

# Design and Construction of Knowledge base for Verb using MRD and Tagged Corpus<sup>1</sup>

Young-Soog Chae, Key-Sun Choi

KORTERM (Korea Terminology Research Center for Language and Knowledge Engineering)  
Korea Advanced Institute of Science and Technology  
373-1, Kusung-dong, Yusong-gu, Taejeon, 305-701, KOREA  
{yschae, kschoi}@korterm.kaist.ac.kr

## Abstract

This paper represents the procedure of building syntactic knowledge base. This study is to construct basic sentence pattern automatically by using the POS-tagged corpus in balanced KAIST corpus, and electronic dictionary for Korean, and to construct syntactic knowledge base with specific information added to the lexicographer's analysis.

The summary of work process will be as follows:

- 1) Extraction of characteristic verb targeting the high frequency verb from KAIST corpus
- 2) Constructing sentence pattern from each verb case frame structure extracted from MRD
- 3) Making out the noun categories of sentence pattern through KCP examples
- 4) Semantic classification of selected verb suitable for classified sentence pattern
- 5) Description of hyper concept to individual noun categories
- 6) Putting the translated words in Japanese to each noun and verb

## 1. Introduction

The case frame structure, which is necessary for syntactic-semantic analysis, is represented by the semantic co-relation of nouns, verbs or adjectival predicates with declined or conjugated endings. The case frame describes the verb and lists arguments the verb takes and the types of nouns or noun phrases in the argument position, forming the concrete structure. The case frame includes information necessary for syntactic-semantic analysis in that it specifies information such as the semantic roles of the nouns and semantic restrictions.

The verbal ambiguity comes from the differences in the contracted distribution of nouns that co-occur in the internal structure of the same verb. With the same case frame, the verb 'Chi-Da' (hit) has different meanings depending on the distributions of N0, N1 nouns as shown in the following:

*Chi-Da 1 : N0-Ka N-Ul Chi-Da*  
*N0=[human]*  
*N1=[salt, pepper, vinegar...]*  
*Meaning : to sprinkle*

*Chi-Da 2: N0-Ka N1-Ul Chi-Da*  
*N0=[truck, car, carriage...]*  
*N1=[human, animal...]*  
*Meaning : to be hurt*

To resolve the verbal ambiguity, we classify verbs into intransitive verb, transitive verb, and intransitive-transitive verb according to the grammatical characteristics of verbs, making the case frame, to which the list of argument nouns is added. The sentence pattern dictionary just lists

specific examples of the arguments and describes the hyper concept. The syntactic knowledge base we construct differs from the sentence pattern dictionary in that it not only describes the hyper concept but also construct the list of all the nouns that co-occur with the verb. The reason why we include the list of all the nouns that co-occur with the verb is that even in the internal structure of the noun that shares the same list of nouns, the foreign counterpart of the verb may correspond to other verbs. This is because the semantic relations of the co-occurrence of verb and noun vary from one language to another. Such a semantic distinction of verb and noun may be used to resolve ambiguity at the step of syntactic analysis, and to generate correct words in the machine translations, and to extract information pattern in the information retrieval. There has been many researches to construct better case frame since the case frames of high quality that have sufficient and correct information for analysis play an important role in the natural language processing.

This study is to construct basic sentence pattern by using the balanced text corpus, POS-tagged Corpus, and Electronic Dictionary for Korean automatically, and to construct syntactic knowledge base with specific information added to the lexicographer's analysis.

## 2. Syntactic Knowledge Base

### 2.1. Essential Elements of Syntactic Knowledge Base

Knowledge as a dictionary has two kinds: dictionary for human being and dictionary for machine. The dictionary for human being can have relatively simple

<sup>1</sup> \* This paper has been supported by AITrc and The Korean Ministry of Science and Technology.

information because it presupposes people's experience, knowledge, and inference. The dictionary for machine must describe all the information that human being uses to understand and infer because such a human knowledge and inference are not available. Human being understands meanings through combination of various kinds of elements. The elements that the dictionary for machine must contain notation, phonetic information, morphological information, syntactic information, semantic information, information of combinative relation, information of importance degree, information of polysemy selection, etc. To construct the syntactic knowledge base, we primarily focus on grammatical and semantic information.

