Processing Korean Cleft Construction in a Typed Feature Structure Grammar

Jong-Bok Kim[1] and Jaehyung Yang[2]

[1] School of English, Kyung Hee University, Seoul, 130-701, Korea
jongbok@khu.ac.kr
[2] School of Computer Engineering, Kangnam University, Kyunggi, 446-702, Korea
jhyang@kangnam.ac.kr

Abstract. Korean has at least three main types of cleft constructions: predicational, identificational, and eventual. The paper tires to provide a constraint-based analysis and tries to implement in the LKB system to check the feasibility of the analysis. This paper shows how a typed feature structure grammar, HPSG, together with the notions of 'type hierarchy' and 'constructions', can provide a robust basis for parsing three types of Korean cleft constructions. We show that this system induces robust syntactic structures as well as enriched semantic representations for real-time applications such as machine translation, which require deep processing of the phenomena concerned.

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

1 Introduction

Cleft constructions are employed to mark a certain constituent as a discourse prominent element. Korean also has at least three main types of clefts:

(1) a. Predicational:
[John-i __ $_i$ senthaykha-n kes-un] [i chayk]$_i$-i-ta
John-NOM select-MOD KES-TOP this book-COP-DECL
'What John selected is this book.'

b. Identificational:
[i chayk]$_i$-i palo [John-i __ $_i$ senthaykha-n kes-i-ta]
this book-NOM very John-NOM select-MOD KES-COP-DECL
'This book is what John selected.'

c. Eventual
kuttay [John-i cip-ey o-n] kes-i-ess-ta
the moment John-NOM home-LOC come-MOD KES-COP-PAST-DECL
'It is at the very moment that John came home.'

The predicational cleft in (1)a consists of a cleft clause with the missing object coindexed with the precopula expression *i chayk* 'this book' whereas the

identificational one in (1)b has the nominative phrase *i chayk* coindexed with the missing object in the following cleft clause. The eventual one starts with a complete pivot clause followed with *kes-i-ta*. All these three types has a pivot or highlighted expression like the English cleft constructions. The structure of these three types of clefts can be schematized as follows:

(2) a. Predicdational: [[$_S$ __ $_i$] KES]-TOP XP$_i$-COP-DECL
b. Identificational: XP$_i$-TOP[[$_S$ __ $_i$] KES]-COP-DECL
c. Eventual: [$_{NP}$ [$_S$ 'clause'] KES]-COP-COP

One main difference that while the cleft clause in the predicational and identificational has a missing element semantically linked to the XP, the eventual cleft clause has no missing element. Here the XP in precopular position is somehow semantically related to the content of the clause-like modifier of *kes* in subject position. The *kes* phrase typically hosts the topic marker, and so describes given information against which the precopular XP presents new information.

This paper aims to review the basic properties of these three different types of Korean cleft constructions and provide a constraint-based analysis. In particular, we will show that these Korean cleft constructions are closely related to Korean copula constructions.

2 Basic Properties of the Cleft Constructions

2.1 Predicational and Identificational Cleft

[rewrite]

1. there is a missing element in the cleft clause; linked to the XP.

topic discussion here

2. the missing element can be either an argument or adjunct in the predicational one, but not the identificational one.

3. pred: xp can be case marked, but not xp

4. both observes the island?

5. both can be unbounded?

As observed earlier, the cleft clause in the two types of cleft constructions include a gap element. In this respect, we can observe that Korean clefts behave like relative clauses, differently from topic constructions which can be either gapped or gapless:

(3) a. ku chayk-un [John-i __ ilk-ess-ta]
the book-TOP John-NOM read-PAST-DECL
'This book, John read.
b. [John-i __ ilk-un] ku chayk
John-NOM read-MOD the book
'the book that John read'
c. [John-i __ ilk-un kes]-un palo i chayk-i-ta
John-NOM read-MOD KES-TOP very this book-COP-DECL
'What John read is this very book.'

(4) a. kkoch-un [cangmi-ka yepputa]
flowers-TOP rose-NOM pretty
'As for flowers, roses are pretty.'
b. *[cangmi-ka yeppu]-n kkoch
roses-NOM pretty-MOD flowers
c. *[cangmi-ka yeppu-n kes]-un kkoch-i-ta
rose-NOM pretty-MOD KES-TOP flower-COP-DECL

The contrast between the gapped examples in (3) and the gapless examples in (4) indicate that both predicational and identificational clefts are like relative clauses, rather than topic clauses.

In the predicational copula, we can observe that the focused XP can be either an argument or an adjunct. The adjunct element can have a semantic case or postposition:

(5) a. [John-i Mary-lul manna-n kes]-un [kongwen-(eyse)]-i-ta
[John-NOM Mary-ACC meet-MOD KES]-TOP park-at-COP-DECL
'It was at the park that John met Mary.'
b. John-i Mary-lul manna-n kes-un [tosekwan-(eyse)]-i-ta
John-NOM Mary-ACC meet-MOD KES-TOP library-at-COP-DECL
'Where John met Mary is (at) the library.'

The postposition or semantic marker of the focused expression is optional. An adverbial element also can be focused as long as it is categorically an adveribal nominal:[3]

(6) John-i Mary-eykey senmwul-ul cwu-n kes-un [ecey]-i-ta
John-NOM Mary-DAT present-ACC give-MOD KES-TOP yesterday-COP-DECL
'It is yesterday when John gave Mary a present.'

