

# Parsing Empty Categories in Korean\*

Sungki Suh

The Korean sentences involving empty subject(s) raise a serious question of indetermination in parsing. The fact that such sentences are parsed without conscious processing difficulty suggests either no reanalysis is involved there or the reanalysis, if any, is automatic. Both possibilities are pursuable and the empirical coverage is satisfactory in both cases. For conceptual reasons, however, we pursue the possibility in which no reanalysis is assumed and a phrase structure permitting an empty category in the matrix subject position is posited. Considering the syntactic properties of Korean, to postulate such a structure is reasonable from the syntactic viewpoint as well as from the parsing perspective.

## 1. Introduction

The universality of the human parser has been one of the most important issues in the recent proposals on natural language processing. Many proposals, however, are based on the analysis of the English data and hence it is necessary to investigate the parsing phenomena from languages other than English if we want to verify the universality of such proposals. In this respect, examining Korean would be meaningful since Korean is typologically very different from English. If the parsing phenomena from the two totally different types of languages can be accounted for by certain principle(s), that can be a good motivation for hypothesizing a universal parser equipped with such principles.

This paper concerns the processing of empty categories (EC) in Korean. In particular, ECs traditionally classified as empty pronoun (*pro*) are considered. First it will be shown that sentences containing *pro* raise the problem of indetermination in parsing. Then two proposals will be made to cope with such a problem: One is to assume that reanalysis is always

---

\* This work was supported by a grant from the Language Research Institute, Seoul National University (1995).

syntactic structure not observed in English-type languages. Both proposals show comparable empirical coverage and make correct predictions. But we will pursue the latter possibility for conceptual reasons. Finally we will return to the issue of the universality of the human parser.

## 2. The Problem with an Empty Subject

The following is a typical example showing the problem of indetermination.

- (1) Inho-ka Mina-lul mannassta-ko malhayssta  
       -Nom       -Ac           met-Comp said  
       'Inho said that he met Mina.'

Without a specific context where there is a prominent noun acting as a discourse topic, the subject of the matrix predicate and that of the embedded predicate must be coreferent in (1); i.e., the interpretation should be 'Inho said that he (himself) met Mina.' Traditionally, (1) has been considered to have a structure like (2).

- (2) Inho<sub>1</sub>-ka [e<sub>1</sub> Mina-lul mannassta-ko] malhayssta

The problem with (2) from the parsing perspective is that the parser cannot build a structure until the matrix verb, which is the last word of the sentence, is encountered. Notice that, although the complementizer on 'mannassta-ko' signals that there will be upcoming lexical items (such as the matrix verb), that doesn't guarantee that the first three words make up a structure like (2), since it is also possible that they constitute a clause proposed to the sentence-initial position, as shown in the following.<sup>1</sup>

- (3) [Inho-ka Mina-lul mannassta-ko]<sub>1</sub> [Yengi-ka t<sub>1</sub> malhayssta]

---

<sup>1</sup> It is reasonable to assume that if the parser makes structural decision between the two possibilities, (i) and (ii) below, it will choose (i) over (ii) since the former is the simpler structure at the moment.

(i) [Inho-ka Mina-lul mannassta-ko] ...

(ii) [Inho<sub>1</sub>-ka [e<sub>1</sub> Mina-lul mannassta-ko] ...]

Note that pursuing the simplest structure upon facing structural ambiguity has been considered as a basic property of the human parser since Frazier (1978) proposed the principle of Minimal Attachment as in the following:

(iii) *Minimal Attachment: Attach incoming material into the phrase marker being constructed using the fewest nodes consistent with the wellformedness rules of the grammar.*

'Yengi said that Inho met Mina.'

The parser, therefore, cannot determine the structure until it sees the next word 'malhayssta'. When it is encountered, the parser realizes that the previous lexical items do not constitute a clause. Hence, at that point, 'Inho-ka' should be placed in the matrix subject position and an empty pronoun should be inserted in the embedded subject position.

Such a problem of indetermination becomes more serious when the parser deals with sentences involving degree 2 embedding or more such as (4).

- (4) Inho-ka hoysa-lul kumantunta-ko      hyeng-eykey malhayssta-ko  
       -Nom job-Acc           quit-Comp brother-to           told-Comp  
 cwucanghayssta  
 claimed  
 'Inho claimed that he told his brother that he would quit his job.'

- (5) Inho<sub>1</sub>-ka [e<sub>1</sub> [e<sub>1</sub> hoysa-lul kumantunta-ko] hyeng-eykey  
 malhayssta-ko] cwucanghayssta

- (6) [Inho-ka hoysa-lul kumantunta-ko] ...

Suppose that (5) is the structure for the string (4). Here, it is simply impossible, up to the point of seeing the verb 'malhayssta-ko', to determine which clause 'Inho-ka' belongs to or how many empty pronouns should be posited. Only when the matrix verb 'cwucanghayssta' is encountered, the parser can determine the status of 'Inho-ka' as a matrix subject and the number of empty pronouns posited in (5).

