# Semi-Automatic Practical Ontology Construction by Using a Thesaurus, Computational Dictionaries, and Large Corpora

**Sin-Jae Kang and Jong-Hyeok Lee**
Div. of Electrical and Computer Engineering, Pohang University of Science and Technology
San 31 Hyoja-Dong, Nam-Gu, Pohang 790-784
Republic of KOREA
sjkang@postech.ac.kr, jhlee@postech.ac.kr

## Abstract

This paper presents the semi-automatic construction method of a practical ontology by using various resources. In order to acquire a reasonably practical ontology in a limited time and with less manpower, we extend the Kadokawa thesaurus by inserting additional semantic relations into its hierarchy, which are classified as case relations and other semantic relations. The former can be obtained by converting valency information and case frames from previously-built computational dictionaries used in machine translation. The latter can be acquired from concept co-occurrence information, which is extracted automatically from large corpora. The ontology stores rich semantic constraints among 1,110 concepts, and enables a natural language processing system to resolve semantic ambiguities by making inferences with the concept network of the ontology. In our practical machine translation system, our ontology-based word sense disambiguation method achieved an 8.7% improvement over methods which do not use an ontology for Korean translation.

## 1 Introduction

An ontology is a knowledge base with information about concepts existing in the world or domain, their properties, and how they relate to each other. The principal reasons to use an ontology in machine translation (MT) are to enable source language analyzers and target language generators to share knowledge, to store semantic constraints, and to resolve semantic ambiguities by making inferences using the concept network of the ontology (Mahesh, 1996; Nirenburg *et al.*, 1992). 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 general, to build a high-quality semantic knowledge base, manual processing is indispensable. Previous attempts were mostly performed manually, or were developed without considering the context of a practical situation (Mahesh, 1996; Lenat *et al.*, 1990). Therefore, it is difficult to construct a practical ontology with limited time and manpower resources. To solve this problem, we propose a semi-automatic ontology construction method, which takes full advantage of already existing knowledge resources and practical usages in large corpora. First, we define our ontology representation language (ORL) by modifying the most suitable among previously developed ORLs, and then design a language-independent and practical (LIP) ontology structure based on the defined ORL. Afterwards, we construct a practical ontology by the semi-automatic construction method given below.

We extend the existing Kadokawa thesaurus (Ohno & Hamanishi, 1981) by inserting additional semantic relations into the hierarchy of the thesaurus. Uramoto (1996) and Tokunaga (1997) propose thesaurus extension methods for positioning unknown words in an existing thesaurus. Our approach differs in that the objects inserted are not words but semantic relations.

Additional semantic relations can be classified as case relations and other semantic relations. The former can be obtained by converting the established valency information in bilingual dictionaries of COBALT-J/K (Collocation-Based Language Translator from Japanese to Korean) and COBALT-K/J (Collocation-Based Language Translator from Korean to Japanese) (Moon & Lee, 2000) MT systems, as well as from the case frame in the Sejong electronic dictionary[1]. The latter can be acquired from concept co-occurrence information, which is extracted automatically from a corpus (Li *et al*., 2000).

The remainder of this paper is organized as follows. We describe the principles of ontology design and an ORL used to represent our LIP ontology in the next section. In Section 3, we describe the semi-automatic ontology construction methodology in detail. An ontology-based word sense disambiguation (WSD) algorithm is given in Section 4. Experimental results are presented and analyzed in Section 5. Finally, we make a conclusion and indicate the direction of our future work in Section 6.