Meanwhile, the identicational cleft does not allow the PP adjunct to be focused, regardless of the presence of the postposition:

(7) a. *[kongwen-(eyse)]$_{PP}$-ka [John-i Mary-lul manna-n kes]-i-ta
park-at-NOM [John-NOM Mary-ACC meet-MOD KES]-COP-DECL
b. *[tosekwan-(eyse)]-ka John-i Mary-lul manna-n kes-i-ta
library-at-NOM John-NOM Mary-ACC meet-MOD KES-COP-DECL

[3] However, true adverbs cannot be focused:

(i) a. John-i talli-n kes-un [chenchenhi]-i-ta
John-NOM run-MOD KES-TOP slowly-COP-DECL
'(lit.) The way John ran was slowly.'
b. *[chenchenhi]-ka John-i talli-n kes-i-ta
slowly-NOM John-NOM run-MOD KES-COP-DECL

As noted here, neither the predicational nor identificational cleft allows a true adverb to be focused.

c. *[ecey]-ka John-i Mary-eykey senmwul-ul cwu-n kes-i-ta
yesterday-NOM John-NOM Mary-DAT present-ACC give-MOD KES-COP-DECL

The two types of clefts also allow long distance dependency:

(8) a. [John-i [Mary-ka __ cohahanta-ko] sayngkakha-nun kes]-un
John-NOM Mary-NOM like-COMP think-MOD KES-TOP
i kulim-i-ta
this picture-COP-DECL
'What John thought Mary likes is this picture.'

b. i kulim-i [John-i [Mary-ka __ cohaha-n-ta-ko]
this picture-NOM John-NOM Mary-NOM like-PRES-DECL-COMP
sayngkakha-nun kes]-i-ta
think-MOD KES
'This picture is what John thought Mary likes.'

What is clefted in both examples is the object of the embedded clause. The cleft example cannot be an adjunct in the embedded clause.

(9) a. [John-i [Mary-ka ku chak-ul ilkessta-ko]] sayngkakha-n ecey
John-NOM Mary-NOM the book-ACC read-COMP think-MOD yesterday
'the time when John thought Mary read the book'

b. [John-i [Mary-ka ku chak-ul ilkessta-ko] sayngkakha-nun kes]-un
John-NOM Mary-NOM the book-ACC read-COMP think-MOD KES-TOP
ecey-i-ta
yesterday-COP-DECL
'The time when John thought Mary read the book was yesterday'

In both relative and cleft examples, the relativized and cleft adjunct is linked to the higher main clause, not to the embedded clause.

We can observe that just like relative clauses, the cleft observes the CNPC (complex noun phrase constraint) but not the Wh-island:[4]

(10) a. [John-i __ piphanha-n kes-un] ku nonmwun-i-ta
John-NOM criticize-MOD KES-TOP the article-COP-DECL
'What John criticized is the article.'

b. *[John-i [[__ ssu-n] salam-ul]] piphanha-n] kes-un ku nonmwun-i-ta
John-NOM write-MOD person-ACC criticize-MOD KES-MOD the article-COP-DECL
'(lit.) What John criticized the person who wrote __ was the article.'

(11) a. John-i nwu-ka ku chayk-ul sass-nunci kwungkumhay hayessta
John-NOM who-NOM the book-ACC buy-Q wonder do
'John wondered who bought the book.'

b. John-i nwu-ka __ sass-nunci kwungkumhay ha-n kes-un ku chayk-i-ta
John-NOM who-NOM buy-Q wonder do-MOD KES-TOP the book-COP-DECL
'What John wondered who bought __ is the book.'

[4] See Kang 2006 for a different view in this matter.

This indicates that the interrogative clause the verb *kwungkuwmhayha-* selects is not an NP but a CP clause so that it can participate in the unbounded dependency.

Eventual Cleft Constructions Unlike the preceding two cleft constructions, KES can nominalize a whole preceding S, highlighting an event, as in (12).[5]

(12) a. ku ttay sako-ka na-n kes-i-ya
that moment accident-NOM happen-PAST KES-COP-DECL
'The fact is that at that moment, an accident happened.'

b. ku yeca-ka John-ul manna-n kes-i-ya
that woman-NOM John-ACC meet-PAST KES-COP-DECL
'The fact is that [that woman met John].'

The clausal focus of *kes* here is all presented as new information, as can be attested by the fact that these examples can be an appropriate answer to the question in (13):

(13) mwusun il-i-ni?
what thing-COP-Q
'What happened?

Any phrase within the event cleft can have a narrow focus interpretation with a phonological prominance on it.

(14) a. kuliko nase HYUNG-I os-ul twici-nun kes-i-ess-ta
b. kuliko nase hyung-i OS-ul twici-nun kes-i-ess-ta
c. kuliko nase hyung-i os-ul TWICI-NUN kes-i-ess-ta

One thing we can observe is that this event cleft needs to be introduced by a temporal and reasonal adverbial element:

[5] KES may present new information relative to a topic which is already given in the context:

(i) Today is a holiday, so why did Chelsoo go to school?
a. chelwsu-nun [onul hyuil-i-nci moll-ass-ta]
Chelsoo-TOP [today holiday-COP-COMP not.know-PAST-DECL]
'Chelsoo did not know that today is a holiday.'
b. chelwsu-nun [onul hyuil-i-nci moll-ass-ten kes]-i-ess-ta
Chelsoo-TOP [today holiday-COP-COMP not.know-PAST-RETR KES]-COP-PAST-DECL
'The fact is that Chelsoo did not know that today is a holiday.'

It is possible to present the new information as a canonical VP, as in (a), or with a bit more drama using the KESform in (b). Intuitively, these are VP-focus examples which present a noteworthy fact, about a given individual. Typically, it is possible to present the whole content as new, e.g., a nominative marker is also possible on *chelswu* in (b).

(15) a. ku ttay sako-ka na-n kes-i-ta
b. kuliko nase hyung-i os-ul twici-nun kes-i-ess-ta
c. imal-ey holangi-ka nwumwul-ul hulli-nun kes-i-ess-ta

Further, unlike the other two KES cannot be replaced by a common noun.