The observations thus far suggest the following: Unless we hypothesize that the parser has an unlimited lookahead device, (which is not adopted by any parsing model), structures such as (2) or (5) cannot be computed without some sort of reanalysis procedure. This follows from the common assumption that the human parser computes structure in a non-delay, incremental fashion. By incremental parsing, we mean that the parser tries to build grammatical structure as each lexical item is received, and hence in (4), at the point of receiving the third item 'kumantwunta-ko', the parser computes a clausal structure with the lexical items received thus far, as in (6) above. However, such a structure is not compatible with the subsequent string, and thus reanalysis is inevitable.

It is of importance to note that since sentences such as (4) do not display

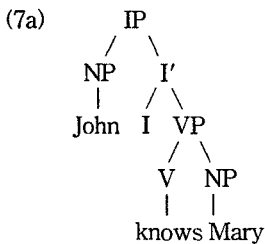
any indication of a garden path sentence, the reanalysis involved in computing (4), if any, must be unconscious, i.e., creating no conscious processing difficulty.

Suh (1994) claims that the processing of (1) or (4) indeed involves reanalysis and such reanalysis is characterized as typical unconscious reanalysis. In the following, the analysis of Korean examples in Suh (1994) will be reconsidered. Then we consider an alternative view, originally from Suh (1992), which proposes a syntactic structure different from that in (2) or (5).

### 3. Structure Addition as Unconscious Reanalysis

Assuming that human parsing is serial<sup>2</sup>, reanalysis component is necessary in the human parser due to the ambiguities observed in natural languages. From Frazier (1978), it has been observed that certain types of reanalyses do not cause conscious processing difficulty, and how to characterize such reanalyses has been an important issue in psycholinguistic research. (cf: Gorrell 1995, Marcus, Hindle & Fleck 1983, Pritchett 1992, Weinberg 1993) What is common in those proposals seems to be that reanalysis characterized as simple 'structure addition' is unconscious. The following examples are relevant to this point.

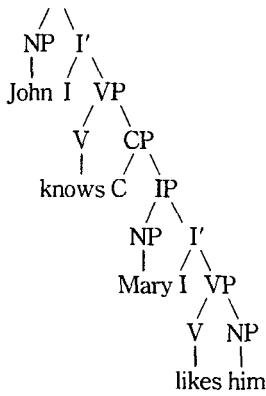
(7) John knows Mary likes him.



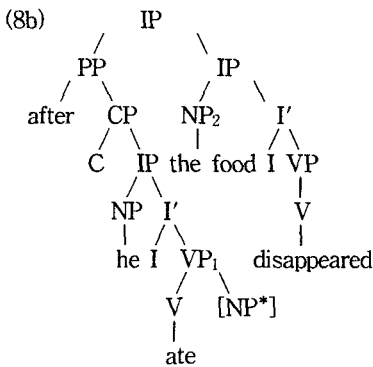
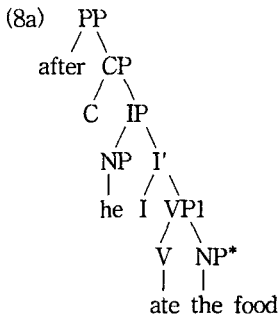
(7b) IP

---

<sup>2</sup> This assumption is not uncontroversial, although most recent proposals on the human sentence processing mechanism assume parsing is serial. Sometimes the difference between serial and parallel parsing in psycholinguistics is not so clear since many parallel models are not really parallel in a strict sense. See Kurtzman (1985) and Gorrell (1987) for the related issue.



(8) #After he ate the food disappeared from the table.



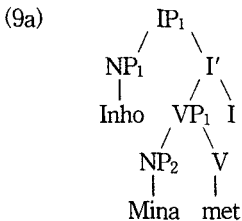
Given that the parser tries to compute the minimal grammatical structure as each lexical item is encountered, 'Mary' in (7) and 'the food' in (8) will be attached initially as the object of 'knows' and 'ate', as in (7a)/(8a). The

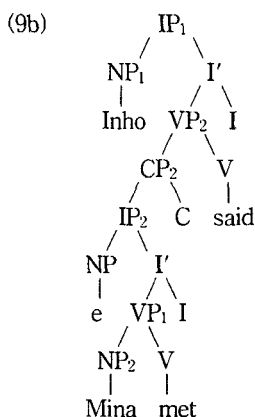
subsequent string, however, requires that they be attached as a subject of a complement clause in (7) and a subject of a matrix clause in (8), respectively. Although both instances involve a misanalysis of a subject as an object, processing difficulty is observed only in one case: namely, when the subject of a matrix clause is misanalyzed.

Such a contrast can be accounted for by employing the notion of 'structure addition': The reanalysis from (7a) to (7b) can be done by simply adding clausal structure (= CP/IP structure) above the initially misanalyzed NP 'Mary'. On the other hand, the reanalysis from (8a) to (8b) cannot be done by simple structure addition. Crucially, the NP\* node posited under VP<sub>1</sub> in (8a) must be deleted in order to derive (8b). Such deletion, which is contrasted with structure addition in nature, is considered to be responsible for the conscious processing difficulty in (8).

