

# Parsing Korean Comparative Constructions in a Typed-Feature Structure Grammar

Jong-Bok Kim, Jaehyung Yang, and Sanghoun Song\*

Kyung-Hee University, Kangnam University, University of Washington

**Jong-Bok Kim, Jaehyung Yang, and Sanghoun Song. 2010. Parsing Korean Comparative Constructions in a Typed-Feature Structure Grammar.** *Language and Information* 14.1, 1–24. The complexity of comparative constructions in each language has given challenges to both theoretical and computational analyses. This paper first identifies types of comparative constructions in Korean and discusses their main grammatical properties. It then builds a syntactic parser couched upon the typed feature structure grammar, HPSG and proposes a context-dependent interpretation for the comparison. To check the feasibility of the proposed analysis, we have implemented the grammar into the existing Korean Resource Grammar. The results show us that the grammar we have developed here is feasible enough to parse Korean comparative sentences and yield proper semantic representations though further development is needed for a finer model for contextual information. (Kyung Hee University, Kangnam University, University of Washington)

**Key words:** Korean comparatives, contextual-dependent, direct, reduced, Korean Resource Grammar

## 1. Types of Korean Comparative Constructions

Comparison constructions, involving comparing two participants in terms of the degree of some gradable property relating to them, are encoded differently in each language. Korean also employs quite different morphological and syntactic properties from languages like English and even Japanese. As illustrated in the following two main types of comparatives in (1), Korean uses the optional comparative marker *te* ‘more’, the postpositional standard marker *pota* ‘than’ as basic elements (cf. Jhang 2001, Choe 2008, Kim and Sells 2009, 2010):

---

\* We thank three anonymous reviewers for helpful comments. Our thank also goes to Incheol Choi, Chan Chung, and Jae-Woong Choe for comments and suggestions. This work was supported by the Korea Research Foundation (Grant No. 2009-A00065).

- (1) a. Phrasal Comparative (PC):  
 tongsayng-i            hyeng-pota            chayk-ul (te) manhi  
 younger.brother-NOM older.brother-than book-ACC more many  
 ilkessta  
 read  
 'The younger brother read more books than his older brother.'
- b. Clausal Comparative (CC):  
 tongsayng-i            [[hyeng-i            — ilk-un]            kes-pota]  
 younger.brother-NOM older.brother-NOM            read-MOD kes-than  
 chayk-ul (te) manhi ilkessta  
 book-ACC more many read  
 'The younger brother read more than what his older brother read  
 (or his older brother did).'

As illustrated here, phrasal comparatives (PC) involve two compared nominals whereas clausal comparatives (CC) have core clausal properties. In the PC (1a), the 'standard' of comparison expression *hyeng* 'older.brother' combines with the postpositional standard marker *pota* 'than'. The target of comparison *tongsayng* 'younger.brother' functions as the subject and the comparative morpheme is realized as an optional adverb *te* 'more'.<sup>1</sup> Unlike this PC example, the standard of comparison in the CC example (1b) is clausal with a syntactic gap. The gapped element in the clause functions as the object of the embedded predicate *read* and is followed by the noun *kes* which can be replaced by a canonical noun like *chayk* 'book'.<sup>2</sup>

With the strong motivation for capturing the truth conditionally identical meaning between phrasal and clausal comparatives as illustrated in (1), the standard analysis of comparatives has assumed that phrasal comparatives are derived from clausal sources through deletion rules. For example, the phrasal comparative in (1a) has the LF structure like the following as a source sentence (cf. Bresnan 1973, Bhatt and Takahashi 2007, Pancheva 2006 for English and Park 2008 for Korean):

- (2) The younger brother read [[*d*-many] books] than the older brother  
 read [~~*d*-many books~~] (*d* = degree)

The transformational rules (e.g., comparative deletion and ellipsis) delete the *d*-many books in the comparative clause, in addition to eliding the remaining constituents in the comparison under identity.

To see if all Korean comparatives can be grouped into these two clausal and phrasal types, we first extracted comparative sentences from the illustrated examples for the verbal and adjectival entries in the Sejong Electronic Dictionary:

<sup>1</sup> Choe (2008) provides a detailed discussion about the status of the optional *te*.

<sup>2</sup> As we will see in due course, the more correct interpretation for clausal comparatives seems to be a free-relative clause like interpretation.

## (3) Sejong Electronic Dictionary Entries

	total entry #	sample Ss #	NP- <i>pota</i> Ss	CM <i>pota</i> Ss
va entries	4,389	14,816	196	6
vv entries	15,181	52,981	298	35
Total	19,570	67,797	486	41

As indicated, from the 67,797 sample sentences for the adjectival (va) and verbal (vv) lexical entries, we extracted total 486 comparative sentences including an NP-*pota* ‘expression’ and total 41 sentences where *pota* is used as a comparative marker as in (10). We analyzed these 486 sentences and identified the following additional types which are hard to be classified either as PC or CC examples:

## (4) Unreduced:

kopseymha-nun kes-i tesseymha-nun kes-pota  
multiplication-do-MOD thing-NOM addition.do-MOD thing-than  
elyepta  
difficult  
‘Doing multiplication is more difficult than doing addition.’

## (5) Missing only a measure expression:

cwasek-ul 134 sek-pota te nulliessta  
seat-ACC 134 seat-than more increase  
‘(We) increased the seats more than 134.’

## (6) Missing the head noun:

ku chinkwu ima-nun na-pota te pantulkelye  
the friend forehead-TOP I-than more shiny  
‘The friend’s forehead is more shiny than mine.’

## (7) S expected type:

ceypwum-uy cil-i wulli-ka kitayha-n kes-pota  
product-GEN quality-NOM we-NOM expect-MOD kes-than  
mihuphata  
insufficient  
‘The product’s quality is less sufficient than we expected.’

## (8) NP expected type:

ku senswu-nun kitay-pota mihuphata  
the athlete-TOP expectation-than insufficient  
‘The athlete is less qualified than the expectation.’

- (9) Temporal Comparison:  
 mwulka-ka caknyeon-pota manhi ollassta  
 prices-NOM last.year-than many increased  
 'Prices are raised more than last year.'
- (10) Comparative marker:  
 cwunghwancatul-un inkun ciyeok-uy pota khun pyongwon-ulo  
 serious.patient-TOP nearby area-GEN more big hospital-to  
 hwusongtoyessta  
 sent  
 'Patients in serious conditions were sent to a bigger hospital nearby.'
- (11) Metalinguistic comparatives:  
 ku kes-un inkan-uy uysa-i-laki-pota hanul-uy  
 the thing-TOP human-GEN intention-COP-NMLZ-than sky-GEN  
 ttus-i-ta  
 intention-COP-DECL  
 'That is not human beings' intention, but is rather the heaven's intention.'
- (12) Superlative:  
 nwukwu-pota ne-lul salanghanta  
 who-than you-ACC love  
 'I love you more than anyone.'

All these types given above behave differently from PC or as CC types. For example, (4) is different from the CC in the sense that the clause contains no syntactic gap. This kind of example compares the degree properties of two events. Measure comparatives in (5) are different from either PC or CC types in the sense that they just describe the degree of an individual's property on the given scale with a measure expression. Examples like (6) are also peculiar since the friend's forehead is syntactically compared with not my forehead but me, which is not possible in English:

- (13) a. \*The friend's forehead is more shiny than me.  
 b. \*The friend's forehead is more shiny than I.