(16) a. ku ttay sako-ka na-n kes/*iyu-i-ta
b. kuliko nase hyung-i os-ul twici-nun kes/*swunkan-i-ess-ta
c. imal-ey holangi-ka nwumwul-ul hulli-nun kes/cangso-i-ess-ta

No island: no relativization

(17) a. os-ul twici-n hyung
b. chakhan haksayng-i-n hyung
c. *kuliko nase os-ul twici-nun kes-i-n hyung

No tense? Future tense?

2.2 Categorical Status of KES

The expression KES is traditionally taken to be a bound noun in terms of morphosyntactic category in the sense that it canonically combines with either a specifier or a sentential complement (cf. see section 4):

(18) a. *(i) kes-kwa *(ce) kes
this thing-CONJ that-thing
b. *(nay) kes-i *(ne) kes-pota khu-ta
my thing-NOM your thing-more big-DECL
'(Lit.) My thing is bigger than your thing'
c. *(John-i talli-nun) kes-ul moll-ass-ta
John-NOM run-MOD KES-ACC not.know-PAST-DECL
'(We) didn't know that John was running.'

As noted here, KES in (18)a and b refers to a nonhuman entity. When KES refers to an individual, its reference value can be a nonanimate or a nonhonored human being as in *elin kes* 'a childish one'. The expression KES in (18)c refers to the proposition denoted by the clause.[6]

The noun KES can also be associated with a gap in the relative clause. However, there is a contrast between an argument and a putative adjunct gap:

(19) a. John-i __ $_i$ ilk-un kes$_i$-ul peli-ess-ta
John-NOM read-MOD KES-ACC discardPAST-DECL
'(I) discarded what John read.'
b. *[[John-i Mary-lul manna-n] kos/*kes-eyse] pwul-i na-ss-ta
John-NOM Mary-ACC meet-MOD place/KES-LOC fire-NOM happen-COP-DECL
'Where John met Mary, there happened a fire.'

[6] In the literature, KES in (18)c has been treated as a complementizer. Cf. Jhang 1994 and Sohn 2004.

The contrast here indicates that even though KES can be linked to an argument in the relative clause, it cannot be linked to a putative adjunct gap: if it can, (19)b with KES would be grammatical. The observations given here hint that there are at least two different types of KES: one referring to an entity (or thing) and the other functioning as an event denoting pronoun.

There is enough evidence that indicates KES here is just a nominal expression. One simple fact can come from the possibility of replacing KES by a common noun:

(20) a. [John-i __ cohaha-nun kes/konchwung]-un camcali-i-ta
'John-NOM like-MOD KES/insect-TOP dragonfly-COP-DECL
'The insect that John likes is dragonfly'

b. camcali-ka [John-i __ cohaha-nun kes/konchwung]-i-ta
'dragonfly-NOM John-NOM like-MOD KESinsect-COP-DECL
'Dragonfly is the insect that John likes.

Though there exist pragmatic conditions on the replacement, nothing blocks its replacement by an appropriate common noun, as illustrated here.

A canonical complementizer like *-ko* does not host a grammatical case marker (NOM or ACC), whereas all phrases headed by KES in the predicational cleft do:[7]

(21) a. [John-i ku sasil-ul molunta-ko-(*lul)]$_{CP}$ malhayessta
John-NOM the fact-ACC not.know-COMP-ACC said

b. [John-i __ sa-n kes]$_{NP}$-i palo i chayk-i-ta
John-NOM buy-MOD KES-NOM very this book-COP-DECL

The plural marking can be attached to a nominal element, but not a true verbal element. The possibility of attaching the plural marker *tul* to the cleft clause clearly indicates that KES in the cleft is a nominal projection:

(22) a. John-i [haksayng-tul-i ku chayk-ul ilkessta-ko]-(*tul) mitessta
John-NOM student-PL-NOM the book-ACC read-COMP-PL believe
'John believed that students read the book.'

b. [John-i __ ilk-un kes-tul-un] i chayk-tul-i-ta
John-NOM read-MOD KES-PL-TOP this book-PL-COP-DECL
'What John read is these books.'

c. i chayk-tul-i palo [John-i __ ilk-un kes-tul]-i-ta
this book-PL-NOM very John-NOM read-MOD KES-PL-COP-DECL
'These books are what John read.'

[7] In the identificational cleft, no grammatical case marking (NOMor ACC) is allowed due to the morphological properties of the copula (cf. Kim et al. 2004):

(i) a. i chayk-i nay chayk-(*i/*lul)-i-ta
this book my book-NOM/ACC-COP-DECL

b. i chayk-i John-i sa-n kes-(*ul)-i-ta
this book-NOM John-NOM buy-MOD KES-ACC-COP-DECL

The plural marking, which cannot follow a complementizer as in (22)a, can appear with the KES in the predicational or identificational one.

Coordination data also indicate that KES is a noun-like expression. The conjunction marker *-wa/kwa* conjoins only NPs, not Ss:

(23) a. [sensayngnim]$_{NP}$-kwa [haksayngtul]$_{NP}$-i hamkkey ttena-ss-ta
teacher-CONJ student-NOM together leave-PAST-DECL

b. *[John-un chayk-ul ilk]$_{S}$-kwa [Mary-nun nolay-lul pwulessta]$_{S}$
John-TOP book-ACC read-CONJ Mary-TOP song-ACC sang
'(Int.) John read books and Mary sang a song.'

We can notice that predicational and identificational clefts both allow nominal coordination:

(24) a. [John-i sa-n kes]-kwa [Mary-ka ilk-un kes]-un
John-NOM buy-MOD KES-CONJ Mary-NOM read-MOD KES-TOP
motwu kacca-i-ta
all fake-COP-DECL
'What John bought and what Mary read are all fake.'

b. i chayk-tul-i [John-i sa-n kes]-kwa [Mary-ka ilk-un
this book-PL-NOM John-NOM buy-MOD KES-CONJ Mary-NOM read-MOD
kes]-tul-i-ta
KES-COP-DECL
'These books are what John bought and what Mary read.'

