Adjunction Structure Formalism for English-Korean Machine Translation System

Jong-Hyun Kim & Hee-Sung Chung

In this paper, we present a knowledge representation which is focused on the correspondence of adjunction structure between English and Korean. We acquire the knowledge available in machine translation, from the linguistic structures and data in the two languages. The adjunction stucture is the constituent representing the relation between the modifier and the modifiee. Our approach is based on the unified method of knowledge description to formalize the correspondence between the two languages. We verify the effectiveness of our method for the purpose of constructing the English-Korean machine translation system.

1. Introduction

We are constructing an English/Korean machine translation system. In this paper, we present a knowledge representation system to describe the correspondence of adjunction structure between English and Korean. The linguistic data is drawn from the class of 'adverbials', including both lexical adverbs and prepositional phrases (PP). We classify the types of correspondence under the framework of unification-based grammar formalism, and utilize the linguistic criteria for the classfication of the corresponding types. In English grammar, the term 'adverbial' covers adjunct, subjunct, disjunct, and conjunct (Quirk et al. (1985)). The coverage of 'adjunction' in our approach is about the structure of modifier-modifiee. As the coverage of modifier-modifiee is a broad subject, we hereby deal with the 'adverbials' which most commonly overlap with the 'modifier'. The adverbials are classified in Korean as well, although not always by the same criteria available in English. The explicit classification and formalization of the adverbials, however, has not been given adequately as yet. We

describe generally the correspondence of the adverbials between the two languages, by designing the knowledge representation within adjunction structure (Chung (1989a), Kim (1988)).

In acquiring the corresponding knowledge, one language can be viewed as the metalanguage to the other language: an adjunction structure in Korean reflects an adjunction structure in English, with the requirements of some matching modifications to get the correspondence and also vice versa. Thus, describing the structure of the adverbials and getting the correspondence is incorporated into the frame of adjunction in our approach. In English, the adverbials denoting space or time, which work as space-time 'chunks', are more readily realized by way of the prepositional phrase. Nevertheless, they can be realized in several ways with regard to part-of-speech. In Korean, the ambiguities in the multiple selection of part-of-speech are far fewer.

In our approach, we are concerned with knowledge acquisition and representation of the adverbials, and also with the knowledge usages of those. We construct the knowledge base system to represent the correspondence in the process of translation. We focus on the problems of how to enlarge and apply the declarative knowledge represented in adjunction structure for the correspondence between the adverbials in English and those in Korean. A preliminary research for deciding the subcategorization of the significant verbs has been worked out (Chang (1989)). The classification of the correspondence of adjunction, in parallel with that of subcategorization, highlights the role of the evolving lexicon which will be used in our system.

In this paper, we first present the outline of knowledge representation in adjunction structure. Second, according to the structural analysis of 'adjunct', we prove that the one-to-many correspondence between the English source words and the Korean target words is reduced ultimately to a one-to-one correspondence. Third, we show an example of translation by the knowledge representation method. Describing the correspondence is based on the explicitness and readability of the knowledge description implicated in the logical framework of the adjunction structure.

2. Adjunction Structure and its Knowledge Representation

When it comes to the matter of knowledge representation, the task of

knowledge acquisition should also be considered. To expand the present knowledge base, we must acquire and systematize heuristic knowledge. The knowledge representation for machine translation is the structure of linguistic data and its processing algorithm.

In this section, we present the outline of knowledge representation in adjunction structure, which is necessary to describe the structure of modifier-modifiee. Let us examine the following expressions.

(2-1) a. the meaning of this sentence
b. the book on the table
c. the road to Lincoln
d. two years before the war
e. a man from the electric company

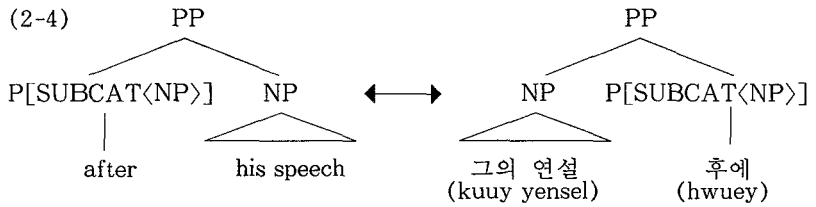
(2-2) a. the car outside the station
b. the car which was standing outside the station
c. The car was standing outside the station.

(2-3) a. In Chicago, he studied metaphysics.
b. He lived in Chicago.

A prepositional phrase is the most common type of postmodification in English. As shown in (2-1), there is a wide range of prepositions distributed among the prepositional phrase postmodifiers. The prepositional phrase (hereafter noted as PP) occurs as an optional element in (2-3a), and also occurs as an obligatory element in (2-3b). In the Head-driven style approach, the preposition in a PP is recognized as a head category which drives the local tree describing the structure of PP. We distinguish between the optional status and the obligatory status of the PPs in their category definition by way of feature structure. We therefore label the optional PP clause element or PP postmodification element as 'adjunct', even though we may call it 'modifier' in some cases.

In most cases, a PP in English is found to correspond to a postpositional phrase (also hereafter PP) in Korean, but in some cases, correspond to a different category. In a PP phrase, the head-final ordering in Korean corresponds to the head-first ordering in English by mirror image as in (2-4).

(2-4)



As the framework of knowledge representation for the correspondence, we use the local tree called adjunction structure and the principle manipulating each specified adjunction structure, in parallel with those in complementation structure (Pollard & Sag (1987)). From the viewpoint of grammar rules, each local tree is described declaratively as the constraints which are to be settled in the feature system. The Adjunct Feature (ADJUNCT) Principle in (2-5a) corresponds to the procedural knowledge which decides the usage of each distinctively specified local tree.

For further application in the various types of modifiers like the relative clauses and prepositional phrases, not to mention lexical adverbs, the symbolic local tree in (2-6a) is adjoined to another local tree as in (2-6b). The value of part-of-speech of head category is shared within endo-centric construction, which is true of the structure within the head projection path, according to the Head Feature Principle (Pollard & Sag (1987)). The local trees in (2-6) are for Korean, where a modifier occurs in front of the modifiee. If a modifier is positioned after the modifiee, as is common in English, it is necessary to convert the linear order between the daughters in the local tree.

In (2-6a) and (2-6b), the value of the adjunct feature is driven to unify with its head sister by the Adjunct Feature Principle, in the local tree where there is no execution of Subcategorization (SUBCAT) Principle in (2-5b). Now, to prevent readers from getting misled, we mention in advance that a category defined as to the adjunct feature is rendered an adjunct category, just as a category defined as to subcat feature is rendered a head category. In other words, one is the adjunct daughter category participating in adjunction structure, and the other is the feature name 'adjunct'. To avoid ambiguity beween them, another feature name might be used instead of the feature name 'adjunct'. The adjunct feature is both a kind of Head feature and a kind of category-valued feature. By the Head

feature principle in (2-5c), the value of adjunct feature is shared within the head projection path.

