

Lexical and Semantic Incongruities between the Lexicons of English and Korean

Yae-Sheik Lee* †
Kyungpook National University

Yae-Sheik Lee. 2001. Lexical and Semantic Incongruities between the Lexicons of English and Korean. *Language and Information 5.2*, 21–37. Pustejovsky (1995) rekindled debate on the dual problems of how to represent lexical meaning and on the information that is to be encoded in a lexicon. For natural language processing such as machine translation, these are important issues. When a lexical-conceptual mismatch occurs in translation of corresponding words from two different languages, the appropriate representation of their meanings is very important. This paper proposes a new formalism for representing lexical entries by first analysing observable mismatches in comparable pairs of nouns, verbs, and adjectives in English and Korean. Inherent mis-interpretations and mis-readings in each pair are identified. Then, concept theories such as those presented by Ganter and Wille (1996) and Priss (1998) are extended in order to reflect the cognitivist view that meaning resides in concept, and also to incorporate the propositions of the so-called ‘multiple inheritance’ system. An alternative to the formalism of Pustejovsky (1995) and Pollard & Sag (1994) is then proposed. Finally, representative examples of lexical mismatches are analysed using the new model. (Kyungpook National University)

1. Introduction

Cross-linguistic lexical differences pose significant obstacles to natural language processing applications such as multilingual information retrieval and machine translation. Of those differences, this paper deals primarily with the following types inherent in pairs of corresponding English and Korean words: (i) “lexical-conceptual” mismatches observable in such pairs as *friend* and *chinkwu*; (ii) “aspectual” differences shown by pairs of verbs in the past tense such as *bribed* and *noymwulcwuessta*; (iii) “polysemous” differences triggered by pairs of verbs and nouns such as *begin* vs. *sicakhata* and *magazine* vs. *capci*. To date no in-depth analyses is known to have been done to capture such differences in terms of the information encoded in the relevant lexical entries¹. It is not implausible to assume that such differences are due to information dissimilarities that the entries

* This work was supported by Korea Research Foundation Grant (KRF-99-041-A00019). I am really thankful for valuable comments and corrections contributed by two anonymous reviewers. I also benefited greatly from the comments and the feedback from the audiences at the 2000 ICKL held in Prague. Of course, I’m solely responsible for the remaining errors.

† Department of English Education, Kyungpook National University, Taegu 702-701 Korea, E-mail: yaesheik@knu.ac.kr

1. Actually Lee, C-M. et al (1998) tried to account for the so-called frame alternations shown by locative verbs like *load*, *fill*, *paint*, and others within the framework of Pustejovsky (1995). To this end, they extended Pustejovsky’s lexical representational schema by simply adding the case structure which encodes case-frames.

contain. If this premise is true, an appropriate description of lexical entries is an integral part of an explanation of the differences.

The purpose of the current analysis is, accordingly, to propose a formalism for representing lexical entries in which meaning dissimilarities can be captured. To do this, the following three tasks will be pursued: First, data will be analysed which show the above mentioned types of lexical mismatches between English and Korean. The data will mainly consist of nouns and verbs. Second, formal concept theories such as those of Ganter and Wille (1996), and Priss (1998) will be extended to aid the development of the new model. An extended version of their formal concept theory is required because the meaning representation of lexical items in this analysis adopts the so called 'multiple inheritance' system as do most of the relevant previous analyses done by Pustejovsky (1995) and those in Briscoe, T. et al (1993). Third, the formalism of Pustejovsky (1995) will be modified to reflect artificial and natural cognitive abilities as per Reboul (2000). Thus the proposed 'new' formalism will be free from most of the problems which Pustejovsky (1995) had in terms of accounting for lexical mismatches between English and Korean.

The remainder of the paper is organized as follows: in Section 2, relevant data are observed to show inherent lexical dissimilarities. In Section 3, discussion is centered on an extension of the formal concept theories of Ganter & Wille (1996), and Priss (1998). In Section 4, a new formalism for representing lexical entries is proposed along with an explanation of relevant lexical incongruities. In Sections 5 and 6, a number of implications and conclusions are stated based on the analysis.

2. Data

2.1 Conceptual Mismatches

The following pairs of corresponding words in English and Korean exemplify semantic or conceptual incongruities². Such conceptual differences may cause even Koreans who are proficient in English to make errors in their translation or interpretation.³ A greater understanding of these conceptual mismatches would diminish such errors in the two languages.

(1) Conceptual mismatches in nouns

a.	English	Korean
	<i>friend</i>	<i>chinkwu</i>
	<i>brother</i>	<i>namdongsayng</i>
	<i>uncle</i>	<i>oysamchon</i>
b.	<i>behavior</i>	<i>hayngdong</i>
	<i>glass</i>	<i>can</i>
	<i>chair</i>	<i>uyca</i>
	<i>bathroom</i>	<i>hwacangsil</i>

2. The meanings of some pairs of words, e.g., 'friend' and 'chingu' stand in a subset relation because they are in a common category of meaning, whereas other pairs, e.g., 'glass' and 'can', do not have a subset relationship because their meanings do not relate to a common domain. For example, the word 'glass' in English bears several senses other than the sense of a drinking vessel. Hence, the conceptual difference between two words in the current analysis refers to the mismatch in terms of a common domain of meaning.

3. Lee and Suk (2001) showed that the errors attributable to conceptual mismatching made up more than half of the misused expressions found in nine 7th grade middle school English textbooks of the National Curriculum in Korea.

idea *saynkak*

(2) Conceptual mismatches in verbs

a.	English	Korean
	<i>wear/put on</i>	<i>ipta</i>
	<i>break</i>	<i>kkayta/pwuswuta</i>
	<i>tear</i>	<i>ccayta/ccicta</i>
	<i>cut</i>	<i>kkushta/ckluta</i>
	<i>boil</i>	<i>talita</i>
b.	<i>take/ride</i>	<i>thata</i>
	<i>take</i>	<i>mekta</i>
	<i>crunch</i>	<i>pwuswuta</i>
	<i>hoist</i>	<i>olita</i>

The data given in the (a)s of (1) and (2) show that English nouns and verbs denote broader conceptual categories than those of their Korean counterparts. Conversely, those nouns and verbs given in the (b)s show just the opposite. For example, as shown in (3) below, the meaning that the English word *friend* denotes also includes well-known and likable people. In contrast, *chinkwu* in Korean denotes a set of people who are also of **similar age** as well as being well-known and likable. In English, *behavior* refers to only **habitual** actions, whereas in Korean, *hayngdong* doesn't denote such habitual actions, but only some actions which may or not be habitual.

