A Computational Approach to the Representation of 'Prototypical' and 'Extended' Properties

Chang-in Lee
Department of Linguistics
Rice University
Houston, Texas

This paper proposes a way of representing some verbs which have combinations of 'extended' participants. The computational representation of these combinations is approached with a lexicographic indication described in Webster's Dictionary, Zadeh's (1987) 'fuzzy set' and a metaphorical pointer.

I. The Current Dynamic System

The current system (Lee 1987, 1988) includes three phases (morphological, syntactic and semantic) in comprehending a given input sentence. The processes for morphological, syntactic and semantic decoding are performed by an algorithm which searches through dictionary-like structures set up between each pair of levels (Lamb 1985, 1987). That is to say, the system recognizes the essential structural differences between the linguistic information and the computer procedure. Thus, the linguistic information is not incorporated into the
Fig. 1. (The System Architecture)
Data Flow ⇒
Dictionary Interaction →
program itself. This system architecture is shown in Figure 1. The given
English or Korean sentence first flows into the morphological decoder,
and morphological analysis (i.e. classification of roots and suffixes) is
completed with the aid of Morphemic Supplements I (and II). The
result of the morphological analysis is then passed into the next phase,
a syntactic decoder which produces (a) syntactically parsed tree(s)
according to Lexemic Supplements I and II. Finally, the syntactically
parsed tree goes up to the semantic decoder which identifies the orga-
nization of a proposition: Process, P-1, P-2, Tense, Aspect, etc. In
order to represent concepts of a proposition, a provisional conceptual
representation is checked to see if the connection between a process
and its participants is conceptually well-formed. A proposition which
is conceptually ill-formed (e.g. %I ate the table) is rejected after com-
parison of the semantic properties of participants with those that a
process presupposes.

II. Extension of the Current System

Based on the theoretical assumption that each process has its own
natic properties (Lamb 1987, Ikekami 1980, Hayes 1978), the current
system checks only 'normal' or 'prototypical' arguments that a verb
would take as participants. For example, the system accepts a P-1
which is 'Liquid' for the English verb 'flow' since the verb normally
takes a P-1 with the property 'Liquid.' The participant, however, can
also have the property 'Abstract' (e.g. Ideas are flowing, Conversation
flows smoothly, Wealth flows from the industry to economy, etc), or the
property 'Human Being' (e.g. The crowds flow), or the property 'Hair'
(e.g. Her hair flows over her shoulder). To take an example from
Korean, the system accepts a P-1 which is 'Human Being' or 'Animal' and a P-2 which is Edible-Solid for the verb mek-'to eat.' These normal combinations come from the general semotactics of the verb mek-'to eat'; The process mek- presupposes a participant that has the property 'Animal' or 'Human Being' as the P-1 and a participant that has the property 'Edible-Solid' as the P-2. We can, however, find some sentences that have the property 'Abstract' as the P-2 (e.g. na-nun kkwum-ul mek-ko sa-n-ta 'I keep on dreaming,' na-nun yok-ul mek-ess-ta 'I got blamed,' ku-nun nai-lul manhi mek-ess-ta 'He is quite old'). In order to accept these examples, semotactic codes for a specific verb need to be extended.

The purpose of this paper is to propose a way of implementing some extended semotactics of a verb. The extended semotactic codes of a verb are indicated with a lexicographic mark (e.g. 1,2,3, etc.), compatibility value (e.g. 1, 0.8, 0.6, etc.) and a metaphorical pointer (i.e. ↑) at the level of conceptual representation.

II. Algorithm for Implementing Extended Semotactics

A. Lexicographic Indication

In the previous section, some examples with 'extended' interconnections between a verb and (a) participant(s) were shown (e.g. Ideas are flowing, Crowds flow, etc.). In this section, we take an example with English verb have in order to explain its usage. Webster's Dictionary describes meanings of have as follows, in this order:

1) 'to possess' Ex. I have a book.
2) 'to feel compulsion, obligation in regard to ___'
   Ex. I have a letter to write.
      I have a task to do.

12) 'to bear, beget' Ex. She is going to have a baby.
13) 'to eat' Ex. I am going to have bananas.

This paper attempts to represent the participants, whether prototypical or extended, by means of sets. Then S1 and S2 for the English verb 'have' can be represented as shown in Figure 2.

\[
\begin{array}{ccc}
\text{have} & \text{S1} & \text{S2} \\
\text{HB} & \text{Con. Obj} \\
\text{HB} & \text{Abstract} \\
\text{HB} & \text{HB} \\
\text{HB} & \text{Edible} \\
\end{array}
\]

Fig. 2.

