Corpus-based Ontology Learning for Word Sense Disambiguation

Sin-Jae Kang

School of Computer and Communication Engineering, Daegu University
and Advanced Information Technology Research Center (AITrc)
Nairi-Ri 15, Jinrang, Kyungsan, Kyungpook, 712-714
Republic of KOREA
sjkang@daegu.ac.kr

Abstract

This paper proposes to disambiguate word senses by corpus-based ontology learning. Our approach is a hybrid method. First, we apply the previously-secured dictionary information to select the correct senses of some ambiguous words with high precision, and then use the ontology to disambiguate the remaining ambiguous words. The mutual information between concepts in the ontology was calculated before using the ontology as knowledge for disambiguating word senses. If mutual information is regarded as a weight between ontology concepts, the ontology can be treated as a graph with weighted edges, and then we locate the least weighted path from one concept to the other concept. In our practical machine translation system, our word sense disambiguation method achieved a 9% improvement over methods which do not use ontology for Korean translation.

1. Introduction

An ontology is a knowledge base with information about concepts existing in the world, their properties, and how they relate to each other. An ontology is different from a thesaurus in that it contains only language independent information and many other semantic relations, as well as taxonomic relations. In this paper, we propose to use the ontology to disambiguate word senses.

All approaches to word sense disambiguation (WSD) make use of words in a sentence to mutually disambiguate each other. The distinctions between various approaches lie in the source and type of knowledge made by the lexical units in a sentence. Thus, all these approaches can be classified into AI-based, knowledge-based, or corpus-based approaches, according to their sources and types of knowledge (Ide, 1998). AI-based WSD methods (Dahlgren, 1988) use a semantic network, or frames containing information about word functions and the relation to other words in individual sentences; or preference semantics, which specifies selectional restrictions for combinations of lexical items in a sentence. The difficulty with handcrafting the knowledge sources is the major disadvantage of AI-based systems. Knowledge-based methods (Resnik, 1995a; Yarowsky, 1992) have utilized machine-readable dictionaries (MRD), thesauri, and computational lexicons, such as WordNet. Since most MRDs and · thesauri were created for human use and display inconsistencies, these methods have clear limitations. Corpus-based methods (Dagan, 1994; Gale, 1992) extract statistical information from corpora which is monolingual or bilingual, and raw or sense-tagged. The problem of data sparseness commonly occurs in the corpus-based approach, and is especially severe when processing in WSD. A smoothing and concept-based method is used to address this problem.

Our WSD approach is a hybrid method, which combines the advantages of corpus-based and knowledge-based methods. We use our semi-automatically constructed ontology as an external knowledge source and secured dictionary information as context information. First, we apply the previously-secured dictionary information to select the correct senses of some ambiguous words with high precision, and then use the ontology to disambiguate the remaining ambiguous words.

The remainder of this paper is organized as follows. In the next section, we describe the semi-automatic ontology construction methodology briefly. The ontology learning is explained in
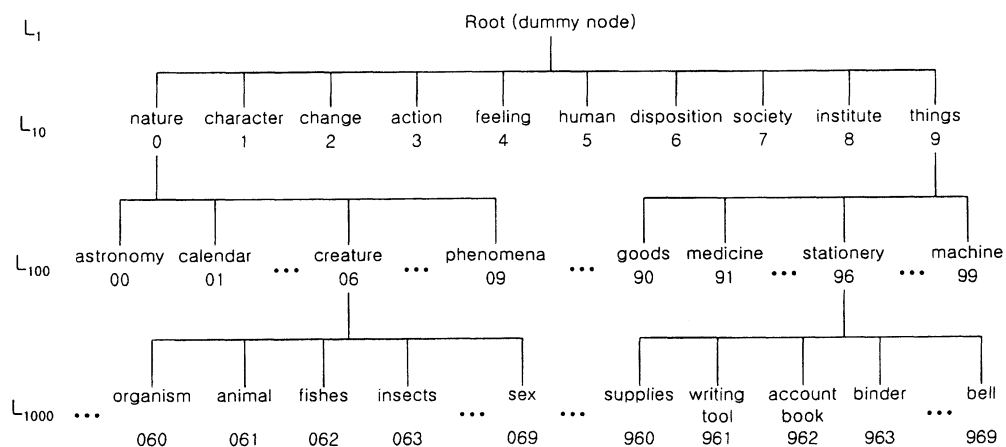
Fig. 1 Concept hierarchy of the Kadokawa thesaurus

Section 3. An ontology-based word sense disambiguation algorithm is given in Section 4. Experimental results are presented and analyzed in Section 5. Finally, we conclude and indicate the direction of our future work in Section 6.

