Language, Information and Computation (PACLIC12), 18-20 Feb, 1998, 163-171

Using Case Prototypicality as a Semantic Primitive

Dan-Hee Yang, Ik-Hwan Lee**, Mansuk Song**

Yonsei University, Seoul, Korea

A lexicon is essential to provide a proper semantic analysis of linguistic expressions. Thus, much work has been done to build one automatically. However, the results were not enough to define word meaning, but no more than a kind of thesaurus or taxonomy. In this paper, we present a new method that is not only more practical and efficient for implementing on computers, but also more reasonable in view of psycholinguistics. We first note that componential analysis is not real in our cognitive system. We define word meaning by a set of Case prototypicalities. We show that this knowledge can be constructed by supervised machine learning on the basis of Case particles and the collocational information from a large corpus.

1. INTRODUCTION

The goal of computational linguistics is to create computational models of language in enough detail so that we can write computer programs to perform various tasks involving natural language, using the notions of algorithms and data structures from computer science. The ultimate goal is to specify models that are similar to human performance in the linguistics tasks of reading, writing, hearing, and speaking (Allen 1995). To build such a computational model, many researches of computational linguists or computer scientists in the area of natural language processing (NLP) have taken plausible theories of language-related disciplines such as linguistics, psycholinguistics, the philosophy of language, and cognitive science. Especially in the semantic analysis of NLP, the meaning of words has been represented, mainly based on the semantic feature hypothesis, by the result of componential analysis, and the process of semantic analysis was substantially dependent on the information of argument structure and selectional restrictions under the thematic role theory.

To make practical working systems in accordance with such theories, we need a lexicon of componential analysis of all words, argument structures that each verb requires, and selectional restrictions which noun phrases should satisfy to meet the required thematic roles. There are various methods of the semantic lexical representation. However, most of them include a taxonomic classification of concepts or words. One of critical bottleneck problems is how to construct a complete and reasonable taxonomy for NLP. Constructing this knowledge manually is in the domain of the linguists' work. On the other hand, computer scientists or computational linguists are responsible for its automatic construction. There have been many approaches to the automatic extraction of linguistic knowledge from large corpora and machine-readable dictionaries in NLP.

This study focuses on automatically building a lexicon among the three types of linguistic knowledge and present a new method that is not only more practical and efficient for implementing on computers, but also more reasonable in view of psycholinguistics. First of all, we take a careful look at the problems of existing linguistic theories, and synthesize some of them in a way different from those adopted in the theories of almost all NLP. Then, we extract triples [noun, Case particle, verb] from a large corpus, and we assign Cases to a word by supervised machine learning on the basis of Case particles and the collocational information. Finally, we define the word meaning by a set of Case prototypicalities. In the following we will look at linguistic theories related to the word meaning: namely, lexical field, componential analysis, lexical decomposition, and acquisition of meaning. In addition, we will see that Case particles have a very important role in the acquisition of word meaning in computers.

* Department of Computer Science, ** Department of English, Yonsei University, Seoul 120-749, Korea.
E-mail: {dhyang, mssong}@december.yonsei.ac.kr, ihlee@bubble.yonsei.ac.kr

2. LINGUISTIC THEORIES ABOUT WORD MEANING

Most extensive works on word meaning have been carried on by generative semanticists such as Lakoff (1965), McCawley (1968), and Postal (1970). Although they assume that primitive semantic elements, atoms from which word meaning are composed, represent some kinds of mental constructs or concepts, most of them fail to note how these concepts connect with what we talk about, their role in articulating the objective significance of language (Chierchia 1990). Let us now look at the lexical field theory underlying the theories of word meaning.

2.1. Structure of Lexical Field

J. Trier (1934) and W. Porzig (1934) explain the structure of lexical field by the constitutional relation of words (Lee 1995). As we see in Figure 1, Trier makes an attempt to establish the semantic relation of words on the basis of the paradigmatic relation. In other words, lexical field is represented by a hierarchical tree diagram. This is a background theory for the componential analysis of modern lexical semantics. In NLP, the traditional semantic nets, one of the knowledge representations used by many researchers, follow this paradigm.