- (3) a. *friend*: well-known, likable
 a'. *chinkwu*: well-known, likable of similar age
 b. *behavior*: doing something, habitual
 b'. *hayngdong*: doing something

As for incongruities in verbs, *put on* or *wear* in English refers to the act of putting on any type of apparel (viz., clothes, shoes, rings, gloves, etc), whereas its Korean counterpart *ipta* only denotes the act of wearing clothes. The English verbs denote a much wider concept than do the Korean verbs. The pairs of verbs given in (2b) above show the opposite phenomenon. That is, English *bake* refers to the act of cooking in the oven by dry heat. In contrast, Korean *kwupta* only denotes a cooking act by dry heat, irrespective of whether it's done in an oven or not.

- (4) a. *put on/wear*: putting on any type of wears
 a'. *ipta*: wearing clothes
 b. *bake*: cooking by dry heat in an oven
 b'. *kwupta*: cooking by dry heat

Conceptual mismatches in adjectives are different again from those inherent in the nouns and verbs shown above. For example,

- (5) Conceptual mismatches in adjectives⁴
 a. English: anxious, nervous, upset
 Korean: pwulanhan, kekchenganun
 b. English: angry, wrathful, irate, irreful, cross (irascible), sulky, etc.
 Korean: hungpwunhan, hwakanan, kyuknohan, etc.

Corresponding words in English and Korean given in (5) denote partially overlapping conceptual domains. For example, English *upset* usually denotes feeling unhappy and angry about something which has happened, whereas Korean *pwulanhan* usually means a disturbed peace of mind due to something which will happen in the near future, but it doesn't necessarily refer to a concept of getting angry or worried. This is evidenced by the example given in (6) below.

- (6) a. Changlay il-i pwulanha-ciman kekchenganun-eps-ta.
 future matter-Nom anxious-but worry-absence-Dec
 '(I'm) anxious, but not worried about what'll happen.'
 b. Kwake il-i pwulanha-ta.
 past matter-Nom upset-Dec
 '(I'm) upset at what happened.'

A comparison of the concepts denoted by English *angry* and by Korean *hungpwunhan*, shows the mismatching dilemma. The adjective 'hungpwunhan' latter sometimes denotes a meaning of getting excited and sometimes a meaning of getting angry. However, the adjective 'angry' never refers to a meaning of getting excited in the positive English sense of the word. Consequently, since such adjectives convey overlapping concepts, an exact matching of them is most often difficult or impossible.

2.2 Aspectual Mismatches

When causative verbs in English are used with the past tense, they always denote accomplished past situations or events. In contrast, their Korean counterparts refer to weak accomplishments in the sense that the accomplishment reading can be cancelled out. This difference might be caused by the conceptual difference encoded in each relevant lexical entry.

- (7) English⁵
 a. *I bribed the mayor, but she refused to accept it.
 a'. I tried to bribe the mayor, but ...
 b. *He started the motor, but it was too cold to start.
 b'. He tried to start the motor, but ...

4. As for conceptual mismatches in adverbs, the following corresponding pairs are indicative. However, an exhaustive list of such adverbial pairs requires more comprehensive research.

(i) Conceptual differences in adverbs

English:	Korean
perhaps	ama, echyemyen
sometimes	kakkum, ttayttaylo
occasionally	kakkum, ttayttaylo
frequently	kakkum, ttayttaylo

It is not easy to account for the corresponding word(s) in Korean which matches with English *sometimes*, *occasionally*, and *frequently*. Furthermore, it is only possible to say that English *perhaps* and Korean *ama* differ from each other in a restricted sense. Consider the examples,

- (ii) a. *Perhaps John must have arrived there by 6.
 b. Ama John-nun yesessi ccum tochakhayss-umeythullimeps-ta
 perhaps -Top 6 o'clock by arrived-must-Dec
 (Literally) Perhaps John must have arrived there by 6.

5. The data in (7) are from Hofmann (1993: 149). The achievement cancellation effect can be observed in such causative verbs as *attach*, *burn*, *chop*, *deceive*, *erase*, *fold*, *grind*, *heat*, *install*, *kill*, *lift*, *mix*, *open*, *pull*, *repair*, *sharpen*, *tear*, *untie*, *wash*, and *wrap up*.

(8) Korean:

- a. Na-nun sicang-eykey noimwul-ul cwu-ess-ciman
 I -Top mayor-to bribe-Acc give-pst-but
 kunuy-nun patci-anh-ass-ta
 she-Top accept-not-pst-Dec
 ‘I bribed the mayor, but she didn’t accept it.’
- b. Ku-nun sidong-lul kel-ess-ciman, kelici-anh-ass-ta.
 he-Top start-Acc put on-pst-but, put on-not-pst-Dec
 ‘I started (he car), but it didn’t start.’

2.3 Polysemous Mismatches

Polysemous incongruities are also apparent in lexical data as (9) below shows. The polysemous lexical items in both languages can be seen to behave differently. The examples are typical types of polysemy.⁶

(9) (i) **Producer vs. product**

- a. ??Ku capci-nun pyencipca-ul haykohay-ss-ta
 the magazine-Top editor-acc fire-pst-Dec
 ‘The magazine fired its editor.’
- b. Mary-nun capci/chayk-ey coffee-lul sos-ass-ta
 Mary-Top magazine/book-over coffee-Acc spill-pst-Dec
 ‘Mary spilled coffee over the magazine/book.’
- c. Ku capci/chayk-nun caymiiss-ta.
 the magazine/book-Top interesting-Dec
 ‘The magazine/book is interesting.’