To this pair of sets, as a first approximation to be revised, let us assign another set (S3) as a lexicographic indication in order to distinguish the degree of usage. If the usage is designated as the 'first' meaning in Webster's Dictionary (Eng.)/Great Korean dictionary (Kor.), it is indicated as '1.' If the usage is designated as the 'second' meaning, it is indicated as '2.' If the usage is designated as the 'third' meaning, it is indicated as '3.' Figure 3 shows this modification.

\[
\begin{array}{ccc}
\text{have} & \text{S1} & \text{S2} & \text{S3} \\
\text{HB} & \text{Con. Obj} & 1 \\
\text{HB} & \text{Abstract} & 2 \\
\vdots & \vdots & \vdots \\
\text{HB} & \text{HB} & 12 \\
\text{HB} & \text{Edible} & 13 \\
\end{array}
\]

Fig. 3.

The Cartesian product of the property sets S1, S2 and S3 for the English verb have is:
S1 x S2 x S3 = {(HB, Con. Obj, 1), (HB, Con. Obj, 2), (HB, Con. Obj, 3), ...
(HB, Abst, 1), (HB, Abstract, 2), ... (HB, HB, 12), ...
(HB, Edible, 13)}

Not all of the members of the Cartesian product of S1 x S2 x S3 for the English verb *have* can be satisfactorily served as semotactic codes. The combinations of P-1 and P-2 properties considered in this paper ((HB, Con. Obj), (HB, Abst), ... (HB, HB), (HB, Edible)) have different lexicographic indications, i.e. semotactic code (HB, Con. Obj) should be indicated as ‘1,’ semotactic code (HB, Abst) should be indicated as ‘2,’ semotactic code (HB, HB) as ‘12,’ etc. The set of possible semotactics considered in this paper is therefore constituted as

R(have) = {(HB, Con-Obj, 1), (HB, Abst, 2), ...
(HB, HB, 12), (HB, Edible, 13)}

Relation (R) is a mathematical expression of a set of semotactic codes. This relation can be interpreted by saying

1) semotactic code (HB, Con-Obj) is indicated as the primary meaning in Webster’s Dictionary, 2) semotactic code (HB, Abst) is indicated as the secondary meaning in Webster’s Dictionary, etc.

As demonstrated by this example, we can constitute a class of semotactic codes for a specific verb with different indications (e.g. 1, 2, 3 etc.) of membership. This constitution is made by assigning set 3 to (an) existing property set(s) on the basis of usage defined in Webster’s Dictionary (Eng.)/Great Korean Dictionary (Kor.).

B. Fuzzy Sets

The first approximation of assigning lexicographic indications based on order in Webster’s Dictionary can break down since there is no
one-to-one correspondence between the order in the Webster's Dictionary and the importance of usage. That is, the meaning described as the secondary or the tertiary meaning in Webster's Dictionary is not always considered as less important than the primary meaning. In order to revise the first approximation, we assign different grades to each semotactic code using the notion of fuzzy sets.

A fuzzy set is defined as "a class of objects with a continuum of grades of membership" (Zadeh 1987: 29). Such a set is characterized by a membership function which assigns to each object a grade of membership ranging between zero and one. The conventional interpretation of the statement 'John is young' is that John is a member of the set which includes young men. However, considering that the class of young men is a fuzzy set, that is, there is no sharp transition from being young to not being young, a subjective indication can be assigned to the numerical ages. For example, the compatibility value of the numerical ages 22, 28 and 35 with the fuzzy set labeled 'young' might be 1, 0.7, and 0.2 respectively. The notion of compatibility refers to a subjective indication of the extent to which the above age values fit one's conception of the label 'young.'

As an application of the notion of fuzzy sets, a set of semotactic codes can be constituted with different grades of membership. As with fuzzy sets, a grade of membership ranging between zero and one is assigned as a subjective indication. In the determination of semotactic membership in the set for a verb, there is no sharp transition from being an exact member to not being an exact member of the set. This gradual transition can result from an inescapable phenomenon in natural language, i.e. the metaphorical senses of the sentence, contextual or cotextual information. For example, the semotactic code (HB Ed-Sol) for the Korean verb *mek* 'to eat' can be given a compatibility
value of '1' while the semotactic code (HB Abstract) can be given a compatibility value lower than '1,' e.g. 0.8, 0.6. (because of the unusual combinations possible between the process mek and its participants). Let us recall the description of meanings of the have in page 72 and form a set of semotactic codes for the verb as shown below:

\[ R(\text{have}) = \{(\text{HB, Con. Obj}), (\text{HB, Abst}), \ldots (\text{HB, HB}), (\text{HB, Edible}) \} \]

Each member of the set is not given the same weight or the same consideration as a compatibility value. The degree of weight or consideration can vary according to frequency of use. The order described in Webster's Dictionary, however, is not necessarily based on usage frequency. Therefore, the order given in the Dictionary does not play a decisive role in assignment of the compatibility value to each member.

Even if the semotactic code (HB, Abst) for the verb have is established for the secondary meaning (e.g. I have a task to do, I have a letter to write, etc.) as expressed in the Dictionary, it does not deserve to receive the same weight as the semotactic code (HB, Edible). It is less frequent usage than I had some apples. Similarly, the semotactic codes (HB, HB) and (HB, Edible) are established for the twelfth and the thirteenth meanings, respectively. Even though they come far down the list as meanings of have in the Dictionary, they cannot receive very low compatibility values such as 0.4 or 0.2. This is because usages such as She had a baby or She had some rice are very common in everyday conversation and are widely accepted expressions. Therefore, the semotactic codes (HB, HB), (HB, Edible) are given the compatibility value of 0.8 in this paper as a subjective indication in spite of the fact that they come late in the Dictionary listing of meanings.