## 2 Ontology Design

### 2.1 Basic Principles

Although no formal principles exist to determine the structure or content of our ontology, we can suggest some principles underlying our methodology. Firstly, an ontology for natural language processing (NLP) must provide concepts for representing word meanings in the lexicon and store selectional constraints of concepts, which enable inferences using the network of an ontology (Onyshkevych, 1997). These inferences can assist in metaphor and metonymy processing, as well as word sense disambiguation. For these reasons, an ontology becomes an essential knowledge source for high quality NLP, although it is difficult and time-consuming to construct. Secondly, an ontology can be effortlessly shared by any application and in any domain (Gruber, 1993; Karp *et al*., 1999; Kent, 1999). More than two different ontologies in a certain domain can produce a semantic mismatch problem between concepts. Further, if you wish to apply an existing ontology to a new application, it will often be necessary to convert the structure of the ontology to a new one. Thirdly, an ontology must support language independent features, because constructing ontologies for each language is inefficient. Fourthly, an ontology must have capabilities for users to easily understand, search, and browse. Therefore, we define a suitable ORL to support these principles.

### 2.2 Ontology Representation Language

Many knowledge representation languages are built specifically to share knowledge among different knowledge representation systems. Five types of ORLs were reviewed, such as FRAMEKIT (Nirenburg *et al*., 1992), Ontolingua (Gruber, 1993), CycL (Lenat *et al*., 1990), XOL (Karp *et al*., 1999), and Ontology Markup Language (OML) (Kent, 1999). According to their semantics, FRAMEKIT and XOL adopt frame representation, CycL and Ontolingua use an extended first order predicate calculus, and the OML is based on conceptual graphs (CGs). Excepting FRAMEKIT and CycL, the other ORLs have not yet been applied to build any large-scale ontology.

Among this variety of ORLs, we chose the simplified OML as the ORL of our LIP ontology, which is based on Extensible Markup Language (XML) and CGs. Since XML has a well-established syntax, it is reasonably simple to parse, and XML will be widely used, because it has many software tools for parsing and manipulating, and a human readable representation. We intend to leave room for improvement by adopting the semantics of CGs, because the present design of our LIP ontology is for the specific purpose of disambiguating word senses. In future, however, we must extend its structure and content to build an interlingual meaning representation during semantic analysis in machine translation. Sowa's CGs (1984) is a widely-used knowledge representation language, consisting of logic structures with a graph notation and several features integrated from semantic net and frame representation. Globally, many research teams are working on the extension and application of CGs in many domains.

[1] The Sejong electronic dictionary has been developed by several Korean linguistic researchers, funded by Ministry of Culture and Tourism, Republic of Korea. (http://www.sejong.or.kr)

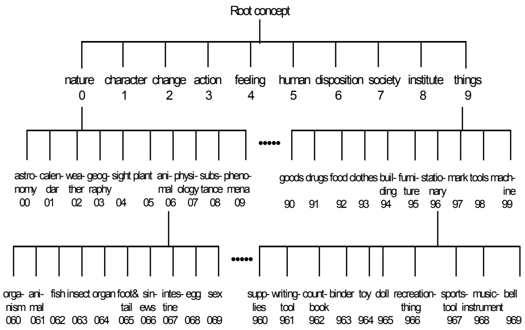

Figure 1. Concept hierarchy of the Kadokawa thesaurus

## 3 Ontology Construction

Many ontologies are developed for purely theoretical purposes, or are constructed as language-dependent computational resources, such as WordNet and EDR. However, they are seldom constructed as a language-independent computational resource.

To construct a language-independent and practical ontology, we developed two strategies. First, we introduced the same number and grain size of concepts of the Kadokawa thesaurus and its taxonomic hierarchy into the LIP ontology. The thesaurus has 1,110 Kadokawa semantic categories and a 4-level hierarchy as a taxonomic relation (see Figure 1). This approach is a moderate shortcut to construct a practical ontology which easily enables us to utilize its results, since some resources are readily available, such as bilingual dictionaries of COBALT-J/K and COBALT-K/J. 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 LIP ontology without additional lexicographic works. In addition, the Kadokawa thesaurus has proven 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 used occasionally 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 Sejong electronic dictionary and the Mikrokosmos ontology (Mahesh, 1996), as shown in Table 1. These semantic relation types cannot express all possible semantic relations existing among concepts, but experimental results demonstrated their usefulness for WSD.

