

# 한국어 분열구문의 전산학적 처리

김종복<sup>o</sup> 양재형  
 경희대학교 강남대학교

jongbok@khu.ac.kr, jhyang@kangnam.ac.kr

## Processing Korean Cleft Constructions in a Typed Feature Structure Grammar

Jong-Bok Kim<sup>o</sup> Jaehyung Yang  
 Kyung Hee University Kangnam University

### 요 약

The expression KES, one of the most commonly used words in the Korean language, has various usages. This expression is also used to express English-like cleft constructions. It appears to provide at two different types of cleft constructions: predicational and identificational. The paper tries to provide a constraint-based analysis of these two types of Korean cleft constructions and tries to implement the analysis in the LKB system to check its feasibility. In particular, the paper shows how a typed feature structure grammar, couched upon HPSG, can provide a robust basis for parsing Korean cleft constructions.

Key words: KES, cleft, computational implementation, predicational, identificational

### 1. Introduction

Cleft constructions are employed to mark a certain constituent as a discourse prominent element. In this respect, Korean has at least two main types of clefts (cf. [1]):

(1) a. Predicational:

[John-i        <sub>i</sub> ilk-un kes-un]  
 John-NOM        read-MOD KES-TPC  
 [kacca]<sub>i</sub>-i-ta  
 fake-COP-DECL  
 'What John read is a fake.'

b. Identificational:

[i chayk]<sub>i</sub>-i palo [John-i        <sub>i</sub>]  
 this book-NOM very John-NOM  
 ilk-un kes-i-ta]  
 read-MOD KES-COP-DECL  
 'This book is what John read.'

The predicational cleft in (1a) consists of a cleft clause with the missing object coindexed with the precopula expression *kacca* 'fake' whereas the identificational one in (1b) has the nominative phrase *i chayk* coindexed with the missing object in the following cleft clause. In these two clefts, the XP is semantically linked to the content of the cleft clause introduced by KES. The structure of these two types of clefts can be schematized as follows:

(2) a. Predicational: [[      ]<sub>i</sub>]-KES]-TOP-XP<sub>i</sub>-COP-DECL

b. Identificational: XP<sub>i</sub>-[      ]<sub>i</sub>]-KES-COP-DECL

As represented here, both of these types have a pivot or highlighted expression like the English cleft constructions.

This paper aims to review the basic properties of these two different types of Korean cleft constructions and provide a constraint-based analysis. We also show a brief summary of the results of implementing this analysis in the LKB (Linguistic Knowledge Building) system ([2]).

### 2. Syntax and Semantics of the Cleft Constructions

The first issue in processing cleft constructions concerns the categorial status of KES. Consider the main uses of KES:

- (3) a. nay kes-i ne kes-pota khu-ta  
 my thing-NOM your thing-more big-DECL  
 '(Lit.) My thing is bigger than your thing.'
- b. [[John-i mek-un] kes/sakwa]-ul  
 John-NOM eat-MOD KES/apple-ACC  
 mek-ess-ta  
 eat-PAST-DECL  
 '(We) ate the thing that John ate.'
- c. [[John-i talli-nun] kes]-ul  
 John-NOM run-MOD KES-ACC  
 moll-ass-ta  
 not.know-PAST-DECL  
 '(We) didn't know that John was running.'

As noted here, KES in (3a) combines with a specifier whereas the one in (3b) combines with the relative clause with one missing argument. In both of these examples, KES refers to a ‘thing’. Meanwhile, KES in (3c) combines with a complete sentence, referring to the event denoted by the clause.<sup>6)</sup> In terms of meaning, we can assume KES to at least the following two lexical entries:

- (4)
- a. 
$$\left[ \begin{array}{l} cn \\ \langle kes \rangle \\ \text{SYN} \left[ \begin{array}{l} \text{HEAD} | \text{POS } \textit{noun} \\ \text{VAL} | \text{SPR} \langle \textit{DetP} \rangle \end{array} \right] \\ \text{SEM} \left[ \begin{array}{l} \text{INDEX } i \\ \text{RELS} \left\langle \left[ \begin{array}{l} \text{PRED } \textit{one-rel} \\ \text{ARG0 } i \end{array} \right] \right\rangle \end{array} \right] \end{array} \right]$$
- b. 
$$\left[ \begin{array}{l} bn \\ \langle kes \rangle \\ \text{SYN} \left[ \begin{array}{l} \text{HEAD} | \text{POS } \textit{noun} \\ \text{VAL} | \text{COMPS} \left\langle \textit{S} \left[ \begin{array}{l} \text{GAP } \langle \rangle \\ \text{IND } \textit{el} \end{array} \right] \right\rangle \end{array} \right] \\ \text{SEM} | \text{INDEX } i \end{array} \right]$$

The lexical entry (4a) means that KES refers to an individual equivalent to the meaning of *one*. The one in (4b) specifies that KES is a bound noun, selecting a saturated sentence. In this case, its INDEX value is identified with that of the sentential complement, insuring that KES denotes an event. One clear example where KES is linked to an event can be found from a perceptive example like the following:

- (5) [John-i sakwa-lul mek-un kes-ul]  
 John-NOM apple-ACC eat-MOD KES-ACC  
 moll-ass-ta  
 not.know-PAST-DECL  
 ‘(He) didn’t know that John ate an apple.’

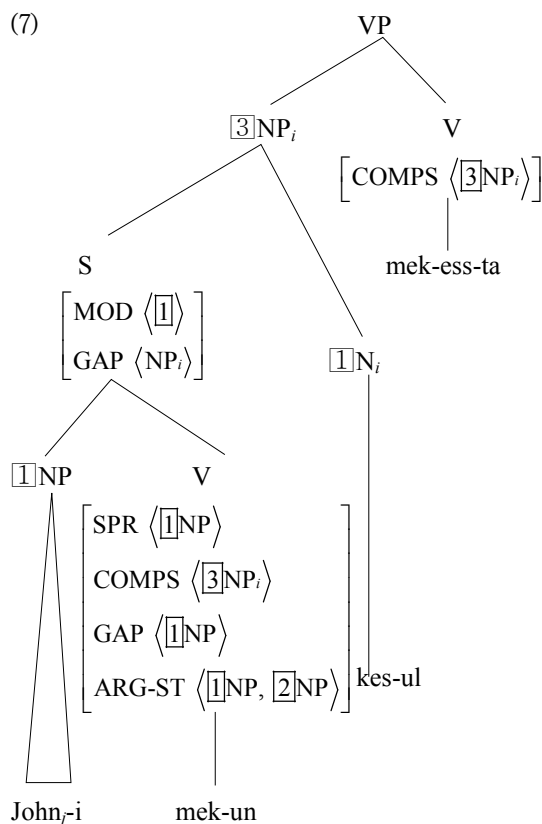
Here the argument of predicate ‘not.know’ is the sentence introduced by KES, similar to the English complementizer *that*. In this sense we can assume that KES denotes an event identified with its sentential complement.

Now let’s consider the following relative clause and cleft example, respectively:

- (6) a. [John-i \_\_\_ mek-un kes/sakwa]-ul  
 John-NOM eat-MOD KES/apple-ACC  
**mek-ess-ta**

- eat-PAST-DECL  
 ‘(We) ate the thing that John ate.’  
 b. [John-i \_\_\_ mek-un kes/kwail]-un  
 John-NOM eat-MOD KES/fruit-TOP  
 sakwa-i-ta  
 apple-COP-DECL  
 ‘What John ate is an apple.’

