit	27.06%	38.82%	38.82%	31.76%	55.29%	34.12%
Day	62.07%	31.72%	28.26%	41.38%	56.55%	46.90%
Detention	65.63%	46.88%	46.88%	56.25%	46.88%	56.25%
Dyke	10.71%	82.14%	82.14%	85.71%	82.14%	85.71%
Facility	25.86%	27.59%	27.59%	27.59%	27.59%	27.59%
Fatigue	76.74%	13.95%	13.95%	23.26%	13.95%	23.26%
Feeling	56.86%	50.98%	50.98%	52.94%	50.98%	52.94%
Grip	15.69%	23.53%	23.53%	23.53%	25.49%	25.49%
Hearth	75.00%	65.63%	65.63%	62.50%	68.75%	62.50%
Holiday	83.87%	90.32%	90.32%	87.10%	90.32%	87.10%
Lady	69.81%	71.70%	71.70%	71.70%	71.70%	71.70%
Material	43.48%	49.28%	49.28%	49.26%	49.28%	49.28%
Mouth	53.33%	20.00%	20.00%	13.33%	51.67%	30.00%
Nation	78.38%	40.54%	40.54%	32.43%	43.24%	32.43%
Nature	45.65%	36.96%	36.96%	17.39%	41.30%	19.57%
Post	1.27%	16.77%	16.77%	16.77%	16.77%	16.77%
Restraint	17.78%	36.67%	36.67%	27.78%	36.67%	27.78%
Sense	37.74%	26.42%	26.42%	50.94%	26.42%	50.94%
Spade	27.27%	6.06%	6.06%	21.21%	6.06%	21.21%
Stress	2.56%	23.08%	23.08%	46.15%	23.08%	46.15%
Yew	17.86%	21.43%	21.43%	35.71%	21.43%	35.71%

performance and the best performance is achieved when we use all the types in combination, but for SENSEVAL-2, antonyms and sisters are sometimes irrelevant to performance improvement.

In order to analyze the different English data results, we investigate the performance of our approach for each word in the SENSEVAL-2 data. Table 3 shows the recall for each word in SENSEVAL-2 data, where *WNB* is a WordNet Baseline, *BR* is a basic relative, and $+h+m$ represents the case when a holonym and a meronym are added to the basic relative. In the table, we observe that some words, such as *sense*, *spade*, *stress* and *yew* are more correctly disambiguated with sisters, while other words, such as *authority*, *child*, *church*, and *circuit* are not. Hence, the contribution of sisters to WSD is dependent on the target words. Antonyms have a negative effect on two words *day* and *child* among 29 words in SENSEVAL-2 data, while antonyms generally improve the performance of the proposed approach on SemCor data. From

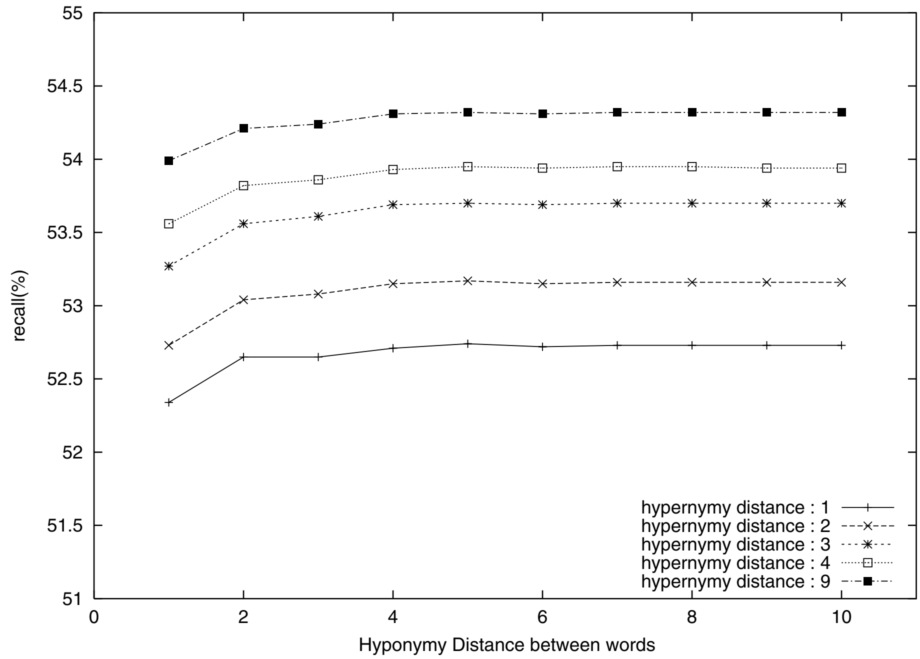

Fig. 7. Performances for hypernymy/hyponymy distances on SemCor.

these experimental results, we also find that the contribution of antonyms also relies on target words. From these observations, we can claim that sisters and antonyms are generally helpful for most words, but not for all words, and that it is desirable to use sisters and antonyms for all words task.