(2-5) a. ADJUNCT Feature Principle (AFP):
The value of the ADJUNCT feature (AD) of an adjunct, unifies with the whole feature structure of its head sister.

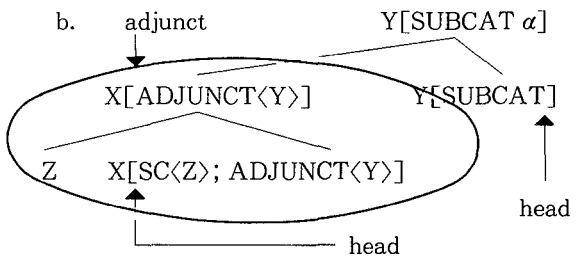
b. SUBCAT Principle (SCP):
The value of the SUBCAT feature (SC) of a head, unifies with the value of the SUBCAT feature of its mother except for the category which unifies with its complement sister.

c. HEAD Feature Principle (HFP):
The values of the HEAD features of a head are identical to the values of the HEAD features of the mother.

(2-6) (α denotes a set of categories, including the null set)

a.

b.



Each of the symbols X, Y, and Z above represents random unspecified feature structure which represents the minimal extension categories. We can start initially with the minimal extension categories, each of which is open to subsumption, and perform a series of unification. In specifying the feature structure that should be created when the category of the 'dummy' feature structure is settled in adjunction structure, each of the category symbols X, Y, and Z includes the value of part-of-speech (hereafter POS)

feature as the least necessary information of the already existent category. The heuristic extension is added, which is necessary and sufficient to describe the sentence structure in each concrete usage. The range of cooccurrence restriction with respect to POS between the modifier and the modifiee is revealed by the equation of the POS values between the adjunct daughter and the head daughter in each local tree. To put it somewhat differently, in the equation of the POS values, there must be no conflicts beween the POS value of the adjunct daughter and the POS value of the head daughter.

(2-7) a. X : POS=/1/
ADJUNCT=Y : POS=/2/

b. (in English;)

/1/ : preposition /2/ : noun	/1/ : preposition /2/ : verb

(in Korean;)

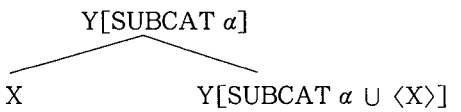
/1/ : postposition /2/ : noun	/1/ : postposition /2/ : verb

(in both Korean and English;)

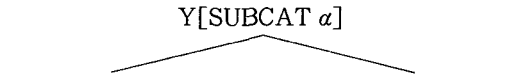
/1/ : relativizer /2/ : noun	/1/ : adverb /2/ : verb

We get the three kinds of local trees in (2-8) which are combined together to compose a whole tree. They are applicable to the structures with head-final ordering; yet, in the structures with head-first ordering, the linear precedence between the daughters in (2-8) needs to be reversed.

(2-8) a. Complementation Structure: M → C, H

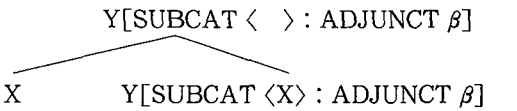


b. Adjunction Structure: M → A, H



c. Complementation Structure: M→C, H

(β denotes a category as a value of ADJUNCT feature)



Each local tree, which corresponds to the structure of a non-terminal category, is one of the three patterns (2-8a), (2-8b), and (2-8c). In the complementation structure of (2-8a), the head category is initialized with the multiset of its subcat attribute, and after popping up one element from the multiset, the returned multiset tracks progress through the head projection path. After a complement matches the first member from the subcat list structure of the head daughter, the remaining undischarged list members of the head is equated with the subcat list of the mother. Namely, a first-rest partitioned set of subcat list structure (which is connected by union; SC → { } ∪ {A1,A2,A3}, SC → {A1} ∪ {A2,A2}, SC → {A1} ∪ {A2} ∪ {A3}), represents a proposition about the connected sets, not just a set. The complementation structure as seen in (2-8c) is a pattern to be settled in itself as an adjunct daughter. And then, to the category given as the value of adjunct feature, certain heuristic extensions about the usage in sentence structure need to be re-entered in addition to the least specified values, as the constraints to be unified with head category. The extension of an initial category is done by partial ordering relation, which is transitive, reflexive and antisymmetric.

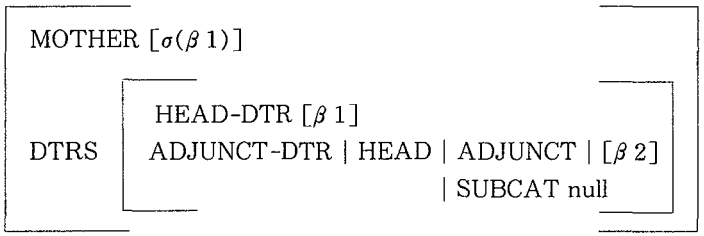
In a complementation structure, the head selects its complement sister by the subcat principle. In an adjunction structure, however, an adjunct selects its head sister, not vice versa, in our approach. In both Korean and English, if we assign the adjunct feature to the feature structure of a head category, it becomes difficult to substitute a head category with a minimal feature structure. The process of specifying the adjunct feature value of the modifier, not of the modifiee, is based on the unification under the substitution s=

{x←f(f(f(...)))}, where s(x) and s(f(x)) are both equal to f(f(f...)).

The adjunct feature principle is applied to the adjunct-head structure of pattern (2-8b). It is necessary to define the constraint blocking the application of the adjunct feature principle in a complementation structure as in (2-8c). So, in (2-9a), we revise the adjunct feature principle to describe just the structure of adjunct-head. The saturation (i.e. [subcat 〈 〉]) of the subcat value is the necessary condition in the application of the adjunct feature principle. In other words, the category, which has the null set value of the subcat feature and a non-null set value of the adjunct feature, is the adjunct daughter in the local tree of the adjunct-head structure. Simply put, we can state the constraint as in (2-10a) and (2-10b).

(2-9)

a. ADJUNCT Feature Principle:



(σ is a substitution such that σ(β 1) = σ(β 2), and σ(β 1) is the unification of β 1 and β 2.)