If the cleft clause in predicational cleft is a CP, we would not expect such a coordination. Further evidence can be found from floating quantifier properties. The antecedent of a floating quantifier (or floated numeral classifier) needs to be within the same clause as illustrated in (25)a. Interestingly, we allow a floating quantifier outside the cleft clause:

(25) a. namca-tul-un [yeca-tul-i sakwa-lul mek-ess-ta-ko]
men-PL-TOP women-NOM apple-ACC eat-PAST-DECL-COMP
sey myeng-i sayngkakhay-ess-ta
three CL-NOM think-PAST-DECL
'As for men, three thought women ate apples.'

b. John-i sa-n kes-i **sey kay-ka** kacca-i-ta
John-NOM buy-MOD KES-NOM three CL-NOM fake-COP-DECL
'As for the things John bought, three are fake.'

The antecedent of *say kay-ka* in (25)b must be in the same clause: if KES were a C, it could not serve as its antecedent since it is in the different clause. The only way appears to treat the KES here as the head of the cleft clause, placing KES and the floating quantifier in the same clause.

What we have seen so far tells us is that regardless of its uses, KES is a nominal element. In particular, its use in cleft clauses is externally nominal

though the clauses have verbal properties in terms of internal syntax. However, the difference comes in semantics: when KES combines with a DetP specifier or a relative clause, it refers to a referential individual. Meanwhile, when it combines with a saturated sentence, it is linked to an event, whose data we will see more in the next section.

3 Syntax and Semantics of the Cleft Constructions

3.1 Predicational and Identificational

The observations we have seen in the previous section have shown us that the cleft clause exhibits nominal properties externally though it displays verbal properties internally. With the aim of implementing the analysis for computational purposes, the challenges are thus how we capture these mixed properties with less stipulations.

We have seen that the KES in cleft is an N in terms of morphosyntactic category. Based on the observations we have made so far, we introduce two different types of KES: one as a common noun and the other as a bound noun. When it functions as a common noun, it refers to a referential individual. Meanwhile, when it serves as a bound noun, it functions as a pure pronoun whose reference is determined by a context or by a governor such as the predicate selecting the phrase headed by KES. These two possibilities of what KES can refer to can be represented as the following lexical entries:[8]

(26) a.
$$\begin{bmatrix} \textit{cn} \\ \langle \text{kes} \rangle \\ \text{SYN} \begin{bmatrix} \text{HEAD} \mid \text{POS } \textit{noun} \\ \text{VAL} \mid \text{SPR } \langle (\text{DetP}) \rangle \end{bmatrix} \\ \text{SEM} \begin{bmatrix} \text{INDEX } i \\ \text{RELS} \left\langle \begin{bmatrix} \text{PRED } \textit{one\_rel} \\ \text{ARG0 } i \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

b.
$$\begin{bmatrix} \textit{bn} \\ \langle \text{kes} \rangle \\ \text{SYN} \begin{bmatrix} \text{HEAD} \mid \text{POS } \textit{noun} \\ \text{VAL} \mid \text{COMPS } \langle \text{S} \begin{bmatrix} \text{XARG } i \\ \text{INDEX } \textit{e1} \end{bmatrix} \rangle \end{bmatrix} \\ \text{SEM} \mid \text{INDEX } \textit{e1} \vee i \end{bmatrix}$$

[8] For the semantics, we follow the representation of MRS (Minimal Recursion Semantics). See Bender et al. 2002, Copestake 2002, Copestake et al. 2005, and Kim 2006 for details.

The lexical entry (26)a means that KES refers to a referential index with the meaning of *one* when it combines with its specifier DetP or occurs as the head in the relative clause. This kind of treatment diverges from the traditional view treating KES as only a bound noun. Corpus data reveal that KES can be used like a common noun in various contexts:[9]

(27) saylowun siswul-ul hanta-ko, yatan-i-ess-nuntey. kes-to chwisotoy-ess-e.
new operation-ACC do-COMP lowsy-COP-PAST-but, KES-also cancel-PAST-DECL
'People were talking about the new way of operation, but it was also canceled.' (Sejong Corpus)

We thus assume that KES is a common noun when it combines with a DetP or with a relative clause. The KES in the cleft is also a common noun since it combines with a gapped relative clause. In such a case, KES is semantically similar to the English pronoun *one*.[10]

Meanwhile, in (26)b, KES is a bound noun combining with a fully saturated sentence. In this case, its INDEX value is identified either with that of the sentential complement or with the XARG's value.[11] This will ensure that when KES combines with a fully saturated sentence, it is linked either to an event or its subject (mainly in the IHRC construction).[12]

Now let's consider the following relative clause and cleft example:

(28) a. [John-i __ mek-un kes/sakwa]-ul **mek-ess-ta**
John-NOM eat-MOD KES/apple-ACC eat-PAST-DECL
'(We) ate the thing that John ate.'

b. [John-i mek-un kes/kwail]-un **sakwa-i-ta**
John-NOM eat-MOD KES/fruit-TOP apple-COP-DECL
'What John ate is an apple.'

The only difference between the examples in (26) and (28) is that the clause the noun KES combines with has an argument gap. In both cases, KES can be

[9] Such uses are dominant in spoken texts but similar words with the same phonological environment do not behave alike.

[10] Though it may be premature to link KES with the substitute pronoun *one*, there are many cases where the two behave similar (cf. Quirk et al. 1985):

(i) a. The one with chocolate frosting has cream filling.
b. Is this the one you want to meet?
c. Do you want these ones?
d. These donuts look delicious; I think I will choose this one.