Suh (1994) claims that the contrast between structure addition and structure deletion is also observed in parsing Korean. In particular, he points out that sentences such as (9), repeated from (1) above, involve reanalysis characterized as structure addition and hence do not cause conscious difficulty.

- (9) Inho-ka Mina-lul mannassta-ko malhayssta  
 -Nom -Acc met-Comp said  
 'Inho said that he met Mina.'





Assuming the strategy of Minimal Attachment, the first three items in (9) will be structured initially as one clause, as in (9a). However, as soon as the next input item ('malhayssta') is encountered, 'Inho-ka' is expelled from the clause headed by 'mannassta-ko'. As seen in (9b), this procedure amounts to building clausal structure (= VP<sub>2</sub>, CP<sub>2</sub>, and IP<sub>2</sub>) above VP<sub>1</sub>. Since the reanalysis from (9a) to (9b) patterns with typical structure addition, no garden path results as expected.

The above explanation based on the idea of unconscious reanalysis applies to other similar cases in Korean, as pointed out in Suh (1994). Also, the notion of structure addition as unconscious reanalysis has been proved to be efficient in accounting for the data from other languages as well.<sup>3</sup> In terms of empirical coverage, then, there may be nothing to be desired. However, the above analysis doesn't seem to be really appealing conceptually. Notice that sentences like (9) are more frequently used than their counterparts containing an overt pronominal, such as (10) or (11).

(10) Inho-ka [caki-ka Mina-lul mannassta-ko] malhayssta

(11) Inho-ka [ku-ka Mina-lul mannassta-ko] malhayssta

(11) sounds awkward for the intended meaning probably because 'ku' in such a structure strongly prefers the disjoint reading (with the matrix subject). For this reason, what is actually used in spoken Korean for the intended meaning is (9) and (10) rather than (11). Between the two, (9) sounds more natural and is preferred, at least in spoken Korean. Given this,

<sup>3</sup> See Gorrell (1995) and Weinberg (1993).

the following question naturally arises: Is it reasonable that sentences with the more frequently used structure are always processed via reanalysis whereas sentences with the less frequently used one are processed without reanalysis? The answer could be yes or no, depending on how trivial the cost of reanalysis is.

Even though there is a growing consensus that the contrast between conscious and unconscious reanalysis is real, it is not clear whether unconscious reanalysis is completely costless; surely it doesn't create conscious processing difficulty, but that doesn't necessarily mean the cost of reanalysis is zero.<sup>4</sup> After all, this seems to be a very tricky issue, but one thing is clear: The idea that it is the most frequently used structural type such as (9) which always causes reanalysis doesn't consist with our intuition.

Given such conceptual inadequateness, we will pursue an alternative structure for (9) below, which does not require reanalysis in the processing of it.

---

<sup>4</sup>In fact, there have been some experimental results showing that unconscious reanalysis can be associated with measurable processing cost. For instance, the eye movement study in Rayner & Frazier (1987), which focused on the following structural types, suggests that the supposed unconscious reanalysis involved in the processing of sentences such as (7) results in some processing cost.

- (i) The zoologist observed that the behavior of the animals was distinctly abnormal.
- (ii) The zoologist observed the behavior of the animals to be distinctly abnormal.
- (iii) The zoologist observed the behavior of the animals to be distinctly abnormal.
- (iv) The zoologist observed the behavior of the animals in several settings.

Note that reanalysis as in (7) is required in the processing of (ii) and (iii) type, given that 'the behavior of the animals' is initially misanalyzed as the object of 'observed' due to the principle of Minimal Attachment. On the other hand, (i) and (iv) are parsed without reanalysis: The presence of complementizer 'that' in (i) signals that 'the behavior of animals' is the subject of the (upcoming) complement clause. In (iv), the initial structural decision based on Minimal Attachment is compatible with the subsequent string. Hence, in both (i) and (iv), no reanalysis is necessary.

The experimental results from Rayner & Frazier (1987) are the following: The null complementizer sentence (ii) and infinitival sentence (iii) took longer to read overall than the overt complementizer sentence (i) or the nominal complement sentence (iv). Also, there was some indication of disruption in processing in the disambiguating region (= the words following 'the behavior of the animals') of (ii) and (iii). Such results suggest that, although the reanalysis involved in (7) or (ii)/(iii) does not cause garden path, there is nevertheless a good possibility that it is associated with some processing cost.



#### 4. The Alternative Structure and a Name-like Empty Category

As originally proposed in Suh (1992), the following is another possibility for the syntactic structure of (9).

(12)  $e_1$  [Inho<sub>1</sub>-ka Mina-lul mannassta-ko] malhayssta

The motivation for proposing (12) is from the parsing perspective, but, as will be shown below, it can also be justified from the theoretical syntactic viewpoint.