### 2.1.1. Syntax Information

In the description of verbs, most basic is the information of sentence pattern, which tells us whether it is intransitive, transitive, or intransitive-transitive. This information generally does not play a crucial role in resolving ambiguity, but can resolve ambiguity of intransitive-transitive verbs such as the following:

*Chi-Da 3: N0-Ka Chi-Da*

*N0=[rainy wind, typhoon, snowstorm...]*

*Meaning : to blow*

*Chi-Da 4: N0-Ka N1-Ul Chi-Da*

*N0=[truck, car, carriage...]*

*N1=[human, animal...]*

*Meaning : to be hurt*

It is difficult to consider *Chi-Da 3* and *Chi-Da 4* the same verb because the latter has the related word such as 'chi-i-da' while the former has no related word, and because the latter is transitive while the former is intransitive. The dictionary of morphological analysis lists under the same headword the words that have the same forms. However, the syntactic-semantic processing requires more specific distinctions. We describe specific meanings by presenting the list of nouns that fill in each argument, which is used as basic data for Dictionary for translation of Sentence.

What is important in the construction of case frame is how many arguments are allowed. This is closely related to how the essential argument differs from the optional argument. For example, the verb *Pi-Kyo-Ha-Da* in principle has three arguments.

*Pi-Kyo-Ha-Da:*

*N0-Ka N1-Lul N2-Wa Pi-Kyo-Ha-Da*

In the above example, the three arguments are essential and the postpositions are intrinsically determined by the verbal predicate. Therefore, the verb does not allow the following case patterns regardless of the characteristics of nouns.

*\*N0-Ka N1-E-Ke N2-Lul Pi-Kyo-Ha-Da*

*\*N0-Ka N1-Lul N2-Ka Pi-Kyo-Ha-Da*

It is very subjective to determine whether the argument is essential or optional. Many researchers tend to construct the sentence patterns with only essential arguments, not

optional arguments, since they presuppose that the optional argument is not directly related to the meaning of the verb. In the following case, however, the optional argument must be presented to distinguish N2 of *Chi-Da 5* from *Chi-Da 6*.

*Chi-Da 5: N-Ka N1-Lul N2-E Chi-Da*

*N0=[man]*

*N1=[salt, pepper, vinegar...]*

*N2=[soup, beef, fish...]*

*Meaning: to sprinkle*

*Chi-Da 6: N0-Ka N2-E N-Lul Chi-Da*

*N0=[man]*

*N1=[vinegar]*

*N2=[work]*

*Meaning : to sprinkle*

*Chi-Da 6* is classified as idiomatic expressions because the number of the nouns that can appear in N1 and N2 is limited and the word order is not free like *Chi-Da 5*. The idiomatic expressions are generally not included in dictionary for syntactic analysis because they are considered a few exceptional linguistic phenomena. This study includes them because they are 255 out of 3,005 sentence patterns (i.e. 11.78%) and they can be used as the basic data for semantic analysis.

### 2.1.2. Semantic Information

In the dictionary for syntactic-semantic analysis, semantic information importantly dealt with is the description of meanings of nouns. In the case of describing the meanings of nouns, it is common to use the semantic characteristics in order to get the meaningful clarification and simplicity. It is possible to classify the frame of semantic characteristic by means of hyper-lower concept, association and internal and external meanings. The common method is the classification employing association based upon the concept. The most frequently used one in the thesaurus of concept-base is the method using 'is-a' relation. In the sentences such as 'This is a pen', 'He is a teacher', 'She is a girl', the nouns before 'is a' is viewed as hyper relation, and the nouns after a lower one. For this reason, the nouns are classified hierarchically. That is, 'pen' is the lower concept of 'this' and 'teacher', 'girl' is the lower concept of 'he, she'. It is a relatively simple and clear method and used commonly for its advantage of hierarchy and structure of knowledge. However, the clarification is based on the concept, the semantic commonness shown in the lowest node is weaker than the classification by associations, so that semantically different words could be seen on the same node. The satisfied thesaurus has not been able to come out due to the difficulty of maintaining the equality with hyper concept as we go down to the lower concept in terms of the following - abstract noun, which does not clearly indicate the semantic boundary or each language has a lot of individual words which are included in the same concept, the meaning of the same formative nouns changed by verbs.