Table 1. Sematic relation types in the LIP ontology

| Types | Relation Lists |
|---|---|
| Taxonomic relation | is-a |
| Case relation | agent, theme, experiencer, accompanier, instrument, location, source, destination, reason, appraisee, criterion, degree, recipient |
| 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 |

Two approaches are used to obtain these additional semantic relations, which will be inserted into the LIP ontology. The first imports relevant semantic information from existing computational dictionaries. The second applies the semi-automatic corpus analysis method (Li *et al.*, 2000). Both approaches are explained in Section 3.1 and 3.2, respectively.

Figure 2 displays the overall constructing flow of the LIP ontology. First, we build an initial LIP ontology by importing the existing Kadokawa thesaurus. Each concept inserted into the initial ontology has a Kadokawa code, a Korean name, an English name, a timestamp, and a concept definition. Although concepts can be uniquely identified by the Kadokawa concept codes, their Korean and English names are

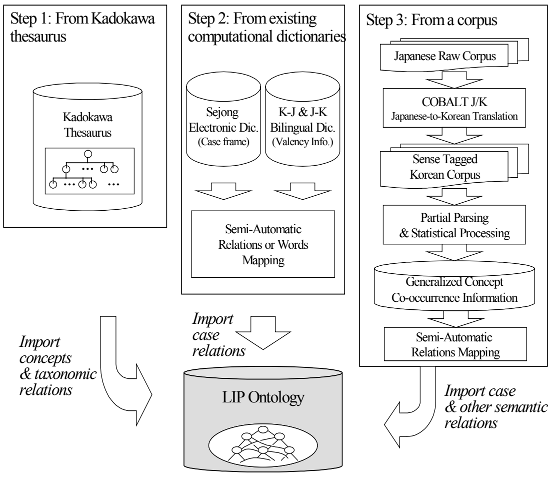
Step 1: From Kadokawa thesaurus
Kadokawa Thesaurus
Step 2: From existing computational dictionaries
Sejong Electronic Dic. (Case frame)
K-J & J-K Bilingual Dic. (Valency Info.)
Semi-Automatic Relations or Words Mapping
Step 3: From a corpus
Japanese Raw Corpus
COBALT J/K Japanese-to-Korean Translation
Sense Tagged Korean Corpus
Partial Parsing & Statistical Processing
Generalized Concept Co-occurrence Information
Semi-Automatic Relations Mapping
Import concepts & taxonomic relations
Import case relations
LIP Ontology
Import case & other semantic relations

Figure 2. Ovreall constructing flow of the LIP ontology

inserted for the readability and convenience of the ontology developer.

## 3.1 Dictionary Resources Utilization

Case relations between concepts can be primarily derived from semantic information in the Sejong electronic dictionary [2] and the bilingual dictionaries of MT systems, which are COBALT-J/K and COBALT-K/J.

We obtained 7,526 case frames from verb and adjective sub-dictionaries, which contain 3,848 entries. Automatically converting lexical words in the case frame into the Kadokawa concept codes by using COBALT-K/J (see Figure 3[3]), we extracted a total of 6,224 case relation instances.

The bilingual dictionaries, which contain 20,580 verb and adjective entries, have 16,567 instances of valency information. Semi-automatically converting syntactic relations into semantic relations by using specific rules and human intuition (see Figure 4), we generated 15,956 case relation instances. The specific rules, as shown in Figure 5, are inferred from training samples, which are explained in Section 4.1. These obtained instances may overlap each other, but all instances are inserted only once into the initial LIP ontology.