##### 4.1. The Motivation for Positing (12)

The immediate advantage with adopting (12) as the syntactic structure for (9) is that the parser can incorporate each word into a present constituent structure as soon as it is encountered. Moreover, no reanalysis procedure is necessary in building a structure like (12); the parser builds up the embedded clause first, and structures the matrix clause next simply by positing an EC in the matrix subject position when the main verb is encountered. Note that such an order of structure building is inevitable from the head-final property of the Korean language. Even when the matrix subject is overt as in (10)/(11), the matrix clause is built after the embedded clause is done, since the matrix verb always appears in the sentence-final position.

There seem to be other cases where postulating a structure analogous to (12) is motivated. Consider the following.

(13) Inho-ka Mina-lul mannassta-ko malhayssta  
       -Nom       -Acc           met-Comp said

(14)  $e_1$  [Inho<sub>2</sub>-ka Mina-lul mannassta-ko] malhayssta  
       'Someone(=e) said that Inho met Mina.'

(15) Inho<sub>1</sub>-ka [ $e_2$  Mina-lul mannassta-ko] malhayssta  
       'Inho said that someone (=e) met Mina.'

(13) is an instance where the empty subject is not coreferential with the overt subject 'Inho-ka'. In an appropriate context containing a discourse topic, it is possible that the empty subject is coreferential with the discourse

topic. Logically, we have two possibilities here: It is interpreted either as (14) or as (15). In actual conversation, however, it is very difficult to interpret the string (13) as (15), even if there is a prominent NP acting as a discourse topic. Note that if (13) is interpreted as (15) at all, a long pause should be placed after 'Inho-ka'.

Meanwhile, given a discourse topic, (13) is naturally interpreted as (14) without any prosodic cue. Such a contrast between (14) and (15) suggests that the parser usually builds a structure like (12) or (14) from the string (9). Obviously, a structure like (15) is a marked case; that is, only when a special cue such as prosodic information or particular context is available, the parser builds a structure like (15).<sup>5</sup>

#### 4.2. Syntactic Arguments for Positing (12)

Thus far, I have shown that from the parsing perspective, it is desirable to postulate a structure in which an empty category is placed in the matrix subject position. Below I will argue that positing a phrase structure analogous to (12) is necessary from the syntactic viewpoint, and representations like (12) are legitimate. A potential problem with (12) is that Binding Condition C in Chomsky (1981) is violated. However, particular syntactic properties of the Korean language lead us to conclude that (12) is not an

---

<sup>5</sup>The contrast between structures like (14) and (15) becomes clearer when we consider sentences involving more degree of embedding:

- (i) hyeng-i Inho-ka Mina-lul manna-nun-kes-ul poassta-ko hayssta  
 brother-Nom -Nom -Acc meet-Comp-thing-Acc saw-Comp said
- (ia) e [hyeng-i [Inho-ka Mina-lul manna-nun-kes]-ul poassta-ko] hayssta  
 'Someone said that my brother saw Inho meet Mina.'
- (ib) hyeng-i [e [Inho-ka Mina-lul manna-nun-kes]-ul poassta-ko] hayssta  
 'My brother said that someone saw Inho meet Mina.'
- (ic) hyeng-i [Inho-ka [e Mina-lul manna-nun-kes]-ul poassta-ko] hayssta  
 'My brother said that Inho saw someone meet Mina.'

(i) is an instance of degree-2 embedding sentence where there are only two overt subject NPs and the empty subject is coreferential with none of the NPs within the sentence. In this case, positing structure (ia) doesn't require any particular context or prosodic cue; an NP in the previous sentence, for instance, can easily be construed as the antecedent of 'e'. Meanwhile, postulating structure (ib)/(ic) seems really difficult. In fact, the speakers I consulted with answered that interpreting (i) as (ib)/(ic) is almost impossible even if particular context is provided. Such a contrast between (ia) and (ib)/(ic) clearly tells us which structure the parser prefers.

instance of Condition C violation.

#### 4.2.1. Binding of the Reflexive and the Empty Category

(16) A : Inho<sub>1</sub>-ka ecey eti-ey kassess-ci  
 -Nom yesterday where-to went-Q

'Where did Inho go yesterday?'

B1 : e<sub>1</sub> kukcang-ey kassesse  
 movie theater-to went

'He went to the movies.'

B2 : \*caki<sub>1</sub>-ka/-nun kukcang-ey kassesse  
 self-Nom/-Top movie theater-to went

'He went to the movies.'

(16) is a part of a conversation containing reflexive 'caki' (or an EC). B can answer A's question by using an EC as a subject of the sentence. (It is also possible to use a topic-marked subject, 'Inho-nun' or 'ku-nun'.) But, as shown in B2, it is not possible to use a reflexive as a subject. Such a contrast suggests that reflexive 'caki' must be bound by its antecedent within a sentence while an EC may be bound by its antecedent from outside of the sentence.<sup>6</sup> Based on this observation, let us consider the following conversation.

---

<sup>6</sup> It has sometimes been pointed out that there are cases where 'caki' is bound by its antecedent from outside of the sentence. The following is such an instance from Yang (1986).

A : Inho<sub>1</sub>-ka kukos-ey salam-ul ponayss-ni  
 -Nom there-to man-Acc sent-Q

'Did Inho send somebody there?'