(2-10)

a. If, in the definition of a category, the SUBCAT list is non-empty and the ADJUNCT list is non-empty, the category is in itself the head of M → C, H

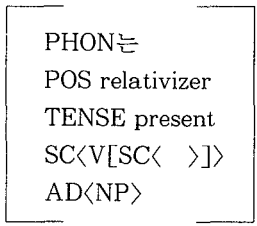
b. If, in the definition of a category, the SUBCAT list is empty and the ADJUNCT list is in itself the head of M → A, H

The optional status of the adjunct daughter in an adjunct-head structure is revealed by the motivation that the incentive for equation is met by the POS feature of the adjunct category, rather than the POS feature of the head category in a local tree. One critical point in assigning the adjunct feature value to adjunct daughter is how we can acquire the lexical specifica-

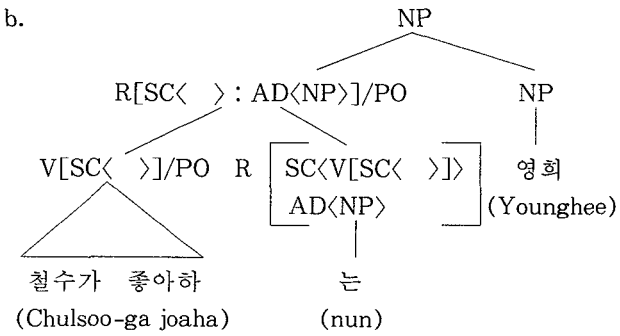
tion of certain (sometimes ad-hoc) English-specific or Korean-specific constraints definable within the adjunct feature value. In Korean, it is not misleading to identify distinct morphological words corresponding to the head in any adjunction structure. For example, in Korean, the relativizer is analyzed in terms of the lexically specified adjunct value. The analysis of the relativization structure in Korean is exemplified in (2-11).

Explaining the dependency with no 'long-distance' statement requries the declarative description of the inheritance (also called 'slash') feature using the sort of feature-passing and feature-discharging technique. In another paper, we are going to present the method to classify the structure correspondence of the relative sentences between the two languages. As the issue is also related with the analysis of unbounded dependency construction (an analysis in English was shown in Pollard & Sag (1990), Chap.4), it is beyond the scope of this paper. The analysis of unbounded dependency constructions is based on a global consequence of a linked series of feature correspondence in a set of local trees.

(2-11) a.



b.



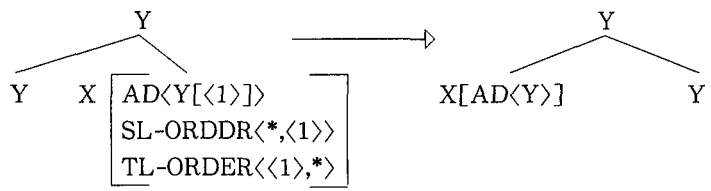
In making the correspondence of the adjunction structures between the two languages, the ordering restrictions can be accounted for more concise-

ly if constraints on order occurrence is imposed on the corresponding structures. There are two possible ways to order the corresponding structures. One is to reverse the order in either the head-first ordering or the head-final ordering, as shown in (2-12). The other is to maintain the order in either the head-first ordering or the head-final ordering, as shown in (2-13). The TL-ORDER represents the ordering in the target language, and SL-ORDER represents the ordering in the source language. The symbol '*' represents the current category. The number '⟨1⟩' represents the indexing of the adjunct value in an adjunct category. The order of the elements in the TL-ORDER represents that the first element is positioned in front of the second element in the generation of corresponding adjunction trees. The correspondence from the relativization structure in English to the one in Korean is true of the reversal case as seen in (2-12a). In the order correspondence of complementation structure, most cases are the reversal as seen in (2-12a), while it is the non-reversal case in (2-13b) in the production of subject NP plus VP.

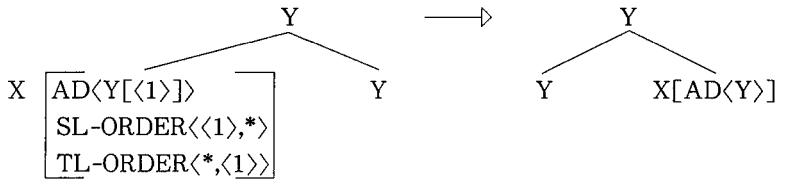
The specific order restrictions in each language are not trivial, especially when related with the configuration between free order and fixed order. In the enumerations of the canonical order correspondence between the two languages as seen in (2-12) and (2-13), we distinguish between the ordering in source language (SL-ORDER) and the ordering in target language (TL-ORDER). The whole set of SL-ORDER has to be acquired heuristically from the exhaustive partial orderings of the source language. At this point, the knowledge of TL-ORDER works meta-heuristically as to the SL-ORDER knowledge. Now, a point not to be overlooked is that all the pertinent information in analyzing sentence structure in a language is not always essential for the transfer procedure between two languages. It is a confusion that we would equate the transferring procedure with the syntactic and semantic analysis in a language.

(2-12) Reversing :

a. (head-first) → (head-final)

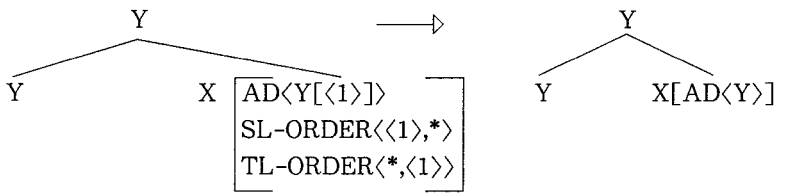


b. (head-final) → (head-first)

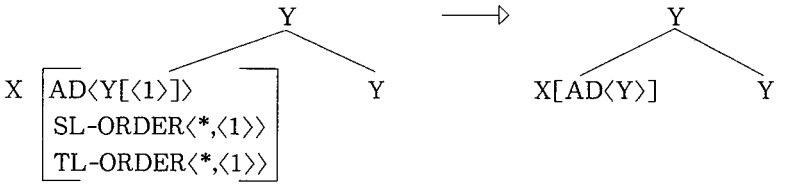


(2-13) Non-Reversing :

a. (head-first) → (head-first)

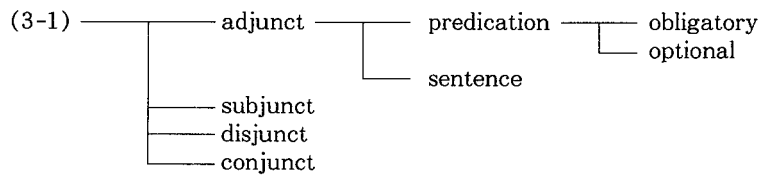


b. (head-final) → (head-final)



3. The Analysis of Complement and Adjunct in English Sentences

The adverbial elements differ considerably from the other elements of clause structure (S, V, O and C). The term 'adverbial' covers adjunct, subjunct, disjunct, conjunct (Quirk et al. (1985)), according to their differences in the realization of semantic roles. Unlike the other adverbials, in Quirk's classification, that which is called adjunct is similar in its use to the other sentence elements, S, C and O. The subtypes of adverbials, each of which is either lexical or phrasal, are classified as in (3-1) and the sample sentences are in (3-2).



(3-2) a. He put it on the table.
b. She kissed her mother on the cheek.
c. She kissed her mother on the cheek on the platform.
d. Resentfully, the workers have stood by their leaders.
e. Evidently, he doesn't object.

An obligatory predication adjunct 'on the table' in (3-2a) resembles an object both in its necessity for verb complementation and in its relative fixed position: the optional status of adjuncts in SVO clauses, as in (3-2c), can be tested by observing that the relation of the verb remains constant irrespective of the presence or absence of the adjunct.