2. Ontology Construction

To construct a practical ontology for WSD, we developed two strategies. First, we introduced the same number and grain size of concepts of the Kadokawa thesaurus (Ohno & Hamanishi, 1981) and its taxonomic hierarchy into the ontology. The thesaurus has 1,110 semantic categories and a 4-level hierarchy as a taxonomic relation (Fig. 1). Semantic categories in level L_1, L_{10}, and L_{100} are further divided into 10 subclasses. The root node is merely a dummy node. Noun and verb categories coexist in the same taxonomic hierarchy of the Kadokawa thesaurus. Verb categories mainly correspond to the code 2xx, 3xx, and 4xx in level L_{1000}. This approach is a moderate shortcut to construct a practical ontology and easily enables us to utilize its results, since some resources are readily available, such as bilingual dictionaries of COBALT-J/K (Collocation-Based Language Translator from Japanese to Korean) (Park et al., 1997) and COBALT-K/J (Collocation-Based Language Translator from Korean to Japanese) (Moon & Lee, 2000), which are machine translation systems developed by POSTECH (Pohang University of Science and Technology, Korea). In these bilingual dictionaries, nominal and verbal words are already annotated with concept codes from the Kadokawa thesaurus. By using the same sense inventories of these MT systems, we can easily apply and evaluate our ontology without additional lexicographic works. In addition, the Kadokawa thesaurus proved to be useful for providing a fundamental foundation to build lexical disambiguation knowledge in COBALT-J/K and COBALT-K/J MT systems (Li et al., 2000).

The second strategy to construct a practical ontology is to extend the hierarchy of the Kadokawa thesaurus by inserting additional semantic relations into its hierarchy. The additional semantic relations can be classified as case relations and other semantic relations. Thus far, case relations have been occasionally used to disambiguate lexical ambiguities in the form of valency information and case frame, but other semantic relations have not, because of the problem of discriminating them from each other, making them difficult to recognize. We define a total of 30 semantic relation types for WSD by referring mainly to the SELK (Sejong Electronic Lexicon of Korean) (Hong & Pak, 2001) and the Mikrokosmos ontology (Mahesh, 1996), as shown in Table 1.

400

Table 1. Semantic relation types in the ontology

Types of Semantic Relation	Relation Lists
Taxonomic relation	is-a
Case relation	agent, theme, experiencer, accompanier, instrument, location, source, destination, reason, appraisee, criterion, degree, receipient
Other semantic relation	has-member, has-element, contains, material-of, headed-by, operated-by, controls, owner-of, represents, symbol-of, name-of, producer-of, composer-of, inventor-of, make, measured-in

These semantic relation types cannot express all possible semantic relations existing among concepts, but experimental results demonstrated their usefulness for WSD. Table 2 shows the number of semantic relations semi-automatically inserted into the ontology from computational dictionaries and large corpora.

Table 2. The number of ontological relation instances

Types	Number
Taxonomic relations	1,100
Case relations	112,746
Other semantic relations	2,093
Total	115,939

