


Article

Word Sense Disambiguation Using Prior Probability Estimation Based on the Korean WordNet

Minho Kim ¹  and Hyuk-Chul Kwon ^{2,*}¹ Department of Software, Catholic University of Pusan, Busan 46252, Korea; minho@cup.ac.kr² School of Computer Science and Engineering, Pusan National University, Busan 46241, Korea

* Correspondence: hckwon@pusan.ac.kr

Abstract: Supervised disambiguation using a large amount of corpus data delivers better performance than other word sense disambiguation methods. However, it is not easy to construct large-scale, sense-tagged corpora since this requires high cost and time. On the other hand, implementing unsupervised disambiguation is relatively easy, although most of the efforts have not been satisfactory. A primary reason for the performance degradation of unsupervised disambiguation is that the semantic occurrence probability of ambiguous words is not available. Hence, a data deficiency problem occurs while determining the dependency between words. This paper proposes an unsupervised disambiguation method using a prior probability estimation based on the Korean WordNet. This performs better than supervised disambiguation. In the Korean WordNet, all the words have similar semantic characteristics to their related words. Thus, it is assumed that the dependency between words is the same as the dependency between their related words. This resolves the data deficiency problem by determining the dependency between words by calculating the χ^2 statistic between related words. Moreover, in order to have the same effect as using the semantic occurrence probability as prior probability, which is used in supervised disambiguation, semantically related words of ambiguous vocabulary are obtained and utilized as prior probability data. An experiment was conducted with Korean, English, and Chinese to evaluate the performance of our proposed lexical disambiguation method. We found that our proposed method had better performance than supervised disambiguation methods even though our method is based on unsupervised disambiguation (using a knowledge-based approach).

Keywords: word sense disambiguation; Korean WordNet; knowledge-based model; data mining; information extraction



Citation: Kim, M.; Kwon, H.-C. Word Sense Disambiguation Using Prior Probability Estimation Based on the Korean WordNet. *Electronics* **2021**, *10*, 2938. <https://doi.org/10.3390/electronics10232938>

Academic Editor: Arkaitz Zubiaga

Received: 31 August 2021

Accepted: 23 November 2021

Published: 26 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The present paper addresses lexical disambiguation occurring in the semantic analysis phase of the natural language analysis process that includes cases of ambiguity. In natural language processing, lexical disambiguation refers to the determination of the correct semantic meaning for a word that has multiple meanings (hereafter referred to as an ambiguous word) by evaluating the meaning in its context [1]. Lexical disambiguation, which is the same as morphological analysis and syntactic analysis, is essential in natural language processing and plays an important role in various application areas. In machine translation, lexical disambiguation is critical to select the correctly translated word for a given word. For example, the English verb ‘build’ can be translated into Korean as construct, build, produce, establish, or develop, and the word that is the most correct should be selected from among these. In information search systems, lexical disambiguation of a query word can provide the high-quality information that a user needs. For example, if a query word inputted by a user is court, the search engine should present the results by categorizing the information into courthouse-related and palace-related suggestions. In addition, it is important to resolve semantic ambiguity in text mining for documents in specialized fields such as medical documents [2,3].

Lexical disambiguation has been a primary interest since the 1950s when natural languages began to be processed by computers. Its study has been conducted based on the following two lexical disambiguation methods. The first is a method based on knowledge bases such as machine-readable dictionaries. The second is a method based on statistical information extracted from large amounts of corpus data. In particular, since the 1990s, studies based on large amounts of corpus data have been actively conducted. In this method, a problem of word sense ambiguity has been simplified as a statistical classification problem in machine learning so that traditional machine learning techniques (for example, case-based learning, decision tree, and Bayesian classifier) are applied to solve the problem. Lexical disambiguation through machine learning is divided into supervised and unsupervised disambiguation, depending on whether a corpus consisting of individual sense-tagged words (hereafter referred to as a sense-tagged corpus) is used for the learning [4].

In lexical disambiguation, supervised disambiguation using a large amount of sense-tagged corpus has shown better performance than other lexical disambiguation methods. However, the construction of a large sense-tagged corpus takes high cost and time, which is a drawback. On the other hand, while implementing unsupervised disambiguation is relatively easy, their performance is usually not satisfactory. In particular, Korean lacks language resources such as machine-readable dictionaries and sense-tagged corpora, compared with English. Therefore, in order to overcome the limitations of these linguistic resources in minority languages such as Korean and Vietnamese, it is urgent to study a method for clarification of vocabulary [5].

In this paper, a novel, unsupervised disambiguation method that shows better performance than existing knowledge-based lexical disambiguation and unsupervised lexical disambiguation methods, without the need of a large sense-tagged corpus, is proposed. Generally, the reason for the low accuracy of knowledge-based lexical disambiguation and unsupervised lexical disambiguation is the lack of the semantic occurrence probability of ambiguous words and the data deficiency problem that occurs while dependency between words is being determined. The novel, unsupervised disambiguation method proposed in this paper uses prior probability estimation based on the Korean WordNet [6], which takes advantage of the semantic characteristic that all words share the same semantic characteristics with their related words (hypernym, hyponym, and coordinate term). Thus, it is assumed that a dependency between words is the same as a dependency between related words so that a data deficiency problem is solved by determining the dependency between words by calculating the chi-square statistic between related words. Moreover, in order to have the same effect when using the semantic occurrence probability as using prior probability, which is used in supervised disambiguation, semantically related words of ambiguous vocabulary are obtained and utilized as prior probability data.

The present paper is organized as follows: In Section 2, existing studies on lexical disambiguation are summarized. The lexical disambiguation method using related words in the Korean WordNet, which is proposed in this paper, is explained in Section 3. In particular, a solution to the data deficiency problem using the expansion of related words and a method of using semantically related words as the prior probability is explained in detail. In Section 4, the experimental method and results are described. Finally, in Section 5, conclusions and future research are discussed.

2. Related Study

Lexical disambiguation has been a major concern since natural language began being processed with computers in the 1950s, but its research achievement has been insufficient compared with studies on morphological analysis. In morphological analysis, part-of-speech tagging accuracy is generally more than 95% with respect to all vocabularies in a total corpus. On the other hand, in lexical disambiguation, sense-tagged accuracy for frequently used specific words is only 80% to 90%.

In early research on lexical disambiguation, studies based on knowledge bases such as machine-readable dictionaries were conducted actively. The study of Lesk [7] can be referred to as a representative example. Lesk identified the meaning of an ambiguous word according to the multiplicity between the words used in the definition of the ambiguous word in a dictionary and the words used in the definition of neighboring words of the ambiguous word in a dictionary. This method had the advantage that it did not require high-cost language resources, and implementation was relatively simple. It did suffer from a severe data deficiency problem. However, that occurred due to its requirement of a highly accurate match between words, only showing a low accuracy of 50% to 70%. To minimize this problem, Luk [8] proposed a method of extracting common words as a definition concept from the Longman modern English dictionary and then extracting statistical information regarding the definition concept from the Brown Corpus to remove ambiguity. However, this method did not provide a fundamental solution to the data deficiency problem.

A study based on a knowledge base opened a new era by utilizing WordNet, a lexical semantic network developed for lexical disambiguation. WordNet calculates the shortest path between meanings in a word sense disambiguation study using a lexical–semantic network [6,9,10]. The similarity or semantic relationship type between meanings is found using the hierarchy’s distance from the highest meaning. Resnik [10] proposed a method of measuring the semantic similarity of nouns in the IS-A hierarchy relationship in WordNet for the use of lexical disambiguation. Agirre et al. [11] defined conceptual density that calculated a distance between words using the semantic relationships of WordNet to calculate the conceptual density between co-occurrence words within the context that included ambiguous words, thereby determining the meaning of the ambiguous word. Mihalcea et al. [12,13] proposed a technique to remove the word sense ambiguity by obtaining the co-occurrence statistical data between the two words and then measuring the semantic density between the two words through WordNet and removing the word sense ambiguity based on the rank. Other than the similarity calculation between senses or concepts based on WordNet, additional studies on lexical disambiguation using WordNet can be found, such as Pederson et al. [14] and Ganesh et al. [15]. Ted proposed a method using a modified dictionary-based algorithm of Lesk that was applicable to WordNet. Ganesh proposed a method that determined the word sense by choosing the synset (synonym set) of the gloss that had the highest similarity once the similarity between the words in the context that included the target word and gloss in WordNet was calculated using cosine and Jacquard similarity. Such WordNet-based lexical disambiguation techniques have an advantage that mitigates the data deficiency problem by expanding an ambiguous word and co-occurrence words used with the ambiguous word.

A graph-based word sense disambiguation method using WordNet is also one of the widely studied methods [13,16–18]. Such methods convert an input sentence into a graph format that has a synset of WordNet as a basic unit and calculates the semantic similarity of the global context rather than the local context using lexical chains. A lexical chain is a sequence of related words in writing that is referred to as a unit that represents consistent meaning in context or a whole paragraph. That is, rather than calculating semantic similarity between words in a local context, the semantic similarity between lexicon chains, or a lexicon chain and a word, is calculated so that information that is more accurate can be obtained for lexical disambiguation. In graph-based lexical disambiguation, well-known graph-based technologies are used to structure the graph, thereby determining the optimal lexicon chain. The graph-based lexical disambiguation method showed the best performance among the methods of lexical disambiguation utilizing WordNet, but it had the drawback that it took a long time to determine the optimal lexicon chain when the graph structure was complicated.

In the case of the Korean language, a large-scale lexical semantic network such as WordNet did not exist in the early days of research on lexical disambiguation, and so studies based on statistical-based lexical disambiguation were conducted. Since 2000,

several lexical–semantic networks have been developed, and thus studies based on lexical–semantic networks have been conducted to overcome word sense ambiguity. Heo et al. [19] proposed a lexical disambiguation model utilizing mutual information extracted from the Korean Noun Concept Network (ETRINET), a compound noun sense-tagged dictionary and raw corpus.