B : ani, caki<sub>1</sub>-ka cikcep kass-e  
 no self-Nom in person went-Dec

'No, he himself went there.'

However, examples like this do not seem to be real evidence against the above observation that 'caki' should be bound within a sentence, since such a usage of 'caki' is possible only when it conveys a contrastive meaning. As shown in the above translation, employing 'caki' has the effect of emphasizing that it is Inho himself, not some other person, who went there. It is precisely such a situation which allows 'caki' to be bound by its antecedent from outside of the sentence. Therefore, it seems reasonable to consider that there are two types of 'caki', pure reflexive and contrastive (or emphatic) pronominal, and only the latter may violate Principle A. Throughout the discussion, we will exclude 'caki' used as a contrastive pronominal.

- (17) A : Inho<sub>1</sub>-ka ecey mwuelako malhayss-ni  
 -Nom yesterday what said-Q  
 'What did Inho say yesterday?'

B : caki<sub>1</sub>-ka Mina-lul top-keyss-ta-ko haysse  
 self-Nom -Acc help-will-Dec-Comp said  
 'He said he would help Mina.'

- (18) caki<sub>1</sub>-ka [<sub>e<sub>1</sub></sub> Mina-lul top-keyss-ta-ko] haysse

- (19) e<sub>1</sub> [caki<sub>1</sub>-ka Mina-lul top-keyss-ta-ko] haysse

In (17B), there must be one null subject since it is a degree 1 embedding sentence and yet it has only one (overt) subject NP. Although there are logically two possibilities, (18) and (19), for the syntactic structure of (17B), the binding properties of 'caki' and the EC observed in (16) strongly suggest that only (19) is the legitimate structure for (17B). Notice that (19) is structurally analogous to (12) or (14) in the previous section. From the observation thus far, it follows that positing an EC in the matrix subject position is not a surprising but a reasonable proposal from the syntactic viewpoint. One thing should be clarified, however: What is the nature of the EC in (12) or (19) and how can we classify it in the system of empty categories?

#### 4.2.2. The Nature of the EC and Binding Condition C

A potential problem with a structure like (12) is that Binding Condition C in Chomsky (1981) is violated: A name is bound (by an EC). However, the following examples suggest that in Korean, Binding Condition C does not apply in the same way as in English.

- (20) Inho<sub>1</sub>-ka Inho<sub>1</sub>-lul piphanhayssta  
 -Nom -Acc criticized  
 'Inho criticized himself.'

- (21) ku-uy hyeng-i Inho<sub>1</sub>-ka Inho<sub>1</sub>-lul piphanhayssta-ko malhayssta  
 his brother-Nom -Nom -Acc criticized-Comp said  
 'His brother said that Inho criticized himself.'

In (20) and (21), a name (= Inho) is coindexed and c-commanded by another

name, yet the sentences are grammatical. Such sentences are not only grammatical, but also preferred sometimes for disambiguating the sentence, as in (21): It is possible in (21) to replace 'Inho' in the embedded object position with a reflexive 'caki' (or a pronoun 'ku'). In that case, however, the sentence becomes ambiguous since 'caki' is a long-distance anaphor and hence able to be coreferential with 'ku-uy hyeng' as well as 'Inho'. Therefore, in order to make the sentence unambiguous, it is necessary to use the name twice, as in (21).

The violation of Principle C similar to the above case can also be observed from other Asian languages, as pointed out by Lasnik (1991) and Lasnik and Uriagereka (1988). Note that, however, that does not mean that Principle C is not effective in Korean-type languages: There are cases where Principle C must be obeyed. Consider the following.

(22) \*ku<sub>1</sub>-ka Inho<sub>1</sub>-lul piphanhayssta  
       he-Nom       -Acc       criticized  
       'Inho criticized himself.'

(23) \*ku<sub>1</sub>-ka Yengi-ka Inho<sub>1</sub>-lul piphanhayssta-ko malhayssta  
       -Nom       -Nom       -Acc       criticized-Comp       said  
       'Inho<sub>1</sub> said that Yengi criticized him<sub>1</sub>.'

(22) and (23) show us that if a name is c-commanded by a pronoun coindexed with it, the sentence becomes ungrammatical. The contrast between (22)/(23) and (20)/(21) suggests that there is still a restriction on the distribution of names in Korean: A name cannot be bound by a pronoun.

In Lasnik (1991), it was proposed that Principle C should be relaxed in order to handle the binding phenomena such as the above. The following was proposed as a replacement of Principle C.

(24) A less referential expression may not bind a more referential one.<sup>7</sup>

---

<sup>7</sup>In Lasnik (1991), the hierarchy of referentiality is established by comparing the feature matrix [+a(naphoric)], [+p(ronominal)], and [+r(eferential)]. The feature [+r] has been added to the system mainly for distinguishing epithets from pure R-expressions or pronouns. Thus, if we compare a pronoun (= [+p, -r]), an epithet (= [+p, +r]), and a name (= [-p, +r]), the pronoun is the lowest and the name is the highest in the referential hierarchy. Below, it will be shown that the name-like EC in Korean should be as high as overt names in the hierarchy of referentiality.