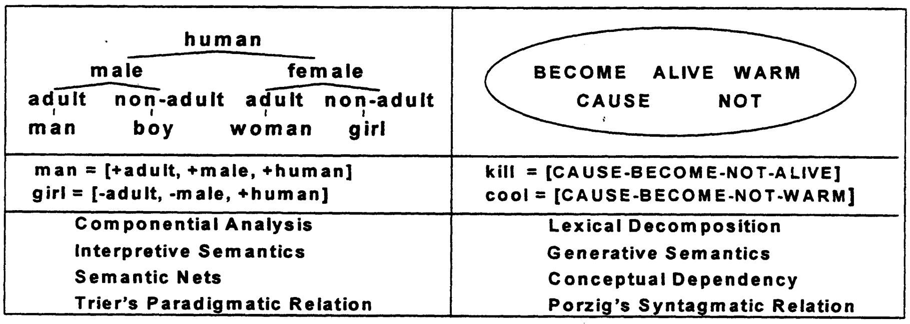

Figure 1. The representation of the meaning of words

On the other hand, Porzig makes an effort to reveal the sense relation among words on the basis of the syntagmatic relation. He introduces the concept of encapsulation. For example (Lee 1995), English 'kick' and 'punch' are translated into French as follows:

a. **kick**: *donner un coup de* ***pied*** ('to strike with the **foot**')
b. **punch**: *dooner un coup de* ***poing*** ('to strike with the **fist**')

In this case, between 'kick' and 'foot' and 'punch' and 'fist', respectively, there is an essential semantic relation, which has a collocational characteristic and lies at the root of the syntagmatic relation. Accordingly, we regard this relation as a fundamental sense relation between words. As far as the above example is concerned, he asserts that the sense of 'with the foot' has been encapsulated in the single term 'kick'. This holds for the relation between 'with the fist' and 'punch'. We now consider the question: how can we pick up the meaning of words?

2.2. Acquisition of Meaning

Eve Clark's (1973) semantic feature hypothesis is based upon a definitional view of word meaning. In other words, the meaning of a word consists of a set of necessary and invariant semantic features. Children acquire the meanings of words within her theory by first acquiring the more general superordinate features. Acquisition then goes from the more general to the more specific. The first features which are acquired are those which are perceptually salient to the child. The most primitive categories involve movement, shape, size, sound, taste, and texture (Ingram 1989). On the other hand, Nelson (1974) insists, by the functional core concept theory, that we acquire meaning as we recognize the functions of objects by some other non-linguistic means (Park 1996). The major difference is that Nelson emphasizes the role of functional semantic features such as *ROLL, SPATTER, MOVE,* etc when children acquire the meaning of words.

Rosch et al's (1973) prototype theory is an approach developed to account for the representation of meaning in adult language. The proposal is that the meaning of words is not a set of invariant features, but rather a set of features which capture family resemblance. Some objects will be most typical of the word's meaning by sharing more of the word's features than others. Certain features, then, will be more important in determining class membership than others, but none are required by all members (Ingram 1989, Taylor 1995). Bowerman (1978) suggests that children use both perceptual and functional features. One type of categorization does not necessarily replace the others over time; that is, the different kinds of classification can be used simultaneously. Therefore, one type of categorization is not necessarily more primitive than another. From these claims, Bowerman concludes that the representation of meaning as features and prototypes needs to be incorporated into a single model (Ingram 1989, Tayler 1995, Park 1996). These theories explain how humans pick up the word meaning.

3. PROBLEMS OF PREVIOUS APPROACHES

Disambiguating word senses and grasping Cases are the main task of semantic analysis in NLP. To achieve this task, much work has been done on automatically acquiring linguistic knowledge such as verb patterns, thesauri and the selectional restrictions with semantic features (Velardi 1991, Li 1996, Resnik 1994, Armstrong 1993, Hindle 1990, Pedersen 1995). For the representation of word meaning, Velardi (1991) classifies four types of semantics: conceptual semantics, surface semantics, technical semantics, and naive semantics. He suggests that surface semantics be more adequate for the purpose of an extensive codification on computers because it can be induced from word cooccurrences in texts. However, it seems that the representation is similar to that of result of semantic analysis of a sentence according to Chomsky's ST-model with a view of interpretive semantics in NLP. So there seems to arise a problem in that building the knowledge requires considerable levels of semantic analyses.