Like other tasks in the field of natural language processing, deep learning-based supervised learning models show good performance in resolving word sense disambiguation [20–23]. However, these models are expensive to build the training data because they require a large corpus containing semantic information for word senses. Therefore, using external resources such as WordNet, knowledge-based word sense disambiguation is a good approach for word sense disambiguation [24–26]. In this study, the relationship between ambiguous words and co-occurrence words within a context is determined using a Korean lexical–semantic network. Moreover, to have the effect of using prior information in supervised disambiguation, semantically related words of ambiguous vocabulary are obtained and utilized as prior probability data.

3. Lexical Disambiguation Using the Korean Lexical Semantic Network

This section explains the unsupervised lexical disambiguation method using the Korean Lexical Semantic Network (KorLex), which is proposed in this paper. Generally, supervised disambiguation shows better performance than unsupervised disambiguation but requires a large-scale, sense-tagged corpus. In this paper, a sense-tagged corpus, which involves a high development cost, is not used. In its place, a morph-tagged corpus of 5 M word phrases is used. Moreover, richer statistical information is exploited through the expansion of semantically related words by utilizing KorLex (Korean Lexical Semantic Network), and prior probability is estimated by calculating semantically related words of ambiguous words.

3.1. Analysis of Relationship between Words Using the Korean Lexical Semantic Network (KorLex)

The Korean Lexical Semantic Network (KorLex) was developed using WordNet as a reference model and included approximately 130,000 synsets and about 150,000 word senses. A synset is a set of synonyms that share the same word sense. In this paper, a word that has more than two synsets in the Korean Lexical Semantic Network is considered as an ambiguous word. For example, in Korean, sagwa is an ambiguous word that has two synsets, sagwa 1 meaning ‘apology’ and sagwa 2 meaning ‘fruit of apple tree (apple)’. To distinguish such ambiguous words, semantic relation words are used. Semantic relation words are words in semantic relations in the hierarchy of the Korean Lexical Semantic Network. Semantic relation words are called hypernym, hyponym, and coordinate terms, depending on the relationship. Figure 1 shows the relation words of sagwa 2 in the Korean Lexical Semantic Network.

Relation words in a hierarchy of the Korean Lexical Semantic Network have the same characteristic. In particular, coordinate terms in relation words have the same co-occurrence words. For example, sagwa (apple) and boksunga (peach) are hyponyms of gwail (fruit), and so are related with mukda (eat) and masitda (delicious). Sagwa (apology) and gamsa (appreciation), however, are hyponyms of inji (recognition), that has no relationship to mukda (eat) or masitda (delicious). Thus, word sense ambiguity can be removed by identifying the relationship of the semantic relation words of ambiguous and co-occurrence words in the local context. Figure 2 shows the relationship between coordinate terms of Sagwa and words in the local context in the Korean Lexical Semantic Network.

The most basic method used to analyze the relationship between two words is to determine the frequency of co-occurrence of the two words. That is, finding how often two words are used in a local context will be a measure to determine the relationship between the two words. However, because of words that are used normally, regardless of the meaning of the ambiguous word, the frequency of the co-occurrence cannot determine the relationship between two words. To overcome this, various kinds of statistical

approaches are used, such as information-theoretic measures, likelihood measures, statistical hypothesis tests, and coefficients of association strength. Among them, we use the chi-square independence test, which is known to be easy to interpret and effective in finding collocations [27–29].

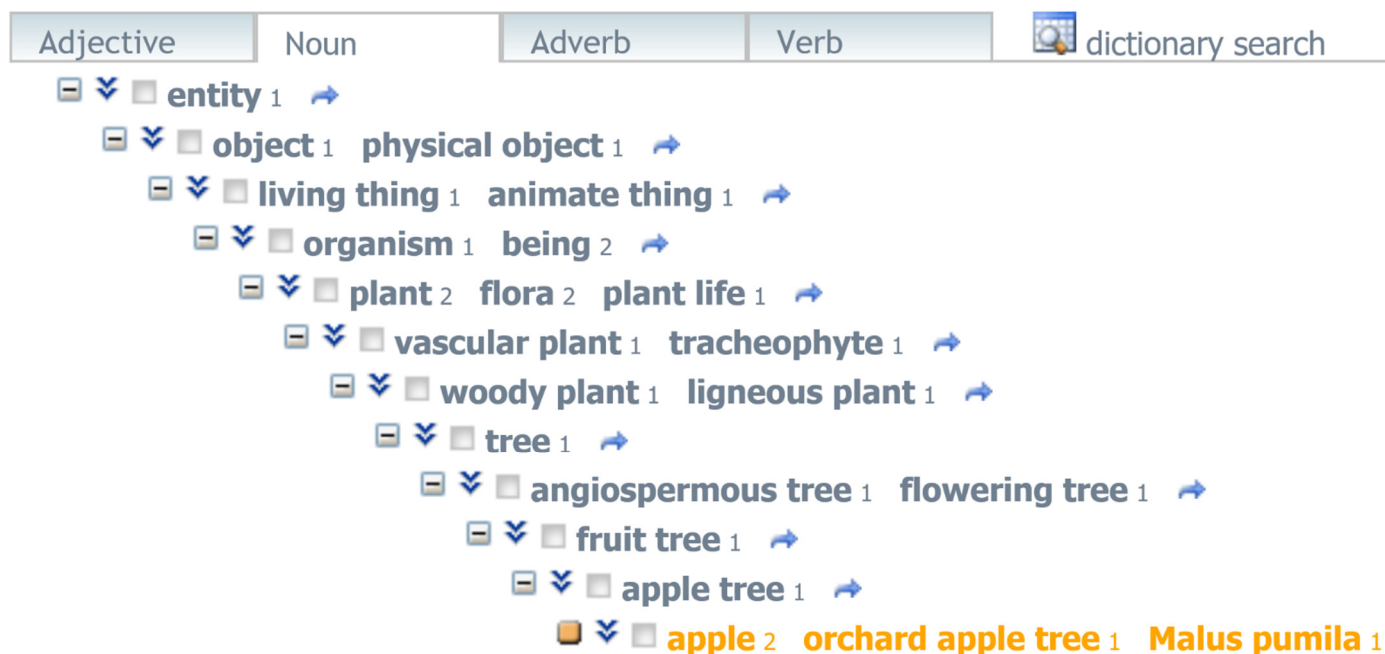


Figure 1. Example of Korean Lexical Semantic Network.

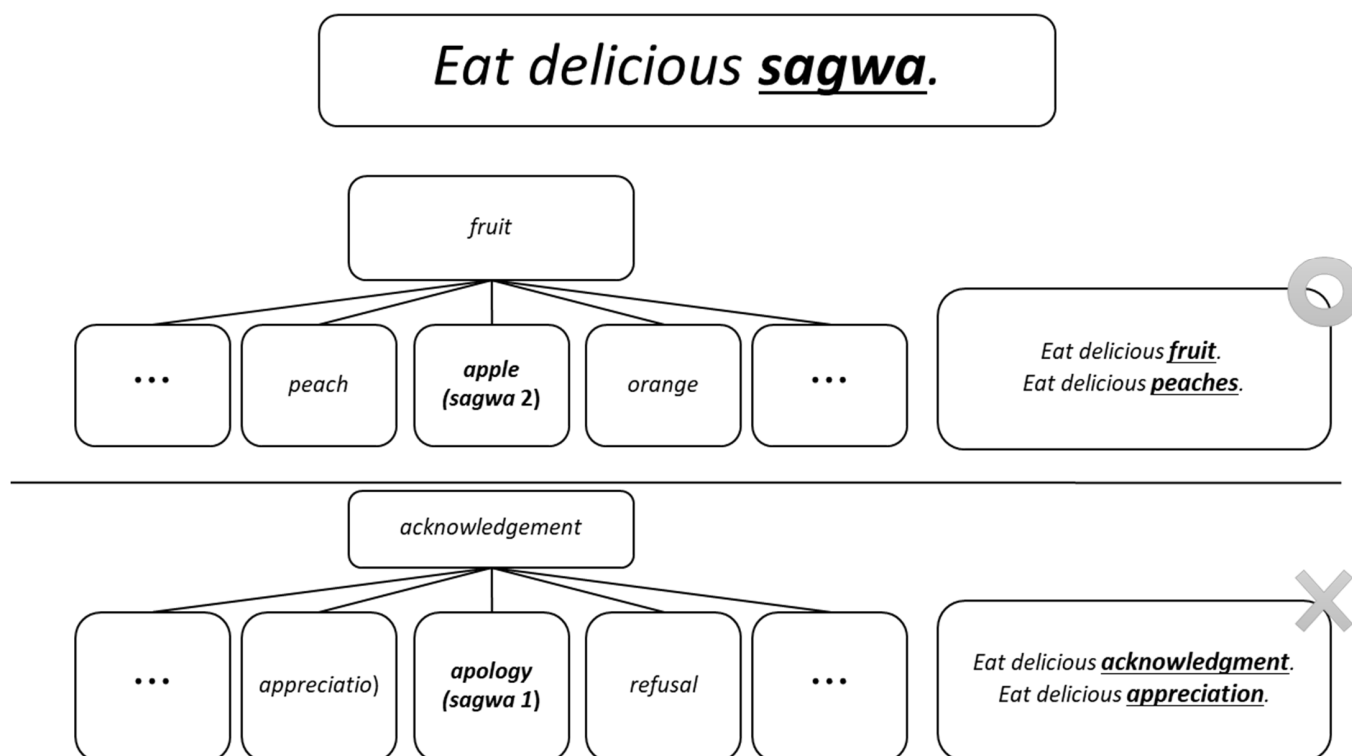


Figure 2. Coordinate terms of sagwa and relationship between words appearing in the local context.