4.2. Experiment 2: Contributions of distant hypernyms/hyponyms

In this experiment, we have examined the impact of varying distances between a sense and hypernyms/hyponyms for WSD. The experiments were conducted with distances ranging from 1 to 10 with increments of 1, where hypernyms at distance 2 from a sense include its parents and grandparents, and hyponyms at distance 2 include its children and grandchildren. The combinations of relative types that gave the best performance in the previous experiment are repeated in this experiment. The experimental results are presented in Figs. 7 and 8.

For SemCor data, far hypernyms and hyponyms as well as near ones are valuable relatives, as shown in Fig. 7, while for SENSEVAL data, only near hypernyms and hyponyms are useful, as shown in Fig. 8. We can find the reason of the different results in Table 4, which shows the performance of the proposed method for each word in SENSEVAL-2 data with or without far hypernyms/hyponyms. Each result is acquired with the BR + h + m + a relatives of Table 2. In this table, *1, 10* means that hypernyms at distance 1 and hyponyms at distance 10 are used. Far hypernyms/hyponyms contribute to most words but not to all words. For example, far hypernyms are helpful for the word *channel*, but are useless for the word *bum*. Nevertheless, considering the

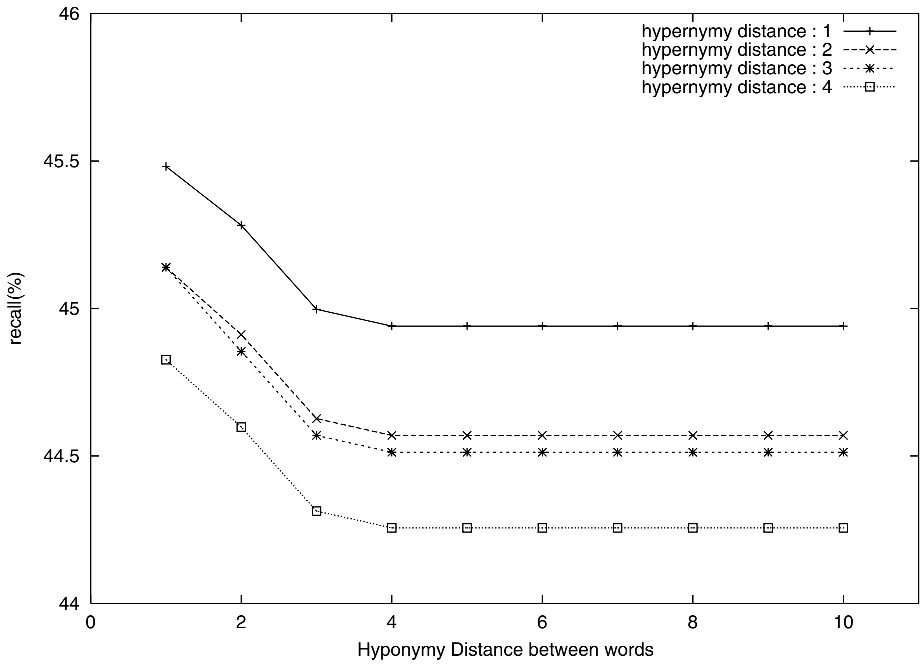

Fig. 8. Performances for hypernymy/hyponymy distances on SENSEVAL-2.

results of SemCor, it is desirable to utilize far hypernyms/hyponyms in order to disambiguate all content words in a general domain. [15]

4.3. Comparison with other works

Agirre and Rigau (1996) and Fernandez-Amoros et al. (2001a) have also evaluated their unsupervised approaches with SemCor. Both approaches have tried to disambiguate senses of nouns based on conceptual density. Although their approach used different versions of WordNet and SemCor, the difference is not significant, making it possible to compare our approach with theirs.

Agirre and Rigau (1996) employed WordNet 1.4 and evaluated their approach on four files (*br-a01*, *br-b20*, *br-j09*, *br-r05*) in SemCor, which contains 1256 occurrences of polysemous words. On the other hand, our approach uses WordNet 1.7.1 and tested the approach on the same files of the current version of SemCor, which has 1338 occurrences of polysemous words. Experimental results show that our method is clearly better than Agirre and Rigau (1996) in all measures, as shown in Table 5.