On the other hand, the thesaurus by associations easily secures the similarities between words in the same the lowest node for its base of human association, whereas, it has problems in terms of low efficiency because the hyper

and lower concepts have difficulty in forming structuring formation. No matter what kinds of methods can be used, the difficulty is followed in forming thesaurus itself, so that there are many limitations in using thesaurus to eliminate and the ambiguity of verbs, the applied field of thesaurus. The meaning of nouns can describe the specific meanings by the co-relations with verbs and, in that case, semantic characters of plural should be given about one noun. Also, it is difficult to apply each meaning to nouns appropriate for the system of thesaurus and even if it is done, it could be subjective. For one example, noun 'water' can be analyzed as the following four meanings.

Word	Meaning	Example
<i>Mul 1</i>	compound	Ku-Nun Mul-UI Ma-Sin_Da. (He drinks water.)
<i>Mul 2</i>	Color	Os-E Pa-Ran Mul-I Dul-Ess-Da. (The clothe was tinged with blue.)
<i>Mul 3</i>	Influence	Ku A-I-Nun Na-Ppun Mul-I Dul-Ess-Da. (The Child is affected badly.)
<i>Mul 4</i>	Season	Po-Do-Ka I-Jei Kkwu-Mul-I-Da. (The grapes are out of season.)

However, It can be different depending on individuals, whether the 'water' of *Mul 2* and *Mul 3* should be explained by dividing into two meanings or they are used by adding the comparative degree to one meaning.

But if the substitution is applied to the example above in terms of giving the substitution with foreign language, the semantic differences of *Mul 1* to *Mul 4* become obvious. Like the following, 'water' only in the meaning of [A compound of Oxygen and Hydrogen] is corresponded with the substitute words of Japanese *Mi-Jwu* and the words from *Mul 2* to *Mul 4* seems to correspond with other substitute words apiece.

Word	Example	Japanese Translation
<i>Mul 1</i>	Ku-Nun Mul-UI Ma-Sin_Da.	彼が水を飲む
<i>Mul 2</i>	Os-E Pa-Ran Mul-I Dul-Ess-Da.	服が青く染まった
<i>Mul 3</i>	Ku A-I-Nun Na-Ppun Mul-I Dul-Ess-Da.	その子は悪に染まった
<i>Mul 4</i>	Po-Do-Ka I-Jei Kkwu-Mul-I-Da.	葡萄の旬が終わりに近づいている

This means the manner of expressions of physical meanings is different in each language, and becomes the useful tool to understand the meaning of Korean language objectively through the substitute word relation with foreign language, which choose the form as well. In particular, the reason Japanese is selected as the subject in this study is that Japanese has similar syntactic traits with Korean language and well developed semantic classification of varied vocabulary such as characteristic word, Chinese character, borrowed word.

The meaning of verb changes on the noun and semantic characteristics of the noun. Particularly, it is generally considered the verbs represent the meaning of plural in case of characteristic word as well as high frequency word.

*Chi-Da 7: N0-Ka N1-Lul Chi-Da*

*No=human*

*N1=Musical instrument*

*Chi-Da 8: N0-Ka N1-Lul Chi-Da*

*No=human*

*N1=curtain*

In case of N1 of *Chi-Da 8*, there is an occasion in the light of the selection of the substitute words of verbs, even though the nouns have the same semantic characteristics in Korean.

Argument Noun	Substituted Verb	Korean Meaning of Substituted Verb
Cke-Tun	しめる	Nae-Ri-Da (Close)
Byeng-Pwung	めぐらす	Dwu-Ru-Da (Surround)
Bal	張る	Nae-Ri-Da (rope)
Kum-Jwul	張る	Kul-Da (rope)

This kind of phenomenon takes place commonly. When we construct syntactic-semantic dictionary for translation, there is limitation in choosing the substitute words with only semantic characteristic of nouns. And we need to make a list of all nouns with each verb. Each language has different semantic relations between nouns and verbs, and the matching language pattern as a result will be different. Therefore, including all the categories completes the basic requirement.

### 3. Extraction of Case frame from Corpus and MRD

#### 3.1. Extraction of Case frame from Corpus

Among the word Corpus formed by support of Ministry of Culture and Tourism and Ministry of Science and Technology, they were successfully used to extract POS-tagged Corpus made in 1997 and 11,600,000 words were created. The selected Objects are verbs extracted from demonstrative verb and general verb. We extract postposition of subjective case before extracted verb and nouns with postposition of subjective case as well.