Comparatives in (7) do not express the sentential complement of the standard clause, whereas comparatives in (8) have only a verbal noun which may also take a sentential complement. Examples like (9) also can form an independent type in that the standard of comparison just denotes a temporal point.

Another intriguing property of Korean comparative constructions, as given in (10), is that the standard marker *pota* can also be used as a comparative expression, meaning *more*. The multi-function property of *pota* allows us to have different uses of *pota* in the same sentence (Kim and Sells 2009):

- (14) a. calinkopi-pota pota hyenmyengha-key sopihan-ta  
 miser-than more wise-COMP consume-DECL  
 ‘(He) consumes more wisely than a miser.’
- b. wuli-ka sayngkakha-yess-ten kes-pota pota caymi-iss-ta  
 we-NOM think-PAST-MOD kes-than more interesting-be-DECL  
 ‘It was more interesting than we thought.’

The two uses of *pota* here are obviously different: the first *pota* in each (a) case is a postpositional marker attached to the standard of comparison whereas the second one in each case, used as a comparative marker, functions as an adverb.

One additional type of Korean comparatives we have identified from the corpus is meta-linguistic comparatives as given in (11). These meta-linguistic comparatives, different from regular comparatives by their semantics, are canonically nominalized by the suffix *ki*. The final type, distinguished from the rest, is the one where the comparative marker *pota*, attached to an indefinite pronoun like *nwuku* ‘who’, induces a superlative meaning, as exemplified in (12).

These varied types indicate the complexity of comparative constructions. In what follows, we first provide a syntactic analysis that can parse their structures and discuss the semantics.

## 2. Parsing the Structure

### 2.1 Clausal Comparatives

As noted in the previous section, clausal comparatives allow the complement of the postpositional *pota* to consist of a gapped clause with the noun *kes*.<sup>3</sup> In fact, a rich set of empirical data indicates that the clause-like complement is a free relative NP headed by *kes*. Previous literature has assumed that the noun *kes* is a complementizer introducing a CP (e.g., Lee 2002, Park 2009). However, evidence undermines this assumption. For example, the complement clause of *pota* can occur only in the NP position:<sup>4</sup>

- (15) a. [nay-ka mantu-n kes]-ul phalassta  
 I-NOM make-MOD kes-ACC sold  
 ‘(He) sold the one that I made.’
- b. i chayksang-i [nay-ka mantu-n kes]-pota khuta  
 This desk-NOM I-NOM make-MOD kes-than big  
 ‘This desk is bigger than the one I made.’

The fact that the object of the verb *phalassta* ‘sold’ can be only an NP implies that the clause headed by *kes* here is also an NP. The identical expression appears

<sup>3</sup> Much of the syntactic analysis for Korean comparatives set forth here follows that of Kim and Sells (2009).

<sup>4</sup> This is another difference from Japanese comparatives allowing plain clauses to be the complement of *yoru* ‘than’.

in the complement position of *pota* as in (15b). In addition, *kes* in clause-like comparatives can be replaced by a common noun as in (16a) and further be preceded by a determiner as in (16b):

- (16) a. Yongho-nun [Chelswu-ka sa-n chayksang]-pota pissan  
 Yongho-TOP Tom-NOM buy-MOD desk-than expensive  
 kes-ul sa-ass-ta  
 thing-ACC buy-PAST-DECL  
 'Yongho bought a more expensive desk than what Chelswu bought.'
- b. Yongho-nun [Chelswu-i sa-n ku kes]-pota pissan  
 Yongho-TOP Chelswu-NOM buy-MOD the thing-than expensive  
 kes-ul sa-ass-ta  
 thing-ACC buy-PAST-DECL  
 'Yongho bought a more expensive book than the one Chelswu bought.'

If *kes* in comparatives were simply a complementizer, such a behavior would not be expected: no complementizer can be replaced by a common noun or combine with a determiner.

The functional noun *kes* in Korean is canonically referring to a nonanimate entity or an event. Consider two main uses where the expression *kes* itself appears (Kim and Sells 2007):

- (17) a. [nay] kes-i [ne] kes-mankhum khu-ta  
 my thing-NOM your thing-as big-DECL  
 '(Lit.) My thing is as big as yours.'
- b. [[Yongho-ka \_\_ mek-un] kes]-ul mek-ess-ta  
 Yongho-NOM eat-MOD kes-ACC eat-PAST-DECL  
 '(We) ate the thing that Yongho ate.'
- c. [[Yongho-ka talli-nun] kes]-ul moll-ass-ta  
 Yongho-NOM run-MOD kes-ACC not.know-PAST-DECL  
 '(We) didn't know that Yongho was running.'

The noun *kes* in (17a) combines with a determiner (a specifier) whereas in (17b) it combines with a relative clause. In both examples, *kes* has a meaning like 'thing'. In (17c) it combines with a complete clause, referring to the event denoted by that clause. But note that *kes* cannot refer to a person:<sup>5</sup>

- (18) a.\*[[Yongho-ka \_\_ manna-n] kes]-ul manna-ss-ta  
 Yongho-NOM meet-MOD kes-ACC meet-PAST-DECL  
 '(We) met the person Yongho met.'

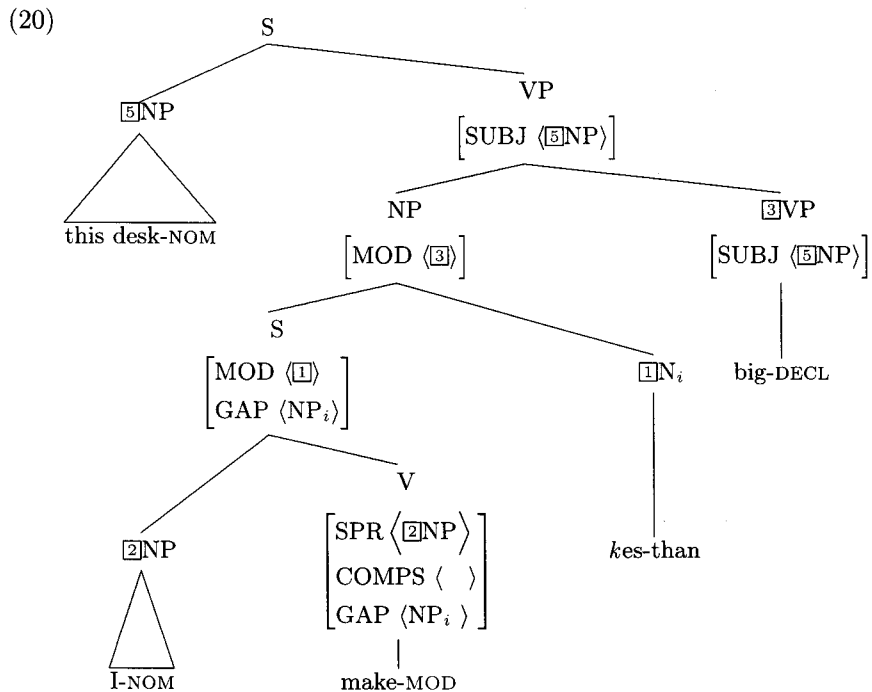
<sup>5</sup> See Kim (2008) and Kim and Sells (2007) for an account of *kes* and comparison with traditional analyses.

- b. ce \*kes/salam-i sacang-i-ta. (referring to a man passing by)  
 that thing/person-NOM president-COP-DECL  
 ‘That man is president (of the company).’