(ii) **aspectual verbs showing complement alternations⁷**

- a. ??Mary-nun chayk-ul sicakhay-ss-ta.
 Mary-Top book-Acc begin-pst-Dec
 ‘Mary began a book.’
- b. Mary-nun chayk-ul ilk-ki sicakhay-ss-ta.
 Mary-Top book-Acc read-ing begin-pst-Dec.
 ‘Mary began reading a book.’

The English versions of the sentences in (9i, a) and (9ii, a) both sound natural and grammatical. The Korean versions appear less so to an English reader because of the linguistic structure of the sentences. Such different polysemous behaviors can also be due to differences in the information encoded in the lexical entries of polysemous lexical items. The polysemous behaviors will be accounted for in this paper by portraying relevant lexical entries differently.

3. Conceptual Semantics

Cognitive linguists or semanticists such as Lakoff (1988) and Langacker (1998) share the same view on meaning, i. e., that meaning resides in concept. Most formal concept

6. For more comparisons of polysemous phenomena in English and Korean, see Lee (2000).

7. From the sentence *Mary began the book*, the following interpretations can be made:

- (i) Mary began to read the book.
 (ii) Mary began to write the book.

The one chosen is dependent on the ongoing discourse. If Mary is an author, then the second interpretation is more likely.

theorists such as Bartsch (1998) and Ganter & Wille (1996) contend that a concept is a pair of two sets of attributes or properties and of objects to which the attributes apply. The following information reflects the formal definition of concept by Ganter & Wille (1996) and Priss (1998):

(10) (i) **Formal Context:**

$K = \langle G, M, I \rangle$, where G and M stand for sets of objects and properties or attributes, respectively. I is a relation between G and M . That is,

$I = G \times M$. *gIm* means that the object g has the attribute m .

(ii) **Formal Concept:**⁸

A formal concept C of the context $\langle G, M, I \rangle$ is a pair $\langle A, B \rangle$ with $A \subseteq G$, $B \subseteq M$, $A' = B$ and $B' = A$.

Note: $A' = \{m \in M: gIm \text{ for all } g \in A\}$; $B' = \{g \in A: gIm \text{ for all } m \in B\}$

(iii) **Subconcept:**

If $\langle A_1, B_1 \rangle$ and $\langle A_2, B_2 \rangle$ with $A_1, A_2 \subseteq G$, and $B_1, B_2 \subseteq M$ of a context $\langle G, M, I \rangle$, and $A_1 \subseteq A_2$ or $B_2 \subseteq B_1$, then $\langle A_1, B_1 \rangle$ is a subconcept of $\langle A_2, B_2 \rangle$, or $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$.

The relation \leq is the **hierarchical order** of the concepts.

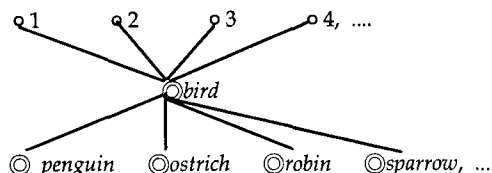
The definition given in (10ii) shows that a concept can only be talked about under the assumption that there is a relevant context. The context determines the incidence relations (I) between objects and attributes. From the sub-concept definition of (10iii) plus the definition of concept of (10ii), the conclusion can be reached that all of the concepts in a given context form a conceptual hierarchy or lattice. If the idea is adopted that lexical items in a language are names for those concepts, then the lexicon of a language can be taken as an extremely large but complex conceptual lattice. Given that each individual concept in a language is described, the overall lexicon-picture of a language is theoretically possible.

Despite such a promising aspect of the above formal concept theory, it has a number of problems. First, according to the definition of sub-concept given in (10iii), a subconcept to a superconcept should inherit all the attributes or concepts of the superconcept. However, this inheritance rule cannot always be applicable as the figure in (11) shows. That is, all the stereotypes or the set of typical attributes of the concept 'bird' cannot be inherited to peripheral sub-concepts like 'penguin' and 'ostrich'. This phenomenon is referred to in this paper as the **Stereotype Inheritance** problem. For example, one of the typical properties of 'bird', namely, that it can fly, doesn't apply to the concepts 'penguin' and 'ostrich'. The formal concept theory needs to be revised to account for our intuition that penguins or ostriches are birds.

8. From the definition of the formal concept given in (10ii), the following proposition follows:
If $\langle G, M, I \rangle$ is a formal context, $A, A_1, A_2 \subseteq G$ and B, B_1 and $B_2 \subseteq M$, then

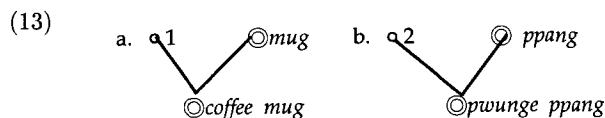
- | | |
|---|---|
| 1) $A_1 \subseteq A_2 \Rightarrow A_2' \subseteq A_1'$ | 1') $B_1 \subseteq B_2 \Rightarrow B_2' \subseteq B_1'$ |
| 2) $A \subseteq A''$ | 2') $B \subseteq B''$ |
| 3) $A' = A'''$ | 3') $B' = B'''$ |
| 4) $A \subseteq B' \Leftrightarrow B \subseteq A' \Leftrightarrow A \times B \subseteq I$ | |

(11) **Stereotype Inheritance Problem**⁹



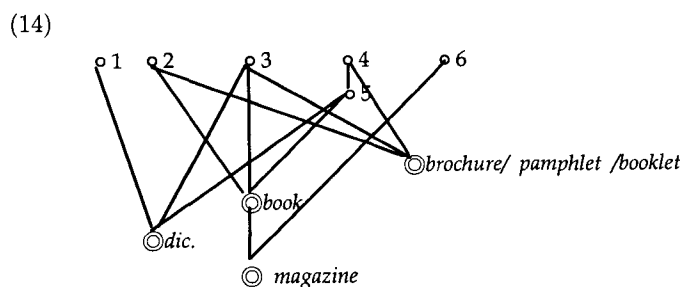
Second, if a complex concept α is composed of several super concepts, then they do not play an equal role in forming the complex concept α . As the examples of (12) and (13) show, the concepts ‘coffee-serving’ and ‘mug’ make a different contribution to the formation of the concept ‘coffee-mug’. To determine what the entities denoted by *coffee-mug* are, we rely more on the concept ‘mug’. This phenomenon is referred to (in this paper) as the **Inequable Contribution** problem.