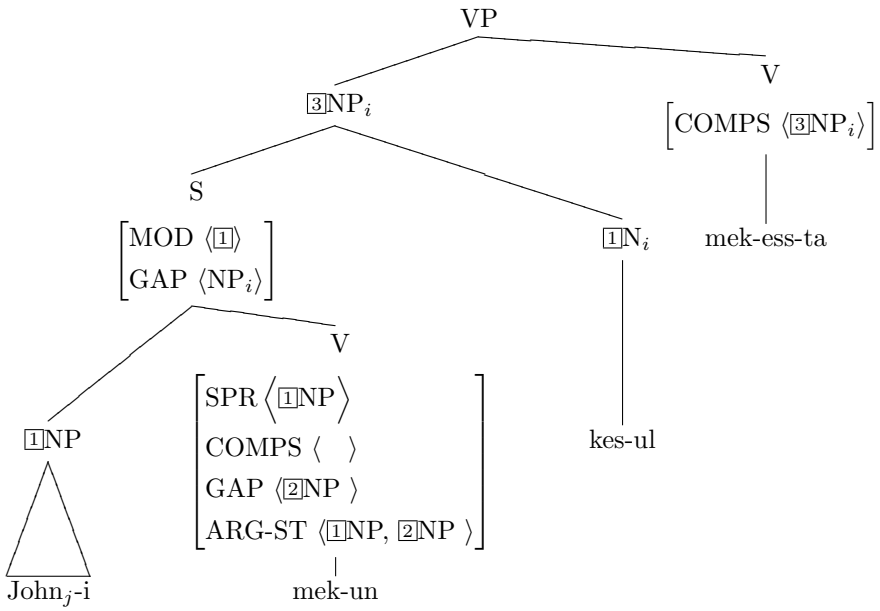
It can refer to an entity as well as a human; it can be pluralized; it can be a member of the set in the given context.

[11] The XARG (external argument) is linked to the subject's index value of all predicates (including both unaccusative and unergative verbs). See Bender 2002 and Kim 2006 for the function of this value.

[12] As an anonymous reviewer points out, there can be examples where the embedded semantic argument in the IHRC is an object, given that this argument is more discourse-prominent.

replaced by a common noun, implying that KES may have a similar semantics too. Let's consider the structure of (28)a first:

(29)



As we noted here, the noun KES in relative clauses is a common noun referring to an individual. Since the verb *mek-un* 'ate' also requires its object to be a referential individual, there is no mismatch between the two requirements. To observe how we obtain the semantics correctly, let us consider the MRS (minimal recursion semantics) representations when the gapped relative clause combines with its head noun KES:[13].

(30)
$$\begin{bmatrix} \text{INDEX } i \\ \text{RELS} \left\langle \begin{bmatrix} \text{PRED } \textit{one\_rel} \\ \text{ARG0 } i \end{bmatrix}, \begin{bmatrix} \text{PRED } \textit{eat\_rel} \\ \text{ARG0 } \textit{e1} \\ \text{ARG1 } j \\ \text{ARG2 } i \end{bmatrix}, \begin{bmatrix} \text{PRED } \textit{name\_rel} \\ \text{ARG1 } j \\ \text{CARG } \textit{john} \end{bmatrix} \right\rangle \end{bmatrix}$$

This simply means that there is an entity 'i' which the person named John eats. This index value is the patient that the person(s) referred to by the unrealized pronoun eat(s).

[13] See Kim (2004) and Kim and Yang (2004) 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 is rather

How about the cleft one? Before we provide its structure, consider the lexical entry for the copula *i-ta*. We have seen that the predicational and identificational copula is different in several respects as represented in the lexical entries:[14]

(31) a. Predicational:

$$\begin{bmatrix} \langle \text{i-ta} \rangle \\ \text{ARG-ST} \left\langle \text{NP}_i, \text{XP}_j \left[\text{PRD} + \right] \right\rangle \\ \text{SEM} \mid \text{RELS} \left\langle \begin{bmatrix} \text{PRED} & \textit{predicative\_rel} \\ \text{ARG1} & i \\ \text{ARG2} & j \end{bmatrix} \right\rangle \end{bmatrix}$$

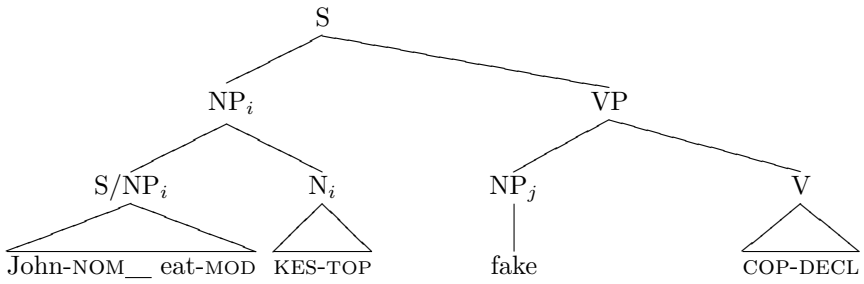
b. Identificational:

$$\begin{bmatrix} \langle \text{i-ta} \rangle \\ \text{ARG-ST} \left\langle \text{NP}_i, \text{NP}_j \right\rangle \\ \text{SEM} \mid \text{RELS} \left\langle \begin{bmatrix} \text{PRED} & \textit{identity\_rel} \\ \text{ARG1} & i \\ \text{ARG2} & j \end{bmatrix} \right\rangle \end{bmatrix}$$

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 (32) for the predicational cleft:

(32)



This structure, including the cleft clause as the subject and the predicative expression, will then induce the meaning similar to (31)a. The predicative expression 'fake' will predicate of this nominal element, inducing a semantic representation like the following:[15]

[14] Unlike Korean, Japanese allows the precopula expression to be GCASE (grammatical case marked).