The restriction on what *kes* can refer to hold in comparative construction, too: it cannot refer to a person:

- (19) Yongho-nun [Chelswu-ka manna-n \*kes/salam]-pota chakha-n  
 Yongho-TOP Chelswu-NOM meet-MOD kes/person-than honest-MOD  
 salam-ul mannassta  
 man-ACC met  
 ‘Yongho met a more honest man than the person Chelswu met.’

As seen here, to compare the properties of two persons, the head of the clausal comparative must be a personal common noun. Based on these observations, we assume that clausal-like comparatives are basically a nominal clause, as represented in the following structure for the sentence (15b), meaning this desk is bigger than the one I made:



As given in the structure here, *kes-than* is combining with a sentence with a missing element. This sentence, like a relative clause, misses the object of *mantu-n* ‘make-MOD’, as represented in the GAP value. This missing element at first is coindexed with the head noun *kes*. This complex NP, consisting of an object-missing S and the head *kes-than*, functions as a modifier to the gradable predicate *khu-ta* ‘big’.

The structure thus assumes that clausal comparatives are in fact NP-phrasal comparatives.

As we have seen, there are also cases where *kes* clauses with no syntactic gap, as illustrated by further examples in (21):

- (21) a. Yongho-nun [Mary-ka talli-n kes]-pota te ppali  
 Yongho-TOP Mary-NOM run-MOD *kes*-than more fast  
 kel-ess-ta  
 walked-PAST-DECL  
 'Yongho walked faster than Mary ran'.  
 b. [wuli-ka ka-nun kes]-i [haksayng-tul-i o-nun kes-pota]  
 we-NOM go-MOD *kes*-NOM student-PL-NOM come-MOD *kes*-than  
 phyenha-ta  
 convenient-DECL  
 'For us to go is more convenient than for students to come.'

Within the relative clause analysis we adopt here, such gapless examples are expected when considering Korean also has amount relative clauses:

- (22) [ku senswu-ka talli-n sokto]-nun hankwuk kilok-i-ta  
 the athlete-NOM run-MOD speed-TOP Korean record-COP-DECL  
 'The speed at which the athlete ran is a national record.'

The head noun *kes* in (21) can be replaced by a noun like *cengto* 'degree', *sokto* 'speed', or *kil* 'way':

- (23) a. Yongho-nun [Mary-ka talli-n cengto/sokto]-pota te ppali  
 Yongho-TOP Mary-NOM run-MOD degree/speed-than more fast  
 kel-ess-ta  
 walk-PAST-DECL  
 'Yongho walked faster than the speed that Mary ran'.  
 b. [wuli-ka ka-nun kil]-i [haksayng-tul-i o-nun  
 we-NOM go-MOD way-NOM student-PL-NOM come-MOD  
 pangpep-pota] phyenha-ta  
 way-than convenient-DECL  
 'For us to go is a more convenient way than for students to come.'

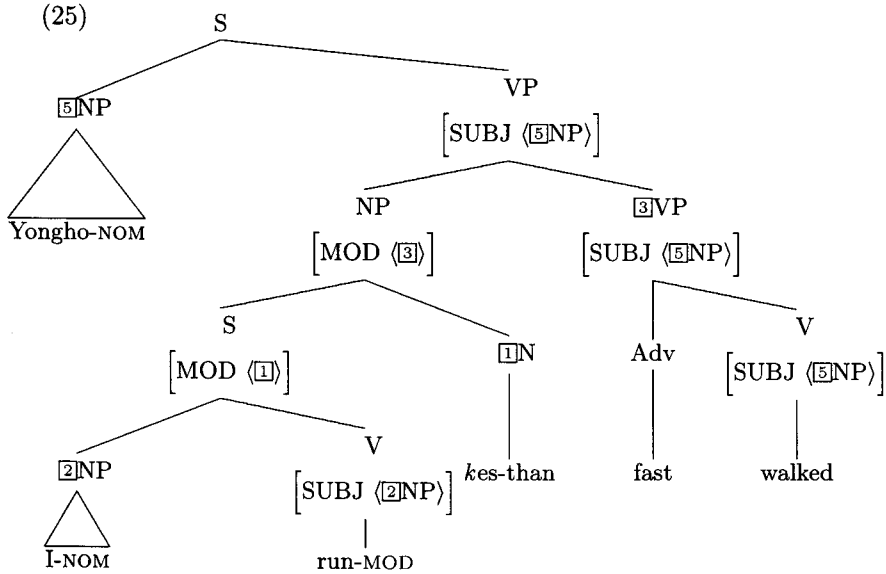
In fact, all the clause-like comparatives with no overt gap can be reinterpreted as amount or degree relative clause. We can notice that even the S-operator clausal examples we have seen in (7) also can be interpreted as a degree relative clause:

- (24) a. Yongho-nun [Bill-i ttokttokha-n kes/cengto]-pota te  
 Yongho-TOP Bill-NOM smart-MOD *kes*/degree-than more  
 ttokttokhata  
 smart  
 'Yongho is smarter than Bill.'



- b. Yongho-nun Bill-i sayngkakha-n kes/cengto-pota te  
 Yongho-TOP Bill-NOM think-MOD kes-than more  
 ttokttokhata  
 smart  
 ‘Yongho is smarter than Bill thought (he is).’

Given that adverbial relative clauses are gapless modifying structures, the un-gapped clause comparatives would assign the following structure for sentences like (21a):



The structure indicates that the gapless clause ‘I run’ modifies the head noun *kes* which can be reinterpreted as a certain degree. This NP-*pota* then is modifying the predicate *fast walked*.<sup>6</sup>

## 2.2 Phrasal Comparatives: Coordination and Modification

As noted earlier, the standard marker *-pota* can be attached only to a nominal element:

- (26)\*Yongho-ka yeppukey-pota sinnakey nolayha-yess-ta  
 Yongho-NOM beautifully-than joyfully sing-PAST-DECL  
 ‘(intended) Yongho sang joyfully rather than beautifully.’

The phrasal comparatives thus allows only an NP-*pota* phrase. This NP-*pota* phrase has rather flexible distributional possibilities:

- (27) a. swuphil-i cengmal sosel-pota caymi-iss-ta  
 essay-NOM really novel-than interesting-be-DECL  
 ‘Essays are really more interesting than novels.’

<sup>6</sup> See section 3 for the discussion of semantics for such an example.

- b. *sosel-pota cengmal swuphil-i caymi-iss-ta*  
 novel-than really essay-NOM interesting-be-DECL  
 'Essays are really more interesting than novels.'

As illustrated here, in general, there is no precedence constraint between the target NP and the standard expression, and further the two need not be adjacent to each other.

Note that the host NP of the marker *pota* can be semantic-case marked:<sup>7</sup>

- (28) a. *tosekwan(-eyse)-pota chaykpang-eyse kongpwu-ka te cal toynta*  
 library-at-than bookstore-at study-NOM more well become  
 'Studying at a bookstore is better than studying at the library.'
- b. *tosekwan(-\*ulo)-pota chaykpang-eyse kongpwu-ka te cal toynta*

Although the locative postposition is optional in the *than* XP, when it is present, it must be identical with the one in the associate (target) NP. An intriguing property is that when the standard phrase is locative, the possibility of scrambling disappears:

- (29)\**chaykpang-eyse tosekwan(-eyse)-pota kongpwu-ka te cal toynta*  
 bookstore-at library-at-than study-NOM more well become

Such a constraint can be further observed even with no semantic case in the standard expression:<sup>8</sup>

- (30) a. *i os-un paykhwacem-pota sicang-eyse cal pali-n-ta.*  
 this clothes-TOP dept. store-than market-at well sell-PRES-DECL  
 'The clothes sell well at the department store rather than at the market.'
- b. *\*i os-un sicang-eyse paykhwacem-pota cal palinta.*  
 this clothes-TOP market-at dept. store-than well sell

Another language particular property of the NP-*pota* phrase is that Korean allows more than one NP-*pota* phrase. In such case too, these standard expressions must be adjacent:

- (31) a. *yenge-pota cwungkwuke-pota hankwuke-ka elyep-ta*  
 English-than Chinese-than Korean-NOM difficult-DECL  
 '(lit.) Korean is more difficult than English and Chinese.'
- b. *\*yenge-pota hankwuke-ka cwungkwuke-pota elyep-ta*  
 English-than Korean-NOM Chinese-than difficult-DECL

<sup>7</sup> Following Kim (2004), we take canonical postpositions to be just nominal suffixes, implying that there is no category of PP distinct from NP.