- (12) a. The magazine has just fired its sports editor.
 a'. Mary spilt coffee over the magazine.
 a''. The magazine really got me upset.
 b. The coffee-mug is a little heavy for me.
 b'. #I really enjoy the coffee-mug.
 (Intended Reading: I really enjoy the coffee in the coffee-mug.)



Note: o1: ‘coffee’s being served’; o2: ‘in the shape of a crucian’

The figure in (14) and the data in (15) show that concepts which usually are categorized as the concept book are composed of different concepts or properties. Within the framework of Ganter & Wille (1996) and Priss (1998), this cannot be accounted for appropriately.¹⁰



9. Note that ‘o’ and ‘⊙’ stand for non-lexicalized and lexicalized concepts, respectively. o1, o2, o3 and o4 stands for the concepts of ‘can fly’, ‘have two wings’, ‘have a beak’, and ‘have feathers’, each.
 10. Note that o1, o2, o3, o4, o5 and o6 stand for *to be referred to*, *to be read*, *written information*, *paper*, *paper bound with glue*, and *organization*, respectively.

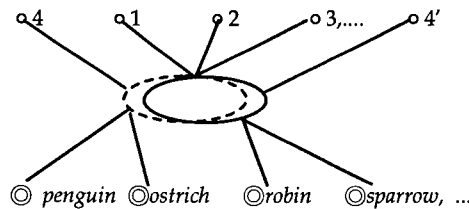
- (15) a. A dictionary is a book in which the words and phrases of a language are listed alphabetically...
- b. A magazine is a publication,
 (*Collins Cobild English Dictionary* 1995)
- c. brochure \Rightarrow book \Rightarrow publication, (*WordNet 1.7 Browser*)

It is customary to take a dictionary, a brochure, or booklet as a kind of book. In other words, they are roughly acceptable as a book. To deal with these data, **Similar Concept** is defined as follows:

- (16) **Similar Concept (SC, hereafter):**
 Concept $\langle A1, B1 \rangle \approx$ Concept $\langle A, B \rangle$ of Context $K = \langle G, M, I \rangle$ with $A, A1 \subseteq G$, and $B, B1 \subseteq M$ iff
 - (i) $B \cap B1 \neq \emptyset$
 - (ii) $\#(B \cap B1) / \#(B)$ or $\#(B \cap B1) / \#(B1) \geq n - x/n$,
 where $\#(B)$ or $\#(B1) = n$ and x depends.
 Note: $\langle A1, B1 \rangle \approx \langle A, B \rangle$ is read as "concept $\langle A1, B1 \rangle$ is a similar concept to Concept $\langle A, B \rangle$ ".

According to the above definition of SC, the concepts 'dictionary', 'magazine', 'brochure' and 'pamphlet' are all SC's to each other. In figure (20), the solid oval and the dotted oval show that they are similar concepts.

- (17) 'bird' and its similar concepts



Note: $o4$ and $o4'$ stand for concept 'cannot fly', 'can fly', respectively. $o1$, $o2$, and $o3$, and the others refer to attributes or super-concepts of the typical concept of 'bird'.

In terms of the concept 'bird' (the solid oval), the concepts 'penguin', and 'ostrich' are similar sub-concepts to concept 'bird'. This means that the concepts 'penguin', 'ostrich', 'robin', and 'sparrow' are similar concepts to each other. The definition of similar sub-concept is given in (18).

- (18) **Similar sub-concept (hereafter SSC):**
 Concept $\langle A1, B1 \rangle$ is a similar sub-concept to Concept $\langle A, B \rangle$ of Context $K = \langle G, M, I \rangle$ with $A, A1 \subseteq G$, and $B, B1 \subseteq M$ iff

$$\exists C [C \in \mathcal{C}(\langle G, M, I \rangle) \wedge C \subseteq \langle A, B \rangle \wedge \langle A1, B1 \rangle \approx C]$$
 where $\mathcal{C}(\langle G, M, I \rangle)$ refers to the set of all concepts in the formal context K .

As the above examples show, the definitions of subconcept, SC and SSC explain why people put concepts which show the **Stereotype Inheritance** problem under the same super-concept. This means SC's can come under a same superconcept in terms of the conceptual lattice. Under the assumption that the lexicon of a language consists of

lexical entries which are represented as chunks of concepts, the lexicon can be thought as a conceptual lattice. In this sense, the relations of subconcept, SC and SSC are necessary to account for the organization of the lexicon. On the other hand, these relations might imply that a fluent bilingual speaker recognizes the concepts carried by two corresponding words as similar concepts. For example, to those who are fluent in English and Korean, *friend* and *chinkwu* carry two similar concepts.

To solve the **Inequable Contribution** problem, the following definition of **Base Concept** is proposed:

(19) **Base Concept (BC, henceforth)**

Concept C is a Base Concept to a complex concept C' if C is an immediate superconcept to C and plays an essential role to determine the entities of C.

For example, the concepts 'mug' and 'ppang' are the BC to the concepts 'coffee mug' and 'pwunge ppang', respectively. In contrast, all the immediate super-concepts to the complex concepts given in (17) are BC's because they play an equally essential role to determine their entities¹¹. With regard to concepts denoted by such causative verbs as *break*, *kill*, *build*, etc., they denote causative events which can be subdivided into process and result states. According to Pustejovsky (1995), the process is the head sub-event to the causative event. This researcher takes it to be the BC to the concept of 'causative event', and reflects such information in the lexical entry.