Case frames

| | | |
|---|---|---|
| *Cilmwunha-ta* (question) | agent | *Salam*(person) |
| *Cungkasikh-ta* (increase) | theme | *Kakyek* (price) |
| *Kyeyhoykha-ta* (plan) | theme | *Pepan* (bill) |
| *Seywu-ta* (construct) | theme | *Saep* (business) |

Case relations

| | | |
|---|---|---|
| 346 | agent | 5 |
| 743 | theme | 171 |
| 344 | theme | 419 |
| 394 | theme | 369 |

Figure 3. Example of conversion from case frames in the Sejong dictionary

Valency information

| | | |
|---|---|---|
| 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) |

Case relations

| | | |
|---|---|---|
| 381 | theme | 449 |
| 071 | accompanier | 06\|05\|5 |
| 217 | destination | 002 |
| 712 | agent | 5 |

Figure 4. Example of conversion from valency information in the bilingual dictionaries

## 3.2 Corpus Analysis

For the automatic construction of a sense-tagged corpus, we used the COBALT-J/K, which is a high-quality practical MT system developed by POSTECH in 1996. The entire system has been used successfully at POSCO (Pohang Iron and Steel Company), Korea, to translate patent materials on iron and steel subjects. We performed a slight modification on COBALT-J/K so that it can produce Korean translations from Japanese texts with all nominal and verbal words tagged with the specific concept codes of the Kadokawa thesaurus. As a result, a Korean sense-tagged corpus, which has two hundred and fifty thousand sentences, can be obtained from Japanese texts. Unlike English, the Korean language has almost no syntactic constraints on word order as long as a verb appears in the final position. So we defined 12 local syntactic patterns (LSPs) using syntactically-related words in a sentence. Frequently co-occurring words in a sentence have no syntactic relations to homographs but may control their meaning. Such words are retrieved as unordered co-occurring words (UCWs). Case relations are obtained from LSPs, and other semantic

[2] The Sejong electronic dictionary has sub-dictionaries, such as noun, verb, pronoun, adverb, and others.
[3] The Yale Romanization is used to represent Korean lexical words.

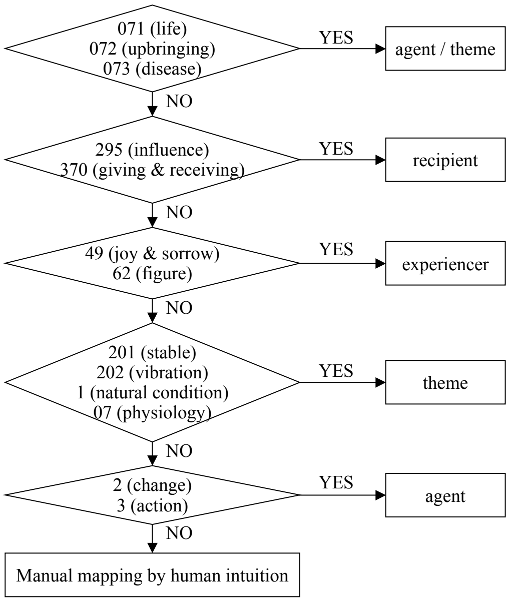

Figure 5. Example of subject relation mapping rules with governer concept codes

relations are acquired from UCWs. Concept co-occurrence information (CCI), which is composed of LSPs and UCWs, can be extracted by partial parsing and scanning. To select the most probable concept types, Shannon's entropy model is adopted to define the noise of a concept type to discriminate the homograph. Although it processes for concept type discrimination, many co-occurring concept types, which must be further selected, remain in each LSP and UCW. To solve this problem, some statistical processing was automatically applied (Li *et al.*, 2000). Finally, manual processing was performed to generate the ontological relation instances from the generalized CCI, similar to the previous valency information. The results obtained include approximately about 3,701 case relations and 1,650 other semantic relations from 9,245 CCI, along with their frequencies. The obtained instances are inserted into the initial LIP ontology. Table 2 shows the number of relation instances imported into the LIP ontology from the Kadokawa thesaurus, computational dictionaries, and a corpus.