<sup>8</sup> To some speakers, examples like (29) as well as (30b) are acceptable, but most of the non-linguist speakers have a clear contrast between canonical and scrambled ones.

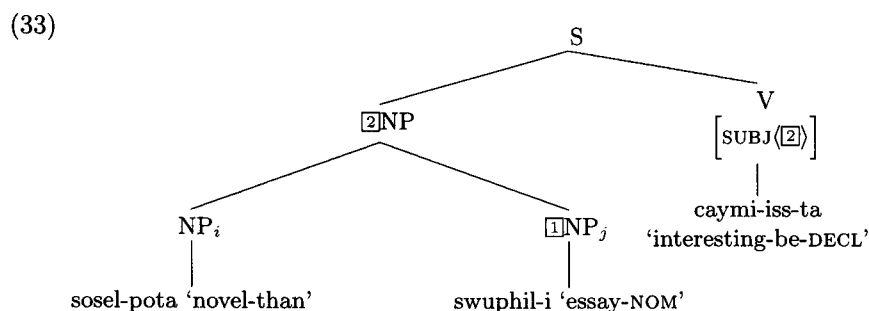
This again indicates that NP-*pota* forms a constituent with the associate NP that follows it.

This contrast indicates that the *pota*-phrase cannot be scrambled freely. Also consider further examples:<sup>9</sup>

- (32) a. *te manhun salamtul-i [sosel-pota] swuphil-ul ilkessta*  
 more many people-NOM novel-than essay-ACC read  
 ‘More people read essays than novels.’
- b. *[sosel-pota] swuphil-ul te manhun salamtul-i ilkessta*  
 novel-than essay-ACC more many people-NOM read
- c. *??\*swuphil-ul [sosel-pota] te manhun salamtul-i ilkessta*  
 essay-ACC novel-than more many people-NOM read
- d. *\*swuphil-ul te manhun salamtul-i [sosel-pota] ilkessta*  
 essay-ACC more many people-NOM novel-than read

This ordering restriction suggests that there should be a configuration where the two compared individuals are combined. The most natural position is the standard and the compared parameter in adjacent positions.<sup>10</sup>

Based on the observations that the simple NP-*pota* prefers to combine with the associate NP when it is immediately followed, we assume they form a coordination-like structure as illustrated in the following:



As indicated here, the standard expression combines with the associate NP, forming a bigger NP.<sup>11</sup> This analysis, assuming the existence of base-generated phrasal comparatives, thus treats the ‘standard’ and compared phrase as an NP coordination.<sup>12</sup>

<sup>9</sup> Examples like (31c) and (31d) can be better with a long pause after the associate *swuphil-ul*. See (??) for further discussion.

<sup>10</sup> The bad scrambled examples here can be improved a lot if there is an intervening expression between the associate NP and the scrambled expression.

<sup>11</sup> The coordination marker *-wa* ‘and’ behaves similar to *pota* in many respects: they attach only to an NP, can follow the associate NP, can have multiple identical phrases in order. See Kim and Sells (2009).

<sup>12</sup> The feature DEG ensures that the standard phrase modifies only a gradable predicate. As we have seen in before, there are cases where the comparative marker *te* ‘more’ is obligatory.

$$(34) \left[ \begin{array}{l} n\text{-than-nmod} \\ \text{SYN} \left[ \begin{array}{l} \text{HEAD | POS } n\text{oun} \\ \text{MOD} \left\langle \begin{array}{l} \text{DEG +} \\ \text{POS } n\text{ominal} \\ \text{IND } j \end{array} \right\rangle \end{array} \right] \end{array} \right]$$

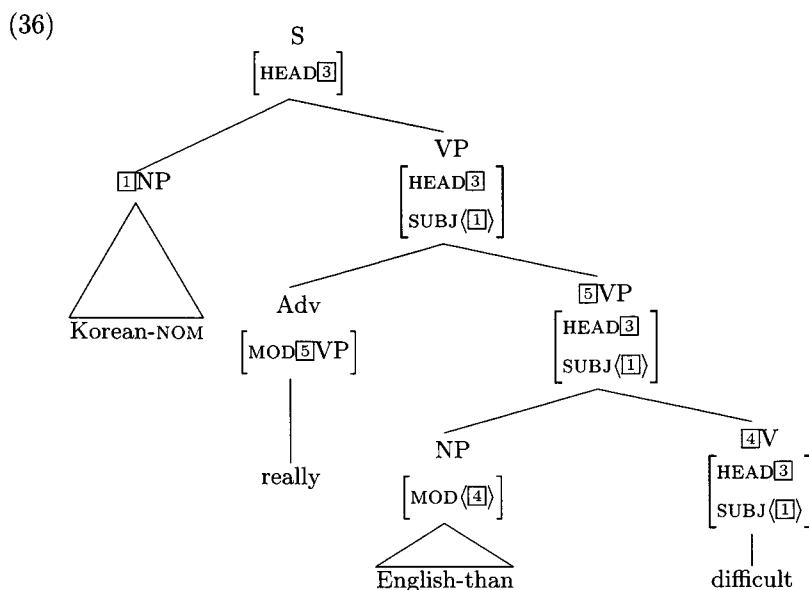
We can observe that there are many contexts where the NP-*pota* and its associate NP are not adjacent:

- (35) a. hankwuke-ka (cengmal) yenge-pota elyep-ta  
 Korean-NOM really English-than difficult-DECL  
 'Korean is really more difficult than English.'
- b. yenge-pota cengmal hankwuke-ka elyep-ta  
 English-than really Korean-NOM difficult-DECL  
 'Korean is really more difficult than English.'

In order to capture such flexible, distributional possibilities of the standard of comparison NP-*pota* expression in a surface-oriented grammar, we assume that the phrase can syntactically modify a verbal element, as represented in the following:

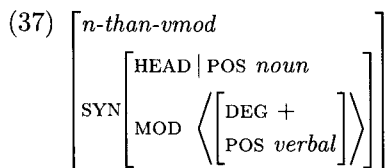
- 
- (i) a. [pyongso-pota [samsip pwun-i \*(te) keli-ess-ta]]  
 normal-than 30 minutes-NOM \*(more) take-PAST-DECL  
 'It took 30 more minutes than usual.'
- b. [nam-pota [(te) mek-ess-ta]]  
 others-than \*(more) eat-PAST-DECL  
 '(He) ate more than others.'

Our assumption is that the comparative marker *te* changes the nongradable predicates *keli-ess-ta* 'take' and *mek-ess-ta* 'eat' into gradable ones so that the standard can be semantically interpreted. If a given predicate is itself gradable, the comparative marker *te* is therefore optional in comparative constructions.