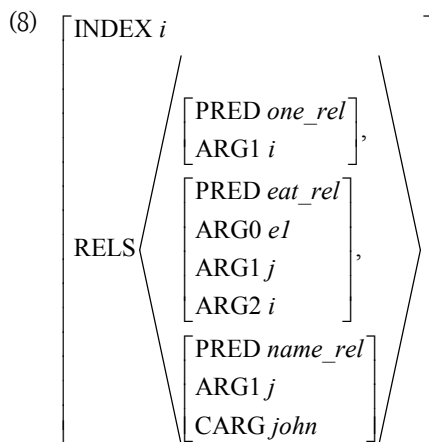
The only difference we can observe here is the predicate. The sentence with a relative clause has a transitive verb whereas the one with a cleft clause has a copula. As we have seen earlier, both the relative and cleft clause are identical in the sense that each clause has a syntactic gap here. In both cases, KES can be replaced by a common noun, indicating that there is no semantic difference. Let’s consider the structure of (6a) with a relative clause first:



As we noted here, the noun KES in relative clauses is a common noun referring to an individual: this information is passed up to the NP projection (the index value ‘i’). Since the verb *mek-un* ‘ate’ also requires its object to be a referential individual, there is no mismatch between these two requirements. To observe how we obtain the semantics correctly, let us consider the MRS (minimal recursion semantics) representations of this NP.<sup>7)</sup>

<sup>6)</sup> In the literature, KES in (3c) has been treated as a complementizer. Cf. [3] and [4].

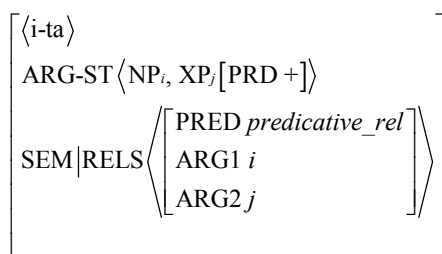
<sup>7)</sup> Minimal Recursion Semantics, developed by [5], is a framework of computational semantics designed to enable semantic composition using



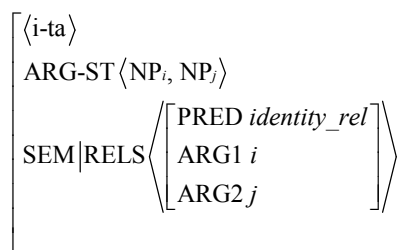
This semantic representation simply means that there is an individual 'i' which the person named John eats. This index value is linked to the ARG value of 'one' coming from KES. This index value is the semantic head information visible at the NP level.

How about the cleft sentence? Before we provide its structure, consider the lexical entry for the copula *i-ta*. We have seen that there are two different cleft constructions. This leads us to assume that there are two different copula uses: predicational and identificational (also see ([1])). The difference of these two different copulas is represented in the following lexical entries:

(9) a. Predicational Copula:



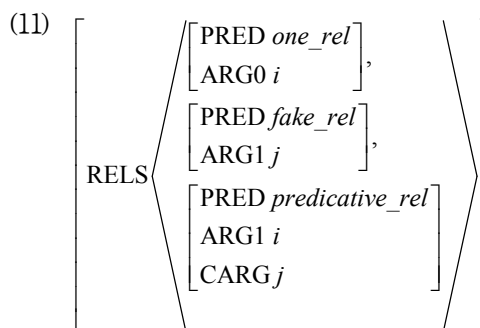
b. Identificational Copula:



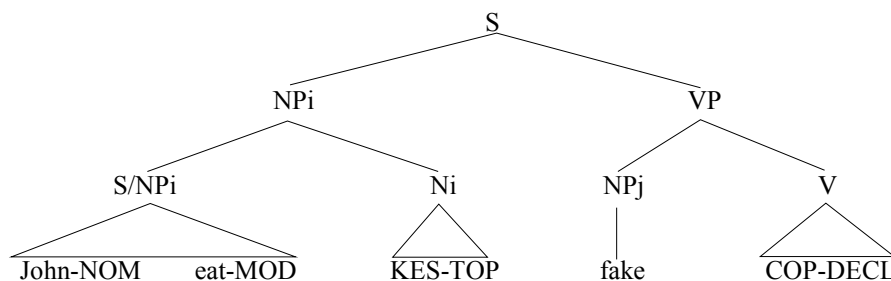
The predicational copula requires its second argument to carry the positive PRD feature, ensuring that this expression predicates of the first argument (subject). The semantics also reflects this. Meanwhile, the identificational copula requires the INDEX value of the first argument is in the *identity\_rel* with that of the second argument. This lexical specification implies that the two expressions here have identical referential types.