## 4 Ontology Application

The LIP ontology is applicable to many NLP applications. In this paper, we propose to use the ontology to disambiguate word senses. All approaches to 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.

Table 2. Imported relation instances

| Types | Number |
|---|---|
| Taxonomic relations | 1,100 |
| Case relations | 19,459 |
| Other semantic relations | 1,650 |
| Total | 22,209 |

Our WSD approach is a hybrid method, which combines the advantages of corpus-based and knowledge-based methods. We use the LIP ontology as an external knowledge source and secured dictionary information as context information. Figure 6 shows our overall WSD algorithm. First, we apply the previously-secured dictionary information to select correct senses of some ambiguous words with high precision, and then use the LIP ontology to disambiguate the remaining ambiguous words. The following are detailed descriptions of the procedure for applying the LIP ontology to WSD work.

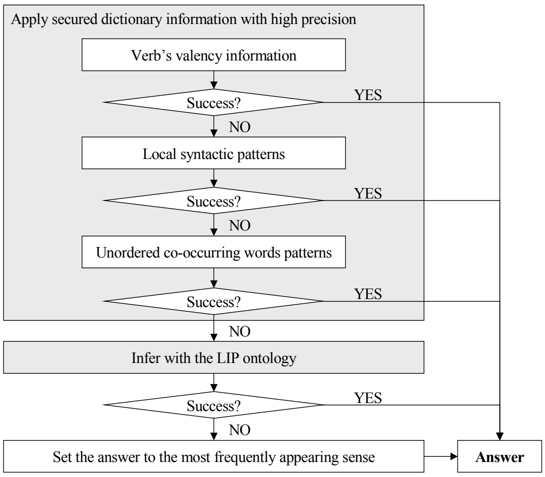
Apply secured dictionary information with high precision
Verb's valency information
Success?
YES
NO
Local syntactic patterns
Success?
YES
NO
Unordered co-occurring words patterns
Success?
YES
NO
Infer with the LIP ontology
Success?
YES
NO
Set the answer to the most frequently appearing sense
Answer

Figure 6. The proposed WSD algorithm

### 4.1 Measure of Concept Association

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

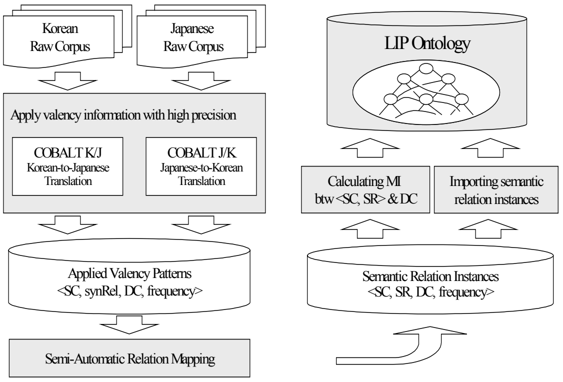

Figure 7. Construction flow of ontology training data

between random variables (Church & Hanks, 1989). Resnik (1995) 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, we bind *source concepts (SC)* and *semantic relations (SR)* 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:

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

The MI between concepts in the LIP ontology must be calculated before using the ontology as knowledge for disambiguating word senses. Figure 7 shows the construction process for training data in the form of <*SC (governer), SR, DC (dependent), frequency*> and the calculation of MI between the LIP 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 using the specific rules (see Figure 5) and human intuition. As a result, we extracted sufficient training data from the Korean raw corpus: KIBS (Korean Information Base System, '94-'97) is a large-scale corpus of 70 million words, and the Japanese raw corpus, which has eight hundred and ten thousand sentences. During this process, more specific semantic relation instances are obtained when compared with previous instances obtained in Section 3. Since such specific instances reflect the context of a practical situation, they are also imported into the LIP ontology. Table 3 shows the final number of semantic relations inserted into the LIP ontology.