In sum, the above theory of formal concept (by Ganter & Wille (1996), and Priss (1998)) is extended by the definitions of SC, SSC, and BC. Such an extension makes the formal theory of concept free of the two conventional problems of Stereotype Inheritance Problem and Inequable Contribution Problem. Specifically, in natural language use, entities categorized by an SC to a typical concept are also allowed to be named after the name of the typical concept. For example, even though penguins and ostriches belong to an SC to the typical concept of 'bird', they are allowed to be named after 'bird' as expected. Similarly, 'dictionary', 'brochure', 'pamphlet', and 'booklet' are all allowed to be named by the typical concept 'book'.

4. Representation of Lexical Entries

4.1 A New Representational Formalism

This section will mainly focus on the problem of how to represent lexical entries in order to answer the question of what types of information are to be encoded in the lexicon. Reboul (2000), claims that Pustejovsky's (1995) model of representing lexical entries falls short of a natural cognition model because it is devised to serve as an artificial cognition model. Pustejovsky's program basically enriches the semantic representation of words within the HPSG framework of Pollard and Sag (1994) by establishing the mechanisms of Argument Structure, Event Structure, and Qualia Structure in lexical entries. Qualia Structure can be maximally given the following four different types of properties: Constitutive, Formal, Telic, and Agentive. The establishment of these mechanisms might be a solution to the problem of how to represent lexical entries. Pustejovsky borrows these four qualia from Aristotelian essential properties for understanding things, stuff and nature.¹² Qualia

11. The BC defined in (22) is similar to Ontological Meaning as per Hwang (2001). He proposed subdividing the whole meaning of a word into two sub-meaning-constituents. One is Ontological Meaning which tells what the object or entity the word in question refers to. The other is Attributive Meaning which informs us how the object denoted by the word looks or functions.

12. Aristotle claimed that to answer the question of what it is, we need to cite the essence of the entity in question or the cause of its being what it is. Furthermore, there are four ways in which we cite

Structure serves as a source for multiple meanings of individual words. For example, the sentence *Mary began the book* can have two interpretations: ‘began to write’ and ‘began to read’ the book. We can get the first reading from Agentive quale and the second reading from Telic quale of the Qualia Structure of *book*. Consider the following matrix for *book* and *begin*.

(20) a. lexical entry for *book* in English

$$\left[\begin{array}{l} \textit{book} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = x : \textit{info} \\ \text{ARG2} = y : \textit{physobj} \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \textit{info} \cdot \textit{physobj} - \textit{lcp} \\ \text{FORMAL} = \textit{hold}(y, x) \\ \text{TELIC} = \textit{read}(e2, w, y) \\ \text{AGENT} = \textit{write}(e1, x, y) \end{array} \right] \end{array} \right]$$

b. lexical entry for *begin* in English

$$\left[\begin{array}{l} \textit{begin} \\ \text{EVENTSTR} = \left[\begin{array}{l} \text{E1} = e1 : \textit{transition} \\ \text{E2} = \textcircled{2} : \textit{process} \\ \text{RESTR} = \langle \quad \circ \quad \alpha \end{array} \right] \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = \textcircled{1} \left[\begin{array}{l} \textit{animate} \\ \text{FORMAL} = \textit{physobj} \end{array} \right] \\ \text{ARG2} = \textcircled{2} \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = P(\textcircled{2}, \textcircled{1}) \\ \text{AGENTIVE} = \textit{begin-act}(e1, \textcircled{1}, \textcircled{2}) \end{array} \right] \end{array} \right]$$

The problem with Pustejovsky’s method of representing lexical entries as above, is that it is not sufficient to model natural cognition, as Reboul (2000) points out. In some sense, this insufficiency means that Pustejovsky’s program cannot capture the above mentioned lexical or conceptual differences in lexical items between languages. Aside from an inability to account for such mismatches, even if we assume his generative mechanisms¹³, his model cannot show how to distinguish one reading from among possible alternative readings, for example, as in *begin the book*. The information in the lexical entries of (20a and b) is insufficient to tell us how to choose a reading of *begin to write*, from *begin the book*. To do this, we would need pragmatic information or world knowledge that the subject of the verbal phrase is an author.

The model in (21) is proposed as an improved alternative to Pustejovsky’s method of representing lexical entries. The work of Lee (2000) and Lee & Kim (2001) has been drawn upon in its development.

cause: (i) material (ii) formal (iii) efficient, and (iv) final.

To be more specific about each cause, what Aristotle meant by **material cause** was that certain aspects of an object’s being are caused by its matter. **Formal cause** is an aspect caused by the structure of an object. **Efficient cause** changes the structure of the matter in a fixed direction or way. **Final cause** explains the reason why particular objects with particular structures allow organisms to function in various ways that allow them to grow, maintain, and reproduce themselves. See Prasada (1999: 121–125) for more details about the four causes.

13. Pustejovsky (1995) proposed the generative mechanisms of *type-coercion* and *selective binding*, *co-composition*. For more details on generative operations, see Chapter 7 of Pustejovsky (1995).