3. Ontology Learning

To use the ontology in natural language processing (NLP) applications, a scoring mechanism was required to determine whether the governor and dependent concepts satisfy their semantic constraints in the ontology. Therefore, in order to measure concept association, we use an association ratio based on the information theoretic concept of mutual information (MI), which is a natural measure of the dependence between random variables (Church & Hanks, 1989). Resnik (1995b) suggested a measure of semantic similarity in an IS-A taxonomy, based on the notion of information content. However, his method differs from ours in that we consider all semantic relations in the ontology, not taxonomy relations only. To implement this idea, *source concepts (SC)* and *semantic relations (SR)* are bound into one entity, since *SR* is mainly influenced by *SC*, not the *destination concepts (DC)*. Therefore, if two entities, $< SC, SR>$, and DC have probabilities $P(<SC, SR>)$ and $P(DC)$, then their mutual information $I(<SC, SR>, DC)$ is defined as:

$$I(< SC,SR >,DC) = \log_2\left(\frac{P(< SC,SR >,DC)}{P(< SC,SR >)P(DC)} + 1 \right) \quad (1)$$

The MI between concepts in the ontology must be calculated before using the ontology as knowledge for disambiguating word senses. Figure 2 shows the construction process for training data in the form of *<SC (governor), SR, DC (dependent), frequency>* and the calculation of MI between the ontology concepts. We performed a slight modification on COBALT-K/J and COBALT-J/K to enable them to produce sense-tagged valency information instances with the specific concept codes of the Kadokawa thesaurus. After producing the instances, we converted syntactic relations into semantic relations by relying on the specific rules and human intuition (Fig. 3). As a result, we extracted sufficient training
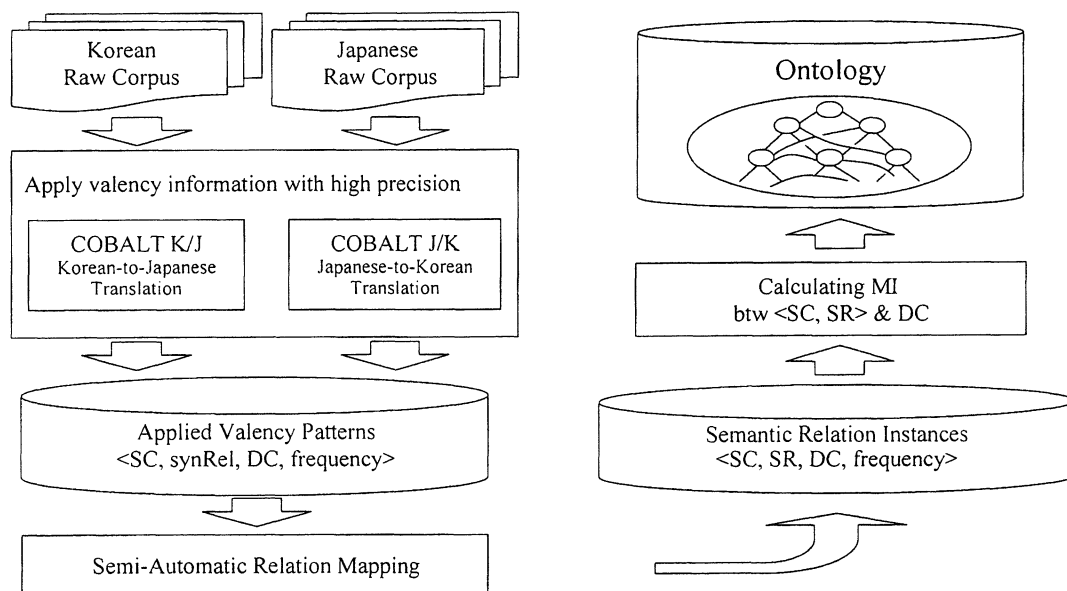
401

Fig. 2 Construction flow of ontology training data

Source Concepts	Syntactic Relations	Destination Concepts
381 (exercise)	*ul/lul* (object)	449 (right)
071 (live)	*wa/kwa* (adverb)	06|05|5 (living thing)
217 (soar)	*lo/ulo* (adverb)	002 (sky)
712 (join)	*i/ka* (subject)	5 (person)
369 (business)	*uy* (modifier)	744 (profit)
290 (bind)	*ey* (adverb)	449 (authority)

By using specific mapping rules and human intuition

381	theme	449
071	companion	06|05|5
217	destination	002
712	agent	5
369	owner-of	744
290	reason	449