[15] The index value of a predicative expression is identified with that of its subject or object that it is predicate of as illustrated by the following:

(33)
$$\begin{bmatrix} \text{RELS} \left\langle \begin{bmatrix} \text{PRED } \textit{one\_rel} \\ \text{ARG0 } i \end{bmatrix}, \begin{bmatrix} \text{PRED } \textit{fake\_rel} \\ \text{ARG1 } j \end{bmatrix}, \begin{bmatrix} \text{PRED } \textit{predicative\_rel} \\ \text{ARG1 } i \\ \text{ARG2 } j \end{bmatrix} \right\rangle \end{bmatrix}$$

The meaning of i 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 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.

We have also seen that even an adjunct can function as the predicative expression in the cleft clause. However, notice that KES cannot function as the head of a putative adjunct relative clause:

(34) a. [[John-i Mary-lul manna-n] kes-nun] ecey-i-ta
John-NOM Mary-ACC meet-MOD KES-TOP yesterday-COP-DECL
'(lit.) When John met Mary was yesterday.'

b. [[John-i Mary-lul manna-n] *kes/kos-eyse] pwul-i na-ss-ta
John-NOM Mary-ACC meet-MOD KES/place-LOC fire-NOM happen-COP-DECL
'The place where John met Mary had a fire.'

We assume that KES in (34)a here, different from the usage in the argument-gapped cleft clause, is combining not with an adjunct relative clause but with a fully saturated complement S. This in turn means that KES here is a bound noun, linked to an event. This event denoting KES clause is a predicative relation with the adjunct *ecey* 'yesterday'. That is, the semantics the analysis generates is something like the following:

(35)
$$\begin{bmatrix} \text{RELS} \left\langle \begin{bmatrix} \text{PRED } \textit{meet\_rel} \\ \text{ARG0 } \textit{e1} \\ \text{ARG1 } j \\ \text{ARG2 } m \end{bmatrix}, \begin{bmatrix} \text{PRED } \textit{yesterday\_rel} \\ \text{ARG0 } \textit{t1} \end{bmatrix}, \begin{bmatrix} \text{PRED } \textit{predicative\_rel} \\ \text{ARG1 } \textit{e1} \\ \text{ARG2 } \textit{t1} \end{bmatrix} \right\rangle \end{bmatrix}$$

In the present anlaysis, the argument-gapped cleft and adjunct gapped cleft are thus different: only the former is treated as a kind of unbounded dependency. This prediction is borne out:

(i)
$$\begin{bmatrix} \langle \text{fake} \rangle \\ \text{SUBJ } \langle \text{NP}_i \rangle \\ \\ \text{SEM} \mid \text{RELS} \left\langle \begin{bmatrix} \text{PRED } \textit{fake\_rel} \\ \text{ARG1 } j \end{bmatrix} \right\rangle \end{bmatrix}$$

As given here, the semantic argument of 'fake' is the index value of the subject it is predicated of.

(36) a. John-i [Mary-ka __ ilkessta-ko] sayngkakha-n kes-un i chayk-i-ta
John-NOM Mary-NOM read-COMP think-MOD KES-TOP this book-COP-DECL
'The one that John thinks Mary read is this book.'

b. John-i [Mary-ka i chaky-ul ilkessta-ko] malha-n kes-un i kos-(eyse)-i-ta
John-NOM Mary-NOM this book read-COMP think-MOD KES-TOP this place-COP-DECL
'The place that John said Mary read this book is in this place.'

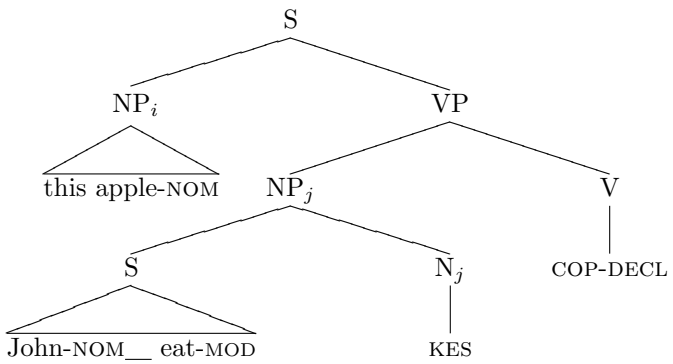
Though the precopular in (36)a expression is linked to the argument gap in the embedded clause, the adjunct precopular one in (36)b modifies only the matrix predicate 'said'.

Now consider the structure of an identificational cleft sentence:

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

A simple tree representation will be something like the following:

(38)



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:

(39)
$$\left[\text{RELS}\left\langle \begin{bmatrix}\text{PRED } \textit{name\_rel}\\ \text{ARG1 } i\\ \text{CARG } \textit{john}\end{bmatrix}, \begin{bmatrix}\text{PRED } \textit{eat\_rel}\\ \text{ARG0 } \textit{e1}\\ \text{ARG1 } i\\ \text{ARG2 } j\end{bmatrix}, \begin{bmatrix}\text{PRED } \textit{one\_rel}\\ \text{ARG1 } j\end{bmatrix}, \begin{bmatrix}\text{PRED } \textit{identity\_rel}\\ \text{ARG1 } i\\ \text{ARG2 } j\end{bmatrix} \right\rangle\right]$$

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, whose data we repeat here:

(40) a. *[kongwen-(eyse)]$_{\text{PP}}$-ka [John-i Mary-lul manna-n kes]-i-ta
park-at-NOM [John-NOM Mary-ACC meet-MOD KES]-COP-DECL

b. *[ecey]-ka John-i Mary-eykey senmwul-ul cwu-n kes-i-ta
yesterday-NOM John-NOM Mary-DAT present-ACC give-MOD KES-COP-DECL

The identificational cleft requires the two expressions to have the identical index value. The index value of either '(at) the park' or 'yesterday' can be identified with the KES in the cleft clause where it can be an referential entity or an event.

3.2 Eventual Cleft

As noted earlier, the event cleft construction is a construction where the entire matrix clause is headed by the nominalizer KES followed by the copula.