From (24), it follows that an R-expression is pronoun-free, provided that a pronoun is less referential than an R-expression. Let us assume that (24) is the core Binding Principle and the difference between English and Korean observed in (20)/(21) is the result of parameterization of the Principle. Then, we can say that if the EC in (12) is classified not as a pronoun/anaphor but as an R-expression, (12) can be considered as a legitimate representation, since the relation between the EC and 'Inho' in (12) does not violate Principle (24).

It seems clear that the EC in question is not anaphoric; it may not be bound by anything within the sentence, as shown in (16)/(19). Then we have two possibilities; the EC is either a pronoun or an R-expression. It is an open question whether Korean has an empty pronoun (= *pro*) or not. However, regardless of that issue, we should admit the existence of non-pronominal & non-anaphoric ECs from the instances like the following.

- (25) cenpan-ey-man e<sub>1</sub> isip cem-ul apsess-ko hwupan-ey-to  
 1st half-in-just twenty points-Acc led-and 2nd half-in-too  
 Lakers<sub>1</sub>-ka keyim-ul aptohayssta  
 -Nom game-Acc dominated

'The Lakers led by twenty points in the first half, and they dominated in the second half, too.'

(25) contains a coordinate conjunction '-ko' (and). The EC in (25), which occurs in the first conjunct, cannot be considered as a pronominal; if we replace the EC with an overt pronoun, the sentence becomes ungrammatical with the intended meaning. Then, what is the nature of this EC? We claim that it is non-anaphoric & non-pronominal, and thus should be an R-expression, i.e., an empty name. We further claim that the same kind of EC occurs in the matrix subject position in sentences such as (12)/(19).<sup>8</sup>

---

<sup>8</sup> An anonymous reviewer pointed out that it might be the case that '-ko' in (25) is a subordinate clause marker rather than a coordinate conjunction. It seems that, in order to decide whether it is used as a subordinate clause marker or not, we should pay attention to the fact that such a usage of '-ko' has something to do with time sequence, as shown in the following.

- (i) e<sub>1</sub> achim-ul mek-ko Inho<sub>1</sub>-nun hakkyo-ey kassta  
 breakfast-Acc eat-Sub.Cl (after) -Top school-to went  
 'Inho went to school after he had breakfast.'

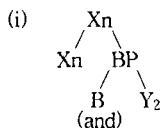
As for the syntax of coordinate structure, there is an interesting proposal supporting our claim. Munn(1992) proposes that in a coordinate structure, the two XPs are in a relation in which the first XP c-commands the second.<sup>9</sup> If

In the sentences involving this kind of ‘-ko’, the event or action described in the ‘-ko’-marked clause must precede that in the matrix clause. Such a restriction, however, does not hold in the sentence below, which contains an EC analogous to that in (25). This in turn suggests that ‘-ko’ in (25) is not a subordinate clause marker.

- (ii) hwupancen-ey-man e<sub>1</sub> kyengko-lul taset pen patass-ko  
 2nd half-in-just warning-Acc 5 times received-and  
 cen kkeyim-ul thonghay Lakers<sub>1</sub>-nun panchik-ul swu-to epsi cecillessta  
 whole game-Acc through -Top foul-Acc too many committed  
 ‘The Lakers received warning 5 times just in the second half, and they committed too many fouls throughout the game.’

In (ii), it is NOT the case that the action mentioned in ‘-ko’-marked clause precedes that in the other clause. Hence it seems reasonable to conclude that (25) or (ii) involves coordination rather than subordination.

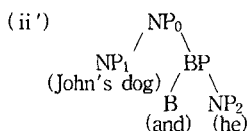
<sup>9</sup>Munn’s (1992) analysis of coordinate structures assumes the following phrase structure.



Munn claims that the coordinate conjunction is a functional head projecting to a maximal projection of category B (= Boolean). The complement of the B head is the second conjunct of the coordinated constituent. The entire BP then is adjoined to the first conjunct of the structure. His main argument for positing such an asymmetrical tree for coordinate structures comes from the binding asymmetry. Consider the following.

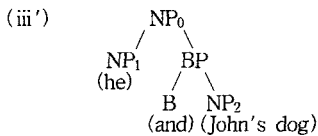
- (ii) John<sub>1</sub>'s dog and he<sub>1</sub>/him<sub>1</sub> went for a walk.  
 (iii) \*He<sub>1</sub> and John<sub>1</sub>'s dog went for a walk.

If we assume a traditional flat structure, in which ‘John’s dog’ and ‘he’ c-command each other, then it is incorrectly predicted that both (ii) and (iii) lead to the violation of Principle C. However, if the structure is asymmetrical as in (i), the contrast between (ii) and (iii) naturally follows: the first conjunct c-commands the second conjunct but not vice versa, and hence only (iii) results in Principle C violation, as the following trees show.



this proposal is correct, then (25) is evidence that in Korean, an EC can appear in the highest position in a sentence, from which it can bind an overt name. Consequently, representations such as (12) are legitimate.