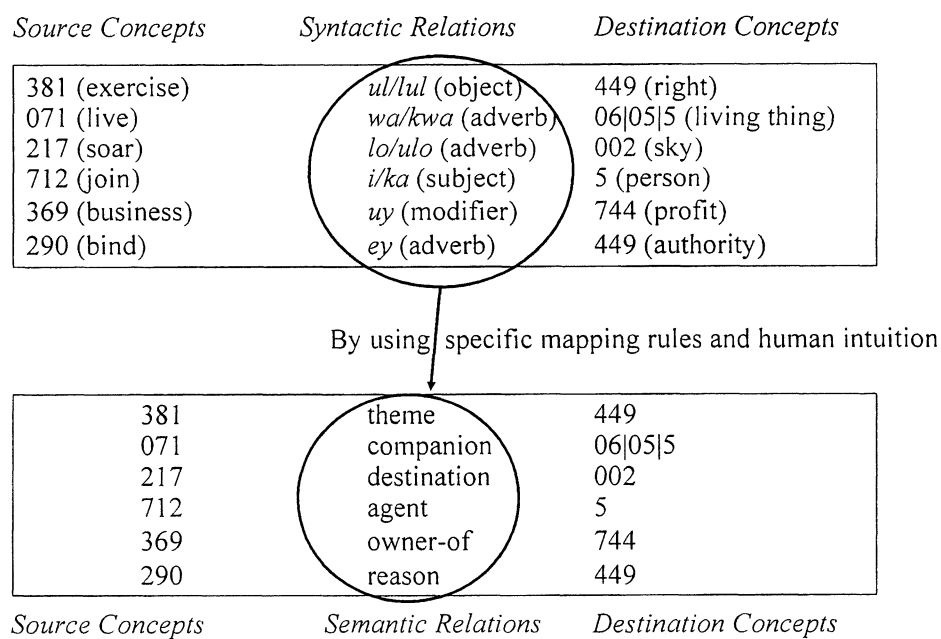*Source Concepts*   *Semantic Relations*   *Destination Concepts*

Fig. 3 Example of conversion from syntactic patterns to semantic patterns

data from the Korean raw corpus, which has 70 million words, and the Japanese raw corpus, which has eight hundred and ten thousand sentences.
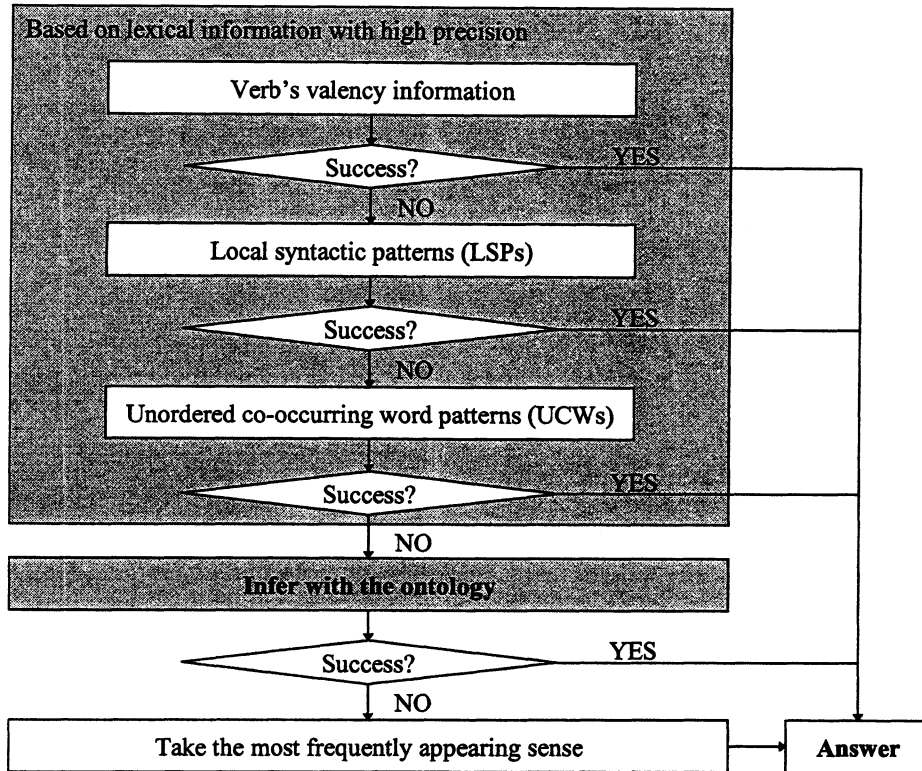
402