Figure 3 shows the relationship analysis between two words using the semantic relation words of an ambiguous word. Assuming that the meaning of an ambiguous word

w_{amb} is s_k and co-occurrence word is v_i , then the chi-square statistic of the two words $x^2(w_{amb} = s_k, v_i)$ is calculated by the following formula according to the relation word r_i of s_k .

$$x^2(w_{amb} = s_k, v_i) = \frac{\sum_{i=1}^l x^2(r_i, v_i)}{l}, \quad (1)$$

Here, if the relation word r_i is an ambiguous word, it would cause a problem when calculating the chi-square statistic. One of the solutions is to exclude ambiguous words from the related words. However, it would be dangerous to exclude ambiguous words since it could, in the worst case, remove all related words. This would also not help to reduce the data deficiency problem. In this study, then, assuming that the semantic frequency of an ambiguous word is the same, the frequency of an ambiguous word w_{amb} that has m meanings is calculated as $c(w_{amb}/m)$.

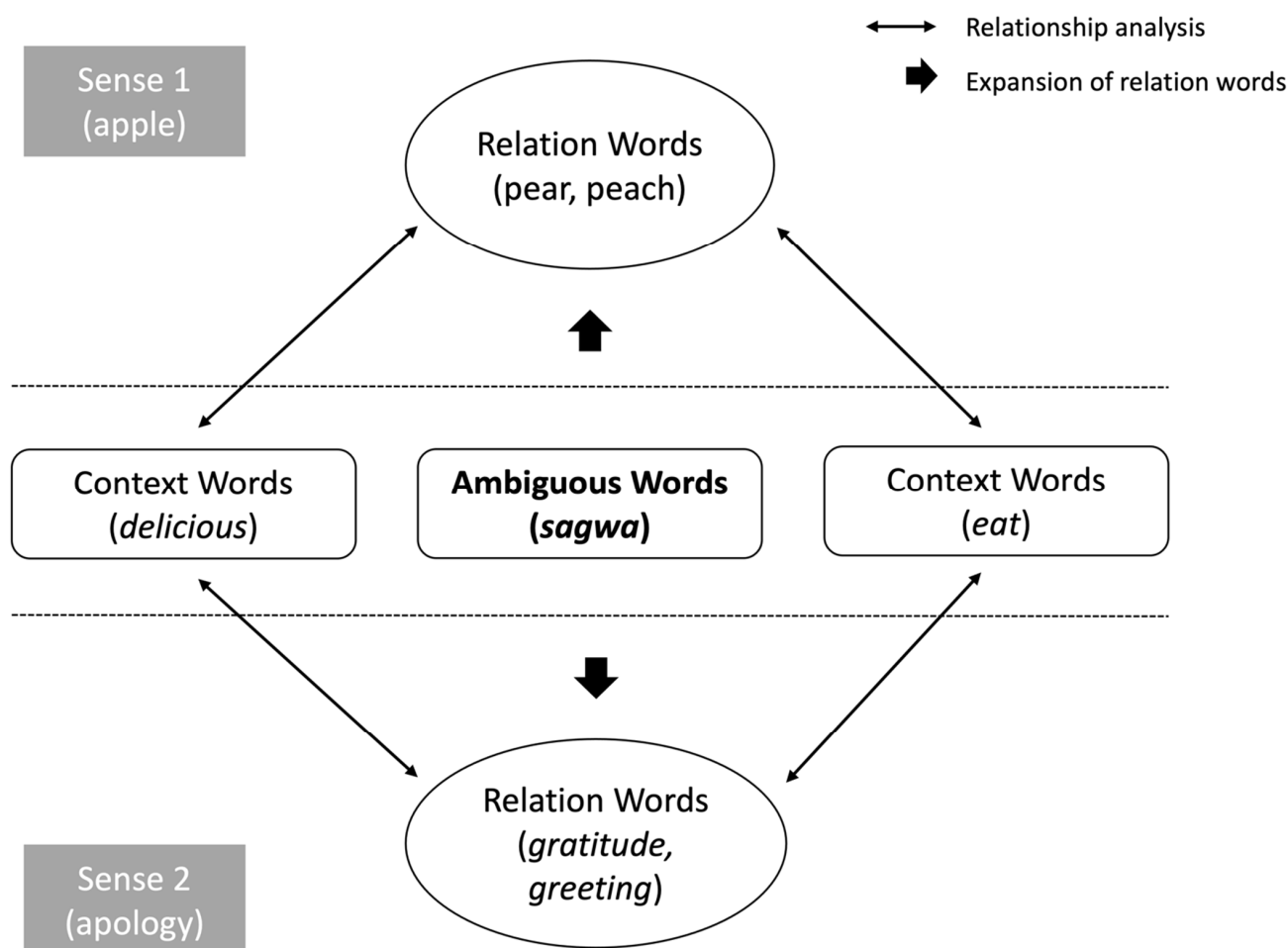


Figure 3. Relationship analysis between two words using the semantic relation words of an ambiguous word.

Table 1 shows an analysis of the relationship between an ambiguous word and the co-occurrence words in a local context in the sentence: ‘Sagwa han gairul megeotda’. (‘I ate an apple’).

Table 1. Coordinate terms of ‘sagwa’ and x^2 value of co-occurrence words in the local context.

Co-Occurrence Word	x^2 Value with Sagwa 1 (Apology)	x^2 Value with Sagwa 2 (Apple)
Han (one)	20.43	50.89
Gai (piece)	20.24	0.69
Meokda (eat)	145.25	0.07

Based on Table 1, various methods that can distinguish the meaning of the ambiguous word ‘sagwa’ can be found. The simplest method is to select the meaning with the largest number of related words in the local context using the chi-square test of independence. A null hypothesis is set that the co-occurrence of two words has no relationship to each other and the hypothesis is tested through the independence test. If the null hypothesis is rejected, then the alternative hypothesis is selected, concluding that the co-occurrence of two words is related to each other.

- Null hypothesis: Two words (w_1, w_2) are not related to each other (independent),

$$P(w_2|w_1) = p = P(w_2|\neg w_1) \quad (2)$$

- Alternative hypothesis: Two words (w_1, w_2) are related to each other (dependent),

$$P(w_2|w_1) = p_1 \neq p_2 = P(w_2|\neg w_1) \quad (3)$$

If the chi-square statistic is above a critical value, the null hypothesis is rejected, concluding that the two words are related to each other. In the chi-square distribution table (Table A1), the critical value is 7.88 when the degree of freedom is 1 and the significant level is 0.005. Table 2 shows the number of semantically related words of the ambiguous word using the chi-square test of independence.

Table 2. Relationship analysis between words using the chi-square test of independence.

Co-Occurrence Word	χ^2 Value with Sagwa 1 (Apology)	χ^2 Value with Sagwa 2 (Apple)
Han (one)	20.43	50.89
Gai (piece)	20.24	0.69
Meokda (eat)	145.25	0.07
Number of related words	3 words	1 word

In Table 2, the number of words related to ‘Sagwa1’ is three while the number of words related with ‘Sagwa2’ is one. Thus, in the sentence ‘I ate an apple’, the ambiguous word ‘Sagwa’ should be ‘Sagwa 1’, which has more related words in the local context. However, this method has several problems. First, when the number of related words is the same, there is no way to resolve the lexical ambiguity. Table 3 shows an analysis of the relationship of the ambiguous word and the co-occurrence words in a local context in the sentence ‘Naneun sagwareul badatda’ (I received an apology).

Table 3. Example of failure of relationship analysis between words using the chi-square test of independence.

Co-Occurrence Word	χ^2 Value with Sagwa 1 (Apology)	χ^2 Value with Sagwa 2 (Apple)
Na (I)	5.47	8.95
Batda (receive)	145.25	0.07
Number of related words	1 word	1 word

As shown in Table 3, the two meanings have the same number of related words, namely, one. Thus, there must be another method to distinguish the meaning of an ambiguous word other than simply using the number of related words. Second, despite the fact that each word’s degree of semantic relation is different, the relationship of the words above the critical chi-square statistic value is the same. That is, some words among the co-occurrence words in the local context may have more weight to determine the meaning of the ambiguous word, but this method cannot reflect this. For example, in Table 3, both of the co-occurrence words, ‘Na’ and ‘Batda’, have a chi-square statistic above 7.88. The co-occurrence word ‘Batda’, however, has a closer relationship with ‘Sagwa1’.

Generally, the larger the chi-square statistic, the greater the relationship between the two words. Therefore, a method of applying the chi-square statistic is to use a sum, average, and multiplication of weight of the chi-square statistic. The multiplication of weight is a factor calculated to show the influence on the meaning based on the ratio of the chi-square statistic assuming that the total influence of all the co-occurrence words on the ambiguous word is one.

As shown in Table 4, the sum, multiplication, average, and multiplication of weight of the chi-square statistic indicate a correct answer for the lexical disambiguation. Among them, using the multiplication of weight showed the best performance in the experiment result. Due to the characteristic of the chi-square statistic, if the frequency of a specific co-occurrence word is significantly over a certain threshold, the chi-square statistic also becomes too large. Because of this, the use of the sum, multiplication, and average of the chi-square statistic can result in an incorrect result where a specific word can decide the outcome. Thus, normalization of the chi-square statistic between 0 and 1 is required. The weight is used for normalization in this study.

Table 4. Semantic determination method using the chi-square statistic.