(21) Scheme of Lexical Entries

$\langle \alpha \rangle$											
CAT	<table style="border-collapse: collapse;"> <tr> <td style="padding-right: 5px;">HEAD [...]</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> <tr> <td style="padding-right: 5px;">ARGSTR [...]</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> </table>	HEAD [...]		ARGSTR [...]							
HEAD [...]											
ARGSTR [...]											
CONT	<table style="border-collapse: collapse;"> <tr> <td style="padding-right: 5px;">REST $\langle \dots \rangle$</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> <tr> <td style="padding-right: 5px;">EVENTSTR [...]</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> <tr> <td style="padding-right: 5px;">MCC</td> <td style="border-left: 1px solid black; padding-left: 5px;"> <table style="border-collapse: collapse;"> <tr> <td style="padding-right: 5px;">MBRS $\langle \dots \rangle$</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> <tr> <td style="padding-right: 5px;">BASE [...]</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> </table> </td> </tr> </table>	REST $\langle \dots \rangle$		EVENTSTR [...]		MCC	<table style="border-collapse: collapse;"> <tr> <td style="padding-right: 5px;">MBRS $\langle \dots \rangle$</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> <tr> <td style="padding-right: 5px;">BASE [...]</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> </table>	MBRS $\langle \dots \rangle$		BASE [...]	
REST $\langle \dots \rangle$											
EVENTSTR [...]											
MCC	<table style="border-collapse: collapse;"> <tr> <td style="padding-right: 5px;">MBRS $\langle \dots \rangle$</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> <tr> <td style="padding-right: 5px;">BASE [...]</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> </table>	MBRS $\langle \dots \rangle$		BASE [...]							
MBRS $\langle \dots \rangle$											
BASE [...]											
SR	<table style="border-collapse: collapse;"> <tr> <td style="padding-right: 5px;">SYNTET [...]</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> <tr> <td style="padding-right: 5px;">ANTSET [...]</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> <tr> <td style="padding-right: 5px;">HYPERSET [...]</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> <tr> <td style="padding-right: 5px;">...</td> <td style="border-left: 1px solid black; padding-left: 5px;"></td> </tr> </table>	SYNTET [...]		ANTSET [...]		HYPERSET [...]		...			
SYNTET [...]											
ANTSET [...]											
HYPERSET [...]											
...											
CONTEXT	[BK RESTR $\langle \dots \rangle$]										

To specify the notations in the matrix above, $\langle \alpha \rangle$ stands for the name of the lexical entry which can serve to address the lexical entry in the lexicon. CAT stands for the categorial information of the lexical item α . The description of ARG-STR encodes the information about the argument structure of a predicate and the thematic role of its arguments (See Grimshaw, 1990). The REST value of the CONT contains the semantic specifications of arguments which are similar to denotational properties. When we try to decide what entity the lexical item denotes, we almost wholly rely on the value of REST. Furthermore, MCC contains the set of concepts which are major constituents of the (overall)concept denoted by the lexical item in question. That is, the value of MCC encodes properties conveyed by the lexical item in question, which we can imagine right away when we hear the lexical item. If a lexical item is polysemous as in the case of *magazine* in (12), each value of MCC can provide one reading of the polysemous lexical item in an appropriate context. BASE is a set of MCC's which are necessary to determine the semantic sort of the entities denoted by the lexical item. It reflects the BC given in (19). For example, if MCC is a type of event, the BC of the MCC would correspond to the "HEAD" subevent of the EVENTSTR¹⁴ in terms of Pustejovsky (1995). SR (Sense Relations) bears information about relevant lexical items that stand in some sense relation such as synonym, antonym, hypernym, troponym, etc. The reason that SR is established in the lexical entry is that a particular lexical item idiosyncratically stands in a sense relation with (a) particular lexical item(s). According to Miller (1998), *heavy* and *weighty*, are closely similar in meaning. Additionally, *light* and *weightless* are also synonymous, and conceptually opposite to the meaning of the first two words. However, *heavy* stands in an antonymy relation only with *light*, and *weighty* with *weightless*. Finally, a pragmatic window, CONTEXT, is built into the lexical entry. This is needed for determining the exact reading of an expression with a pragmatically varying meaning such as [*begin* + NP]. Compared with Pustejovsky's lexical entry model, this proposed schema has more flexibility and a greater capacity to account for the polysemous behaviors of the word *begin*.

14. The current paper adopts the same STRUCTURE as Pustejovsky (1995). Thus, it contains information about the subevents and the notion of 'HEAD' event.

In short, the above ‘new’ formalism is deemed to reflect natural cognitive abilities much more than Pustejovsky’s (1995) model vis-a-vis Reboul (2000). The belief is that the proposed model reflects our conceptual intuitions by encoding the major conceptual constituents of the relevant concept and the BASE concept in the MCC. In addition, it encodes SRs (sense relations) in the lexical entries. As mentioned above, knowledge on the sense relations which a lexical items can have with its relevant lexical items is also part of our cognitive ability. Furthermore, the information in the CONTEXT will reflect a possible interaction between lexical knowledge and pragmatic knowledge. However, since the current formalism maintains most of Pustejovsky’s representational practices, at best it remains a hybrid for simultaneously reflecting both natural and artificial cognitive abilities.

4.2 Explanation of Lexical Incongruities

This section will explain the lexical incongruities in some of the English and Korean word samples given earlier where mismatches in meaning occurred. The sample incongruities will be explained in the context of the ‘new’ representational formalism. In terms of conceptual differences, it is more or less easy to reflect these in the lexical entries in question since we can encode them into the REST portion of the CONTEXT. For example, in the following lexical entries for English *friend* and Korean *chinkwu*.

- (22) a. <friend>
 CONT REST < [REL well-known], [REL likable] >
 b. <chinkwu>
 CONT REST < [REL well-known], [REL likable],
 [REL of similar age] >

In terms of the aspectual differences observed in Section 2.4, it is plausible to account for the achievement-cancellability of past tense accomplish-verbs on the basis of lexical difference because it is implausible to treat the past tense in either language differently. The entries in (23) a, and b, show this.