The total of 18,609 verbs were extracted. Some of the extracted verbs are ones adhesive to adjective like *Si-Si-Kol-Kol-Ha-Da* and some of the lower frequency verb not listed in the dictionary like *Hui-Bun-Duk-I-Da* comes from word Corpus. At the same time, the verbs not listed in the dictionary appeared by problems such as spelling errors. However, the verbs, which were intentionally used by writer, are not excluded because they are not listed in the dictionary. The total of 1,954,238 case frame were representatively extracted. When we studied the number of verbs, *Ha-Da* (7.11%) was followed by *Iss-Da* (4.65%), *Ebs-Da* (3.51%), and *Doi-Da* (1.96%). In case of

examining the nouns, objective case + *Ha-Da* (1.53%) was followed by subjective case + *iss-Da* (1.13%) and *Su* + subjective case + *Ubs-Da* (0.25%) was the highest when considering the nouns.

The frequency information was eliminated for integration in this article, although it can be usefully employed in the field of several natural language processing. Moreover, there is a case in that the extraction of case and noun information forming Case frame is not possible because the verbs are placed as predicatives in the sentence. These information, which is not extractable can be used characteristic information of the verb, yet the Case frames were removed for integration.

### 3.2. Extraction of Case frame from MRD

It is presumed that the case frame information extracted from a dictionary comes from the exact examples and words used by general users. Words Corps from case frame information reflects the words in the real conversation, and it can be told that by using the frequency information, the basic information has the high frequency. It is needed to integrate the case frame structured by these two ways above. The case frame from a dictionary, which is the exact information from and the case frame from words Corpus, which the real communication has a little different characters. To integrate the two case frames, the work requires to change them into one. We need to integrate the case frame after changing the nouns used as examples into hyper information by taxonomy. This allows the automatic structure of case frame.

We extracted case frame using 'Korean Dictionary' published by the Korean Language Research Society. Korean Dictionary has 402,305 categories including North Korean, old Korean words which are mainly focused on 45,703 intransitives, transitives, intransitive-, transitives. The original dictionary used signs for publication and they are arranged, however, it is necessary to change the formation to be able to deal them through machine. The machine changes the formation and reads it for works. We should analyze the structure of dictionary for formation and the following three steps are required.

The first step is to remove the special character (':', '-') from headword in the original dictionary, and create new headword. These new headword and id number of word become keys of dictionary categories. In addition, we replace swung dash (~) with headword to sustain the intrinsic information before structuring work.

The second step is to add new tag with special symbol in the dictionary. If it is not clear to add tag only with special symbol, we attach tag with heuristic rules. In the examples above, heuristic rule is that when '--' appears in the headword, it indicates the end of headword and the start of Chinese Character and Origin.

The third step can be called Standard General Markup Language considering its construction generally. It is constructed as SGML : ISO 8879 forms.

#### 3.2.1. First Step of Case frame Structure

The first case frame is constructed as several steps. If target verbs in the dictionary have examples, we name parts of speech after analyzing the morphemes of sentential examples. After naming parts of speech, as

what we did in the word corpus, we make the case frame from case articles and nouns with understanding the case article. Verbs with examples were 8275, and 3899 verbs can extract the case. Some verbs in the examples appear together. In the examples of headword verbs, verbs don't appear as a headword, but we extracted case information for only headword verbs to be clear. We used morphological analyzer and added parts of speech to examples by using tagger. We could generally extract the case frame without parser owing to the simple and clear examples in the dictionary. One verb has various meanings, and these verbs belong to ambiguous word. Information about ambiguous words are specified in the dictionary, and can be profitably used in semantic clarification. In this article, we do not consider the semantic information about ambiguous words for integration, viewing all verbs as the same ones.

#### 3.2.2. Second Step of Case frame Structure

This chapter deal with how to integrate the two case frames. After establishing taxonomy from 'Korean Dictionary', we find genus term from the taxonomy and integrate the case frame after replacing them with nouns. In the case of putting the nouns in itself, various examples appear and will be different according to the level of concept of examples. We change nouns shown in the examples into the highest concept and integrate them after unifying the characters of examples.

The characters of each constructed case frame are as follows. Verbs in total in the dictionary are 405,703. 8275 verbs have examples, and 3899 can extract case frame with analyzing the examples. When the case frame appear, and we ignore a part of the nouns, the number of case frame was 7160. If the noun information in the examples is different, it is viewed as a different case frame that are 10493. The number of case frame will be 9713 if we use the examples with clarification information and replace them.