Co-Occurrence Word	χ^2 Value with Sagwa 1 (Apology)	χ^2 Value with Sagwa 2 (Apple)
Na (I)	5.47	8.95
Batda (receive)	145.25	0.07
Sum	150.72	9.02
Multiplication	794.5175	0.6265
Average	75.36	4.51
Multiplication of weight	0.3672	0.0228

The following formula shows the word sense disambiguation using co-occurrence words of semantically related words in the local context and weight of the χ^2 value.

$$WSD(w_{amb}, C) = \underset{s_k}{\operatorname{argmax}} \prod_{v_j \in C} \frac{\chi^2(w_{amb} = s_k, v_j)}{\chi^2(w_{amb} \neq s_k, v_j)} \quad (4)$$

$$\chi^2(w_{amb} \neq s_k, v_j) = \sum_{i=1}^n \chi^2(w_{amb} = s_i, v_j) - \chi^2(w_{amb} = s_k, v_j) \quad (5)$$

To prevent a resulting value of the above formula from being zero or infinite value because the χ^2 value was zero, a non-observed data frequency was estimated using the Good–Turing frequency estimation. In addition, performance can be different depending on which relation words and which relationships in the Korean Lexical Semantic Network are used. In this study, a relationship that can be used for word sense disambiguation is divided into five relationships and χ^2 is calculated by varying the weight. The five relationships are: ① coordinate term (s), ② hyponym (c), ③ hypernym (p), ④ hyponym of coordinate term (sc), and ⑤ coordinate term of hypernym (ps). In Section 4, the weight (λ) of relation words will be found through experiments.

$$C(x) = \frac{\sum_{i=1}^l \chi^2(r_i \in x, v_j)}{l} \quad (6)$$

$$\chi^2(w_{amb} = s_k, v_j) = \lambda_s C(s) + \lambda_c C(c) + \lambda_p C(p) + \lambda_{sc} C(sc) + \lambda_{ps} C(ps) \quad (7)$$

Furthermore, the data deficiency problem will be solved by normalizing the number of words using the part-of-speech information of the words. Table 5 shows the normalized expression and an example of words.

Table 5. Normalization of words using the part-of-speech information.

Normalization	Word
Counting Unit	Han(1), Doo(2), Sip(10), Bak(100)
Dependent noun on unit	Segi, wol, boon
Numeral	1, 2, 10, 100
Proper noun	Catholic University of Pusan, Minho Kim
General pre-noun	Geu, Yi, Enu

3.2. Expansion of Semantically Related Words to the Ambiguous Word

In Section 3.1, semantic relation words of the ambiguous word were expanded in the hierarchical structure of the Korean Lexical Semantic Network. However, a lack of statistical information due to an ongoing data deficiency problem prevented finding the semantically related co-occurrence words. One of the reasons is insufficient relation words. For example, ‘shinjang’ used as a meaning of kidney is the lowest hyponym in the Korean Lexical Semantic Network; thus, there is no hyponym and only two coordinate terms ‘Kongpat (kidney)’ and ‘Bulgotsepo (flame cell)’. Even using all five relationships in Section 3.1, only 13 related words can be found. To solve such a data deficiency problem, words that are related to an ambiguous word must be expanded.

In this paper, a set of semantically related words of an ambiguous word is created through the chi-square statistic used in Section 3.1. The related words refer to a collocation of two words in a semantic co-occurrence relationship. This is a significant clue to determine the correct meaning of the ambiguous word. First, the collocation words that are in the relationship of collocation with the ambiguous word are found using the chi-square statistic from the Sejong morph-tagged corpus. Then, a set of semantically related words of the ambiguous word is created using the chi-square test of independence to determine which meaning of the ambiguous word is in the collocation relationship with the collocation words found. Table 6 shows a part of a set of semantically related words of the ambiguous word ‘Noon’.

Table 6. Set of semantically related words of ‘noon’.

Rank	Noon (eye)		Noon (Snow)	
	Related Word	χ^2	Related Word	χ^2
1	Glasses	4416.99	Weather service	15,187.17
2	Organ	1681.90	Previous year	2438.09
3	Feel	733.49	Continue	141.42
4	Nose	251.31	Rain	118.95
5	Method	200.78	Fall	78.87
6	cannot	147.87	Rear	43.53
7	Ear	145.29	Appear	30.58
8	Keep an eye	130.47	Start	23.44
9	Touch	97.48	Out of	19.42
10	Mouth	96.83	Day	9.69

Using the semantically related words of the ambiguous word, word sense ambiguity can be removed as shown in Section 3.1. Not only can using the relationship between semantically related words of an ambiguous word and co-occurrence words in the local context be a method for removing word sense ambiguity, but also using the appearance of semantically related words of the ambiguous word in the local context. Word sense ambiguity is removed using the semantic determination formula in Section 3.1. Figure 4 shows the expression of the relationship analysis between words using related words.

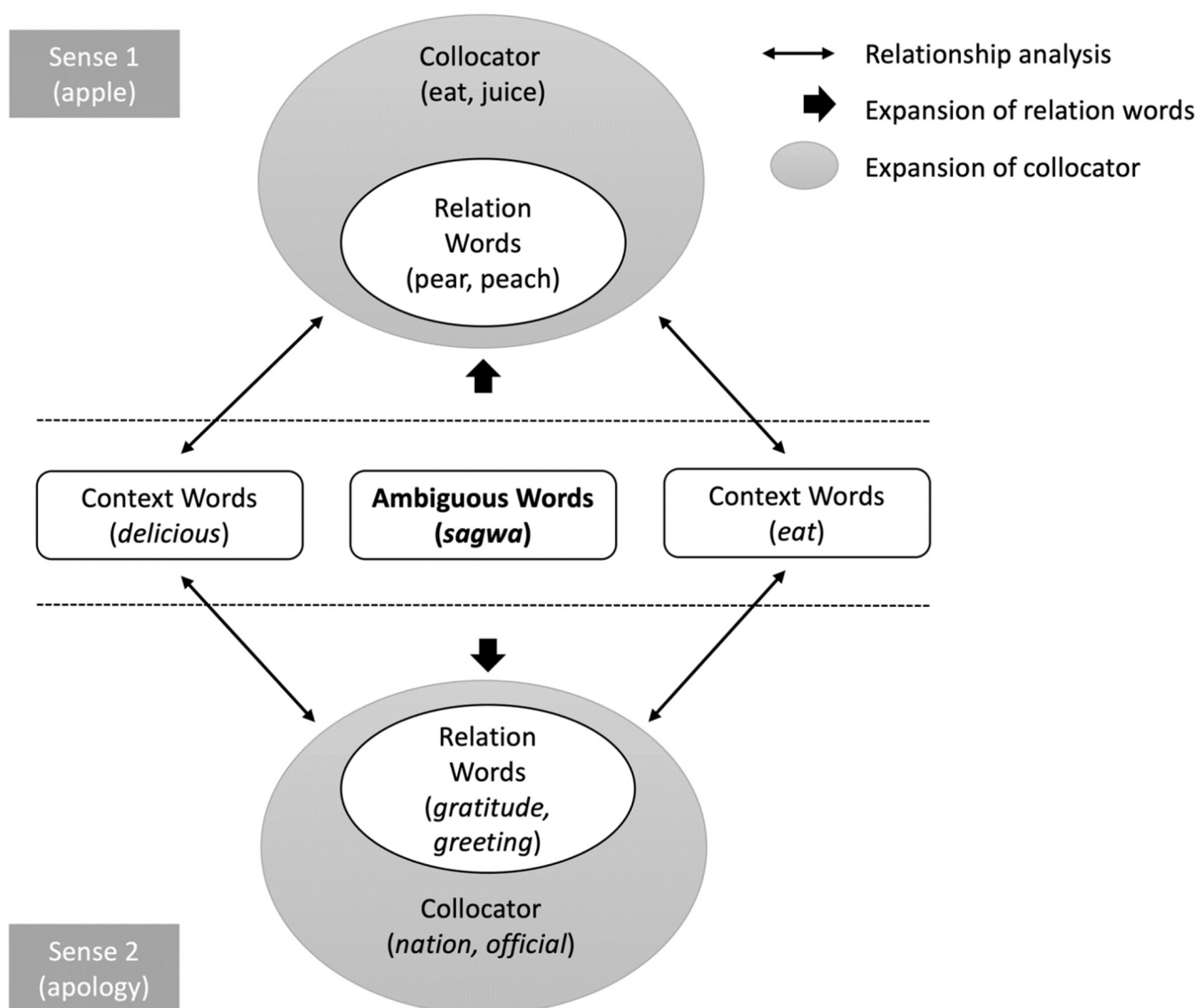


Figure 4. Relationship analysis between words using semantically related words of the ambiguous word.

Moreover, we attempted to overcome the data deficiency problem by expanding the coordinate terms of words that are highly related among the related words of an ambiguous word. In Section 4, we will discuss how many expansions of the related word are needed through an experiment. Word sense ambiguity will be removed using the coordinate terms of the semantically related words of an ambiguous word. Not only using the relationship between the semantically related words of an ambiguous word and the co-occurrence words in the local context but also using the appearance of the semantically related words of the ambiguous word in the local context as shown in Section 3.1 can be a method of removing word sense ambiguity. Figure 5 shows the expression of the relationship analysis between words using the coordinate terms of the semantically related words of an ambiguous word.