- (23) a. lexical entry for *bribe* in English

<bribe>	
CAT	[HEAD verb ARGSTR < [INP], [NP] >]
CONT	[REST < [REL bribe; ACTOR [FORMAL human]; UNDERGOER [FORMAL human]] > EVENTSTR [E1: process; E2: result-state HEAD E2 REST E1 ≤ E2]]
MCC	[MBRS < process, result-state > BASE { E2 }]

b. lexical entry for *noymwulcwuta* in Korean

$\langle \text{noymwulcwuta} \rangle$	
CAT	$\left[\begin{array}{l} \text{HEAD verb} \\ \text{ARGSTR} \langle \text{1NP}, \text{2NP} \rangle \end{array} \right]$
CONT	$\left[\begin{array}{l} \text{REST} \langle [\text{REL bribe}; \text{ACTOR } \text{1}[\text{FORMAL human}]; \\ \text{UNDERGOER } \text{2}[\text{FORMAL human}]] \rangle \\ \text{EVENTSTR} \left[\begin{array}{l} \text{E1: process; E2: result-state} \\ \text{REST E1} \leq \text{E2} \end{array} \right] \\ \text{MCC} \left[\begin{array}{l} \text{MBRS} \langle \text{process, result-state} \rangle \\ \text{BASE} \{ \text{E2} \} \end{array} \right] \end{array} \right]$

The event denoted by *noymwulcwuta* ('bribe') in Korean consists of two sub-events of the bribing process and the resultant state of being bribed. This information is encoded both in the EVENTSTR and MCC. In this sense *bribe* in English is the same as its Korean counterpart *noymwulcwuta*. However, they differ in the sense that *bribe* in English denotes an event which has the resultant state as the head sub-event. In contrast, the event denoted by *noymwulcwuta* in Korean is underspecified in terms of such headness. In other words, the EVENTSTR of *noymwulcwuta* doesn't contain the information about headness. That of *bribe* in English bears the resultant state as the 'head' sub-event. The presence or absence of the explicitly specified information about the 'head' event in the EVENTSTR determines the accomplishment-cancellability because the 'head' sub-event is assumed to be focused and modified by the past tense.

In terms of the different polysemous behaviors by English *magazine* and Korean *capci*, item (24) below shows how differently the lexical entries can be represented. This serves to highlight that the polysemous difference observed in (9ii) above is due to the conceptual difference reflected in the value of REST of the CONTENT.

$\langle \text{magazine} \rangle$	
CONT	$\left[\begin{array}{l} \text{REST} \langle [\text{REL book}], [\text{REL bundle of paper}], \\ [\text{REL written-info}], [\text{REL regularly-} \\ \text{published}], [\text{REL organization}] \rangle \\ \text{MCC} \left[\begin{array}{l} \text{MBRS} \langle \text{written-info, company,} \\ \text{bundle of paper, ...} \rangle \\ \text{BASE} \{ \text{written-info, company, bundle of paper} \} \end{array} \right] \end{array} \right]$
SR	$\left[\begin{array}{l} \text{SYNSET} \{ \text{magazine, ...} \} \\ \text{HYPERSET} \{ \text{periodic publication, ...} \} \end{array} \right]$

b. <capci>

CONT	REST < [REL book], [REL bundle of paper], [REL written-info], [REL regularly-published] >
MCC	MBRS <written-info, company, bundle of paper, ...>; BASE { written-info, bundle of paper }
SR	SYNSESET { capci, ... } HYPERSET { cengkikahaymwul, ... }

The different polysemous behaviors for *magazine* and *capci* can be accounted for structurally. That is, English *magazine* only contains the relation value of [REL organization] in the REST, which enables us to get the reading of the magazine company, whereas its Korean counterpart lacks the value as (24b) shows. The meaning of magazine company has to be expressed by the word *capci* and the suffix *-sa* which means a company.

The lexical entries in (25) a, and b, below account for the different polysemous behaviors in English *begin* and its Korean counterpart *sicakhata*.

(25) a. <begin>

CAT	HEAD verb ARGSTR < [1] NP, [2] NP/VP >
CONT	REST < [REL begin; ACTOR [1][FORMAL animate]; UNDERGOER [2] [FORMAL phyobj/event]] > EVENTSTR [E1: transition; E2: process] REST initial overlap MCC [MBRS <transition, process>] BASE { ... }
CONTEXT	BKROUND REST < [R([1],[2]), ...]>

b. <sicakhata>

CAT	HEAD verb ARGSTR < [1] NP, [2] NP >
CONT	REST < [REL begin; ACTOR [1][FORMAL animate]; UNDERGOER [2] [FORMAL event]] > EVENTSTR [E1: transition; E2: process] REST initial overlap MCC [MBRS <transition, process>] BASE { ... }
CONTEXT	BKROUND REST < [R([1],[2]), ...]>

The main differences between the above lexical entries are highlighted by shadowing them. The first entry denotes that English *begin* takes as its complement an NP or a VP which denotes a physical object or an event, respectively. In contrast, the second entry shows that Korean *sicakhata* only takes an NP referring to an event. This can be evidenced by the following examples:

- (26) a. ??John-un ku chayk-ul sicakhay-ss-ta.
 John-Top the book-Acc begin-pst-Dec
 ‘lit. John began the book.’
- b. John-un ku chayk ilk-ki-ul sicakhay-ss-ta.
 John-Top the book read-nom-Acc begin-pst-Dec
 ‘John began to read the book.’
- c. Kutul-un yenkuk/phati-ul sicakhay-ss-ta.
 they-Top play/party-Acc begin-pst-Dec
 ‘They began the play/party.’

The suffix -*ki* in (26b) is a nominalizer which turns a verbal phrase into a nominal one. Without the nominalized verbal phrase, the acceptability of (26a) drastically worsens. However, (26c) shows that *sicakhata* in Korean can take an NP without the nominalized verbal phrase only if the NP denotes an event such as play, party, movie, or music, etc. Furthermore, those to whom (26a) sounds natural have to determine the unbound predicate R in the REST of the BACKGROUND of the CONTEXT on the basis of the relevant contextual information available to them. This means that the current way of representing lexical entries returns pragmatic meaning to the field of pragmatics, unlike Pustejovsky (1995). In short, the present proposal is considered to be more flexible and capable in accounting for the polysemous behaviors of *begin* and *sicakhata*.

5. Implications and Limitations

The analysis and discussion in this paper showed how the proposed formalism for representing lexical entries can contribute to accounting for lexical-conceptual mismatches between two different languages. The ‘new’ model is considered to have potential in several areas: First, the current framework can be used for capturing the frame alternations by locative verbs in a more efficient way than that of Jackendoff (1990), Lee, C-M. et al. (1998) and Kim (1999) since it encodes argument and event information. It can also reflect the notion of “HEAD” event which is assumed to be responsible for the frame alternations by Jackendoff (1990) and Pustejovsky (1995).¹⁵ Second, since the current

15. The following data exemplify the so-called transitivity alternation. To account for this phenomenon, Pustejovsky (1995) proposes to decompose events denoted by causative verbs into two subevents such as **process** and **state**.