One of the most influential analyses, which conforms to the cognitive reality, was Schank's *conceptual dependence* (CD) which attempt to represent the meaning of action verbs with 11 primitives such as *ATRANS, MOVE,* and *PROPEL*. In his later work, primitives are used as building blocks to construct larger structures (e.g., SCRIPT, MOP, TOP) in order to capture the meaning of verbs (Schank 1977, 1982). His idea and methodology are very systematic and have had a considerable influence on the field of NLP. However, because he follows generative semantics and the lexical decomposition paradigm, he is open to censure due to lack of extensibility or scalability. Pedersen (1995) attempts a system that automatically acquires the meaning of unknown nouns and verbs from the corpora. He follows P. Kay's view (1971) that human lexicons are largely organized as taxonomies of concepts. The acquisition of meaning is defined as locating an existing concept node in a concept hierarchy that defines an unknown word. If there is no node of such a concept, then a node is created and placed into the concept hierarchy. However, the essential of this process is just to construct hierarchical taxonomies.

All existing methods (Hindle 1990, Pereira 1993, Grefenstette 1993) for word classification take no account of the thematic role or Case, but uses only a collocational relation in the form of [Subject Verb Object]. However, Jeong (1993) asserts that just by a statistical processing of the collocational information or mutual information, we can know that the words which belong to the same category are related with each other in some way, but never know what relationship there is among them. If so, what is the inherent flaw of the existing works? In the following we will try to answer this question, basing our account on the cognitive reality.

Table 1. Roles and actions represented in one-word utterance

ROLL OR ACTION	UTTERANCE	CONTEXT
Agent	Dada	Hears *someone* come in
Action or resulting from action	Down	When sits *down* or steps *down* from somewhere
Object affected by action	Ban	When wants *fan* turned off
State of object affected by action	Down	When *shuts* cabinet door
Object associated with another object or location	Poo	With hand on *bottom* after being changed, usually after bowel movement
Possessor	Lara	On seeing *Lauren's* empty bed
Location	Bap	Indicating location of feces on *diaper*

3.1. Cognitive Reality

In general, the cognitive reality is not a major issue to computer scientists or computational linguists. Nevertheless we lay a great stress on it because human language is the highly sophisticated intelligent system that is fundamentally different from animal cries and calls. Furthermore, human language is the essence of human intelligence in itself. Considering such a characteristic, we think that imitation is the best approach for establishing a computational model of human language. Currently, most, if not all, representations of word meaning are based on the Clark-like semantic features. However, the resulting knowledge might not conform to the cognitive reality in that humans, even linguists, cannot naturally enumerate the semantic features of a word, although we have no doubt that they have an obvious concept of the word and that they are able to speak fluently.

Piaget suggests that the ability to represent objects and events is a necessary prerequisite for the acquisition of any system of symbolic representation for knowledge and experience (Piaget 1951). Children use words not just to name objects, but to pick out the roles those objects play in whatever event being described. The frequency of naming some objects suggests that some roles may be more salient than others (Nelson 1974).

Nelson (1974) reports that children appear to name mostly movers (e.g., people, vehicles, animals) and movables (e.g., food, clothing, toys), with a few recipients (people). Greenfield and Smith (1976), in fact, look at the order of acquisition of these different roles. They find that the two children they observed both began by naming movers or agents, then movables or objects affected by an action, and finally places or locations, and possessors or recipients. Table 1 lists the different roles and the actions or states that the child talked about in the order from top to bottom that they emerged in his speech. The roles children pick out appear to be precursors to the semantic roles expressed in the adult utterances (Clark H. and Clark E. 1977). When we ask someone what an atomic bomb is, he may answer: *It can kill tens of thousand people. It can destroy a city completely. It forced Japan to yield in the second world war.* He is not expected to say something like its structure, its color, its hierarchical analysis into the semantic features. Nevertheless, he is willing to think that he knows it and he can use the word well without any problem in a real life. The next question that we take into account is whether building a lexicon by computers is feasible as expected.