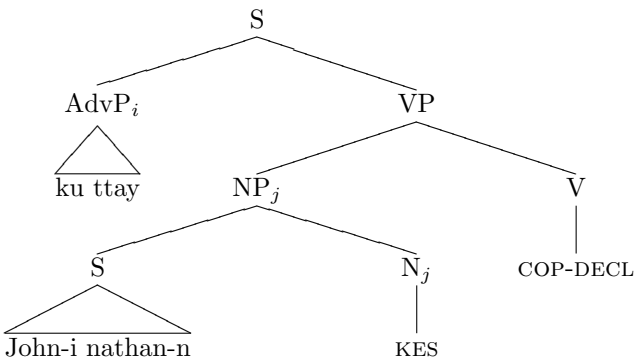
construction of the cleft is more complicated. One tricky fact is that the eventual cleft can be similar to identificational cleft.

(41)
$$\left[\begin{array}{l}\langle \text{i-ta} \rangle \\ \text{ARG-ST} \left\langle \text{XP}_i[\textit{adverbial}],\ \text{S}_{s1}\begin{bmatrix}\text{VFORM -n}\\ \text{NFORM kes}\end{bmatrix}\right\rangle \\ \text{SEM}\,|\,\text{RELS} \left\langle \begin{bmatrix}\text{PRED } \textit{event-related\_rel}\\ \text{ARG1 } i\\ \text{ARG2 } s1\end{bmatrix}, \begin{bmatrix}\text{PRED } \textit{focus\_rel}\\ \text{ARG1 } s1\end{bmatrix}\right\rangle\end{array}\right]$$

A sample tree for event cleft: The subject is realized as a *pro* which can be linked to a temporal point or anything that can cause the event denoted by the cleft-clause. This is similar to English it-clefts like *It is then that Tom ate the big apple.* or *It was that ...*

A simple tree representation will be something like the following:

(42)



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:

(43)
$$\begin{bmatrix} \text{RELS} \left\langle \begin{bmatrix} \text{PRED } \textit{temporal\_rel} \\ \text{ARG1 } k \end{bmatrix}, \begin{bmatrix} \text{PRED } \textit{name\_rel} \\ \text{ARG1 } i \\ \text{CARG } \textit{john} \end{bmatrix}, \begin{bmatrix} \text{PRED } \textit{appeared\_rel} \\ \text{ARG0 } e1 \\ \text{ARG1 } i \end{bmatrix}, \right. \\ \left. \begin{bmatrix} \text{PRED } \textit{one\_rel} \\ \text{ARG1 } e1 \end{bmatrix}, \begin{bmatrix} \text{PRED } \textit{event-related\_rel} \\ \text{ARG1 } k \\ \text{ARG2 } e1 \end{bmatrix}, \begin{bmatrix} \text{PRED } \textit{focus\_rel} \\ \text{ARG1 } e1 \end{bmatrix} \right\rangle \end{bmatrix}$$

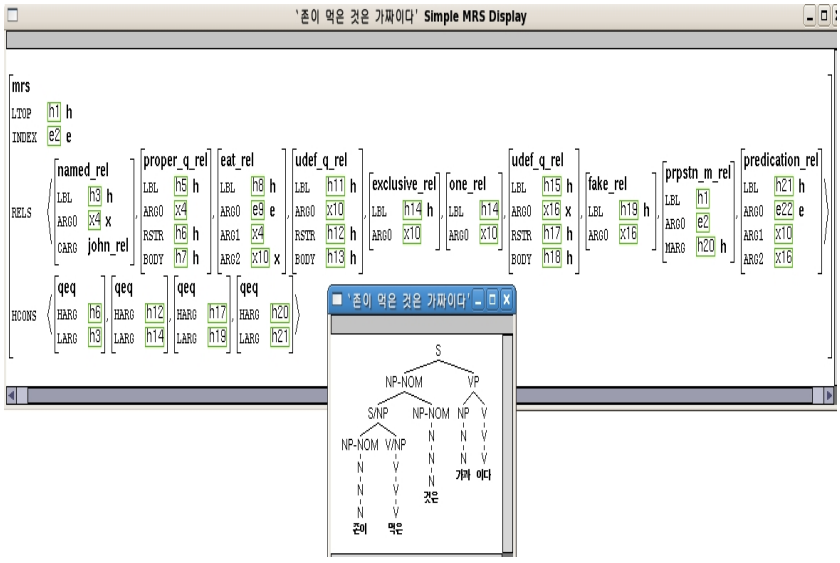
4 Results of the Implementation

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. [**?**] and [**?**]). To test its performance and feasibility, it has been implemented into the LKB (Linguistic Knowledge Building).[16] The test results give the proper syntactic as well as semantic structures for all the coordination patterns from simple binary or ternary to complex patterns we find in the language.

For example, ((44)) is the syntactic and MRS structure for the example .....

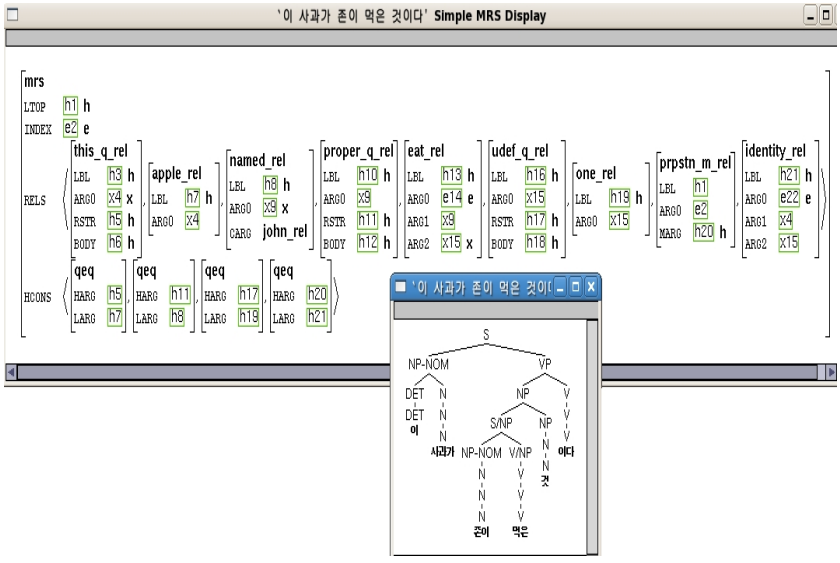
(44) a. This apple is what John ate.
b. What John ate is a fake.
c. John-i sakwa-lul mek-un kes i-ta

(45)



[16] 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.