Fig. 4 Proposed WSD algorithm

4. Word Sense Disambiguation

The ontology is applicable to many fields. In this paper, we propose to use the ontology to disambiguate word senses. All approaches to word sense disambiguation (WSD) make use of words in a sentence to mutually disambiguate each other. The distinctions between various approaches lie in the source and type of knowledge constructed by the lexical units in a sentence.

Our WSD approach is a hybrid method, which combines the advantages of corpus-based and knowledge-based methods. We use the ontology as an external knowledge source and secured dictionary information as context information.

For a given ambiguous word W, Figure 4 describes our overall WSD algorithm. First, the verb's valency information is applied by using the formulas (2) and (3). $S(W)$ denotes a set of word senses of the word W, $SR(V)$ a selectional restriction of a verb V that takes the word W as its argument. C_i and P_j are concept types. $Csim(C_i, P_j)$ in Eq. (2) is used to compute the concept similarity between C_i and P_j, where $MSCA(C_i, P_j)$ is the most specific common ancestor of concept types C_i and P_j. If the matching score of valency information (Eq. (3)) is greater than a threshold, then set the sense of the word W to C_i and exit. Otherwise, LSPs and UCWs are used in order (Li $et\ al.$, 2000).

$$Csim(C_i, P_j) = \frac{2 * level(MSCA(C_i, P_j))}{level(C_i) + level(P_j)} * weight \qquad (2)$$

$$Vsim(S(W), SR(V)) = \max_i(Csim(C_i, P_j)), \qquad (3)$$
$$1 \le i \le n; 1 \le j \le m; C_i \in S(W); P_j \in SR(V)$$

403

If lexical information such as valency information, LSPs and UCWs is unavailable, or fails to match with other words, we use the ontology as a second knowledge base. The roles of the ontology in WSD are as follows. First, if previously-secured information for a concept is not available in a dictionary, the ontology provides extended semantic constraints for the concept. The extended semantic constraints were made in the previous ontology-building phase by other semantic constraints, including the same concept code. Second, if a direct semantic relation between concepts is not available in the ontology, the ontology and its scoring mechanism provide a relaxation procedure, which approximates their semantic association. The following are detailed descriptions of the procedure for applying the ontology to WSD work.

If MI is regarded as a weight between ontology concepts, the ontology can be treated as a graph with weighted edges. All edge weights are non-negative and weights are converted into penalties by the formula below. c indicates a constant, such as a maximum MI value over all possible pairs.

$$Pe(<SC, SR>, DC) = c - I(<SC, SR>, DC) \quad (4)$$

We use the formulas below to locate the least weighted path from one concept to the other. The score functions are defined as respectively:

$$Score^*(C_i, C_j) = \begin{cases} 1 & if\ C_i = C_j, \\ \min_p \left(Pe(<C_i, R_p>, C_j) \right) \\ \qquad if\ C_i \neq C_j\ and\ C_i, C_j\ have\ direct\ relations\ R_p, \quad (5\text{-}1) \\ \min_{C_k \in \{concepts\ in\ C_i \rightarrow C_j\}} \left(Score(C_i, C_k) * Score(C_k, C_j) \right) \\ \qquad if\ C_i \neq C_j,\ and\ C_k\ has\ direct\ relations\ with\ C_j. \end{cases}$$

$$Score^+(C_i, C_j) = \begin{cases} 0 & if\ C_i = C_j, \\ \min_p \left(Pe(<C_i, R_p>, C_j) \right) \\ \qquad if\ C_i \neq C_j\ and\ C_i, C_j\ have\ direct\ relations\ R_p, \quad (5\text{-}2) \\ \min_{C_k \in \{concepts\ in\ C_i \rightarrow C_j\}} \left(Score(C_i, C_k) + Score(C_k, C_j) \right) \\ \qquad if\ C_i \neq C_j,\ and\ C_k\ has\ direct\ relations\ with\ C_j. \end{cases}$$

Here C and R indicate concepts and semantic relations, respectively. It was found from the result that $Score^*$ and $Score^+$ formulas are almost same performance in inferring with the ontology. By applying these formulas, we can verify how well selectional constraints between concepts are satisfied. In addition, if there is no direct semantic relation between concepts, these formulas provide a relaxation procedure, which enables it to approximate their semantic relations. This characteristic enables us to obtain hints toward resolving metaphor and metonymy expressions.