3.2. Building a Lexicon

In practice, the componential analysis and lexical decomposition are problematic whether or not they are done manually or automatically. Before everything, it is extremely difficult, if not perhaps impossible in principle to find a suitable, linguistically universal collection of semantic primitives in which all words can be decomposed into their necessary properties. Even simple words whose meanings seem straightforward are extremely difficult to characterize (Hirst 1987). It may be nonsense if we believe that we can make computers execute such a task well with a current technology. The task requires not only all kinds of knowledge in the world but also the highest intelligence. There is an excessively long distance from the four fundamental rules of arithmetic.

The prime example of scientific taxonomies is the classification of plants and animals based on the proposals made by the Swedish botanist Linnaeus. *Thesaurus of English Words and Phrases* by Mark Roget is representative of the linguistic classification. It took many years for experts to classify them. To make matters worse, a single classification constructed manually or automatically seems least likely to satisfy the ultimate need of NLP because there need to apply a variety of criteria for classifications depending on contexts (Leem 1993). In addition, existing automatic methods have left much to be needed to define word meaning from a corpus.

According to what mentioned so far, it is very important to represent the word meaning in the way how computers can construct them without difficulty. But since we cannot grasp the entity of any relationship just by statistical processing, we assert that supervised machine learning process is indispensable. Furthermore, if the representation is cognitively plausible, it adds luster to what is already brilliant. In the following, we insist that Cases be the very key to this problem.

4. CASE PROTOTYPICALITY AS A SEMANTIC PRIMITIVE

Humans seem to have inconsistent, multi-partial knowledge about words. Furthermore, the knowledge is likely to be as direct as requiring little inference for surface understanding of sentences. Humans instantly grasp the meaning of a word by its thematic role or Case without componential analysis. Only in case that ambiguity occurs, they try to analyze the meaning more deeply. Therefore we should be free ourselves not only from the hierarchical taxonomy, but also from the procedural representation of knowledge bearing inference in mind. Accordingly, it is necessary to have a non-hierarchical and non-procedural representation for word meaning.

A corpus may be regarded as a linguistic model of a real world. Besides, the Case of a word is determined within a sentence. However, we have the previous knowledge about Cases which helps us to understand various situations where the word may be used. If we did not have such knowledge, we can never communicate with each other. If we do not know about X, or Cases which X may be used in a various situation, we cannot use X except for '*What is X?*'. However, we are able to infer the Cases through looking at the usage in various contexts. This is the very acquisition process of the meaning of words. Naturally, Nelson's functional core concept theory seems to be more reasonable and cognitively real.

As we noted earlier, we cannot express the hierarchical knowledge naturally and that it is extremely difficult to build. Since Porzig's field theory is based on the syntagmatic relation, we do not suffer from a hierarchical structure. Hereupon, we take the notion of encapsulation to be represented by Case. Humans can describe a prototype of an object well, but they cannot say the lexicographical definition of an object well. We consider such facts support the prototype theory well. As a result, Case in itself becomes the semantic feature or semantic primitive. We do not insist that Case be the smallest unit of the linguistic meaning, but we think it is the smallest unit in a high level of cognitive process. Case is a highly abstract representation of word meaning, which has a linguistic universality. We do not consider that Case determination is a final phase of semantic analysis. It goes without saying that an in-depth understanding should require a further deep analysis on other representations.

4.1. Representation of Word Meaning

Focusing on the paradigm of Porzig, Nelson, and Rosch et al., we establish three hypotheses as follow:

- Humans grasp the meaning of a word as its Case or role in context at a surface level or intuitive level.
- Words are not represented by the binary semantic features, but by the probabilistic or fuzzy one.
- There is no strict boundary of category, but the membership is the degree of similarity to the prototype (Taylor 1995).

Table 2. Representation of word meaning

◆ **Meaning representation of noun**
$\{ (c_n, z) \mid n\ p\ v \}$ for all p, v, where c_n = Case which may be assigned to the noun, n is a noun, p is a Case particle, v is a verb, and z is a prototypicality[1]
◆ **Meaning representation of verb**
$\{ (c_v, z) \mid n\ p\ v \}$ for all n, p, where c_v is a Case which the verb may require.