Table 3. Final relation instances in the LIP ontology

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

### 4.2 Locate the Least Weighted Path from One Ontology Concept to Other Concept

If we regard MI as a weight between ontology concepts, we can treat the LIP ontology as a graph with weighted edges. All edge weights are non-negative and weights are converted into penalties by the below formula *Pe*. *c* indicate a constant, maximum MI between concepts of the LIP ontology.

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

So we use the formula below to locate the least weighted path from one concept to the other concept. The score function *S* is defined as:

$$S(C_i,C_j)=\begin{cases}1 & \mathit{if}\ C_i=C_j,\\ \min\limits_{p}\big(Pe(<C_i,R_p>,C_j)\big) & \\ & \mathit{if}\ C_i\neq C_j\ \mathit{and}\ C_i\xrightarrow{R_p}C_j,\\ \min\limits_{C_k\in\{C_i\to C_j\}}\big(S(C_i,C_k)+S(C_k,C_j)\big) & \\ & \mathit{if}\ C_i\neq C_j,\mathit{and}\ C_k\xrightarrow{R_p}C_j.\end{cases}$$

Here *C* and *R* indicate concepts and semantic relations, respectively. By applying this formula, we can verify how well selectional constraints between concepts are satisfied. In addition, if there is no direct semantic relation between concepts, this formula provides a relaxation procedure, which enables it to approximate their semantic relations. This characteristic enables us

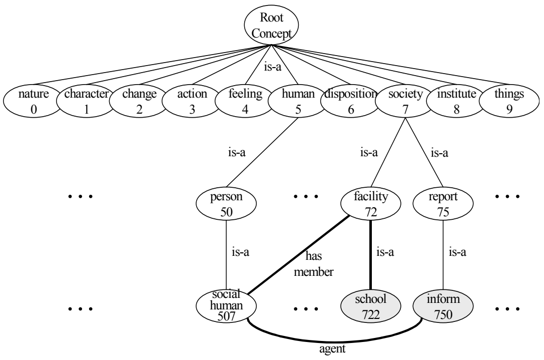

Figure 8. Example of the best path between concepts “school” and “inform” in the LIP ontology

to obtain hints toward resolving metaphor and metonymy expressions. For example, when there is no direct semantic relation between concepts such as “school” and “inform,” the inferring process is as follows. The concept “school” is a “facility”, and the “facility” has “social human” as its members. The concept “inform” has “social human” as its agent. Figure 8 presents an example of the best path between these concepts, which is shown with bold lines. To locate the best path, the search mechanism of our LIP ontology applies heuristics as follows. 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 assign a fixed weight to those edges experimentally. 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 Evaluation

For experimental evaluation, eight ambiguous Korean nouns were selected, along with a total of 404 test sentences in which one of the homographs appears. The test sentences were randomly selected from the KIBS. Out of several senses for each ambiguous word, we considered only two senses that are most frequently used in the corpus. We performed three experiments: 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 the baseline for WSD work. The second, PTN, 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, LIP, shows the results of our WSD method using the LIP ontology. The experimental results are shown in Table 4. In these experiments, the LIP method achieved an 8.7% improvement over the PTN method for Korean analysis. The main reason for these results is that, in the absence of secured dictionary information (see Figure 7) about an ambiguous word, the ontology provides a generalized case frame (i.e. semantic restriction) by the concept code of the word. In addition, when there is no direct semantic restriction between concepts, our search mechanism provides a relaxation procedure (see Figure 8). Therefore, the quality and usefulness of the LIP ontology were proved indirectly by these results.