In our definition of categories, the obligatory predication adjunct is a kind of oblique complement, and the optional predication adjunct or the sentence adjunct is a kind of adjunct. In Section 2, we used the terms 'adjunct' and 'complement' with that implication. Here, we use the term 'adjunct' to refer to the coverage of the adverbials as modifier, except in the case of the obligatory predication adjuncts which are classified into 'complement'. In the structure description of sentences, we distinguish between complementation and adjunction, and adjoin an adjunction structure without the modifications of the depths of the complementation structures. The implication is that adjuncts may be freely omitted, while omission of oblique complements leads to unexpected readings. Although the distinction between oblique complements and adjuncts is somewhat problematic (at least with respect to prepositional phrases), the point is that an oblique object fills a (possibly implicit) argument slot of the verb, whereas an adjunct restricts or modifies the relation expressed by the verb (Colban J. D. (1990 : 361)).

The verbs denoting a stationary position or movement are especially likely to take the obligatory predication adjunct: These fall within the purview of the obliquely subcategorized complements in our approach. Their abbreviated definitions are given in (3-3), (3-4) and (3-5). The order restrictions are accounted for by imposing constraints on the state sets of subcat features that is fundamentarily unordered in some sense. Namely, the order of the elements in the complements list reflects the 'obliqueness hierarchy (Pollard & Sag (1990))'. This hierarchy of subcat list in itself reveals the general ordering. The subject NP begins the list. The more oblique element is to the right. Each member encoding grammatical relations by the relative

positions in the subcat list represents the still unsaturated arguments of the verb predicate. Yet, in the case of adjunction, we show the linear precedence between head and adjunct, by the enumeration of SL-ORDER along with TL-ORDER as seen in (2-12) and (2-13).

(3-3) position (: lack of movement) verb:
V ∈ {be, live, work, keep, meet, stay, stop, work}
P ∈ {above, across, along, behind, beside, between}
ex) They kept it in the sofa.
; V| SC〈NP[sbj], NP[obj], PP[position]〉]
ex) She lives in a cottage.
; V[SC〈NP[sbj], PP[position]〉]

(3-4) movement 'with an end' verb:
V ∈ {lay, put, place, sit}
P ∈ {on, (into), (onto), ...}
ex) He sat the baby on the table.
; V[SC〈NP[sbj], NP[obj], PP[position]〉]

(3-5) movement verb:
V ∈ {drive, fly, run}
P ∈ {into, onto, out of, to, etc}
ex) A bird flew into my bedroom this morning
; V[SC〈NP[sbj], PP[directon]〉]

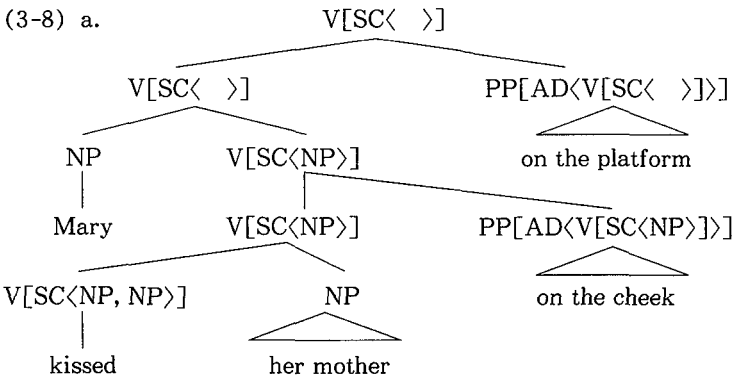
There is a difference between the optional predication adjunct and the sentence adjunct. As shown in (3-6), the most obvious way in which the sentence adjunct 'on the platform' marks itself off from the predication adjunct 'on the cheek' is by its relative freedom to occur at the initial position as well as at the end position in sentence. As defined in (3-7), the head word 'on' which drives the structure of the predication adjunct 'on the cheek' has, as its adjunct value, V[SC〈NP〉] which represents the VP. At the same time, the head word 'on' which drives the structure of the sentence adjunct 'on the platform' has, as its adjunct value, V[SC〈 〉} which represents the S. In other words, the VP category is represented by way of a SUBCAT list of length zero, and S category is represented by way of a SUBCAT list of length one. As the list unification algorithm states, two

lists of same length are unified on the condition that each of the corresponding pair of elements is unifiable. If a unification fails between two lists of different lengths, as a corollary, the whole unification fails. In the local structure of M → A H, the adjunct value of adjunct sister category checks the core representation of the head sister category: POS value∧ SUBCAT value- list. The tree structures of (3-6) are given in (3-8). In (3-8b), the comma (,) is described as a marker to be affixed to the adjunct. We refrain from the analysis of 'comma' in this paper, because there may be another analysis available.

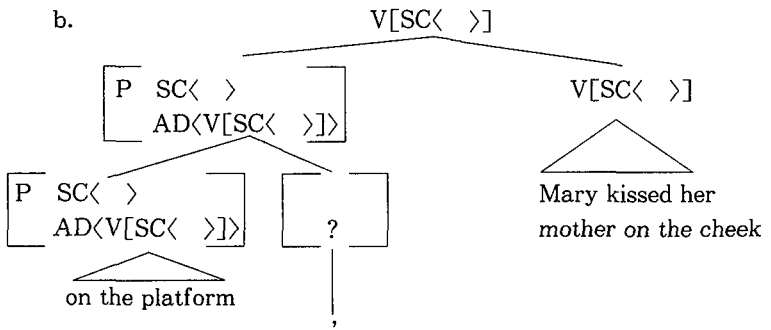
(3-6) a. Mary kissed her mother on the cheek on the platform.
b. On the platform, Mary kissed her mother on the cheek.

(3-7) a. on the cheek
on: P[SC⟨NP⟩; AD⟨V[SC⟨NP⟩]⟩]
b. on the platform
on: P[SC⟨NP⟩; AD⟨V[SC⟨ ⟩]⟩]

(3-8) a.



b.

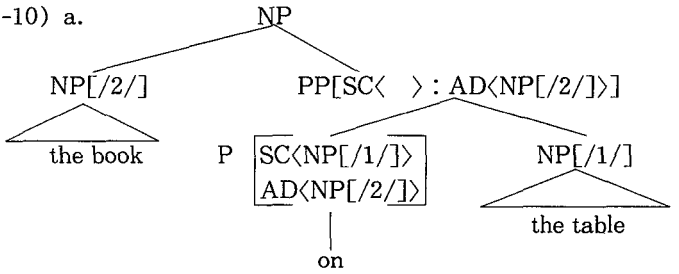


(3-9) a. the book on the table
b. the book is on the table

The above (3-9a) and (3-9b) retain the same meaning, They, however, are realized differently with regard to constituent structure. In (3-9b) the verb 'be' is a head category as a two-place relation, which takes, as the subcat feature value, both subj NP complement and oblique PP complement. In (3-9a) the preposition 'on' takes a NP complement as the subcat value and also takes, as the adjunct value, the NP which the PP 'on the table' modifies.