As in Table 2, we define the meaning of nouns and verbs in terms of a set of Case prototypicalities. The merits of this definition are that since both word meaning and selectional restrictions are represented by Case prototypicalities in the same way, the process of grasping Cases of a word within a sentence reduces to a remarkably plain mechanical task. Consequently, the algorithms for semantic analysis get to be greatly straightforward. We hope to emphasize that the process of acquiring a set of Case prototypicalities has a role of componential analysis.

To acquire this knowledge, we extract triples [noun, Case particle, verb] from a large corpus and automatically assign Cases to a word by applying supervised machine learning algorithms. Then, regarding Cases as semantic primitives, we represent the meaning of a word in a corpus in terms of a set of Case prototypicalities. More strictly speaking, we find a set of Cases which each verb can require and each noun is used for, in company with its prototypicality. Now, how can we implement this idea on computers? The clue to the solution of the problem lies in Case particles and the collocational information. Let's take them into consideration in the following.

Table 3. N : M relationship between Case particle and deep Case

CASE PARTICLE	DEEP CASES FOR WHICH IT CAN BE USED
-lo 'to, as, into, for, toward, of, from, etc.'	***ORIENTATION, PATH, GOAL, ATTRIBUTE, QUALIFICATION, PURPOSE, INSTRUMENT, MATERIAL, CAUSE***
-ey 'at, to, for, by, etc.'	***LOCATION, PATIENT, CAUSE, INSTRUMENT, AGENT, BENEFACTIVE,***

4.2. Acquisition of Word Meaning

4.2.1. Case Particles

Case particles follow a noun and determine its role in a sentence. They are similar to prepositions in English. So they are sometimes called postpositions. Their main usage is to manifest the grammatical relations of words within a sentence. Case particles are classified into seven major types. Nominative particle follows a subject. Objective particle follows an object, and the like. Adverbial particles are used in a variety of ways depending upon the preceding nouns and predicates. There are about 30 adverbial particles and two particles may be used as a combined one.

As we see in Table 3, a Case particle in Korean manifests surface Case in the N : M relationship so that one Case

[1] the degree that the word is close to an exemplar of the Case.

particle can be used to represent several (deep) Cases[2] and vice versa. This phenomenon occurs in Japanese and English in the similar way. 'に' and 'to' are used for denoting ***ORIENTATION*** in '北に ある', 'lies *to* the north' respectively; ***GOAL*** in 'ロンドンに 歩いて行く', 'walked to London', and so forth. Thus the order of words in a sentence in Korean is not important except peculiar situations. Figure 2 shows the two parts of information we are interested in for determining Case.

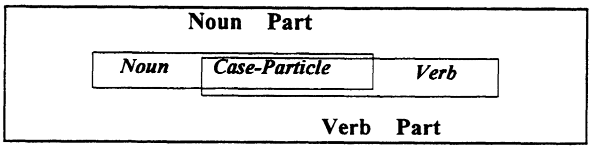

Figure 2. The information for determining Case

We classify Korean Case particles into three types according to the Case determinability of the above information as in Table 4 (Nam 1993).

Table 4. The classification of Case determinability

CLASS	DEFINITION AND EXAMPLE
1	Only the verb part determines its Case regardless of the noun part. (e.g., *-lo sayngkakhata* 'think as', *-lo alhta* 'sick with', *-lo sungcinhata* 'be promoted to', *-lo cawukhata* 'be filled with', *-lo kwusengtoyta* 'consist of', *-lo caluta* 'cut with')
2	The noun part selects one of possible meanings (Cases) of the verb if only the verb part cannot do. (e.g., *mwul-lo salta* 'live on water', *mwul-lo pyenhata* 'be turned into water')
3	Other part does. (e.g., *ceyil hankangkyo-lo kata* 'go to the First Han River Bridge' or 'go by way of the First Han River Bridge' is possible)

As a result, we can grasp Case without considering the semantic relationship between a noun and a verb in case of Class 1 as referring to Case particle.