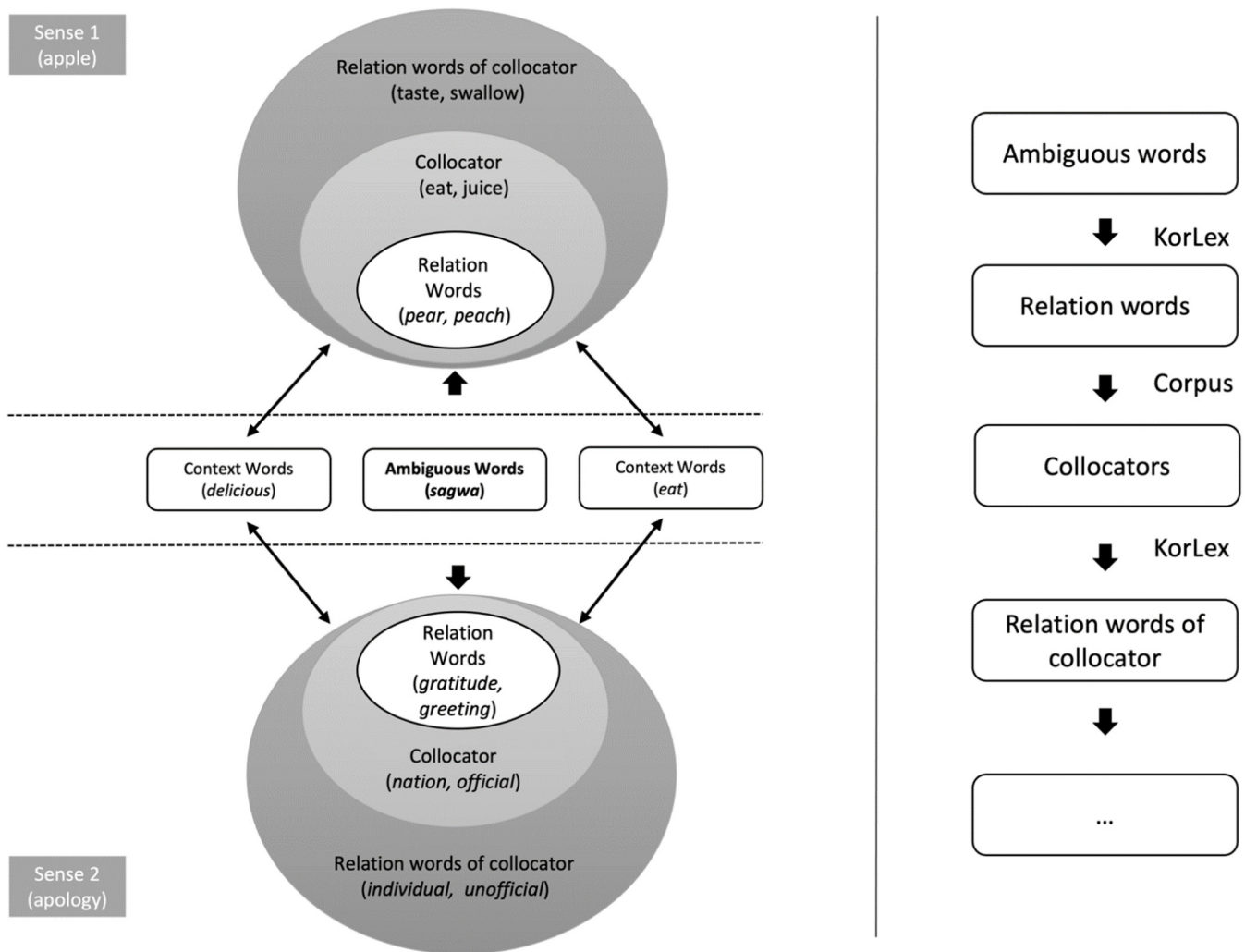


Figure 5. Relationship analysis between words using coordinate terms of semantically related words of the ambiguous word.

3.3. Estimation of Prior Information Using Semantically Related Words of an Ambiguous Word

In supervised disambiguation, a disambiguation corpus, a corpus that was classified into meaning in a specific context of all appearances of the ambiguous word, is used as the learning data. A naïve Bayesian classifier is a statistical theory that is applied in natural language processing and lexical disambiguation. The naïve Bayesian classifier identifies the meaning using adjacent words of the ambiguous word in a large-scale context. Adjacent words provide useful information to identify the meaning of the ambiguous word so that statistical inference can be applied using the co-occurrence frequency information of these adjacent words. The naïve Bayesian classifier uses Bayesian decision rules to minimize error probability when determining class.

Assuming that c is words used for a contextual feature in a context where the ambiguous word w appears in a corpus, the decision rule of the naïve Bayesian classifier that solves the word sense ambiguity based on the contextual feature is as follows:

$$\begin{aligned} \hat{S} &= \operatorname{argmax}_{s_i \in \text{Senses}(w)} P(s_i | c_1, \dots, c_m) = \operatorname{argmax}_{s_i \in \text{Senses}(w)} \frac{P(c_1, \dots, c_m | s_i)}{P(c_1, \dots, c_m)} \\ &= \operatorname{argmax}_{s_i \in \text{Senses}(w)} P(s_i) \prod_{i=1}^m P(c_i | s_i) \end{aligned} \quad (8)$$

In the above formula, $P(c_1, \dots, c_m | s_i)$ and $P(s_i)$ are calculated by maximum-likelihood estimation from the learning corpus of disambiguation. Here, $P(c_1, \dots, c_m | s_i)$ is posterior probability and $P(s_i)$ is prior probability.

Generally, the reason for the high performance of probability models, such as the naïve Bayesian classifier, in supervised disambiguation is due to the large influence of prior probability, which is the semantic probability of words. Most ambiguous words have two or more meanings, but only one or two meanings are actually used frequently in our daily lives. Thus, if we know the semantic word's prior probability in advance, it would significantly increase the lexical disambiguation performance.

Moreover, in this paper, in order to realize the same effect of using prior information in supervised disambiguation, semantically related words of the ambiguous vocabulary are obtained and utilized as prior information. Using prior information, word sense ambiguity can be solved even under cases where words that are strongly related to a specific meaning in the local context are not found or semantically related words cannot be found due to the lack of statistical information caused by data deficiency.

In this paper, the semantic prior probability of an ambiguous word can be calculated using the weight of the semantically related words of the ambiguous word, as indicated in the following formula. The prior probability of meaning s_k of an ambiguous word is assumed to be a ratio of the frequency of $RW(w_{amb})$ and the related word $RW(w_{amb} = s_k)$ of meaning s_k .

$$P(s_k) = \frac{c(w_{amb} = s_k)}{c(w_{amb})} = \frac{c(RW(w_{amb} = s_k))}{c(RW(w_{amb}))} \quad (9)$$

4. Experiment and Evaluation

4.1. Experiment Environment

In this paper, the 'Sejong morph-tagged corpus (approximately 5 M word phrases)', a 21st Century Sejong Project deliverable, was used to extract statistical information. Nouns, adjectives, and verbs were extracted from the Sejong morph-tagged corpus and the co-occurrence frequency of all the words was found in the dictionary.

In order to compare the lexical disambiguation method proposed in this paper with other studies, experiments were conducted using the Korean learning data called SENSEVAL-2. SENSEVAL is a contest for word sense disambiguation technology under the sponsorship of ACL SIGLEX and EURALEX. It has been held every three years since 1998. Two Korean teams participated in the second contest. The target words in SENSEVAL-2 for Korean learning data were 'mal', 'noon', 'son', 'baram', 'geori', 'jari', 'euisa', 'mok', 'jeom', and 'bam'. The detailed data composition can be found in Appendix A.

The evaluation measure for lexical disambiguation methods in this paper was accuracy.

The accuracy can be obtained as follows:

$$Accuracy(\%) = \frac{\text{the number of ambiguous words whose meanings were correctly distinguished}}{\text{the number of ambiguous words}} \quad (10)$$

4.2. Experiment Method

A window size of context was considered when co-occurrence words appeared in the local context of the ambiguous word used for the lexical disambiguation. A window size refers to the number of words on the right and left sides of the ambiguous word. As the window size became larger, accuracy also increased rapidly, eventually being stabilized. In this paper, considering the size of the statistical dictionary, five was selected as a basic value of the window size.

The performance also depends on which relation words are used in the Korean Lexical Semantic Network. In this study, a relationship that can be used for word sense disambiguation was divided into five relationships, as shown in Section 3.1 and x^2 was calculated by varying the weight. The weight of the coordinate terms was fixed to 1.0 while varying the weights of other related words in the experiment.

As shown in Figure 6, when the weight of the coordinate term is one, the best accuracy was obtained if the weight of the hyponym was 0.8 and the weight of the hypernym was 0.2. Furthermore, the accuracy was higher when there was no expansion of the hypernym and the coordinate term of the hypernym. In this study, weights of relation words were set as follows.

$$x^2(w_{amb} = s_k, v_j) = \lambda_s C(s) + \lambda_c C(c) + \lambda_p C(p) + \lambda_{sc} C(sc) + \lambda_{ps} C(ps) \quad (11)$$

$$(\lambda_s = 0.5, \lambda_c = 0.4, \lambda_p = 0, \lambda_{sc} = 0.1, \lambda_{ps} = 0)$$

In addition, when the coordinate terms of the semantically related words of an ambiguous word were expanded, the range of related words to be expanded was changed in the experiment. Figure 7 shows a change in accuracy according to changes in the expansion range of the coordinate terms of the semantically related words of an ambiguous word.