- (i) a. They sank the boat. a'. The boat sank slowly.
 b. John broke the window. b'. The window broke.

Pustejovsky contends that the above transitivity alternation is attributable to the underspecification of headness in the relevant event structure. In other words, either of the two subevents can be syntactically realized. For example, if a situation to be expressed is a state, the intransitive frame occurs. Similarly, Jackendoff (1990) suggests that the conceptual structures denoted by the above causative verbs encode the optional (underlined) part of the external cause of the entire cause-event as (ii) shows:

- (ii) break
 V
 < ———NP_j >

formalism is an extension of HPSG, it could contribute to extending the semantic portion of the HPSG formalism. Third, greater elaboration of the proposed formalism might lead to the development of a comprehensive formalism on which a multi-purpose machine-readable lexicon data base can be built since it encodes information on both natural and artificial cognitive abilities.

However, the paper has a number of recognized limitations. In terms of the manner of representing lexical entries, it has not mentioned how MCCs (to a concept) are determined. With regard to complex concepts and their component concepts, no discussion was made about what it means for a complex concept to consist of its constitutive concepts. It is obvious that the set intersection is not the only possible link to constitutive concepts (to a complex concept). In the case of the complex concept "book", the set intersection operation over its constitutive concepts 'written information' and 'bound paper with glue' cannot bring the complex concept into being. Instead, the relation 'hold' can. These limitations are problems that await further study.

6. Concluding Remarks

A plausible and tenable account of lexical and conceptual differences between English and Korean has been made in this paper. The discussion extended the formal concept theory of Ganter and Wille (1996), and Priss (1998) with the newly defined notions of SC, SSC, and BC, and proposed an alternative way to represent lexical entries as per Pustejovsky (1995). The extended formal concept theory can overcome the two traditional problems of 'Stereotype Inheritance' and 'Inequable Contribution'. Furthermore, compared with Pustejovsky (1995), the proposed method of representing lexical entries is free from most of his model's problems, and is more efficient in reflecting lexical differences between English and Korean. However, the limitations of the research prevent a more comprehensive formalism from being proposed that could eventually lead to a machine-readable lexicon data base. Further research on the issue is required.

References

- Bartsch, R. 1998. *Dynamic Conceptual Semantics*. Stanford: CSLI Publication.
- Briscoe, T. et al (eds.) 1993. *Inheritance, Defaults, and the Lexicon*. Cambridge: Cambridge University Press.
- Collins Cobild English Dictionary. 1995. Harper Collins Publishers.
- Ganter, B. and R. Wille 1996. *Formal Concept Analysis: Mathematical Foundation*. Berlin: Springer.
- Grimshaw, J. 1990. *Argument Structure*. MA: MIT Press.
- Hwang, H-S. 2001. *Meaning of Morphological Units and Word Foundation in Korean*. Seoul: Wulin Publishing Company.
- Hofmann, T. R. 1993. *Realms of Meaning: An Introduction to Semantics*. London: Longman Publishing.
- Jackendoff, R. 1990. *Semantic Structure*. Cambridge: MIT Press.

[EVENT CAUSE([]I, INCH([STATE BE([]J, [PROPERTY BROKEN]))]))]

I believe that more elaboration of the above two frameworks also can capture the following frame alternation shown by 'locative' verbs.

- (iii) a. John loaded the truck (with hay). a'. John loaded (the)hay (onto the truck).
 b. John sprayed paint (onto the wall). b'. John sprayed the wall (with paint).

This topic is, of course, worth intensive further study in this line.

- Kim, M. 1999. *A cross-linguistic perspective on the acquisition of locative verbs*. Doctoral Dissertation, U. of Delaware.
- Lakoff, G. 1988. Cognitive Semantics. In *Meaning and Mental Representation eds.* by Eckman, F. R. et al., 119-154, London: Plenum Press
- Langacker, W. R. 1998. Cognitive Grammar: A Basic Introduction, Handout read at Cognitive Linguistics Conference at KNU.
- Lee, Y-S. 2000. Lexical Mismatches between English and Korean: with Particular Reference to Polysemous Nouns and Verbs, *The Language and Information*. Vol. 4, No. 1.
- Lee, Y-S. and Kim, J-B. 2001. Mapping Relations between Lexical Information and Conceptual Structures. Ms.
- Lee, Y-S. and Suk, D. S-Y. 2001. Conceptual Incongruity between English and Korean Lexical Items: with particular Reference to Misuses in English Textbooks in Korea, read at *Harvard ISOKL-2001*.
- Lee, C. et al. 1998. Lexical Semantic Structure for Predicates in Korea read at 10th ESSLLI in Saarbrücken.
- Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. The University of Chicago Press?
- Miller, G. A. 1998. Modifiers in WordNet. in Fellbaum C. ed, *WordNet: an Electric Lexical Database*, 47-68, Cambridge, Massachusetts: the MIT Press.
- Pollard, C. and I. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago:University of Chicago Press.
- Prasada, S. 1999. Names of Things and Stuff: An Aristotelian Perspective. In *Language, Logic and Concepts eds by Jackendoff, R. et al.* 119-146. MA: MIT Press.
- Priss, U. E. 1998. *The Formalization of WordNet by Methods of Relational Concept Analysis, in WordNet: an electronic lexical database*, 179-196. MA: MIT Press
- Pustejovsky, J. 1995. *The Generative Lexicon*. MA: the MIT Press
- Reboul, A. 2000. Words, concepts, mental representations, and other biological categories. In *The Lexicon-Encyclopedia Interface ed.* Peeters B., 55-95. Amsterdam: Elsevier Science Ltd.

Submitted on: May 30, 2001

Accepted on: November 8, 2001