The verb 'be' in (3-9b), which takes two arguments (subj and oblique PP complements), combines with an argument (the PP complement) to produce a VP category, which is in fact an abbreviation of V[SC〈 〉]. The analyses of the tree structures are presented in (3-10) and (3-11).

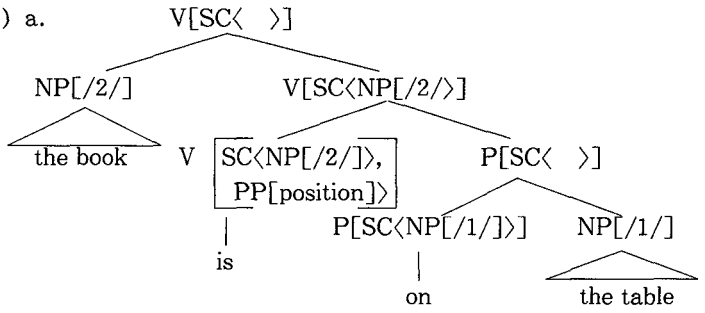
(3-10) a.



b.

PHON on
POS P
SC〈NP[/1/]〉
AD〈NP[/2/]〉

(3-11) a.



b. [PHON on; POS P; SC⟨NP[/1/]⟩] [PHON be; POS V; SC⟨NP[/2/], PP[]⟩]

On the other hand, a particular semantic role does not determine a particular type of realization structure. For example, the sentence adjuncts denoting time-position are realized in several ways in (3-12). The head word 'after' within the adjuncts 'after her speech' and 'after she spoke' is defined in two ways, as seen in (3-13).

(3-12) a. Apparently, he tried to telephone me yesterday.
b. Apparently, he tried to telephone me after his speech.
c. Apparently, he tried to telephone me after he spoke.

(3-13) [PHON after; [POS P; SC⟨NP⟩; AD⟨V[SC⟨ ⟩]⟩] ∨ [POS conjunct; SC⟨V[SC⟨ ⟩]⟩; AD⟨V[SC⟨ ⟩]⟩]]

4. Practical Examples of Correspondence

In a given English-Korean dictionary, let D_e be the set of words in English and D_k be the set of words in Korean. Let's consider the Cartesian product, $D_e \times D_k$, of the two sets. A mapping betwen the two sets may be defined as $I(D_e) = D_k$, which implies the image of D_e is D_k, taking the inverse mapping $I^{-1}(D_k) = D_e$. By generalizing the relation between the two sets, we may consider the word set of the source language to be the domain, and the word set of target language to be the range. Among the partitionings between the two sets, the relation includes three kinds of correspondene: 1) one-to-one 2) one-to-many 3) many-to-many.

Obviously, one-to-one correspondence is isomorphic. The lexical correspondence, however, is not one-to-one but one-to-many (or many-to-many) in most cases, which relations chiefly depend on the synonyms and homonyms of the two languages. As an elementary strategy for the trans-

lation of these correspondence, we adopted a normalization procedure which decomposes one-to-many or many-to-many correspondence into one-to-one correspondence.

We classify the correspondence between the target word (hereafter TWORD) and the source word (hereafter SWORD) according to each concrete usage, and then describe in lexicon the declarative knowledge of the correspondence. By the one-to-one normalization from the apparent one-to-many correspondence, a source word is translated into a target word. The description of a source word usage includes the information of the usage circumstances. As for the translation which is dependant on synonyms or homonyms, we specify the usage circumstance by the representaion of the canonical structures and semantic features.

There are two kinds of ambiguities characteristic of English, but not of Korean. First, in English, we can see multiple realizations in the selection of the part-of-speech of the modifying adverbials, as in the POS value paths in (4-1).

(4-1) a. He did the work in the room.
b. He did the work there.

(4-2) a. 그는 방안에서 그 일을 했다.
(kunun panganeyse ku ilul haissta)
b. 그는 그 곳에서 그 일을 했다.
(kunun ku koseyse ku ilul haissta)

Second, multiple part-of-speech selections of the constituent modified by the same adjunct can be found as in (4-3), which do not exist in Korean.

(4-3) a. the book on the table
b. I will see you on Monday.

Let us examine the following (4-4), (4-5), and (4-6). Among them, (4-4a) corresponds to (4-5a), and (4-4b) corresponds to (4-5b). (4-6a) corresponds to (4-6b). The postposition '| $ | (wiuy)' in (4-4a) is used as the head driving the genitive PP adjunct '책상 위의(chaysangwiuy)', and another postposition '위에(wiey)' in (4-4b) is used as the head driving the

PP complement '책상 위에(chaysangwiey)'. In English, however, the same preposition 'on' is used in both (4-5a) and (4-5b).

(4-4) a. 책상 위의 편지(chaysangwiuy phyenci)
b. 책상 위에 있는 편지(chaysangwiey issnun phyenci)

(4-5) a. the letter on the table
b. the letter which is on the table

(4-6) a. I heard the news on the radio.
b. 나는 라디오에서 그 소식을 들었다.
(nanun latioeyse ku sosikul tulessta)

In (4-5), the preposition 'on' in English can be used as the head both in a PP adjunct modifying NP and in a PP complement. Also, in (4-6a), the preposition 'on' is used as the head in a PP adjunct modifying VP. However, in Korean, distinct postpositions are used according to the status of the PP driven by a postposition. In the category definition of '위에(wiey)' in (4-4b), the value of adjunct feature is empty. The postposition '위의(wiuy)' denoting the possessive relation in (4-4a) is given only as Noun in the range of POS value of the adjunct feature. In English, the same preposition 'on' is used uniquely, regardless of whether the PP driven by 'on' is a complement of verb or an adjunct of NP or S. In Korean, however, one of those distinct postpositions '위의', '위에', and '에서' is used as the head driving a PP, depending on whether the PP is a complement of verb or an adjunct of NP or VP. Thus the specification of them is shown as in (4-7).