To recapitulate, in Korean, the name-like EC in (12) may lead to the legitimate violation of the conventional Principle C, as overt names do. Therefore, (12) is a grammatical representation in the same way as the following, and postulating structures like (12) enables the parser to process seemingly problematic sentences without any reanalysis.<sup>10, 11</sup>



<sup>10</sup> It is worth noting that the name-like EC here is always construed to the most prominent NP in the sentence, and hence there is only one way to interpret the EC, as shown in the following.

- (i) Inho-uy hyeng-uy chinkwu-ka Mina-lul mannassta-ko malhayssta  
 -Gen brother-Gen friend-Nom -Acc met-Comp said  
 'Inho's brother's friend<sub>1</sub> said that he<sub>1</sub> met Mina.'

Here, the EC posited in the matrix subject position can be construed to neither 'Inho' nor 'Inho's brother'; it should be coreferential with the subject NP, 'Inho's brother's friend'. This is contrasted with the fact that a pronoun can be construed to any NP in the sentence unless Principle B or C is violated.

<sup>11</sup> One question related to the syntactic property of the name-like EC would be whether the EC belongs to the same category as *wh*-trace. If it does, then conventional Principle C will be legitimately violated in Korean sentences involving *wh*-trace. However, the ungrammatical status of the following sentence suggests that it is not the case; that is, strong crossover effect is observed in Korean.

- (i) \*Inho<sub>1</sub>-ka nwukwu<sub>1</sub>-lul piphanhayss-ni  
 -Nom who-Acc criticized-Q  
 'Who<sub>1</sub> did Inho<sub>1</sub> criticize?'

Thus, the answer for the above question seems to be negative. I claim that the name-like EC is different from *wh*-trace in that only the former, being a true R-expression, is constrained by Principle C. Note that, as pointed out by Higginbotham (1983), among others, it is counter-intuitive to classify *wh*-trace as an R-expression since it (or its antecedent) does not have any referentiality. Classifying *wh*-trace as an R-expression is, in fact, from the binding-theoretic motivation and hence theory-internal. If we suppose that *wh*-trace doesn't belong to R-expression, then there is no reason to expect a parallelism between the name-like EC and *wh*-trace in terms of applying Principle C. In other words, if strong crossover phenomenon is accounted for by some principle other than Principle C, the above issue becomes trivial.



- (26) Inho<sub>1</sub>-ka [Inho<sub>1</sub>-ka Mina-lul mannassta-ko] malhayssta  
 -Nom -Nom -Acc met-Comp said  
 'Inho said that he met Mary.'

## 5. Concluding Remarks

In the previous sections, I have shown that it is possible to postulate more than one syntactic structure for the analysis of sentences containing a complement clause and a null subject. One is a conventional structure, in which the overt subject binds the null subject, and the alternative structure places a null subject classified as an empty name in the matrix clause. As for the processing of such sentences, the former assumes unconscious reanalysis procedure, whereas the latter does not require any reanalysis procedure. Empirically, both approaches are satisfactory in the respect that they can account for the lack of garden path in processing sentences containing a null subject. Conceptually, however, the approach where a name-like EC is posited seems more desirable since the nature of unconscious reanalysis and the cost associated with it is not very well known at this time.

Positing a name-like EC is not just empirically driven; theoretically, the existence of such an EC is justified when we consider the syntactic properties of the Korean language.

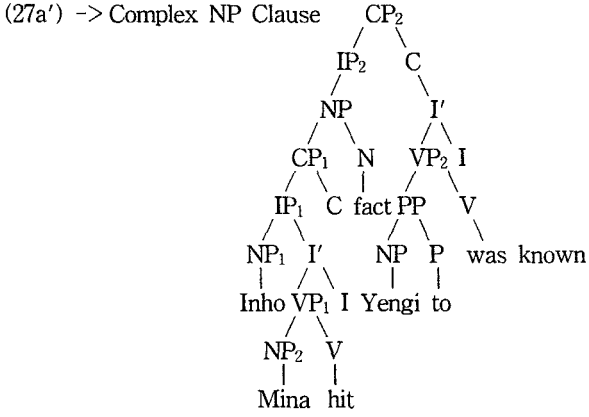
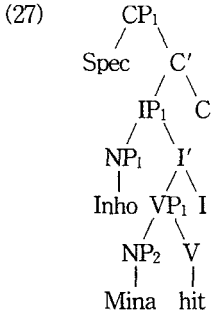
Although this paper supports the analysis in which a language-specific grammatical device is necessitated, that doesn't mean that the analysis based on the universal parsing principles is not worth pursuing. Rather, the supposed universal principles such as structure addition as unconscious reanalysis are surely necessary for dealing with the core Korean data like the following.

- (27a) [Inho-ka Mina-lul ttayli-n] sasil-i Yengi-eykey allyecyessta  
 -Nom -Acc hit-Comp fact-Nom -to was known  
 'The fact that Inho hit Mina was known to Yengi.'