In this structure, the NP-*pota* modifies the verbal predicate *difficult*, forming a modifier structure.

This kind of modification syntactic structure is basically projected from the following lexical entry:



This lexical entry means that the XP-*pota* syntactically functions as a modifier to a verbal element.

An issue may arise from assuming two different functions of the NP-*pota*, one modifying the following associate NP, forming a coordination-like structure, and the other modifying a verbal predicate.<sup>13</sup> This will be a burden to the grammar, but seems to be inevitable when considering the distributional properties of the NP-*pota* as well as its semantic interactions.

### 3. Non-canonical Comparatives and Semantics of the Comparatives

There are two main issues we haven't discussed yet: semantics of the comparative constructions and analysis of the other types not belonging to either phrasal or clausal types. In terms of semantics, phrasal comparatives appear to be similar to

<sup>13</sup> As also questioned by a reviewer, the present analysis assigns two parsing trees for the sentences for the ordering in which the associate immediately precedes the NP-*pota*. As noted, this kind of spurious ambiguity may be an issue in the present analysis.

clausal comparatives. For example, the PC in (1a) and the CC in (1b) will have the identical LF structure:

- (38) [[MORE]] ( $\lambda d$  the younger brother read  $d$ -many books) ( $\lambda d$  the older brother read  $d$ -many books).

In capturing the systematic meaning relationships between PCs and CCs and further computing the semantics in a compositional way, literature has assumed to derive phrasal or non-clausal comparatives from clausal sources (cf. Kennedy 2007). In a compositional analysis, for example, the standard expression, the complement of *than*, denotes a set of degrees compared to the degree in the matrix clause while the comparative morpheme (MORE) denotes a relation between two sets of degrees. The main gist of such an analysis is that the *than*-clause and the main clause provide a predicate of degrees.

However, there are many obstacles to compose the meaning of comparatives in a compositional way. The first issue is the status of the functor 'MORE' that selects two propositional arguments. In languages like Korean, the functor, comparative marker, is not present in syntax always. Unlike the English comparative marker, *te* 'more' is optional in most cases. Within a compositional analysis where the comparative marker *more* is a functor taking two degree-denoting arguments, we need to assume an invisible comparative morpheme, MORE, as in (38).

A second major issue that arises from such a compositional analysis is the existence of many comparative constructions whose interpretations are context-dependent. One such clear instance concerns the head-deletion type as we have seen in (6). Corpus data provide us plenty of such head-deletion examples:

- (39) a. nay yenge sillyek-un Chelswu-pota nasta  
my English ability-TOP Chelswu-than better  
'My English is better than Chelswu's'
- b. i kos-uy choyswu-nun talun kos-pota chelcehi  
this place-GEN prisoner other place-than strictly  
kamsipat-ko issta  
be.watched-COMP exist  
'The prisoners here are more strictly watched than those in other places.'
- c. Chelswu-uy nunglyek-un talun aitul-pota chwulcunghata  
Chelswu-GEN ability-TOP other children-than outstand  
'Chelswu's ability is more excellent than other children's.'

As noted earlier, English counterparts are impossible as in (40):

- (40) a. \*My English is better than Chelswu.  
b. \*'Chelswu's ability is more excellent than other children.'

In such examples, the NP complement of *pota* is not the standard expression. The standard expression just sets the context which will help us to conjecture the target of comparison.

Such a head-deleted comparative construction is not confined to genitive expression. The following examples are all from the Sejong Electronic illustrative examples:

- (41) a. ku anak-un nai-pota nulke pointa  
 the lady age-than old look  
 ‘The lady looks old for her age.’
- b. i ccok-eyse tangki-nun him-i ce ccok-pota nemwu yakhay  
 this side pulling-MOD power-NOM that side-than more week  
 ‘The pulling power in this side is much weaker than the power in that side.’
- c. Chelswu-ka ckwulyek-eyse Yengchel-pota wuwelhata  
 Chelswu-NOM endurance-at Yengcheol-than better  
 ‘Chelswu’s endurance is better than Yengcheol.’

In the example (41a), the NP-*than* means ‘compared to’. There isn’t in fact any thing that is compared. In (41b), the syntactic standard expression is just ‘this side’ and the context provides the proper standard expression, ‘pulling power in this side’. In (41c), what is compared is not the two persons, but the endurance each person has. As such, there are many syntactic environments where comparisons in the language highly depend on context.

As Beck et al. (2004) and Oda (2008) suggest, there are many cases in languages like Japanese where the interpretation of comparatives hinge on context. In such a context-dependent analysis, the standard expression denotes just a set of individuals, functioning as setting a context for comparison. Within this context-dependent, non-degree abstraction analysis, comparatives are assumed to have a similar meaning to the English expression ‘compared to’:

- (42) a. Compared to the old brother, the younger brother is taller/tall.  
 b. Compared to what Yongho read, Chelswu wrote more/many papers.

Given these kinds of paraphrase, the truth conditions of comparatives can be something like the followings:

- (43) a.  $\max(\lambda d \text{ Mary wrote } d\text{-many papers}) > c$   
 $c =$  the number made salient by the utterance context  
 $:$  – the number of papers Yongho wrote
- b.  $\exists d(\text{Mary wrote } d\text{-many papers}) \ \& \ d > c$   
 $c =$  the number made salient by the utterance context

: – the number of papers Yongho wrote

The matrix degree is abstracted over (with more) or stay in situ and existentially bound (with many). The variable  $c$  is a contextually provided degree whose value is provided by the complement of *pota* ‘than’. This means the value of the context variable  $c$  is inferred from the set of individuals denoted by the standard NP-*pota* expression.: the NP-*pota* thus sets the context, providing information about the value of a free variable  $c$ . This context-dependent analysis means that there is no degree movement in the matrix clause.<sup>14</sup>

Adopting this contextual dependent analysis, we treat both uses of the NP-*pota* (combining with an NP associate or modifying a VP) as a modifier whose semantic argument is just the standard expression. In addition, it introduces the contextual background relation *contextual-comparison*, reflecting the NP-*pota* functions as a contextual setter. This supertype NP-*pota* is basically projected from the following lexical entry:

$$(44) \left[ \begin{array}{l} n\text{-}than\text{-}mod \\ \text{SYN} \left[ \begin{array}{l} \text{HEAD} \mid \text{POS } n\text{oun} \\ \text{MOD} \langle \text{XP}[\text{IND } \boxed{1}] \rangle \end{array} \right] \\ \text{SEM} \left[ \begin{array}{l} \text{IND } i \\ \text{RELS} \left\langle \left[ \begin{array}{l} \text{PRED } \textit{pota\_rel} \\ \text{ARG1 } i \end{array} \right] \right\rangle \end{array} \right] \\ \text{CNXT} \mid \text{BKGR} \left[ \begin{array}{l} \text{PRED } \textit{contextual-comparison} \\ \text{ARG1 } \boxed{1} \\ \text{ARG2 } i \end{array} \right] \end{array} \right]$$

The lexical entry, as we have seen, syntactically modifies either a nominal or a verbal element. However, in terms of semantics, the NP projected from this word has an individual index. Notice that in the context information the relation *contextual-comparison* relation is also introduced whose arguments are linked to both the modifying predicate and the standard expression. This supertype has two subtypes realized in syntax: *n-than-nmod* and *n-than-vmod*, depending on what the phrase projected from this word modifies, as we have seen earlier. Consider the former with its semantics:

<sup>14</sup> An alternative parametric view between English type comparatives and Japanese type comparatives are given by Kennedy (2007). The analysis maintains that languages may differ in whether the comparative morphology selects a standard of type  $d$  (degree comparison) or type  $e$  (individual comparison) with assuming two different comparative morphemes (more), one for a clausal and the other for phrasal. An issue for such an analysis is the optionality of the comparative morphology in Korean.



$$(45) \left[ \begin{array}{l} n\text{-than-nmod} \\ \text{SYN} \left[ \begin{array}{l} \text{HEAD} \mid \text{POS } n\text{oun} \\ \text{MOD} \left\langle \begin{array}{l} \text{DEG } + \\ \text{POS } n\text{ominal} \\ \text{IND } j \end{array} \right\rangle \end{array} \right] \\ \text{SEM} \mid \text{RELS} \left\langle \begin{array}{l} \text{PRED } \textit{pota\_rel} \\ \text{ARG1 } i \\ \text{ARG2 } j \end{array} \right\rangle \end{array} \right]$$

The NP-*pota* projected from such a word will combine with its associate NP. In this case, the relation *contextual-comparison* takes these two NPs as its arguments, leading us a clear semantic composition too. The predicate modifying NP-*pota* will be projected from the following:

$$(46) \left[ \begin{array}{l} n\text{-than-vmod} \\ \text{SYN} \left[ \begin{array}{l} \text{HEAD} \mid \text{POS } n\text{oun} \\ \text{MOD} \left\langle \begin{array}{l} \text{DEG } + \\ \text{POS } v\text{erbal} \\ \text{IND } e1 \end{array} \right\rangle \end{array} \right] \\ \text{SEM} \mid \text{RELS} \left\langle \begin{array}{l} \text{PRED } \textit{pota\_rel} \\ \text{ARG1 } i \end{array} \right\rangle \end{array} \right]$$

The NP-*pota* projected from such a word will syntactically modify a predicate. In this case, the relation *contextual-comparison* takes different arguments: one is the modifying predicate and the other is the standard NP expression itself. The interpretation is almost similar to ‘compared to’.

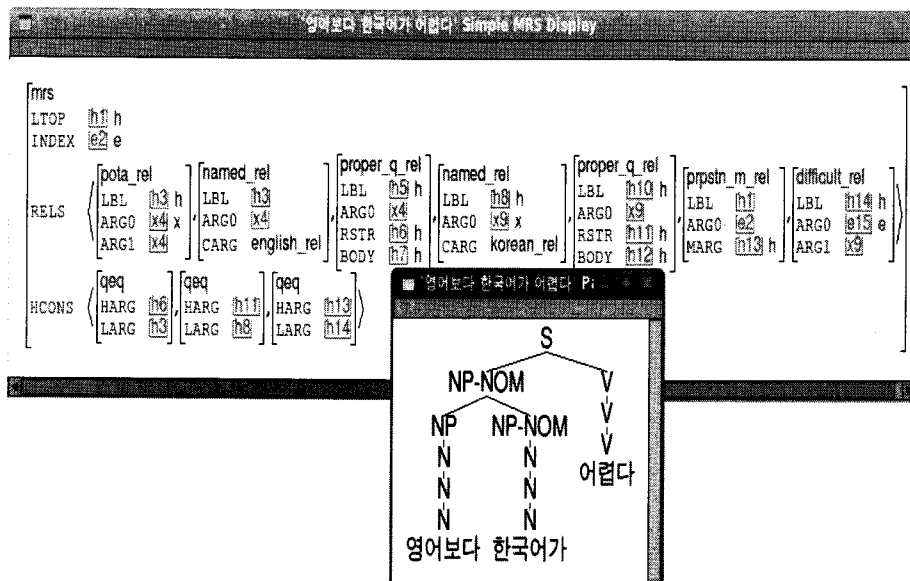
This line of approach assumes that the standards of comparisons are inferred from context, and comparisons are made by pragmatics. This is different from a compositional analysis in which the semantics of comparison is compositionally derived. Though it appears that the analysis leaves the burden of proper meaning composition to context, this way of direction is rather unavoidable when considering highly context-dependent properties of the comparative constructions in Korean, i.e., head-noun deletion comparatives.

#### 4. A Computational Implementation

The analysis we have presented so far has been incorporated in the typed-feature structure grammar HPSG for KRG (Korean Resource Grammar) aiming at working with real-world data (cf. Copestake 2002 for English, Kim and Yang 2004 for Korean.) To check the computational feasibility of the analysis, we have implemented the analysis into the LKB (Linguistic Knowledge Building) system.<sup>15</sup> As

<sup>15</sup> The current Korean Resource Grammar, version 2.0, as of July 2009, has 659 lexical types and 114 phrasal types, 99 grammar rules, 304 inflectional rules, 39,688 lexical entries, and 1198 test-suite sentences, and 77% successful parsing rates.

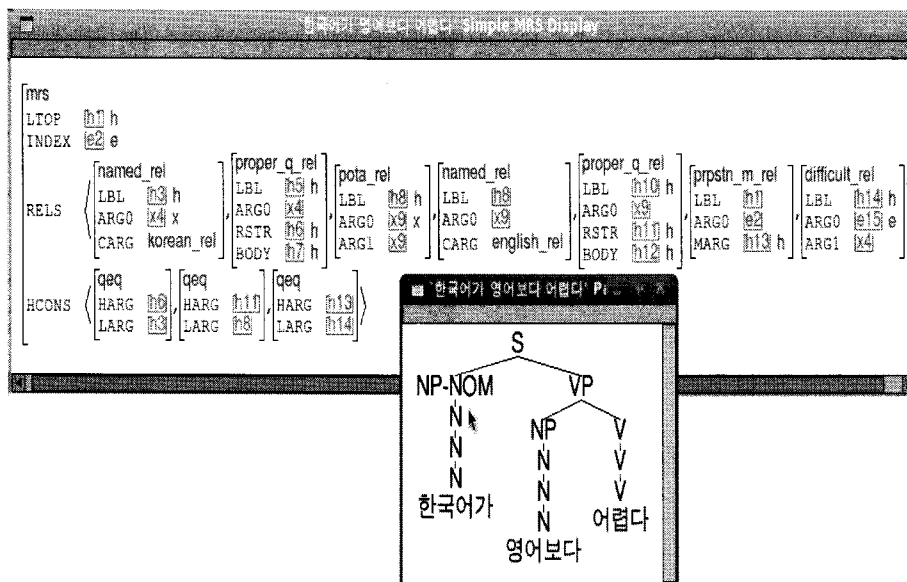
the first step we selected 100 test suite sentences from our 486 sample sentences as well as literature. The test results give us the proper syntactic as well as semantic structures for Korean comparative constructions.



[Figure 1] Parsed Tree and MRS for the phrasal comparative *Korean is more difficult than English*

Figure 1 and 2 are the parsing results of the two main types of phrasal comparative. The small boxes in the figures indicate parsed tree structures whereas the big boxes denote the semantic representations. Consider Figure 1 first. In terms of the syntactic structure, the grammar generates the structure in which the standard phrase *English-than* combines with the subject *Korean*, forming a coordination-like structure. We can notice here that the MRS the grammar generates provides enriched information of the phrase.<sup>16</sup> 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*. The attribute HCONS is to represent quantificational information (see Bender et al. 2002). Each of the types *relation* has at least three features LBL, PRED (represented here as a type), and ARG0. For the proper noun English and Korean, each has two related EPs: *named\_rel* and *proper\_q\_rel*. The meaning of the comparative marker *than* is represented as *pota\_rel*. The relation takes just one argument ( $x_4$ ) whose value is linked to English. This reflects our idea that the

<sup>16</sup> MRS (Minimal Recursion Semantics), developed by Copestake et al. 2005, is a framework of computational semantics designed to enable semantic composition using only the unification of type feature structures. See Bender et al. 2002 and Kim 2006 for its implementation in English and Korean, respectively. The value of the attribute SEM(ANTICS) in our system represents a simplified MRS. ARG0 canonically refers to the index value of the EP (elementary predicate) itself whereas ARG1 or ARG2 refers to the predicate's semantic arguments.