Given these, we then can generate a structure like (10) for the predicational left:

This structure, including the cleft clause as the subject and the predicative expression, will then induce the meaning similar to (9a). The predicative expression 'fake' will predicate of this nominal element, inducing a semantic representation like the following:<sup>8)</sup>



(10)



only the unification of type feature structures. See [5] and [6]. The value of the attribute SEM(ANTICS) in our system represents a simplified MRS. Also see [7] for the analysis of Korean relative clauses. ARG0 canonically refers to the index value of the EP (elementary predicate) itself whereas ARG1 or ARG2 refers to the predicate's semantic arguments. CARG refers to constant arguments whose value can be a name.

The meaning of the copula *i-ta* is relevant to the variable missing in the cleft. This index value and the index value of 'fake' is in the *predicative* semantic relation. This in

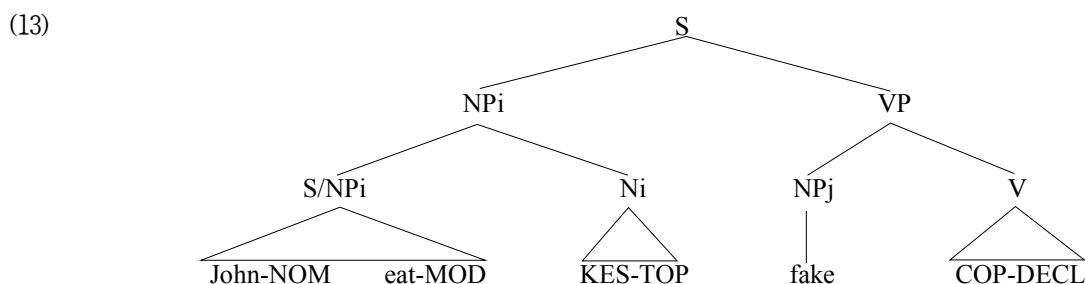
<sup>8)</sup> The index value of a predicative expression is identified with that of the gapped element in the cleft clause.

turn means that as long as the precopular expression can predicate of the cleft-clause subject, there is no categorial restriction on the type of the precopular expression. That's why we allow other than an NP in this position.

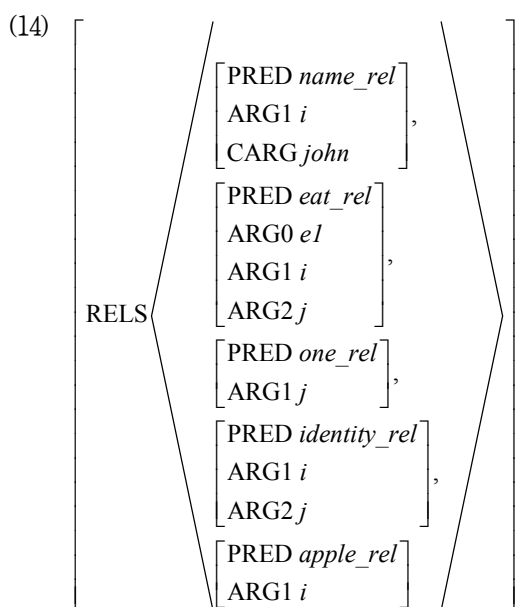
Now consider the structure of an identificational cleft sentence:

- (12) i sakwa-ka John-i mek-un  
 this apple-NOM John-NOM eat-MOD  
 kes-i-ta  
 KES-COP-DECL  
 'This apple i what John ate' in the language.'