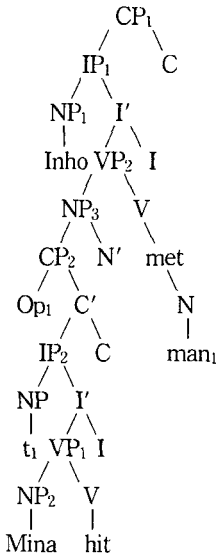
- (27b) Inho-ka [t<sub>i</sub> Mina-lul ttayli-n] salam<sub>1</sub>-ul mannassta  
 -Nom -Acc hit-Rel man-Acc met  
 'Inho met the man who hit Mina.'

The fact that the above sentences involving a relative clause/complex NP

clause do not display any indication of garden path can naturally be accounted for only when we assume that reanalysis characterized as structure addition is unconscious. Notice that the initial three input items in (27a) and (27b) are identical and able to make up a clause. Assuming that the human parser builds the simplest structure at each stage of parsing, the structural commitment made at the point when 'ttayli-n' is encountered will be something like (27a) rather than (27b) since the latter requires computing extra clausal unit and an empty category  $t_i$ . This, however, means that reanalysis is inevitable in parsing (27b) since the subsequent string is not compatible with the earlier structural decision, (by which the initial three items are structured as a clause). Since (27b) is parsed without any conscious difficulty, the reanalysis involved here must be unconscious. The following phrase-structure trees show us that the reanalysis in (27b) indeed patterns with the typical unproblematic reanalysis, i.e., structure addition.



(27b') -> Relative Clause



Notice that it is not necessary to delete anything from the previously computed structure (27) for obtaining the final structure (27b'): it is enough to simply build structures such as VP<sub>2</sub>, CP<sub>2</sub>, and IP<sub>2</sub> above VP<sub>1</sub> node. In other words, the transition from (27) to (27b') is a typical instance of unconscious reanalysis; that is, structure addition.<sup>12</sup>

In sum, what has been characterized as unproblematic reanalysis in English-type languages is also observed in Korean, and the lack of garden path in sentences requiring reanalysis is accounted for straightforwardly by such a notion of unproblematic reanalysis. This implies that such reanalysis module is a central part of the human sentence processing mechanism. After all, the human parser seems to be equipped with both such universal parsing principles and language-specific strategies or devices such as employing particular empty categories.

<sup>12</sup> Meanwhile, when reanalysis characterized as structure deletion is involved in the processing of a sentence, a severe garden path results. See Suh (1993) for the Korean data showing the contrast between structure addition and structure deletion.

## References

- Chomsky, N. (1981) *Lectures on Government and Binding*, Foris, Dordrecht.
- Frazier, L. (1978) *On Comprehending Sentences: Syntactic Parsing Strategies*, Ph.D dissertation, The University of Connecticut, Storrs. ———
- and Fodor, J.D. (1978) 'The Sausage Machine: A New Two-stage Parsing Model,' *Cognition* 6.
- Gorrell, P. (1995) *Syntax and Parsing*. Cambridge University Press.
- (1987) *Studies in Human Syntactic Processing: Ranked-parallel versus Serial Models*, Ph.D dissertation, University of Connecticut, Storrs.
- Higginbotham, J. (1983) 'Logical Form, Binding, and Nominals,' *Linguistic Inquiry* 14.
- Kurtzman, H. (1985) *Studies in Syntactic Ambiguity Resolution*, Ph.D dissertation, MIT.
- Lasnik, H. (1991) 'On the Necessity of Binding Conditions,' in R. Freidin, ed., *Principles and Parameters in Comparative Grammar*, MIT Press.
- and Uriagereka, J. (1988) *A Course in GB Syntax: Lectures on Binding and Empty Categories*, MIT Press.
- Marcus, M. (1980) *A Theory of Syntactic Recognition for Natural Language*, MIT Press, Cambridge.
- , Hindle, D., and Fleck, M. (1983) 'D-theory: Talking about Talking about Trees,' *Association for Computational Linguistics* 21.
- Munn, A. (1992) 'A Null Operator Analysis of ATB Gaps,' *The Linguistic Review* 9.
- Pritchett, B. (1992) *Grammatical Competence and Parsing Performance*, University of Chicago Press.
- Rayner, K., and Frazier, L. (1987) 'Parsing Temporarily Ambiguous Complements,' *The Quarterly Journal of Experimental Psychology* 39A.
- Suh, S. (1994) *The Syntax of Korean and Its Implications for Parsing Theory*, Ph.D dissertation, University of Maryland, College Park.
- (1993) 'How to Process Constituent Structure in Head-final Languages: The case of Korean,' *Proceedings of CLS* 29.
- (1992) 'Constituent Structure Processing in Korean,' *Proceedings of ESCOL* 1991.
- Weinberg, A. (1993) 'Parameters in the Theory of Sentence Processing: Minimal Commitment Theory Goes East,' *Journal of Psycholinguistic*

*Research 22.*

Yang, D. W. (1986) 'hankwuke-uy tayyongsalon' ('A Discussion of Korean Anaphors') *kwukehak* 15.

Department of English  
College of Arts and Humanities  
The Catholic University of Korea  
43-1 Yokkok-Dong, Wonmi-Gu.  
Puchon City, 420-743, Korea