(46)



si-wa kulim-kwa kuliko iyaki 'poem-and, picture-and and story' where the morphological marker *-wa* and the lexical coordinator *kuliko* occur together. In terms of the syntactic structures, the grammar generates only one structure for the NP as given in the output here: *kulim-kwa kuliko iyaki* forms a *tern-conj-ph* and then this resulting phrase will form a *bin-conj-ph* with *si-wa*.[17] We can notice here that the MRS the grammar generates provides enriched information of the phrase. The value of LTOP is the local top handle, the handle of the relation with the widest scope within the constituent. The attribute RELS is basically a bag of elementary predications (EP) each of whose value is a *relation*.[18] Each of the types *relation* has at least three features LBL, PRED (represented here as a type), and ARG0. The INDEX value here is identified with the ARG0 (C-ARG) value of the first *and_rel* within the RELS list here. The L-INDEX value of this relation is identified with the *udef_q_rel* for the noun *poem* that serves as the first conjunct.[19] The R-INDEX value is identified not with any conjunct but with the ARG0 of the other *and_rel* representing the semantics of *kulim-kwa kuliko iyaki* 'picture and story'.

[17] The system does not combine *si-wa* with *kulim-kwa* first since the latter is marked with [COORD *and*], which would violate the constraint on *coord-ph*.

[18] The attribute HCONS is to represent quantificational information. See [**?**].

[19] Korean common nouns do not require a determiner to project an NP. Even though a determiner is not available, we need to express an underspecified quantification on the noun in order to make the semantics compatible with the semantic output of other languages, and to make scope restrictions work. Such a move is essential in deep processing aimed at multilingual applications.

5 Conclusion

We have seen that there are two types of Korean clefts constituted of a cleft clause, focused expression, and a copula. These predicational and identificational cleft inherit many properties from the corresponding copula constructions.

We first have seen that KES is better treated as a nominal element as a morphosyntactic category. It can refer to either an individual or an event; its reference value can be determined either by a context or an external environment such as the main verb. The pronoun KES in the cleft clause refers to an individual entity as in the relative clause. Given these basic assumptions, we can provide a clean analysis of Korean cleft constructions that can be computationally implemented.

References

Bender, Emily M., and Dan Flickenger, and Stephan Oepen. 2002. The Grammar Matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, 8-14, Taipei, Taiwan.

Chung, Chan and Jong-Bok Kim. 2003. Differences between externally and internally headed relative clause constructions. In J.B. Kim and Stephen Wechsler (eds.) *Proceedings of HPSG 2002*, 3–25.

Collins, Peter C. 1991. *Cleft and Pseudo-Cleft Constructions in English.* London and New York: Routledge.

Copestake, Ann. 2002. *Implementing Typed Feature Structures.* Stanford: CSLI Publications.

Copestake, Ann, Dan Flickinger, Carl Pollard, and Ivan A. Sag. 2005. Minimal Recursion Semantics: an Introduction. *Research on Language and Computation* 3.4: 281–332.

Davidse, Kristin. 2000. A Constructional Approach to Clefts. *Linguistics* 38.6: 1101-1131.

Heycock, Caroline and Anthony Kroch. 1999. Pseudo-cleft connectedness: Implications for the LF interface level. *Linguistic Inquiry* 30, 365-397.

Jo, Jung-Min. 2004. *Grammatical Effects of Topic and Focus Information.* Ph.D. Dissertation. University of Illinois, Urbana-Champaign.

Kang, Bosook. 2006. Some pecularities of Korean 'kes' cleft constructions. *Studia Linguistica* 60.3: 251–281.

Kim, Jong-Bok. 2004. *Korean Phrase Structure Grammar* (In Korean). Hankook Publishing.

Kim, Jong-Bok. 2006. Minimal Recursion Semantics: An Application into Korean. *Journal of the Linguistic Association of Korea*, 14.2: 59–85.

Kim, Jong-Bok and Jaehyung Yang. 2004. Projections from Morphology to Syntax in the Korean Resource Grammar: Implementing Typed Feature Structures, *Lecture Notes in Computer Science*, Vol.2945, pp.13-24, Springer-Verlag, 2004.2.

Kim, Jong-Bok, Peter Sells and Michael T. Wescoat. 2004. Korean Copular Constructions: A Lexical Sharing Approach. To appear in M. Hudson, S.-A. Jun and P. Sells (eds.) *Proceedings of the 13th Japanese/Korean Linguistics Conference.* Stanford, CSLI Publications.

Kim, Jong-Bok and Peter Sells. 2007. Nominalizer *kes* and Information Structure. Paper presented in the 12th ISOKL, Harvard University, August 3–5, 2007.

Jhang, Sea-Eun. 1994. *Headed nominalizations in Korean: Relative clauses, clefts, and comparatives.* Ph.D. Dissertation. Simon Fraser University.

Quirk, Randolph, S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A comprehensive grammar of the English language.* London: Longman.

Sohn, Keun-Won. 2004. Kes-clefts, Connectedness effects, and the implications thereof. *Studies in Generative Grammar*, 14: 561–571.