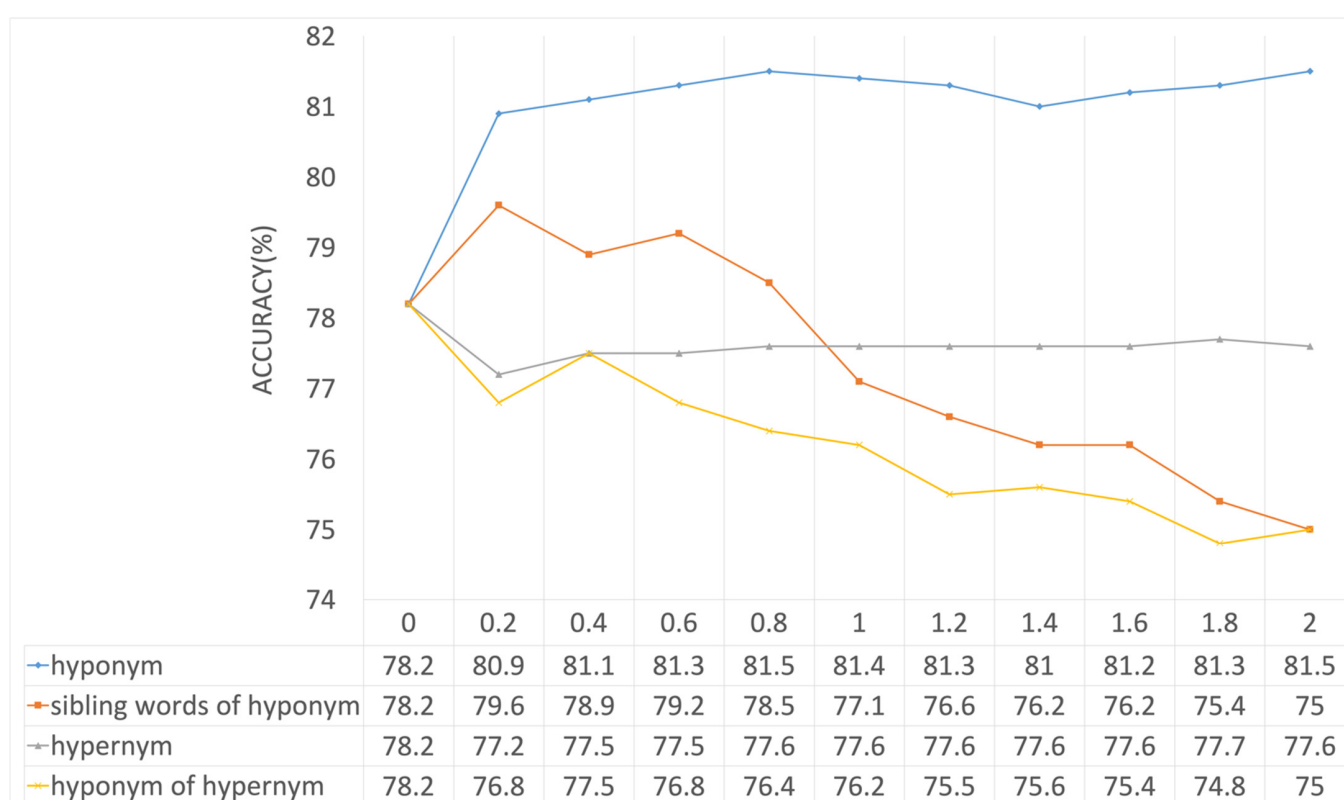


Figure 6. Change in accuracy according to the weight of relation word.

As shown in Figure 7, it is more effective for word sense disambiguation to expand only the words that are highly related rather than expanding the collocation coordinate terms of all the semantically related words. In this study, only the coordinate terms of the collocation words in the top 25% of the semantically highly related words of an ambiguous word were expanded.

To evaluate the performance of the algorithm proposed in this paper, a method of determining the meaning by the most frequent class (MFC) was used as the baseline for comparison of performance. In addition, performance for a basic algorithm and the newly improved algorithm was compared.

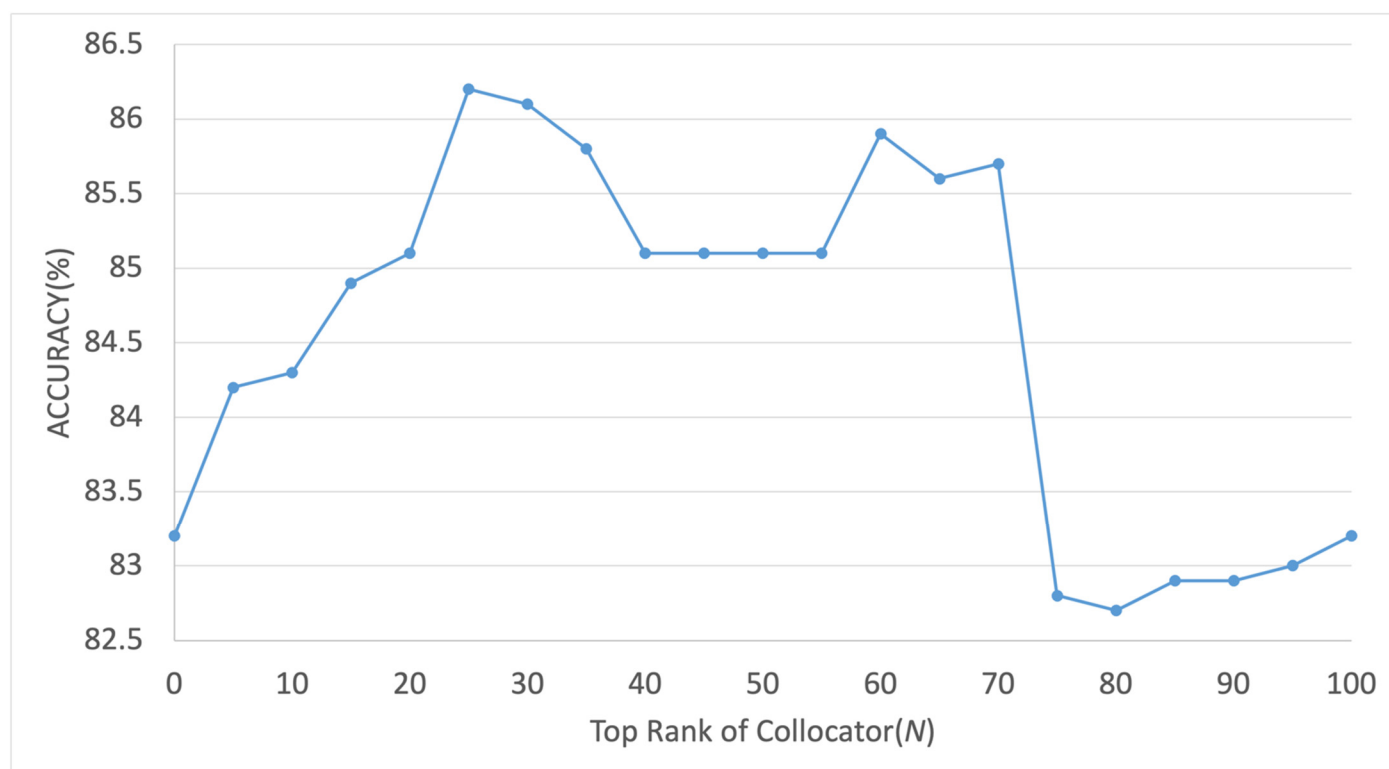


Figure 7. Change in accuracy due to expanded range of coordinate terms of the semantically related words of an ambiguous word.

The basic algorithm is the one that was used previously in the lexical disambiguation system at Busan University that calculated χ^2 using the semantic coordinate terms of the ambiguous word. The improved method for lexical disambiguation in this paper solves the data proficiency problem as follows:

- ① A weight is adjusted according to the types of semantically related words of an ambiguous word so that more information regarding the relation words can be used than in existing methods.
- ② Semantically related words of an ambiguous word and the coordinate terms of the related words are expanded so that more information can be used than in existing methods.
- ③ Using the part-of-speech information of words, normalization is done for words such as numerals and proper nouns.

Table 7 shows a comparison of the performance between the basic and improved algorithms. The average accuracy of MFC was 78.29%, while accuracy of the proposed algorithm was 88.11%. Then, the number ① improvement method was applied to the basic algorithm. Next, the numbers ① and ② improvement methods were applied, and finally, the numbers ①~③ were applied. Accuracy was improved 5.08%, 8.01%, and 9.90%, respectively.

The proposed method showed better accuracy in most ambiguous words than the MFC. However, the MFC had significantly high accuracy for words whose evaluated corpus meaning was biased to one-sided direction such as ‘baram’ and ‘mok’. In particular, the accuracy for ‘mal’ was the lowest in terms of lexical disambiguation. This was because ‘mal’ expressing ‘grain’ or ‘unit of quantity of liquid’ appeared more frequently than ‘mal’ meaning ‘means to express people’s thought and feeling’ that was widely used in general. Table 8 shows the ratios of the meanings of ‘mal’ in the ‘Korean learning data in SENSEVAL-2’.

Table 7. Performance evaluation of the proposed method.

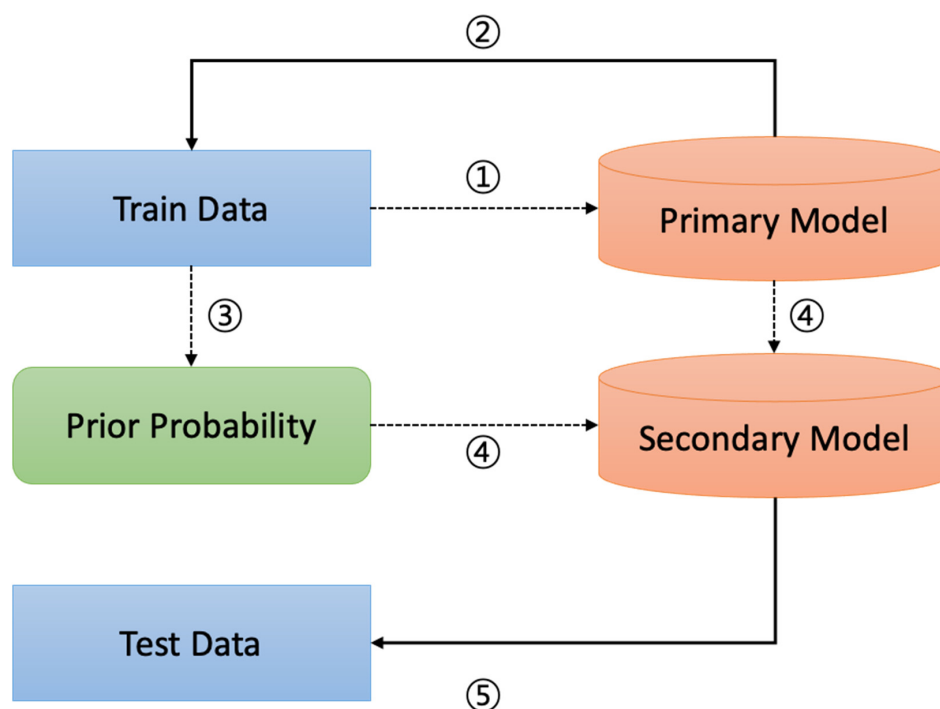
Ambiguous Word	Accuracy (%)				
	MFC	Basic Algorithm	Improvement ①	Improvement ① + ②	Improvement ① + ② + ③
noon	93.98	93.98	93.98	94.74	94.74
son	97.73	93.18	97.73	97.73	98.48
mal	34.65	46.53	54.46	54.46	65.35
baram	98.98	98.98	96.94	96.94	94.90
geori	53.44	47.33	47.33	68.70	74.05
jari	89.11	95.05	96.04	96.04	96.04
euisa	62.42	56.36	87.27	89.70	88.48
mok	99.00	97.00	94.00	94.00	96.00
jeom	89.90	90.91	88.89	89.90	94.95
bam	71.29	77.23	77.23	77.23	77.23

Table 8. Example of ‘SENSEVAL-2 Korean Learning Data’.