With the same method we could extract measure information about the extracted case frame in the POS-tagged corpus. You will see that the extracted case frame in both sides can be shown as the same form in the dimension of measure information. The information which can be obtained after integrating two case frames consists of the number of verbs with the case frame and of replaced examples. The different information has no meaning because it replaces the examples with the highest information to integrate. After integrating them, the verbs with case frame are 105,661, and the number of case frame transposed the examples is 590,796. Each case frame information is not largely overlapped, but it can be shown correlatively.

## 4. Experiments on Clustering of Verbs

Our experiment is based on proper verbs that are used high frequently and extracted from MRD and Corpus. Unlike general Chinese verbs, the proper verbs have an ambiguity all verbs possess, and it is not easy to express as a sentence pattern to distinguish the meaning of verbs because the distribution of meaning differences is similar to the high concept of nouns according to each verb. It is enough to analyze the appropriateness of syntax knowledge base defined in this article.

( eg ) *Chi-Da*

Pattern	Freq.	Ratio	Pattern	Freq.	Ratio
<i>Chi-Da 1</i>	99	1.00	<i>Chi-Da 17</i>	0	0.00
<i>Chi-Da 2</i>	0	0.00	<i>Chi-Da 18</i>	0	0.00
<i>Chi-Da 3</i>	0	0.00	<i>Chi-Da 19</i>	1	0.01
<i>Chi-Da 4</i>	0	0.00	<i>Chi-Da 20</i>	61	0.62
<i>Chi-Da 5</i>	261	2.63	<i>Chi-Da 21</i>	7	0.07
<i>Chi-Da 6</i>	7	0.07	<i>Chi-Da 22</i>	1223	12.34
<i>Chi-Da 7</i>	0	0.00	<i>Chi-Da 23</i>	1463	14.76
<i>Chi-Da 8</i>	0	0.00	<i>Chi-Da 24</i>	635	6.41
<i>Chi-Da 9</i>	0	0.00	<i>Chi-Da 25</i>	899	9.07
<i>Chi-Da 10</i>	0	0.00	<i>Chi-Da 26</i>	2648	26.72
<i>Chi-Da 11</i>	349	3.52	<i>Chi-Da 27</i>	10	0.10
<i>Chi-Da 12</i>	0	0.00	<i>Chi-Da 28</i>	1259	12.71
<i>Chi-Da 13</i>	504	5.09	<i>Chi-Da 29</i>	152	1.53
<i>Chi-Da 14</i>	0	0.00	<i>Chi-Da 30</i>	326	3.29
<i>Chi-Da 15</i>	0	0.00	<i>Chi-Da 31</i>	0	0.00
<i>Chi-Da 16</i>	5	0.05	<i>Chi-Da 32</i>	0	0.00
TOTAL				9909	100.0

*Chi-Da* has the highest ambiguity in Korean, and the total 44 of sentence pattern is listed with 32 general sentence patterns and 12 idiomatic sentence patterns.

According to the table, *Chi-Da* is the verb which has the most sentence patterns not shown in Corpus. The examples of not shown in the tables are as follows.

*Chi-Da 2 (to hit)*      *Jong-Ul Chi-Da*  
*Chi-Da 3 (to be hurt)*    *Tu-Ruk-I Sa-Ram-Ul Chi-Da*  
*Chi-Da 4 (to grow)*    *Dwai-Ji-Ka Sai-Kki-Lul Chi-Da*  
*Chi-Da 8 (to test)*      *Chul-Su-Ka Si-Hem-Ul Chi-Da*

*Chi-Da* is the verb that has the obvious differences of representative and emergent pattern of a lemma shown in the Corpus. The lack problem of appearance can be solved by describing expressive pattern such as *Chi-Rwu-Da* etc in the real communication with a lemma. Besides, when the similar expression exist like *Jong-I Wul-Ri-Da* (The bell ring.) in *Chi-Da 2*, we will be able to solve the lack problem of appearance by describing it with the category of associated word.