Table 4. Experimental results of WSD (%)

| Homograph | Sense | BASE | PTN | LIP |
|---|---|---|---|---|
| *Pwuca* | father & child | 76.9 | 69.2 | 86.0 |
| | rich man | | | |
| *Kancang* | liver | 67.3 | 87.8 | 91.8 |
| | soy sauce | | | |
| *Kasa* | housework | 48.1 | 88.5 | 96.1 |
| | words of song | | | |
| *Kwutwu* | shoe | 79.6 | 85.7 | 95.9 |
| | word of mouth | | | |
| *Nwun* | eye | 82.0 | 96.0 | 92.0 |
| | snow | | | |
| *Yongki* | courage | 62.0 | 74.0 | 82.0 |
| | container | | | |
| *Kyengpi* | expenses | 74.5 | 78.4 | 90.2 |
| | defense | | | |
| *Kyeongki* | times | 52.9 | 80.4 | 95.6 |
| | match | | | |
| Average Precision | | 67.9 | 82.5 | 91.2 |

## 6 Conclusion

In this paper we have proposed a semi-automatic construction method of the LIP ontology and an ontology-based WSD algorithm. The LIP ontology includes substantial semantic relations between concepts, and differs from many of the resources in that there is no language-dependent knowledge in the resource, which is a network of concepts, not words. Semantic relations of the LIP ontology are generated by considering two different languages, Korean and Japanese. In addition, we can easily apply the ontology without additional lexicographic works, since large-scale bilingual dictionaries have words already annotated with concept codes of the LIP ontology. Therefore, our LIP ontology is a language independent and practical knowledge base. You can apply this ontology for other languages, if one merely inserts Kadokawa concept codes for each entry into the dictionary. Our ontology construction method requires manual processing, i.e., mapping from syntactic relations to semantic relations by specific rules and human intuition. However, this is necessary for building a high-quality semantic knowledge base. Our construction method is quite effective in comparison with other methods.

We plan further research on how to effectively divide the grain size of ontology concepts to best express the whole world knowledge, and how to utilize the LIP ontology in a full semantic analysis process.

## Acknowledgements

The authors would like to thank the Ministry of Education of Korea for its financial support toward the Electrical and Computer Engineering Division at POSTECH through its BK21 program.

## References

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

Gruber, Thomas R. 1993. A Translation Approach to Portable Ontology Specification. *Knowledge Acquisition* 5(2):199-220.

Karp, P. D., V. K. Chaudhri, and J. F. Thomere. 1999. XOL: An XML-Based Ontology Exchange Language. *Technical Note 559*, AI Center, SRI International, July.

Kent, Robert E. 1999. Conceptual Knowledge Markup Language: The Central Core. In *the Electronic Proceedings of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management(KAW`99).* Banff, Alberta, Canada, October.

Lenat, D. B. *et al.* 1990. Cyc: toward programs with common sense. *Communications of the ACM*, 33(8):30-49.

Li, Hui-Feng *et al.* 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, Kavi. 1996. Ontology Development for Machine Translation: Ideology and Methodology. *Technical Report MCCS 96-292*, Computing Research Laboratory, New Mexico State University, Las Cruces, NM.

Moon, Kyunghi and Jong-Hyeok Lee. 2000. Representation and Recognition Method for Multi-Word Translation Units in Korean-to-Japanese MT System. *COLING 2000*, pages 544-550, Germany.

Nirenburg, Sergei, Jaime Carbonell, Masaru Tomita, and Kenneth Goodman. 1992. *Machine Translation: A Knowledge-Based Approach*, Morgan Kaufmann Pub., San Mateo, California.

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

Onyshkevych, Boyan A. 1997. An Ontological-Semantic Framework for Text Analysis. *Ph.D. dissertation*, Program in Language and Information Technologies, School of Computer Science, Carnegie Mellon University, CMU-LTI-97-148.

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

Sowa, John F. 1984. *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley Pub., MA.

Takenobu, Tokunaga *et al.* 1997. Extending a thesaurus by classifying words. In *Proceedings of the ACL-EACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources*, pages 16-21, Madrid, Spain.

Uramoto, Naohiko. 1996. Positioning Unknown Word in a Thesaurus by using Information Extracted from a Corpus. In *Proceedings of COLING-96*, pages 956-961, Copenhagen, Denmark.