Word	Meaning No.	Meaning	Number of Data
mal	mal_1	Domestic livestock	11
	mal_2	Marker in the game board such as Yut game or Gonu	0
	mal_3	Means to express people’s thought and feeling	33
	mal_6	End	22
	mal_9	Unit of quantity for grain or liquid	35

4.3. Analysis of Effect of the Prior Probability Estimation

A method of using prior information in the previous studies was developed that included: performing lexical disambiguation on raw corpus using a basic model (hereafter referred to as a primary model) based on unsupervised disambiguation, and applying the extraction of prior knowledge from the previous result to the primary model (hereafter referred to as a secondary model). Figure 8 shows this process.

**Figure 8.** Prior probability estimation method of existing study.

In order to compare with the proposed prior probability estimation, an experiment was conducted using the prior probability estimation method shown in Figure 8. Using statistical information extracted from the learning corpus, a primary model was constructed (using related words and relation words) and ② using the primary model, lexical disambiguation was conducted with regard to learning corpus. ③ Prior knowledge was extracted using the lexical disambiguation result, ④. A secondary model was constructed using the primary model and extracted prior knowledge and ⑤ lexical disambiguation was conducted with regard to the evaluation of the corpus using the secondary model.

Table 9 shows the performance of lexical disambiguation when prior probability estimated using the method in Figure 8 and the prior probability proposed in this study were used.

Table 9. The experiment result of effect analysis of prior probability estimation.

Ambiguous Word	Accuracy (%)			
	Learning Corpus		Evaluation Corpus	
	Primary Model (Related Word, Relation Word)	Primary Model (Related Word, Relation Word)	Secondary Model (Use of Prior Knowledge Extracted from the Learning Corpus Tagging Result)	Existing Model (Prior Knowledge Estimation)
noon	89.87	87.02	92.45	94.54
dari	82.34	81.34	83.84	85.78
bam	82.49	91.09	93.34	98.57
bae	84.47	83.47	86.39	89.21
sagwa	86.25	85.99	83.68	90.41
shinjang	81.64	75.24	83.98	83.46
yeonggi	85.56	77.46	88.69	89.65
indo	75.64	73.99	83.97	87.24
insa	80.24	79.68	84.47	89.45
janggi	86.43	87.87	90.96	92.47
Average	83.49	82.32	87.18	90.14

As shown in Table 9, the use of prior knowledge estimated using semantically related words of the ambiguous word contributed more to the lexical disambiguation than using prior knowledge extracted from the learning corpus tagging results. This result was revealed because the accuracy of the primary model was 83.49% on average so that the secondary model could not be constructed using the accurate prior knowledge.

To determine whether the proposed method in this paper showed the same performance in other languages, we conducted an experiment with English. For the English, the English WordNet was used instead of the Korean Lexical Semantic Network. We evaluate our methods using the SemCor [30], which is an English corpus with semantically annotated texts. The semantic analysis was done manually with WordNet 1.6 senses (SemCor version 1.6) and later automatically mapped to WordNet 3.0 (SemCor version 3.0). The SemCor corpus consists of 352 texts from Brown corpus. Table 10 shows the performance of our model and the performance of the existing model for SemCor.

Existing models to be compared are deep-learning language model-based fine-tuning models for word sense disambiguation. All three models were fine-tuned to 80% of SemCor using the basic model and then evaluated for 20%. As can be seen in Table 10, our proposed method showed almost the same or slightly lower performance than the existing models even though it is not supervised learning. In particular, in the evaluation results for SE13, our model showed the best performance. This is because, in the case of SE13, the number of training data are very small compared with other tasks. It can be seen that the deep

learning-based model performs better if the training data for target word is sufficient, but our proposed method performs better when the training data are insufficient.

Table 10. SemCor test results.

Model	Accuracy (%)				
	SE7	SE2	SE13	SE15	All
BERT [31]	71.9	77.8	76.5	79.7	76.6
RoBERTa [32]	69.2	77.5	77.2	79.7	76.3
ELECTRA [33]	62.0	71.5	73.9	76.0	70.9
Proposed Model	70.1	75.4	78.1	75.1	75.7

4.4. Practical Experiment with the Proposed System

As explained earlier, lexical disambiguation can be utilized as a preprocessing system in various natural language processing application areas such as information search or machine translation. To increase the acceptance of the system as a preprocessing system, it is necessary to increase the performance of the lexical disambiguation and reduce the processing time and minimize the required storage space. In particular, the calculation of the chi-square statistic and semantic prior probability takes significant time. In this study, the dictionary of chi-square statistics between the words and the prior probability dictionary were constructed beforehand. The chi-square statistics and related words were previously obtained through a search method thereby minimizing the processing time of the lexical disambiguation.

To search the chi-square statistic, a certain block unit of indexes was created using the chi-square information. A target block was found using a word pair key to load the block from the file to the memory thereby fetching the chi-square statistic of the target word pair using a binary search. The prior probability information was connected to the word index directly, thereby fetching the prior probability information when a word pair key was searched. Figure 9 shows the aforementioned chi-square statistic and search method for the prior probability information.

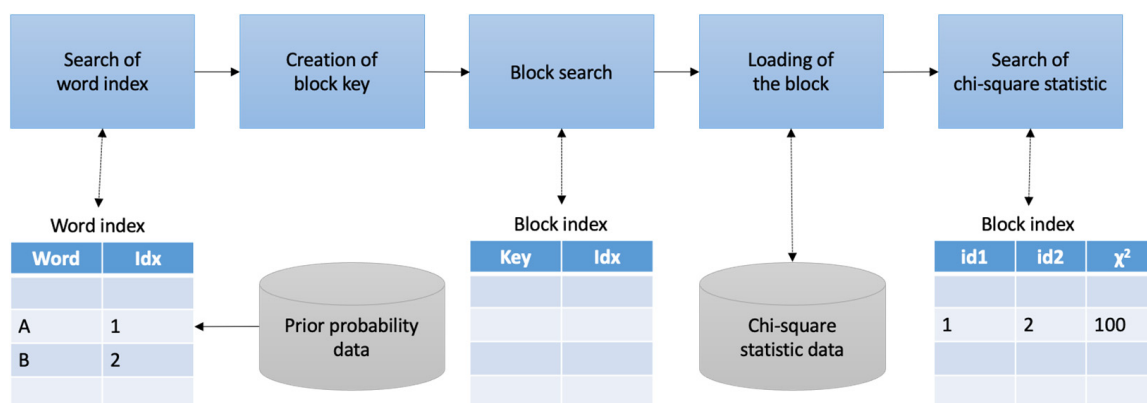


Figure 9. Chi-square statistic and prior probability search method.

For the practical experiment of the lexical disambiguation technology based on the large-scale chi-square statistic and prior probability, speed was measured. Based on the top ranking of the appearance of words in the Sejong semantic-tagged corpus, 200 ambiguous words were extracted and tested over 10,000 sentences to analyze the processing speed and performance of the lexical disambiguation. The average number of meanings of the ambiguous words was 5.7 words.

As shown in Figure 10, when the memory-based search method was not used, the execution time was 350 s but when the memory-based search method was used, the

execution time was 22 s. That is, an average of 450 ambiguous words was resolved per second.

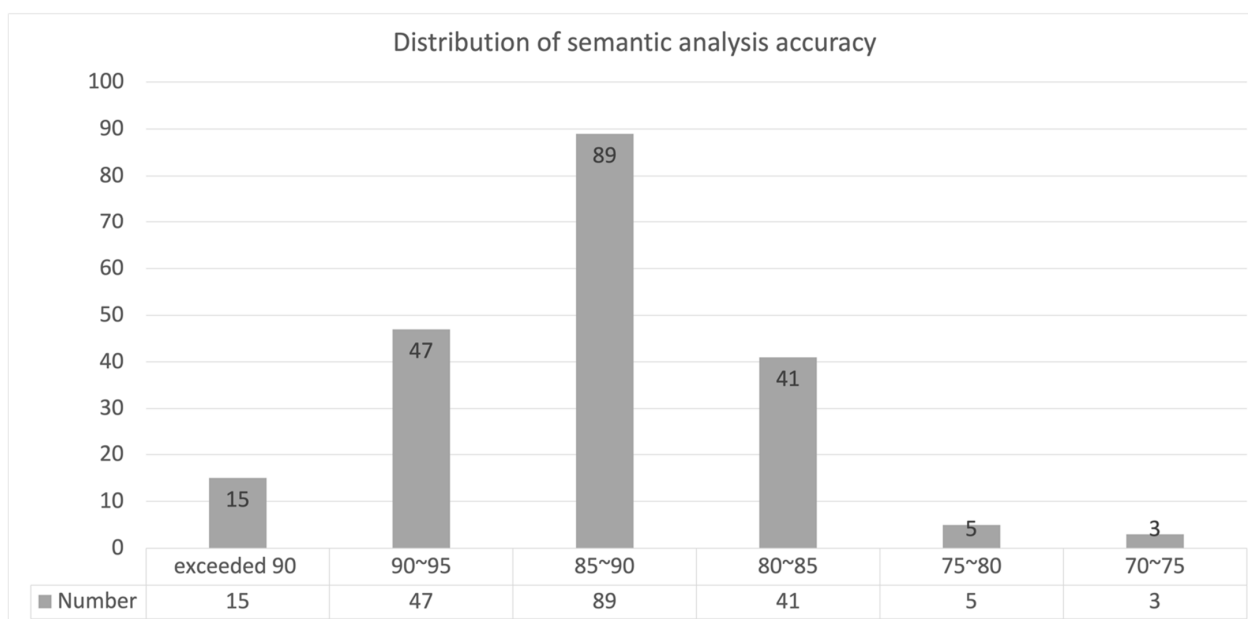


Figure 10. Distribution of semantic analysis accuracy over 200 ambiguous words.