To locate the best path, the search mechanism of the ontology applies the following heuristics. Firstly, a taxonomic relation must be treated as exceptional from other semantic relations, because they inherently lack frequencies between parent and child concepts. So we experimentally assign a fixed weight to those edges. Secondly, the weight given to an edge is sensitive to the context of prior edges in the path. Therefore, our mechanism restricts the number of times that a particular relation can be traversed in one path. Thirdly, this mechanism avoids an excessive change in the gradient.

5. Experimental results

We performed an evaluation of the proposed WSD algorithm using the ontology. This section presents the experimental results. Eight ambiguous nouns and four ambiguous verbs were selected, along with a total of 604 test sentences in which one test noun or verb appears. The test sentences were randomly selected from the raw Korean corpus. Out of several senses for each ambiguous word, we considered only two or three senses that are most frequently used in the corpus.

We performed three experiments for MT system: The first experiment, *BASE*, is the case where the most frequently used senses are always taken as the senses of test words. The purpose of this experiment is to show a baseline in WSD work. The second, *LEX*, uses only secured dictionary information, such as the selectional restriction of verbs, local syntactic patterns, and unordered co-occurring word patterns in disambiguating word senses. This is a general method without an ontology. The third, *ONTO*, shows the results of our WSD method using the ontology. The experimental results are compared with each other in Table 3. In these experiments, the ONTO method achieved a 9% improvement over the LEX method.

Table 3. Experimental results of word sense disambiguation in COBALT-K/J (%)

POS	Lexical word	Sense	BASE	LEX	ONTO
Noun	*Pwuca*	father & child / rich man	65.3	69.2	86.0
	Kancang	liver / soy sauce	66.0	87.8	91.8
	Kasa	housework / words of song	48.0	88.5	96.1
	Kwutwu	shoe / word of mouth	78.0	85.7	95.9
	Nwun	eye / snow	82.0	96.0	94.0
	Yongki	courage / container	62.0	74.0	82.0
	Kyengpi	expenses / defense	74.5	78.4	90.2
	Kyeongki	times / match	52.9	80.4	93.2
Verb	*Nayli-ta*	get off / draw	42.0	72.0	88.0
	Seywu-ta	make (a plan) / build	54.0	88.0	95.4
	Ssu-ta	use / write / put on (a hat)	46.0	86.0	96.0
	Taywu-ta	burn / give a ride	50.0	86.0	92.0
Average Precision			60.1	82.7	91.7

Table 4 shows the applicability and precision for each phase in the WSD algorithm. In ontology phase, the ratio of applicability was 18.1% and precision was 86.4%.

Table 4. Applicability and precision for each phase in the WSD algorithm

Phase	Applicability (%)	Precision (%)
Verb's valency information	34.8	91.6
Local syntactic patterns (LSPs)	9.8	91.4
Unordered co-occurring word patterns (UCWs)	28.2	92.3
Infer with the ontology	**18.1**	**86.4**
Take the most frequently appearing sense	9.1	74.2
Sum / Average Precision according to Applicability	100	89.2

The main reason for these results is that, in the absence of secured dictionary information about an ambiguous word, the ontology provides an extended case frame by the concept code of the word. In addition, when there is no direct semantic constraint between concepts, our search mechanism

provides a relaxation procedure. Therefore, the quality and usefulness of the ontology were indirectly proved by these results.