4.2.2. Building the Training Data

The observation of a corpus shows that there is a proper number of words belonging to Class 1. This information considerably reduces the efforts for preparing a training data. The key point of our approach is to collect the verbs that belong to Class 1 from a corpus as many as possible. First, we extract triples [noun, Case particle, verb] from a corpus. Second, we carefully select verbs which are considered to take a unique Case under Class 1. This step seriously influences the accuracy of the machine learning described in the following section. Finally, we manually assign Cases to them considering the given Case particle. The result is called a training set TSET, which consists of [noun, Case particle, verb, Case]. Note that we will restrict our consideration to the instance that Case particle is '*-lo*' from now on.

4.2.3. Machine Learning Algorithm

If we let *c* be a Case, a verb is represented by [{ (n, f) | *n* is a noun which co-occur with the verb, *f* is the relative frequency of the noun in TSET }, *c*], where *f* is calculated as dividing the frequency which the noun co-occurs with the verb by the total frequency of the noun in the TSET. The set of such verbs is called Instance Set (ISET) which is used as an input of this algorithm. Thus the nouns assume the role of attributes of verbs in learning the concepts of each Case. The other input is a set of hypothesized linear threshold units (HSET), where linear threshold unit (LTU) is an intensional concept representation of a Case. In the beginning, the HSET is set to an empty set initially. The goal of this algorithm is to get the relevant HSET as an output.

Much of the work on threshold concepts has been done within the 'connectionist' or 'neural network' paradigm, which typically uses a network notation to describe acquired knowledge, based on an analogy to structures in the brain. To make the concept of LTU clear in this study, it can be stated as

If $\sum w_i f_i > \beta$ then c_k, where $c_k \in$ { ***GOAL, ATTRIBUTE, INSTRUMENT, ...*** }, w_i is a weight, f_i is the attribute's value of a noun n_i, and β is a threshold.

[2] When we use 'Case', this means deep Case except for 'Case particle'.

To classify a verb, one multiplies each observed attribute's value by its weight, sums the products, and sees if the result exceeds the given threshold. In principle, an arbitrary LTU can characterize any extensional definition that can be separated by a single hyperplane drawn through the instance space, with the weights specifying the orientation of the hyperplane and the threshold giving its location along a perpendicular. For this reason, target concepts that can be represented by linear units are often referred to as being linearly separable (Langley 1996). Accordingly, since the HSET is able to function as a classifier, it can decide to which Case concepts a new verb belongs.

The Case Learner (CL) algorithm in Table 5 is based on the *perceptron revision method* (PRM) (Langley 1996) which is an incremental approach to inducing LTUs using the gradient descent search. However, we modify it slightly to prevent a phenomenon of oscillation. In other words, we reflect a proportion (α) of a previous delta ($\Delta w_i(h-1)$) into a current weight (w_i). Also, we use the *perceptron convergence procedure* (PCP) (Langley 1996) in Table 6 which induces LTUs nonincrementally by applying CL algorithm iteratively to the ISET until it produces an HSET that makes no errors or until it exceeds a specified number of iterations. This algorithm guarantees to converge in a finite number of iterations on LTUs that make no errors on these training data.

Table 5. Case Learner (CL): learning the concepts of each Case

```
Inputs
  ISET: v_t = [{ (n, f) | n is a noun, f is the relative frequency
  of the noun }, c], where v_t is a verb and c is a Case, 1 ≤ t ≤
  the total number of verbs in the TSET.

  HSET: a set of LTUs which are intensional concept repre-
  sentations of each Case
Output
  HSET: a revised one
Parameters
  α : a momentum term that reduces oscillation
  η : a gain term that determines revision rate.
  cf. H: If Σ w_i f_i > β then c_k.   H ∈ HSET
      Δw_i(h): a current delta of weight
      Δw_i(h-1): a previous delta of weight

Procedure CL (ISET, HSET: input; HSET: output)
{
 for each training instance v_t in ISET
 {
    C = c of v_t;
    for each H_j in HSET
    {
       C_k = Case which H_j predicts for v_t;
       if C_k is same as C
       then continue;
       if C_k is negative and C is positive
       then s = 1
       else if C_k is positive and C is negative
            then s = -1;
       for each attribute n_i of v_t
       {
          f_i = the attribute's value of n_i;
          if n_i exists in H_j
          then w_i = the weight for n_i in H_j
          else w_i = f_i;
          Δw_i(h) = sηf_i + αΔw_i(h-1);
          w_i = w_i + Δw_i(h);
       }
       β = β + sη;
    }
 }
 return the revised HSET;
}
```