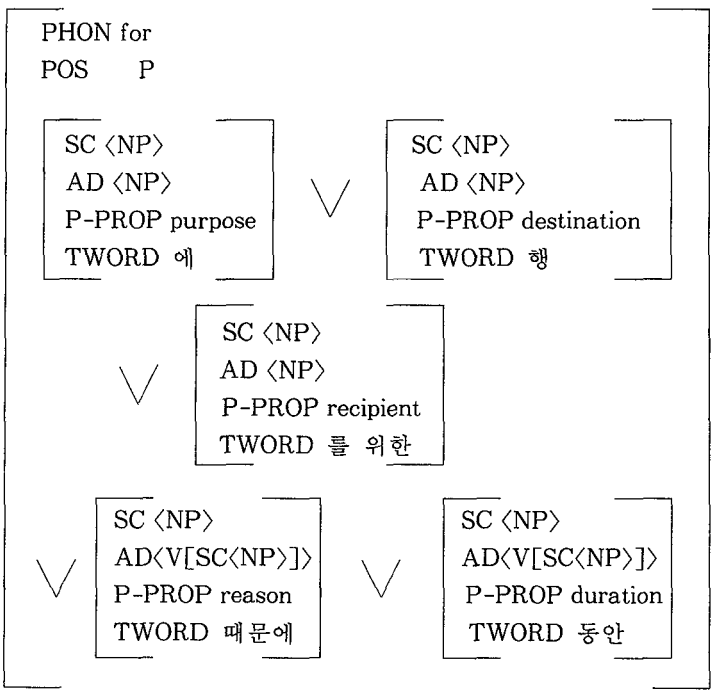
(4-7)

```
[ PHON on
  POS P
  [ SUBCAT ⟨NP⟩
    ADJUNCT ⟨NP⟩
    P-PROP position
    TWORD 위의 ]
  ∨
  [ SUBCAT ⟨NP⟩
    ADJUNCT ⟨ ⟩
    P-PROP position
    TWORD 위에 ]
  ∨
  [ SUBCAT ⟨NP⟩
    ADJUNCT ⟨V[SC⟨NP⟩]⟩
    P-PROP source
    TWORD 에서 ] ]
```

For another example, in the analysis of the preposition 'for' in (4-8), the multiple options for the TWORD are represented according to the classifications of semantic properties. We introduce the PROPerty feature (hereafter PROP) as a semantic constraint to describe the selection of TWORD in the usage of a lexical item. The PROP feature is necessary in selecting a TWORD corresponding to a SWORD. The P-PROP is the abbreviation of preposition-PROP (i.e. in X-PROP, X represents the first letter of the POS value of the category where the PROP is used).

(4-8) a. the best man for the job.(; TWORD 그 일에(ey))
b. This is the train for Seoul.(; TWORD 서울 행(hayng))
c. Here is a gift for you.(; TWORD 너를 위한(lul wihan))
d. I did it for the money.(; TWORD 돈 때문에(ttaemwuney))
e. He's been away for days. (; TWORD 여러 날 동안(tongan))

(4-9)



The subtypes 'subjunct', 'disjunct' and 'conjunct', which are incorporated within the purview of adjunct in our approach, have been undealt up to this

point. The term 'subjunct' is a subtype of the adverbials which have a subordinate role distinct from other predication or sentence adjuncts. By the same analogy, the adverbials called 'disjunct' are syntactically more detachable and in some respects superordinate, in that they seem to have a scope that extends over the sentence as a whole. The adverb 'carefully' in (4-10) is subjunct. The adverb 'clearly' in (4-11a) is a disjunct, and the 'clearly' in (4-11b) is a predication adjunct.

(4-10) Carefully, Leslie greeted the stranger casually.
(4-11) a. Clearly, he solved the problem.
b. He solved the problem clearly.

The usage of the adverb 'clearly' varies according to its relative position in a sentence. The 'clearly' in (4-11a) takes as its adjunct value V[SC〈 〉], and the 'clearly' in (4-11b) takes as its adjunct value V[SC〈NP〉]. The selection of a different TWORD in each usage depends upon the realization of each semantic property. By the use of the adjunct value and property value, we can get the normalization of one-to-one correspondene. As illustrated in (4-12), the source word 'clearly' corresponds to the target word '분명히(pwunmyeonghi)' if the usable property value is 'judgement' with regard to 'degree of truth', and corresponds to the target word '분명하게(pwunmyenghakey)' if the usable property value is 'manner'.

(4-12)

PHON clearly			
POS adverb			
SC〈 〉	∨	SC〈 〉	
AD〈V[SC〈 〉]〉		AD〈V[SC〈NP〉]〉	
A-PROP judgement		A-PROP manner	
TWORD 분명히		TWORD 분명하게	

There can also be found the case that a source word corresponds to one target word, for different type of properties. The disjunct 'technically' in (4-13a) bears the property of 'content', and the subjunct in (4-13b) bears the property of 'viewpoint', yet both correspond to the same TWORD '기술

적으로(kiswulcekulo)', as shown in (4-14). The distinct readings of the two are encoded with the semantic properties.

(4-13) a. Technically, our task is to recycle the waste products.
b. Technically, recycling the waste products will be easy.

(4-14)

PHON technically
POS adverb
SC〈 〉
AD〈V[SC〈 〉]〉

A-PROP content
TWORD 기술적으로

∨

A-PROP viewpoint
TWORD 기술적으로

The subtype called 'conjunct' expresses the speaker's assessment of the relation between two linguistic data, and is more like disjuncts than predication adjuncts in having a relatively detached and superordinate role as compared to other clause elements. For example, the conjunct 'though' in (4-15a) is defined as in (4-15b). Conditional forms encode the anaphoric condition that puts its argument slot into the role of an anteceding utterance. As a circumstance that modifies the relation of the main clause, conditionals are defined as sentence.adverbials.

(4-15) a. Though he is poor, he is satisfied with his situation.

b.

PHON though
POS conjunct
SC 〈V[SC〈 〉]〉
AD 〈V[SC〈 〉]〉
C-PROP concessive
TWORD 음에도 불구하고
(umedo poolkuhago)

5. The Types and Properties for Correspondence

In the problem of the knowledge acquisition bottleneck, with regard to the usage of the adjunct in the sentence structures, we presuppose that the usages are classified by the hierarchy of the subcat value and the adjunct value, which are 'canonical' head features. The phrase structure must be defined by, and hence rigorously conform to, preestablished 'canonical' rule format. It is the lack of adherence to well-defined canonical rule forms that makes it difficult to construct a covering grammar for the translation process. To underscore the role of grammar, let us review the canonical forms to define the phrase structure by the types and properties. A fundamental task of such an establishment is to recognize the individual building blocks of the sentence being translated.

The types for the correspondence of the adjunction structures can be described as in (5-1), except for (5-1b) which is for complementation. We put the combination of both the subcat list and the adjunct list as a type constraint. To distinguish between complement and adjunct, two aspects are intertwined: One is the linguistic analysis of linguistic data, and the other is the division between the head-complement matching and the adjunct-head matching. Any category, which is settled as the head of adjunct, is defined with respect to some of those types. We interpret an undefined value of adjunct feature as simply indicating that the current category does not have a value, rather than the value being 'unknown'. Henceforth, verbs (which never works as modifiers) does not have the combination of both the subcat list and the adjunct list, and have just the subcat list. Namely, while a word (which can work as both modifying adjunct and subcategorizing head) is given selectively the conjunction of SC and AD, the verbs within sentences do not have the definition of adjunct feature. Once we acquire the linguistic knowledge of the constituent structure by the conjunctively defined values, the knowledge retrieved from lexicon designates the put-to-use knowledge capable of recognizing the sentence structure.