(eg) *Ol-Ri-Da*

In case of *Chi-Da* and *Ol-Ri-Da*, which has the distributive differences in sentence patterns, *Chi-Da* is related with relatively limited noun category while *Ol-Ri-Da* is associated with the even distribution of noun boundary. It becomes clear in the translated word in Japanese. For *Chi-Da*, it has total 13 Japanese verbs like, 'あげる', 'かける', '差し上げる', '申し上げる', '建てる', and etc whereas *Wul-Ri-Da* is corresponded with 49 '入れる', '打つ', '討つ', '生む', '起こす', '切る', 'する', '叩く', '造る', '付ける', etc in total. While the *Chi-Da* is weaker in the main semantic boundary than *Ol-Ri-Da* and has stronger idiomatic expression combined with nouns, *Ol-Ri-Da* often makes general expression and therefore, the corresponding Japanese verbs appears less in the sentence.

Pattern	Freq.	Ratio	Pattern	Freq.	Ratio
<i>Ol-Ri-Da 1</i>	1959	17.80	<i>Ol-Ri-Da 15</i>	355	3.23
<i>Ol-Ri-Da 2</i>	624	5.67	<i>Ol-Ri-Da 16</i>	219	1.99
<i>Ol-Ri-Da 3</i>	537	4.88	<i>Ol-Ri-Da 17</i>	640	5.82
<i>Ol-Ri-Da 4</i>	55	0.50	<i>Ol-Ri-Da 18</i>	0	0.00
<i>Ol-Ri-Da 5</i>	596	5.42	<i>Ol-Ri-Da 19</i>	325	2.95
<i>Ol-Ri-Da 6</i>	584	5.31	<i>Ol-Ri-Da 20</i>	35	0.32
<i>Ol-Ri-Da 7</i>	546	4.96	<i>Ol-Ri-Da 21</i>	108	0.98
<i>Ol-Ri-Da 8</i>	717	6.51	<i>Ol-Ri-Da 22</i>	240	2.18
<i>Ol-Ri-Da 9</i>	312	2.83	<i>Ol-Ri-Da 23</i>	66	0.60
<i>Ol-Ri-Da 10</i>	711	6.46	<i>Ol-Ri-Da 24</i>	2	0.02
<i>Ol-Ri-Da 11</i>	28	0.25	<i>Ol-Ri-Da 25</i>	392	3.56
<i>Ol-Ri-Da 12</i>	304	2.76	<i>Ol-Ri-Da 26</i>	542	4.92
<i>Ol-Ri-Da 13</i>	0	0.00	<i>Ol-Ri-Da 27</i>	280	2.54
<i>Ol-Ri-Da 14</i>	25	0.23	<i>Ol-Ri-Da 28</i>	804	7.31
TOTAL				11006	100.0

## 5. Conclusions

This paper explains building procedures for knowledge base of verb with Korean and Japanese. This article forms case frame automatically by MRD and KAIST corpus. The information classification with hyper concept is done first by MRD as a means of automatic construction. After analyzing the morphemes of Corpus and examples in the dictionary, each example shifted into the highest concept by information classification. Then, the automatic construction of case frame is done by integrating information from the examples in the dictionary and the KAIST corpus. After that, we achieved the information base of syntactic knowledge with additional information by hand. The construction of syntactic knowledge base is scheduled to extend to verbs in Korean, noun meaning as a dictionary respectively. Noun meaning as a dictionary is the thesaurus concerning semantic system of nouns considering the usage of verb. The corpus, electric dictionary, and circulated construction system will have structure making up for the problems of automation and manual work. Because *Chi-Da* has high possibility of different emergent frequency by listed categories and contents of corpus, it is very hard to describe the important information to the sentence patterns.

## 6. References

- [Choi 1999] Yong-Suk Choi, Ju-Ho Lee, Key-Sun Choi, "Study on Automatic Construction and Evaluation method of Case frame", Proceedings of Hangul and Korean Information Processing, 1999.
- [Song 1999] Young-Bin Song, Key-Sun Choi, and etc, "Dictionary Making for Disambiguation", Proceedings of Hangul and Korean Information Processing, 1999.
- [Cho 1999] Pyeong-Ok Choi, Cheol-Yeong Ock, Soo-Dong Lee, Jae-Duk Park and Dong-In Park, "A Korean Noun Semantic Hierarch Based on Semantic Feature", International Conference on Computer Processing of Oriental Languages, pp. 211-216, Tokushima Japan, March 24-26, 1999
- [Li 1998] Hang Li and Naoki Abe, "Generalizing Case Frames Using a Thesaurus and the MDL Principle", Computational Linguistics, Volume 24, Number 2, pp. 217-244, June 1998