4.2.4. Defining Word Meaning

The prototypicality of a verb for each Case is calculated by the left part of LTUs or $\sum w_i f_i$ obtained by the learning process and is scaled down between -1 and 1. The prototypicality of a noun for each Case is $w_k / \sum abs(w_i)$ of the corresponding $\sum w_i f_i$. w_k is the weight of the noun, which we regard as prototypicality since it represents the discretability of classifying Cases. Finally, we define the meaning of a word by using the assigned Cases and prototypicalities. Table 4 shows the formal representation or definition of a word that belongs to noun or verb category.

Table 6. PCP for assigning Cases

```
Input, Output
  cf. CL algorithm

Procedure PCP (ISET: input; HSET: output)
{
 HSET = empty sets;
 count = the maximum number of iterations;
 while count > 0
 {
   for each instance v_t in ISET
   {
     for each H_j in HSET
       if H_j incorrectly predicts for v_t
       then go to ERROR;
   }
   return HSET;
   ERROR:
     count = count - 1;
     CL (ISET, HSET);
 }
}
```

4.2.5. Experimental Results

We experimented with the Case particle *-lo* 'to, towards, as, into, for, of, from, with, etc.' which is one of the most complex particles in Korean. To simplify this experiment, we classify three Cases: GOAL CASE for *ORIENTATION, PATH, GOAL* and INSTRUMENT CASE for *INSTRUMENT, MATERIAL, CAUSE* and ATTRIBUTE CASE for *QUALIFICATION, ATTRIBUTE, PURPOSE*. By choosing 30 training verbs considering the frequencies of verbs in YSC IX Corpus[3], the training data lead to about 1,000 triples. Table 7 is a part of the results of this experiment. The accuracy of this results is about 87 percent. According to the word definitions in Table 2, a noun *kang* 'river' is represented by { (GOAL, 0.327), (INSTRUMENT, 0.001), (ATTRIBUTE, -0.173) } and a verb *kata* 'go', by { (GOAL, 0.783), (INSTRUMENT, 0.377), (ATTRIBUTE, 0.183) }.

Table 7. Cases prototypicalities of nouns and verbs

Words \ Cases	GOAL CASE	INSTRUMENT CASE	ATTRIBUTE CASE
kang 'river'	0.327	0.001	-0.173
kil 'road'	0.281	0.130	0.003
yenphil 'pencil'	-0.001	0.190	0.226
pep 'law'	0.041	0.319	0.132
kata 'go'	0.7833	0.377	0.183
sayonghata 'use'	-0.347	0.111	0.121
tayhwahata 'talk'	0.000	0.024	0.000
mantulta 'make'	0.001	0.667	0.476

5. CONCLUSION AND FUTURE WORK

This study demonstrates that words can be defined by a set of Case prototypicalities by using the syntactic relations among words considering the characteristics of Case particle. Furthermore, by the machine learning mechanism, the definition of a word can be automatically induced from a corpus. This result is taken to be the direct knowledge about a word for the syntactic and semantic analyses in NLP. Especially, we emphasize that our approach is not only more practical and efficient for implementing on computers, but also more reasonable in view of psycholinguistics.

To improve our experiment, we have got something more to do. First, we should classify, in more detail, the types of Cases which each Case particle may take. For this, the meaning of each Case should be defined more strictly. Second, we should think out how to discriminate necessary arguments, optional arguments, and adverbials by computers. Third, we should improve the algorithms for automatically assigning Cases to words.

[3] Yonsei Corpus IX: 1.2 million words extracted from books for children, built by Korean Lexicographical Center of Yonsei University.