A simple tree representation will be something like the following:



The lexical constraints of the identificational copula insure that the index value of the subject is identified with that of KES as represented in the expected semantics of this sentence:



Unlike the predicational one, the identificational one requires the identity of two index values. This is why

neither the PP or an adverbial element can be focused in identificational cleft:

- (15) a. [[John-i Mary-lul manna-n]  
 John-NOM Mary-ACC meet-MOD  
 kes-nun] kongwon(eyse)-i-ta  
 KES-TPC park.at-COP-DECL  
 '(Lit.) When John met Mary was at the park.'  
 b. \*[kongwen-(eyse)]<sub>pp</sub>-ka [John-i  
 park-at-NOM [John-NOM  
 Mary-lul manna-n kes]-i-ta  
 Mary-ACC meet-MOD KES]-COP-DECL  
 'It is at the park John met Mary.'

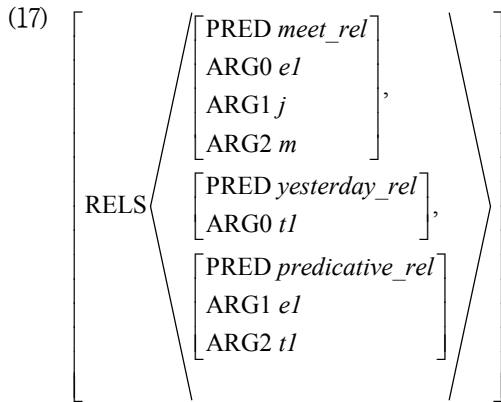
The identificational cleft requires the two expressions to have the identical index value whereas the predicational

cleft only requires the precopular element to be predicative, regardless of its referential property. Like the locative adverbial element, we have seen that a true adverbial-nominal can occur as the focus element in the predicative cleft:

- (16) [[John-i Mary-lul manna-n] kes-un]  
 John-NOM Mary-ACC meet-MOD KES-TPC  
 ecey-i-ta  
 yesterday-COP-DECL  
 '(Lit.) When John met Mary was yesterday.'

There is no syntactic gap in the seemingly cleft-clause: we assume that KES here, combining with a fully saturated S, links to an event denoted by the sentential complement.

This event denoting KES clause is in a predicative relation with the adjunct *ecey* 'yesterday'. That is, the semantics the analysis generates is something like the following:



syntax in the Korean Resource Grammar: implementing typed feature structures. In: Lecture Notes in Computer Science. Volume 2945. Springer-Verlag (2004) 13-24

[8] Kim, J.B.: Korean Phrase Structure Grammar. Hankwuk Publishing, Seoul (2004) In Korean.

### 3. Results of the Implementation

The analysis we have presented so far has been incorporated in the typed-feature structure grammar HPSG for Korean (Korean Resource Grammar) aiming at working with real-world data (cf. [7] and [8]). To test its performance and feasibility, it has been implemented into the LKB (Linguistic Knowledge Building).<sup>9)</sup> The test results give the proper syntactic as well as semantic structures for the two types of Korean cleft constructions.

### References

- [1] Kim, J., Sells, P.: Some remarks on Korean nominalizer *kes* and information structure. *Studies in Generative Grammar* 17 (2007) 479-494
- [2] Copestake, A.: *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford (2002)
- [3] Jhang, S.E.: *Headed nominalizations in Korean: Relative clauses, clefts, and comparatives*. Hankwuk Publishers, Seoul (1995)
- [4] Sohn, K.W.: *Kes-clefts, connectedness effects, and the implications thereof*. *Studies in Generative Grammar* 14 (2004) 561-571
- [5] Copestake, A., Flickinger, D., Pollard, C.J., Sag, I.A.: *Minimal recursion semantics: an introduction*. *Research on Language and Computation* 3 (2005) 281-332
- [6] Bender, E.M., Flickinger, D.P., Oepen, S.: *The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars*. In Carroll, J., Oostdijk, N., Sutcliffe, R., eds.: *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics, Taipei, Taiwan (2002)* 8-14
- [7] Kim, J.B., Yang, J.: *Projections from morphology to*

<sup>9)</sup> The current Korean Resource Grammar has 394 type definitions, 36 grammar rules, 77 inflectional rules, 1100 lexical entries, and 2100 test-suite sentences, and aims to expand its coverage on real-life data.