Another comparison is done with Fernandez-Amoros et al. (2001a). They tested their method on every noun in 171 SemCor documents and reported a 31.3% recall of their system. On the other hand, our approach achieves 52.34% recall when it is evaluated on 186 documents in the current version of SemCor.

[15] SemCor is a part of Brown Corpus, which covers press reportage, fiction, scientific text, legal text and so on.

Table 4
Effects of distant relatives on words in SENSEVAL-2 data

Word	1, 1	1, 10	10, 1	10, 10
Art	51.02%	57.14%	53.06%	58.16%
Authority	36.96%	34.78%	32.61%	31.52%
Bar	49.01%	49.01%	49.01%	49.01%
Bum	73.33%	73.33%	17.78%	17.78%
Chair	84.06%	84.06%	84.06%	84.06%
Channel	17.81%	17.81%	34.25%	34.25%
Child	42.19%	45.31%	50.00%	51.56%
Church	50.00%	50.00%	54.69%	54.69%
Circuit	55.29%	55.29%	49.41%	49.41%
Day	56.55%	53.10%	55.86%	52.41%
Detention	46.88%	46.88%	75.00%	75.00%
Dyke	82.14%	82.14%	25.00%	25.00%
Facility	27.59%	27.59%	29.31%	27.59%
Fatigue	13.95%	13.95%	11.63%	11.63%
Feeling	50.98%	54.90%	50.98%	54.90%
Grip	25.49%	25.49%	21.57%	21.57%
Hearth	68.75%	68.75%	65.63%	65.63%
Holiday	90.32%	48.39%	90.32%	51.61%
Lady	71.70%	71.70%	71.70%	71.70%
Material	49.28%	49.28%	49.28%	49.28%
Mouth	51.67%	51.67%	46.67%	46.67%
Nation	43.24%	43.24%	40.54%	40.54%
Nature	41.30%	39.13%	32.61%	30.43%
Post	16.77%	16.77%	16.77%	16.77%
Restraint	36.67%	35.56%	33.33%	35.56%
Sense	26.42%	30.19%	26.42%	30.19%
Spade	6.06%	6.06%	30.30%	30.30%
Stress	23.08%	23.08%	30.77%	30.77%
Yew	21.43%	21.43%	89.29%	89.29%

For the corpus of SENSEVAL-2 lexical sample task, the proposed approach can be compared with the unsupervised systems that participated in SENSEVAL-2. There are four systems (Kilgarriff, 2001): ITRI-WASPS, UNED-LS-U, CLresearch DIMAP, and IIT-2. ITRI-WASPS Tugwell and Kilgarriff, 2001 was a semi-automatic system which adopted a bootstrapping algorithm with manual patterns. UNED-LS-U (Fernandez-Amoros et al., 2001b) and IIT-2 (Haynes, 2001) used the definition of each word in WordNet, as described in Section 1. CLresearch DIMAP (Litkowski, 2001) used a dictionary containing disambiguation information. Table 6 shows the experimental results [16] for each system. The table shows that our ap-

[16] The answers of each system are publicly available. We extracted nominal parts from the answers and scored them with a scoring program for SENSEVAL-2.

Table 5
Comparison with Agirre and Rigau (1996)

	Coverage	Precision	Recall
Agirre and Rigau (1996)	79.6%	43%	34.2%
Our method	100%	54.33%	54.33%

Table 6
Comparison with other unsupervised systems in SENSEVAL-2

	Coverage	Precision	Recall
ITRI-WASPS	91.73%	55.62%	51.03%
UNED-LS-U	100%	44.50%	44.50%
CLresearch DIMAP	100%	34.32%	34.32%
IIT-2	100%	30.84%	30.84%
Our method	100%	45.48%	45.48%

proach slightly outperforms the best automatic system, UNED-LS-U, [17] except for semi-automatic system ITRI-WASPS.

5. Conclusions

We have proposed a method that determines the sense of a nominal word in a context by selecting a substituent word from WordNet relatives of the nominal word. Since each relative is usually related to only one sense of the target word, our approach identifies the proper sense with the selected relative. The substituent word is selected based on the co-occurrence frequency between the relative and the words surrounding the target word in a given context. We collected the co-occurrence frequency from a raw corpus, not a sense-tagged one that is often required by other approaches. In short, the proposed method disambiguates senses of words only through the set of WordNet relatives of the target words and a raw corpus.

In this research, we have also investigated the characteristics of WordNet that should be taken into account for WSD: In WordNet, (1) not all relatives have a strong relationship among each other, (2) many senses of polysemous words do not have any monosemous relatives, (3) there are many polysemous relatives, and (4) there are many polysemous words.