Fourth, more plausible normalization method is required to adjust prototypicalities calculated by LTUs for each Case and by weights of each noun. Finally, to overcome a difficulty of building the large training data, which is an intrinsic problem of the machine learning, we should try to develop more sophisticated techniques, incorporating both the supervised and unsupervised learning strategies.

ACKNOWLEDGEMENTS

This research was funded by the Ministry of Information and Communication of Korea under contract AB-97-B-0182.

REFERENCES

1. Allen, James (1995). *Natural Language Processing*, 1-2. Redwood: Benjamin/Cummings.
2. Chierchia, Gennaro and Sally McConnell-Ginet (1990). *Meaning and Grammar*, 349-360. Cambridge, MA: MIT Press.
3. Clark, Herbert H.; and Eve V. Clark (1977). *Psychology and Language*. New York: Harcourt Brace Jovanovich.
4. Grefenstette, G. (1993). Automatic Thesaurus Generation from Raw Text Using Knowledge Poor Technique. In *9th Annual Conf. of the UW Centre for the New OED and Text Research.*
5. Hindle, Donald (1990). Noun Classification from Predicate-Argument Structures. In *Proc. of the 28st Meeting of the Association for Computational Linguistics (ACL-90)*.
6. Hirst, Graeme (1987). *Semantic Interpretation and the Resolution of Ambiguity*, 28-29. New York: Cambridge U. Press.
7. Ingram, David (1989). *First Language Acquisition: Method, Description and Explanation,* 398-432. Cambridge U. Press.
8. Jeong, Young-Me (1993). *An introduction to Information Retrieval*, 196-202. Seoul: Gume Trade.
9. Langley, Pat (1996). *Elements of Machine Learning,* 67-94. San Francisco: Morgan Kaufmann.
10. Lee, Ik-Hwan (1995). *An Introduction to Semantics*, 58-225. Seoul; Hanshin.
11. Li, Hang (1996). Clustering Words with the MDL Principle, *cmp-lg/9605014*.
12. Lim, Hong-Bin and Jae-Young Han (1993). *A Study on Classification of Korean Vocabulary*, 1-17. Seoul: National Institute of the Korean Language.
13. Nam, Ki-Shim (1993). *The Usage of Korean Particle*. Seoul: Seogwang Academic Press.
14. Park, E-Do (1996). *The Acquisition of Mother Tongue and Learning of Foreign Language*, 75-115. Seoul: Hanguk Munhwa.
15. Pedersen, Ted (1995). Automatic Acquisition of Noun and Verb Meanings. *Technical Report 95-CSE-10*. Department of Computer Science and Engineering of Southern Methodist U.
16. Pereira, F. ; N. Tishby; L. Lee (1993). Distributed Clustering of English Words. In *Proc. of ACL 93*.
17. Piaget, J. (1951). *Play, Dreams, and Imitation in Childhood.* New York: W.W. Norton.
18. Resnik, Philp Stuart (1994). *Selection and Information: A Class-Based Approach to Lexical Relationships*. Ph.D. Dissertation, Pennsylvania U.
19. Schank, Roger (1982). *Dynamic Memory*, 77-123. New York: Cambridge U. Press.
20. Schank, Roger and Robert Abelson (1977). *Scripts Plans Goals and Understanding*, 11-17. New Jersey: Lawrence Erlbaum.
21. Taylor, John R (1995). *Linguistic Categorization: Prototypes in Linguistic Theory*. Oxford U. Press.
22. Velardi, Paola (1991). Acquiring a Semantic Lexicon for Natural Language Processing, 343-349. In Uri Zernik, eds., *Lexical Acquisition*. Lawrence Erlbaum.
23. Yang, Dan-Hee; Ik-Hwan Lee; and Mansuk Song (1997). Automatically Defining the Meaning of Words by Cases. In *Proc. of the International Conf. on Cognitive Science '97 (ICCS-97)*.
24. Yang, Dan-Hee; Seong-Hyeon Yang; Young-Sin Lee; and Mansuk Song (1997). The Definition and Representation of Word Meanings Suitable for Natural Language Processing. In *Proc. of SOFT EXPO '97 CONFERENCE.*