In order to measure the different grades possible for a semotactic
code, let us assign Set 4 to the existing sets. (S1: P-1 Property, S2: P-2 Property, S3: a Lexicographic Indication). Figure 4 shows this modification.

<table>
<thead>
<tr>
<th></th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>HB</td>
<td></td>
<td>Con. Obj.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>HB</td>
<td></td>
<td>Abst.</td>
<td>2</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.4</td>
</tr>
<tr>
<td>HB</td>
<td>HB</td>
<td></td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>HB</td>
<td></td>
<td>Edible</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4.

The Cartesian product of S1, S2, S3 and S4 is:

\[ S1 \times S2 \times S3 \times S4 = \{(HB, Con-Obj, 1, 1), (HB, Con-Obj, 1, 0.8), (HB, Con-Obj, 1, 0.6), \ldots, (HB, Abst, 1, 1), \ldots, (HB, Abst, 2, 1), (HB, Abst, 2, 0.8), (HB, HB, 12, 0.8), \ldots, (HB, Edible, 13, 0.8), \ldots\} \]

Of the numerous members of this product, only a few will fit our concept of semotactic codes for have, and only these members are implemented, ignoring the rest of the product S1 × S1 × S3 × S4.

\[ R(\text{have}) = \{(HB, Con-Obj, 1, 1), (HB, Con-Obj 2, 1), (HB, Abst, 2, 1), \ldots, (HB, HB, 12, 0.8), \ldots, (HB, Edible, 13, 0.8)\} \]

As demonstrated by the examples, we can constitute a set of semotactic codes for a specific verb with different grades of membership (e.g. 1, 0.8, 0.6, etc.) in addition to different lexicographic indications. This set of codes is made by assigning Set 4 to the existing sets (S1, ⟨S2⟩, S3) based on frequency of use.
C. Metaphorical Expressions with Metaphorical Pointers

This section shows the strategy of assigning the compatibility value in addition to the lexicographic indication cannot deal with metaphorical expressions, e.g. Ideas are flowing, Conversation flows smoothly, Wealth flows from the industry to economy, etc. The reason that participants like ideas conversation, wealth or hair can be associated with the process flow is not because of their own 'prototypical' properties 'HB(Mass)' and 'Abstract' but because of their 'extended' properties 'Liquid.' Similarly, the reason that the participant time can be associated with various processes like waste (e.g. You're wasting my time), save (e.g. You can save your time), spend (e.g. How do you spend your time these days?), borrow (e.g. He's living on borrowed time) or use (e.g. You don't use your time properly) is because the participant 'time' is treated as 'money,' not as 'sequential relations.' Therefore, it is not possible to assign precise quantity to semotactic codes which are based on the applied concept.

In a way of dealing with metaphorical expressions in the process of comprehension, we distinguish the semotactic codes of process into two types—regular and metaphorical, and the distinction is made by using an upward arrow '↑.' Let us take some dictionary entries as an example. The associated participants with the process flow and the semotactic codes for the process will be entered as follows:

- ideas/↑liquid
- crowds/↑liquid
- conversation/↑liquid
- hair/↑liquid
- flow(<liquid>)/flow
- flow(<M↑-liquid>)/move freely
N. Rationale for Lexicographic Indication, Compatibility Value and Metaphorical Pointer

A question might be raised about the need to indicate the degree of mutual relationship between a process and its participants at the level of conceptual representation. If the goal of the system is to comprehend a given input sentence, why not simply extend the number of semotactic codes for a process without any indication? The reason is that the indication by a lexicographic number, compatibility value and a metaphorical pointer is for later application, i.e. machine translation. In machine translation, a conceptual representation with a numerical indication higher than ‘1′ (e.g. ‘5,’ ‘6,’ etc) or with a low compatibility value (e.g. 0.6, 0.4, 0.2, etc.) may require another semantic dictionary before the representation is passed to the target language. Also, the translation of the process can be postponed until a system searches through a dictionary, called the Master Dictionary, of the target language. A conceptual representation with a metaphorical pointer can be reanalyzed in the process of generation based on the properties of participants in source language. Alternatively, a conceptual representation with a high lexicographic number indication, a low compatibility value or a metaphorical pointer could be post-edited by human
beings.

ABBREVIATIONS

Abst: Abstract
Con-Obj.: Concrete-Object
Ed-Sol: Edible-Solid
HB: Human Being
P-1: Participant-1
P-2: Participant-2

REFERENCES


_____. 1987. Notes on dynamic grammar. MS.


요 약

기초적의미와 확장의미의 표현에 관한
컴퓨터 접근 방법

이 창 인
(Rice University)