(5-1) a. the head of the 'PP' postmodifier:

$$\text{TYPE-0} = \begin{bmatrix} \text{SC} \langle \text{NP} \rangle \\ \text{AD} \langle \text{NP} \rangle \end{bmatrix}$$

b. the head of the oblique 'PP' complement:

$$\text{TYPE-1} = \begin{bmatrix} \text{SC} \langle \text{NP} \rangle \\ \text{AD} \langle \ \rangle \end{bmatrix}$$

c. the head of the predication 'PP' adjunct:

$$\text{TYPE-2} = \begin{bmatrix} \text{SC} \langle \text{NP} \rangle \\ \text{AD} \langle \text{V}[\text{SC}\langle \text{NP} \rangle] \rangle \end{bmatrix}$$

d. the head of the sentence 'PP' adjunct:

$$\text{TYPE-3} = \begin{bmatrix} \text{SC} \langle \text{NP} \rangle \\ \text{AD} \langle \text{V}[\text{SC}\langle \ \rangle] \rangle \end{bmatrix}$$

e. the predication 'adverb' adjunct:

$$\text{TYPE-4} = \begin{bmatrix} \text{SC} \langle \ \rangle \\ \text{AD} \langle \text{V}[\text{SC}\langle \text{NP} \rangle] \rangle \end{bmatrix}$$

f. the sentence 'adverb' adjunct:

$$\text{TYPE-5} = \begin{bmatrix} \text{SC} \langle \ \rangle \\ \text{AD} \langle \text{V}[\text{SC}\langle \ \rangle] \rangle \end{bmatrix}$$

The one-to-many correspondence between SWORD and TWORD can be reduced to a one-to-one correspondence by the use of hierarchy types and semantic properties. The semantic properties are used to eliminate the am-

biguity in the hierarchical structure types. For the word that involves the ambiguity, we specify the lexical semantic features and introduce that into the syntactic feature system. Consequently, the constraints of lexical semantic features are described in the partial phrase struture, and plays a role of adjusting the semantic sensitive translation. As shown in (5-2), the one-to-one corresponding pair of 〈SWORD, TWORD〉 is obtained in the intersection of the constraints TYPE and PROPerty.

(5-2) a. The quarrel over pay was the reason for his resignation.
; [TYPE-0 : PROP aboutness: 〈over, 에 대한(e tayhan)〉]
b. They live over the road.
; [TYPE-1 : PROP orientation: 〈over, 너머에(nemeey)〉]
c. They threw a blanket over the door.
; [TYPE-1 : PROP destination: 〈over, 으로(ulo)〉]
d. We camped there over the holiday.
; [TYPE-3 : PROP duration: 〈over, 동안(tongan)〉]

The feature PROP is a kind of Head feature to be shared within the head projection path. In (5-3), some lists of PROPerties available in the prepositions 'at, on, in, to' are shown in the first column. The pairs 〈SWORD, TWORD〉 are exemplified in the intersections of the constraints TYPE and PROP. The blanks among the intersections of the constraints represent no existent usage found. In the divisions of these properties, the number [1] of position [1] or goal [1] represents point, the number [2] represents line (or surface), and the number [3] represents area (or volume).

In classifying the lists of PROPs, the properties in the usages of one source word have to be subdivided enough to recognize each semantic distinction and then get a one-to-one correspondence from one English source word to one Korean target word. We adopted the lists of PROPs based on Quirk's data classification. The whole lists, however, are not shown in this paper for convenience. Additional PROPs will be acquired in our foregoing works, while the efforts to remove some probable redundancies among the lists are being made continuously.

(5-3)

PROPerty	SWORD	type 0 TWORD	type 1 TWORD	type 2 TWORD	type 3 TWORD
position[1]	at	에서의 (eyseuy)	에서 (eyse)	에서 (eyse)	에서 (eyse)
position[2]	on	위의 (wuy)	위에 (wiey)	에서 (eyse)	에서 (eyse)
position[3]	in	안의 (anuy)	안에 (aney)	에서 (eyse)	에서 (eyse)
goal[1]	to	으로의 (ulouy)	으로 (ulo)		
recipient	to		에게 (eykey)		
duration	to			까지 (kkaci)	까지 (kkaci)
reaction	to				으로 (ulo)

6. Translation Example by the Knowledge Representation

As mentioned in Section 2, adjunction structure serves to describe the structure of the modifier and modifiee. Viewed from compositionality, the constituent structure of a word sequence is made by putting together local trees, each of which represents either complementation structure or adjunction structure. A word described declaratively in the lexicon includes, as the constraints, the heuristic knowledge of the word usage. The knowledge must be necessary and sufficient to use the word in the structures of sentences. To select the most suitable TWORD, the knowledge description of the correspondence is defined within lexical items.

In translating a source sentence into a target sentence, the knowledge representation in the adjunction structure is extensively used in generating the corresponding local trees. By the selection of the suitable TWORD for a SWORD, the processing of the lexicon-driven translation is done in the course of getting the whole tree from the partial information obtained from each local tree.

We present an example of the translation processing in the Appendix. The number '⟨3⟩' in TL-ORDER represents the indexing of the adjunct value in the adjunct category. The bracketed numbers '[1], [2]' represent the indexing of the complement in subcat list.

7. Future Works

In previous approaches, the basic researches did not deal extensively with the correspondence between the two languages. That is due to the shortage of linguistic knowledge engineers, as well as the differences in linguistic structures and the vastness of linguistic data which must be dealt with. In our opinion, the acquisition of empirical knowledge of linguistic data is the prerequisite to developing the machine translation system. Our divide-and-conquer approach for constructing the English-Korean machine translation system in progress is as follows.

First, we formalize the knowledge of corresponding structures according to an exact analysis of linguistic data. Second, in using the method of knowledge representation, we focus on preserving the monotonicity of the system. Third, we develop an effective processing algorithm for constructing the English-Korean machine translation system.

Our approach for constructing the system must include the precisement of logical study and experimental study; The former is given by the mathematical formalization, the latter by the correspondence between the two languages. In this paper, we presented the outline of the knowledge representation between the two languages, and showed the procedure for normalizing the correspondence. As a consequence of our research, we have found that the information-based grammar fomalism is more efficient than the other procedural knowledge representation formalisms, in constructing a machine translation system. As our future works for the revision of this paper, the remaining problems include a more precise analysis of the correspondence between the two languages.