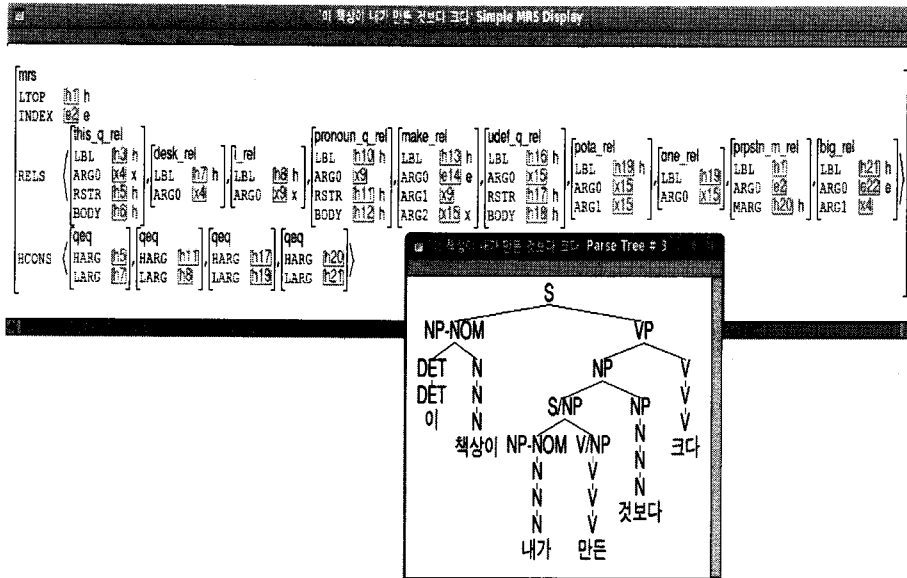
[Figure 2] Parsed Tree and MRS for the phrasal comparative  
*Korean is more difficult than English*<sup>1</sup>

standard expression is interpreted something like ‘compared to English’. The EP *difficult\_rel* takes ‘Korean’ as its argument.

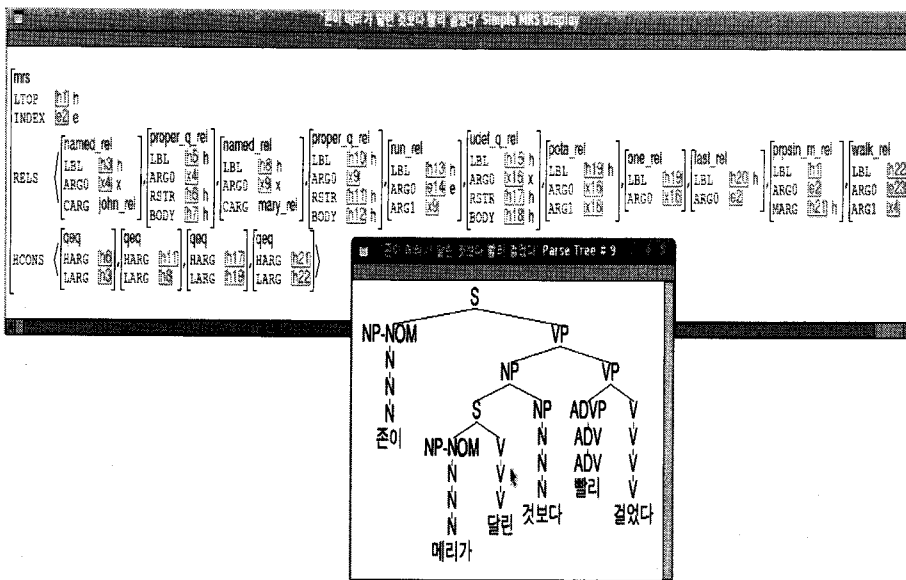
Figure 2 is similar to Figure 1, but the difference comes from the place of *English-than*. In this sentence this follows the subject *Korean* and modifies the predicate *difficult*. The semantic arguments of *pota\_rel* are not different from those in Figure 1: its semantic argument ARG1 is linked to English. (x9). This means that even though the grammar generates two different structures for such a sentence, we have the identical meaning.

Figure 3 and 4 are the results of parsing the gapped clausal comparative *This desk is bigger than the one I made* and the gapless clausal comparative *John ran faster than Mary did*. In Figure 3, the comparative clause ‘[I-NOM\_\_ make-MOD] kes-than’ is parsed as an NP. The expression *kes* is linked to the object whose index value is an individual (x15). This clause-like NP is modifying the predicate ‘big’ in terms of syntax. In terms of meaning, we have something like ‘Compared to the one (desk) that I made, this desk is big.’ Meanwhile, in Figure 4, the clause contains no syntactic gap: what is compared is the degree of two events John’s running and Mary’s running. This is why the [Mary-NOMrun-MODkes-than] must be linked to the event of Mary’s running (e14). These two flexible interpretations of *kes* are independently supported by the general uses of *kes* in non-comparative constructions (cf. Kim and Sells 2008).

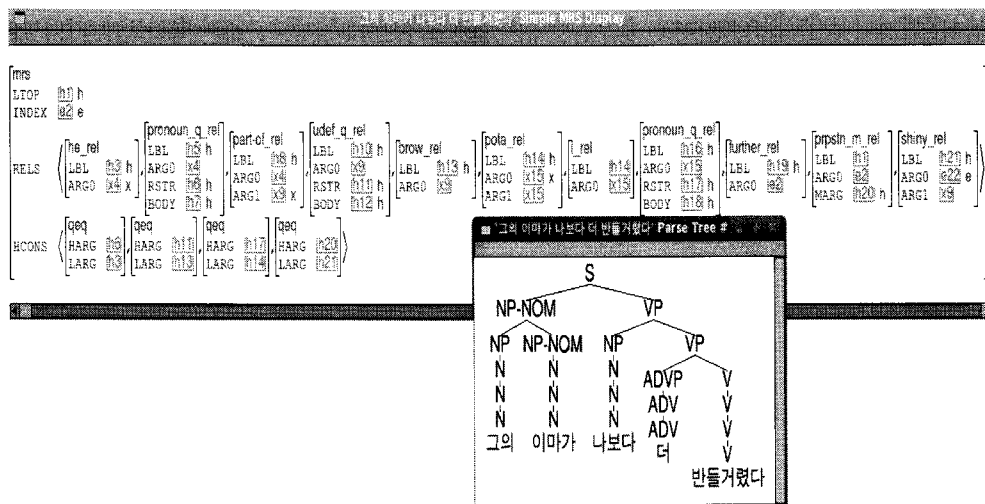
The implementation also can parse a context-dependent comparison example like (6) as given in Figure 5. The small box here once again represents the parsed syntactic structure whereas the big box denotes its MRS meaning representations.