We have tried to reflect these characteristics into the proposed method. As a result, the proposed method (1) handles the relatives individually and thus the relatives do not interfere with each other, (2) makes use of polysemous relatives as well as monosemous relatives, (3) controls polysemous relatives effectively by excluding the polysemous relatives that are not related to the target word in context, and (4) uses a co-occurrence frequency matrix in order to efficiently disambiguate the senses of all target words.

[17] UNED-LS-U, CLresearch DIMAP and IIT-2 are automatic systems.

We tested the proposed method on SENSEVAL-2 data, SemCor data, and Korean data. The experimental results show that the proposed method disambiguates many polysemous words in SemCor data, a small number of words in SENSEVAL-2 data and Korean data effectively, and achieves better performance than the WordNet baseline model. Furthermore, the proposed method appears to outperform other unsupervised approaches when we compare the proposed method using SemCor and SENSEVAL-2 data.

We have also conducted experiments in order to examine which types of relatives are important for WSD and to what extent distant hypernyms/hyponyms contribute to WSD. The results show that most relative types are useful, that sisters and antonyms are not helpful for all words, and that far hypernyms/hyponyms are not useful on all words. However, many words are disambiguated correctly using sisters, antonyms and far hypernyms/hyponyms. Based on these results, we claim that the importance of sisters, antonyms and distant relatives depend on polysemous words or senses.

For future research, we will investigate the dependency between the types of relatives and the characteristics of words or senses in order to devise an improved method that better utilizes various types of relatives for WSD. Since it was difficult to generalize the SENSEVAL-2 data, especially in comparison with the SemCor data, we plan to evaluate our approach on more polysemous words in SENSEVAL-1 data. This will allow us to make finer conclusions on proper relative types for the polysemous words. As an extension to the current approach, we are considering a way to utilize the similarity between definitions of words in WordNet.

References

Agirre, E., Rigau, G., 1996. Word sense disambiguation using conceptual density. In: Proceedings of COLING'96, Copenhagen Denmark, 1996, pp. 16–22.

Edmonds, P., Cotton, S., 2001. SENSEVAL-2: Overview. In: Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems, Toulouse, France, 2001, pp. 1–5.

Fellbaum, C., 1998. An WordNet Electronic Lexical Database. MIT Press, Cambridge, MA.

Fernandez-Amoros, D., Gonzalo, J., Verdejo, F., 2001a. The role of conceptual relations in word sense disambiguation. In: Proceedings of the 6th International Workshop on Applications of Natural Language for Information Systems, Madrid, Spain, pp. 87–98.

Fernandez-Amoros, D., Gonzalo, J., Verdejo, F., 2001b. The UNED systems at SENSEVAL-2. In: Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems, Toulouse, France, pp. 75–78.

Haynes, S., 2001. Semantic tagging using WordNet examples. In: Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems, Toulouse, France, pp. 79–82.

Karov, Y., Edelman, S., 1998. Similarity-based word sense disambiguation. Computational Linguistics 24 (1), 41–59.

Kilgarriff, A., 2001. SENSEVAL-2: Overview. In: Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems, Toulouse, France, pp. 17–20.

Leacock, C., Chodorow, M., Miller, G.A., 1998. Using corpus statistics and WordNet relations for sense identification. Computational Linguistics 24 (1), 147–165.

Litkowski, K., 2001. SENSEVAL-2: Overview. In: Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems, Toulouse, France, pp. 107–110.

Luk, A.K., 1999. Statistical sense disambiguation with relatively small corpora using dictionary definitions. In: Proceedings of the 33rd Annual Meeting of the ACL, June, pp. 181–188.

YCSLA 254
DISK / 18/6/04
ARTICLE IN PRESS
No. of pages: 21
DTD 4.3.1/SPS

Ng, H.T., Lee, H.B., 1996. Integrating multiple knowledge sources to disambiguate word sense: an exemplar-based pproach. In: Joshi, A., Palmer, M. (Eds.), Proceedings of the Thirty-Fourth Annual Meeting of the Association for Computational Linguistics. Morgan Kaufmann, San Francisco, pp. 40–47.

Schütze, H., 1998. Automatic word sense discrimination. Computational Linguistics 24 (1), 97–123.

Tugwell, D., Kilgarriff, A., 2001. SENSEVAL-2: Overview. In: Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems, Toulouse, France, pp. 151–154.

Yarowsky, D., 1992. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. In: Proceedings of COLING-92. Nantes, France, July, pp. 454–460.