6. Conclusion

In this paper we have proposed a corpus-based ontology learning method and an ontology-based WSD algorithm. The ontology, which includes extensive semantic relations between concepts, differs from many resources in that it has no language-dependent knowledge, which is a network of concepts, not words. The ontology can be applied to other languages, if the concept codes, which corresponding to the senses of each headword, are merely inserted into their dictionaries.

In order to learn ontology for WSD, we automatically produced sense-tagged valency information instances from large raw corpus. After producing the instances, we semi-automatically converted syntactic relations into semantic relations, and then the mutual information between concepts in the ontology was calculated. If mutual information is regarded as a weight between ontology concepts, the ontology can be treated as a graph with weighted edges and weights are converted into penalties. By locating the least weighted path from one concept to the other concept, we can verify how well selectional constraints between concepts are satisfied.

The ontology is applied to disambiguate word senses in the form of an ontological graph search. The search mechanism determines whether selectional constraints between concepts are satisfied or not, and includes a relaxation procedure, which enables concept pairs with no direct selectional restriction to approximate their semantic association. This characteristic enables us to obtain hints toward resolving metaphor and metonymy expressions.

The ontology calls for further specific concepts and semantic relations to improve the WSD performance. A further direction of this study will be focused on how to combine the concept of the semantic web (Berners-Lee et al., 2001) and the ontology in NLP applications.

Acknowledgements

References

Berners-Lee, T., Hendler, J., and Lasilla, O. 2001. The Semantic Web. *Scientific American*, May.

Church, K. and Hanks, P. 1989. Word association norms, mutual information, and lexicography, *In Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, pp. 76-83.

Dagan, I. and Itai, A. 1994. Word sense disambiguation using a second language monolingual corpus. *Computational Linguistics*, 20(4):563-596.

Dahlgren, K. G. 1988. *Naive Semantics for Natural Language Understanding*. Kluwer Academic Pub., Boston.

Gale, W. A., Church, K. W., and Yarowsky, D. 1992. A method for disambiguating word senses in a large corpus. *Computers and the Humanities*, 26(5):415-439.

Hong, C. S. and Pak, M. G. 2001. Developing a large scale computational lexical database of contemporary Korean : SELK, *In Proceedings of the 19th International Conference on Computer Processing of Oriental Languages (ICCPOL 2001)*, Seoul, Korea, pp. 20-26.

Ide, N. and Veronis, J. 1998. Introduction to the special issue on word sense disambiguation: the state of the art. *Computational Linguistics*, 24(1):1-40.

Li, H. F., Heo, N. W., Moon, K. H., Lee, J. H., and Lee, G. B. 2000. Lexical Transfer Ambiguity Resolution Using Automatically-Extracted Concept Co-occurrence Information, *International Journal of Computer Processing of Oriental Languages*, World Scientific Pub., 13(1):53-68.

Mahesh, K., and Nirenburg, S. 1996. *Knowledge-based systems for Natural Language Processing*, Memoranda in Computer and Cognitive Science. NMSU CRL Technical Report, MCCS-96-296.

Moon, K. H. and Lee, J. H. 2000. Representation and Recognition Method for Multi-Word Translation Units in Korean-to-Japanese MT System, *In the 18ᵗʰ International Conference on Computational Linguistics (COLING 2000)*, Germany, pp. 544-550.

Ohno, S. and Hamanishi, M. 1981. *New Synonyms Dictionary*, Kadokawa Shoten, Tokyo. (Written in Japanese.)

Park, C. J., Lee, J. H., Lee, G. B., and Kakechi, K. 1997. Collocation-Based Transfer Method in Japanese-Korean Machine Translation, *Transaction of Information Processing Society of Japan*, 38(4):707-718. (Written in Japanese)

Resnik, P. 1995a. Disambiguating noun groupings with respect to WordNet senses. In *Proceedings of the Third Workshop on Very Large Corpora*, Cambridge, MA, pp.54-68.

Resnik, P. 1995b. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In *Proceedings of IJCAI-95*, Montreal, Canada, pp. 448-453.

Yarowsky, D. 1992. Word sense disambiguation using statistical models of Roget's categories trained on large corpora. *The 14ᵗʰ International Conference on Computational Linguistics*, Nantes, France, pp.454-460.