The average accuracy of the lexical disambiguation was 86.3%, which was about 4% lower than using the SENSEVAL-2 data. This was because the average number of meanings was larger than the number of ambiguous words in the SENSEVAL-2 data. Figure 10 shows the semantic analysis accuracy distribution over 200 ambiguous words. More than 90% of accurate semantic analysis was shown in 67 ambiguous words, which accounted for 31% of the total words; 85~90% accuracy was shown in 89 ambiguous words.

5. Conclusions and Future Research

This paper proposed a novel unsupervised disambiguation method that showed better performance than existing knowledge-based lexical disambiguation or unsupervised lexical disambiguation methods without need of a large amount of sense-tagged corpus.

Since the related words in the Korean Lexical Semantic Network have the same characteristics, the meaning of an ambiguous word could be distinguished by determining the relationship between the semantic relation words of the ambiguous word and the co-occurrence words in a local context. Moreover, the performance of the lexical disambiguation method was improved by using more relation word information than existing methods. Weights were adjusted depending on the semantic relation word type of an ambiguous word and expanding the semantically related words of the ambiguous word and coordinate terms of the related words. Finally, numerals and proper nouns were normalized using the part-of-speech information to solve the data deficiency problem, while semantically related words of an ambiguous word were obtained and used as prior information in order to have the same effect of using prior information in supervised disambiguation.

The contributions of this study are as follows: First, lexical disambiguation was conducted using statistical information without a sense-tagged corpus by utilizing KorLex, which is a Korean Lexical Semantic Network. Second, better performance was achieved using only the minimum information (frequency of appearance of a single word, frequency of appearance of co-occurrence, and part-of-speech information) than the existing knowledge-based lexical disambiguation method.

Future research will first include, evaluating additional ambiguous words using other evaluation data to further increase the reliability of the systems. Second, a study on

preprocessing, such as selection constraints, will be done for an analysis that cannot be solved by statistical information due to the data deficiency problem.

Author Contributions: Conceptualization, M.K.; methodology, M.K.; software, M.K.; validation, M.K. and H.-C.K.; formal analysis, M.K.; investigation, M.K.; resources, M.K.; data curation, M.K.; writing—original draft preparation, M.K.; writing—review and editing, M.K. and H.-C.K.; visualization, M.K.; supervision, H.-C.K.; project administration, H.-C.K.; funding acquisition, H.-C.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2020-0-01450, Artificial Intelligence Convergence Research Center(Pusan National University)).

Data Availability Statement: The datasets analysed during the current study are available in the National Institute of Korean Language, <https://corpus.korean.go.kr> (accessed on 22 November 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Table of the chi-square distribution.

Chi-Square (χ^2) Distribution										
Area to the Right of Critical Value										
Degrees of Freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

References

- Ide, N.; Véronis, J. Introduction to the special issue on word sense disambiguation: The state of the art. *Comput. Linguist.* **1998**, *24*, 1–40.
- Kim, S.-K.; Huh, J.-H. Artificial intelligence based electronic healthcare solution. In *Advances in Computer Science and Ubiquitous Computing*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 575–581.
- Kim, S.-K.; Huh, J.-H. Consistency of medical data using intelligent neuron faster R-CNN algorithm for smart health care application. *Healthcare* **2020**, *8*, 185. [[CrossRef](#)] [[PubMed](#)]
- Navigli, R. Word sense disambiguation: A survey. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 1–69. [[CrossRef](#)]
- Le, N.-B.-V.; Huh, J.-H. Applying sentiment product reviews and visualization for BI systems in vietnamese E-commerce website: Focusing on vietnamese context. *Electronics* **2021**, *10*, 2481. [[CrossRef](#)]
- Yoon, A.-S.; Hwang, S.-H.; Lee, E.-R.; Kwon, H.-C. Construction of Korean WordNet. *J. KIISE Softw. Appl.* **2009**, *36*, 92–108.
- Lesk, M. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the 5th Annual International Conference on Systems Documentation*, New York, NY, USA, 1 June 1986; pp. 24–26.
- Luk, A.K. Statistical sense disambiguation with relatively small corpora using dictionary definitions. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, Cambridge, MA, USA, 26–30 June 1995; pp. 181–188.
- Miller, G.A.; Beckwith, R.; Fellbaum, C.; Gross, D.; Miller, K.J. Introduction to WordNet: An on-line lexical database. *Int. J. Lexicogr.* **1990**, *3*, 235–244. [[CrossRef](#)]
- Resnik, P. Disambiguating noun groupings with respect to WordNet senses. In *Natural Language Processing Using Very Large Corpora*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 77–98.
- Agirre, E.; Rigau, G. Word sense disambiguation using conceptual density. *arXiv* **1996**, arXiv:preprint cmp-lg/9606007. [[CrossRef](#)]
- Mihalcea, R.; Moldovan, D. A method for word sense disambiguation of unrestricted text. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics*, College Park, MD, USA, 20–26 June 1999; pp. 152–158.
- Mihalcea, R. Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling. In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, Vancouver, BC, Canada, 6–8 October 2005; pp. 411–418.
- Pedersen, T. A simple approach to building ensembles of naive bayesian classifiers for word sense disambiguation. *arXiv* **2000**, arXiv:preprint cs/0005006.
- Ramakrishnan, G.; Prithviraj, B.; Bhattacharyya, P. A gloss-centered algorithm for disambiguation. In *Proceedings of the SENSEVAL-3, the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, Barcelona, Spain, 25–26 July 2004; pp. 217–221.
- Sinha, R.; Mihalcea, R. Unsupervised graph-based word sense disambiguation using measures of word semantic similarity. In *Proceedings of the International conference on semantic computing (ICSC 2007)*, Irvine, CA, USA, 17–19 September 2007; pp. 363–369.
- Navigli, R.; Lapata, M. Graph connectivity measures for unsupervised word sense disambiguation. In *Proceedings of the IJCAI*, Hyderabad, India, 6–12 January 2007; pp. 1683–1688.
- Agirre, E.; Soroa, A. Personalizing pagerank for word sense disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, Athens, Greece, 30 March–3 April 2009; pp. 33–41.
- Heo, J.; Seo, H.-C.; Jang, M.-G. Homonym disambiguation based on mutual information and sense-tagged compound noun dictionary. *J. KIISE: Softw. Appl.* **2006**, *33*, 1073–1089.
- Scarlina, B.; Pasini, T.; Navigli, R. SenseBERT: Context-enhanced sense embeddings for multilingual word sense disambiguation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, New York, NY, USA, 7–12 February 2020; pp. 8758–8765.
- Bevilacqua, M.; Pasini, T.; Raganato, A.; Navigli, R. Recent trends in word sense disambiguation: A survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, Montreal, QC, Canada, 21–26 August 2021.
- Kohli, H. Transfer learning and augmentation for word sense disambiguation. *arXiv* **2021**, arXiv:2101.03617.
- Chen, H.; Xia, M.; Chen, D. Non-parametric few-shot learning for word sense disambiguation. *arXiv* **2021**, arXiv:2104.12677.
- Pasini, T. The knowledge acquisition bottleneck problem in multilingual word sense disambiguation. In *Proceedings of the IJCAI*, Yokohama, Japan, 11–17 July 2020; pp. 4936–4942.
- Zhimao, L.; Ting, L.; Sheng, L. *Unsupervised Chinese Word Sense Disambiguation Based on Equivalent Pseudowords*; Information Retrieval Laboratory of Computer Science & Technology School, Harbin Institute of Technology: Harbin, China, 2014.
- Rouhizadeh, H.; Shamsfard, M.; Rouhizadeh, M. Knowledge based word sense disambiguation with distributional semantic expansion for the persian language. In *Proceedings of the 2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*, Mashhad, Iran, 29–30 October 2020; pp. 329–335.
- Bordag, S. A comparison of co-occurrence and similarity measures as simulations of context. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, Haifa, Israel, 17–23 February 2008; pp. 52–63.
- Kolesnikova, O. Survey of word co-occurrence measures for collocation detection. *Comput. Y Sist.* **2016**, *20*, 327–344. [[CrossRef](#)]
- Párraga-Valle, J.; García-Bermúdez, R.; Rojas, F.; Torres-Morán, C.; Simón-Cuevas, A. Evaluating mutual information and chi-square metrics in text features selection process: A study case applied to the text classification in PubMed. In *Proceedings of the International Work-Conference on Bioinformatics and Biomedical Engineering*, Granada, Spain, 6–8 May 2020; pp. 636–646.

-
30. Raganato, A.; Camacho-Collados, J.; Navigli, R. Word sense disambiguation: A unified evaluation framework and empirical comparison. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Volume 1, Long Papers, Valencia, Spain, 3–7 April 2017; pp. 99–110.
 31. Du, J.; Qi, F.; Sun, M. Using bert for word sense disambiguation. *arXiv* **2019**, arXiv:1909.08358.
 32. Blevins, T.; Zettlemoyer, L. Moving down the long tail of word sense disambiguation with gloss-informed biencoders. *arXiv* **2020**, arXiv:2005.02590.
 33. Duarte, J.M.; Sousa, S.; Milios, E.; Berton, L. Deep analysis of word sense disambiguation via semi-supervised learning and neural word representations. *Inf. Sci.* **2021**, *570*, 278–297. [[CrossRef](#)]