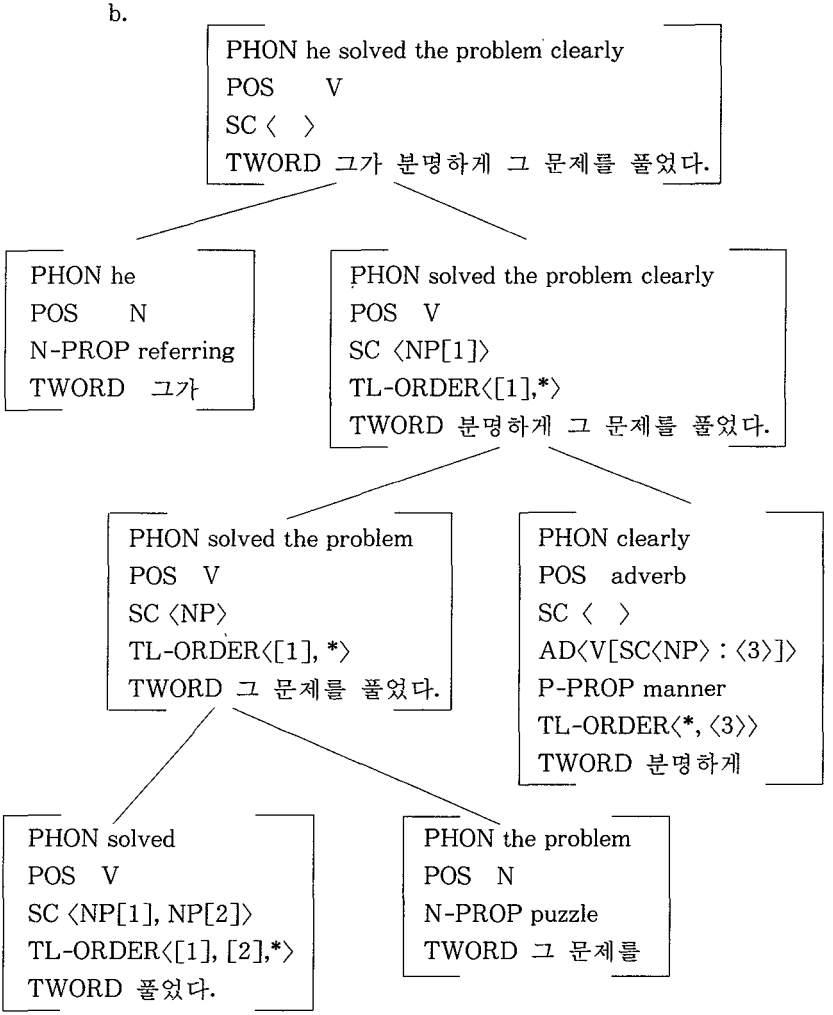
〈Appendix〉

(6-1) a. He solved the problem clearly.

(translate) → 그가 분명하게 그 문제를 풀었다.

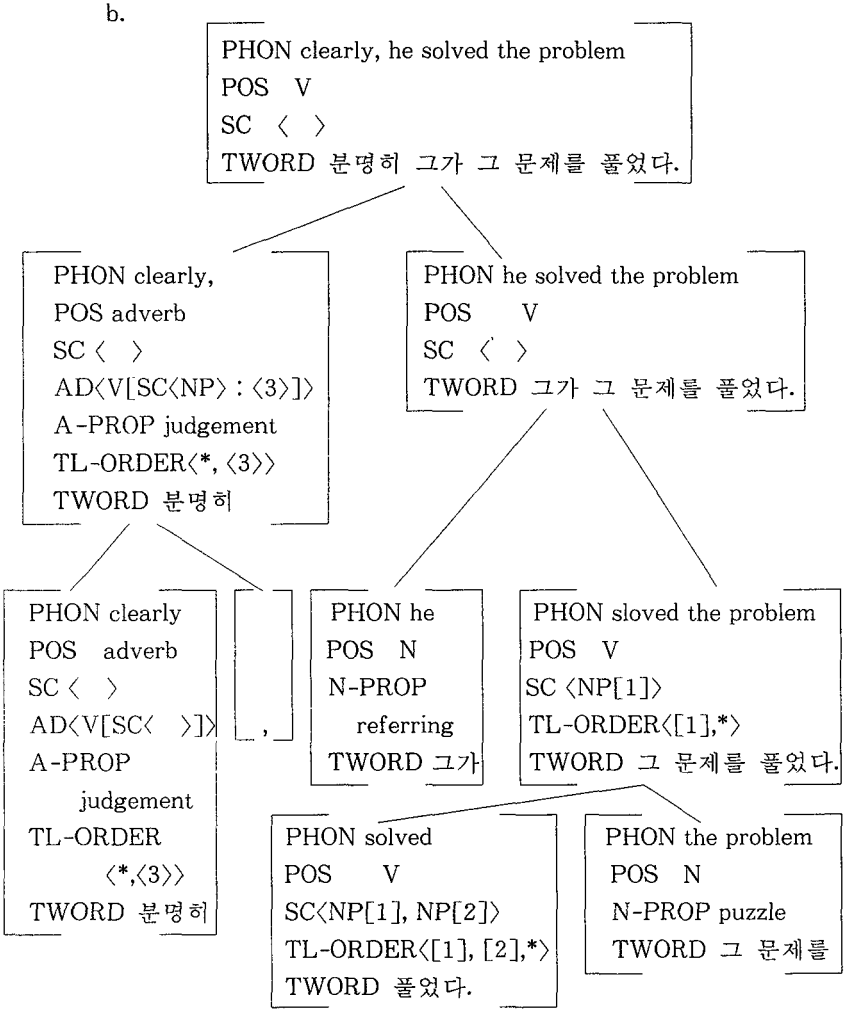
(kuka pwunmyenghake ku mwuncey-lul phyuessta)

b.



(6-2) a. Clearly, he solved the problem.
(translate) → 분명히 그가 그 문제를 풀었다.
(pwumyenghi kuka ku mwuncey-lul phyuessta

b.



References

Chang, S. J. (1980) *A Research Report about the Correspondence of Verbs between Korean and English*, Electronics and Telecommunications Research Institute, Korea.

Chung, H. S. (1986) *Korean Information Processing*, Ph. D. dissertation, University of Tokyo (in Japanese).

Chung, H. S. (1989a) 'Natural Language Processing Using the Unified Knowledge Representation,' *ETRI Journal* Vol 11-1, Electronics and Telecommunications Research Institute(in Korean).

Chung, H. S. (1989b) *Korean/English Machine Translation System*, System Engineering Center, KAIST, Korea (in Korean).

Crow, J. M. (1990) 'Locations Now and Then,' *CSLI Lecture Notes* 22, Stanford University.

Gazdar, G., Klein, E., Pullum, G.K. and Sag, I. (1985) *Generalized Phrase Structure Grammar*, Basil Blackwell.

Kim, J. H. (1988) 'Control and Unbounded Dependency in Korean,' ms. thesis, Seoul National University (in Korean).

Pollard, C. and Sag, I. (1987) *An Introduction to HPSG* (*Information-based Syntax and Semantics Volume 1*), CSLI, Stanford University.

Pollard, C. and Sag, I. (1990) *Information-based Syntax and Semantics Volume* 2, CSLI, Stanford University (in preparation).

Quirk, R., Greenbaum, S., Leech, G. and Svartivik, (1985) *A Comprehensive Grammar of the English Language*, Longman Group.

Jong-Hyun Kim
Department of Linguistics
Seoul National University
San 56-1 Sinlim-dong, Kwanak-ku
Seoul 151-742
Korea

Dr. Hee-Sung Chung
Korea Academy of Industrial Technology
70-6 Yangjae-dong, Seocho-ku
Seoul 137-130
Korea