[Figure 3] Parsed Tree and MRS for the gapped clausal comparative *This desk is bigger than the one I made*



[Figure 4] Parsed Tree and MRS for the gapless clausal comparative *John ran faster than Mary*



[Figure 5] Parsed Tree and MRS for the gapless clausal comparative *His forehead is more shiny than mine*

'krg/comparative/research/sj/comparative' Coverage Profile							
Aggregate	total items	positive items	word string	lexical items	distinct analyses	total results	overall coverage
	#	#	φ	φ	φ	#	%
$5 \leq i\text{-length} < 10$	59	59	5.93	34.70	119.64	55	93.2
$0 \leq i\text{-length} < 5$	41	41	3.73	26.50	12.79	39	95.1
<b>Total</b>	<b>100</b>	<b>100</b>	<b>5.03</b>	<b>31.32</b>	<b>75.31</b>	<b>94</b>	<b>94.0</b>

[Figure 6] Profile of the Two Test Suites

In terms of the syntactic structure, we thus can observe the grammar generates a structure in which the standard phrase *na-pota* 'I-than' modifies the predicate 'more shiny'. The MRS in the bigger box provides a proper *pota\_rel* 'than' semantic relation, leading us to interpret this sentence as 'Compared to me, his forehead is shiny.' The comparison between his forehead and my forehead is construed based on the context.<sup>17</sup>

In addition, as a way of evaluating the computational feasibility of the analysis, we also established a coverage profile for the test suite using the program [incr tsdb()]; the 'baseline' to be parsed with the existing KRG (Korean Resource Grammar) and the 'comparative' to be parsed with the new grammar (Oepen and Carroll 2000). Figure 6 is the resulting profile we obtained: As shown in Figure 6, the overall coverage of 'comparative' is 94% as shown above, which is the same as that of 'baseline', but the resulting readings of 'comparative' are almost twice as

<sup>17</sup> In our implementation, this information is encoded in the feature CONTEXT and visible in the AVM output which we do not provide here because of the limited space.

many as those of ‘baseline’, which means our revised grammar yields the promising parsing results as well as the same results that the previous one does. The unparsed sentences have to do with unwritten grammars for meta-linguistic comparatives and comparative forms of adverbs. Several issues still remain to be tackled: reducing the number of parsed readings and checking the grammar with more data as well as even with negative (ungrammatical) sentences.<sup>18</sup>

In terms of computational implementation, there still are more issues for our analysis to be resolved. However, we can observe that the grammar implemented in the LKB system is feasible enough to extend to more complex data in a process of building a comprehensive KRG.

## 5. Conclusion

We have seen that there exist at least two main types of comparatives: phrasal and clausal comparatives. The proper syntactic and semantic treatment of comparatives constructions has been challenges even to theoretical aspects. Korean is no exception. In analyzing phrasal comparatives, we assume the NP-*pota* can function both as a coordination-like phrase and a VP-modifier. This move is theoretically unavoidable to reflect coordination-like properties like English *than*. Clausal comparatives also have at least two subtypes: gapped and gapless. Clausal comparatives behave like an NP headed by the expression *kes*. There is evidence that *kes* is a nominal element which can refer to either an individual (gapped comparative) or an event (gapless comparative). Of two main approaches in dealing with the semantics, we have assumed that the standards of comparisons are inferred from context, and comparisons are made by pragmatics. This is different from a compositional analysis in which the semantics of comparison is compositionally derived.

Given these basic assumptions, we have built a constraint-based grammar for Korean comparative constructions. The grammar we have built within a typed-feature structure system and well-defined constraints, eventually aiming at working with real-world data, has been implemented in the LKB (Linguistic Knowledge Building) system. We have shown that the grammar can parse the appropriate syntactic and semantic aspects of complex Korean comparatives. Even though the test data set we used in checking the feasibility of the system is limited, the test results show us that the grammar, built upon the typed feature structure system, is efficient enough to build proper syntactic and enriched semantic representations for Korean phrasal and clausal comparatives.

## <References>

- Beck, Sigrid, Toshiko Oda and Koji Sugisaki. 2004. Parametric Variation in the Semantics of Comparison: Japanese vs. English. *Journal of East Asian Linguistics* 13: 289–344.
- 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

<sup>18</sup> Most of the overgeneration the grammar produces is related to parsing cleft constructions including the expression *kes* whose syntactic and semantic complexities have been a big challenge to both theoretical and computational analyses. See Kim and Yang (2009).

- 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.
- Bhatt, Rajesh, and Shoichi Takahashi. 2007. Direct comparisons: resurrecting the direct analysis of phrasal comparatives. In *Proceedings of SALT 17*, 19-36. CLC Publications, Cornell University, Ithaca, NY.
- Bresnan, Joan. 1973. Syntax of the Comparative Clause Construction in English. *Linguistic Inquiry* 4:275-343.
- Chae, Hee-Rahk. 1998. A comparative analysis of *tough*-and comparative construction in English and Korean. *Language Research* 34.1: 33-71.
- Choe, Hyon-Sook. 2008. On the nature of the optionality of the morpheme *te* in Korean phrasal comparatives. *Studies in Generative Grammar*, 18: 719-743.
- 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.
- Corver, Norbert. 2005. Comparative Deletion and Subdeletion. *The Blackwell Companion to Syntax*, 582-637. Blackwell.
- Hankamer, J. 1973. Why There are Two Thans in English. In *Proceedings of Chicago Linguistic Society* 9:179-191.
- Heim, Irene. 2000. Degree operators and scope. In B. Jackson, and T. Matthews (Eds.), *Proceedings of Semantics and Linguistic Theory* 10. Ithaca, NY: CLC Publications, 40-64.
- Jhang, Sea-Eun. 2001. Comparative Constructions in Korean. In *Comparative Korean Studies*, 9: 1-18.
- Kennedy, Christopher. 1997. *Projecting the Adjective: The syntax and semantics of gradability and comparison*. Doctoral dissertation, UC Santa Cruz.
- Kennedy, Christopher. 2005. Semantics of Comparatives. In *the Encyclopedia of Language and Linguistics*, Second Edition. Oxford: Elsevier.
- Kennedy, Christopher. 2007. Standards of Comparison. Papers presented at the Colloque de Syntaxe et Sémantique de Paris.
- 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 Peter Sells. 2007. Some Remarks on Information Structure and 'kes'. *Studies in Generative Grammar* Vol. 17, Num 4, Winter 2007, 479-494.
- Kim, Jong-Bok and Peter Sells. 2009. On the Syntax of Korean Comparatives. *Language and Information* 13.2: 29-45.
- Kim, Jong-Bok and Peter Sells. 2010. A Phrasal Analysis of Korean Comparatives. To appear in *Studies in Generative Grammar*.

- Lee, Jeong-shik. 2002. Ellipsis in Korean comparatives. *Studies in Modern Grammar* 30: 155-179.
- Napoli, Donna. 1983. Comparative Ellipsis: A Phrase Structure Analysis. *Linguistic Inquiry* 14: 675-694.
- Oda, Toshiko. 2008. *Degree constructions in Japanese*. Doctoral dissertation, University of Connecticut, Storrs.
- Pancheva, Roumyana. 2006. Phrasal and clausal comparatives in Slavic. In *Formal approaches to Slavic linguistics 14: the Princeton meeting*, ed. J. Lavine, S. Franks, M. Tasseva-Kurktchieva and H. Filip, 236-257.
- Park, So-young. 2009. Three types of Korean comparatives. *Japanese/Korean Linguistics* 16, CSLI Publications.
- Pinkham, Jessie. 1996. A Computational Approach to the Comparative Construction. In *Proceedings of the 32nd Chicago Linguistics Society Meeting*.
- Shimoyama, Junko. 2008. Clausal comparatives and cross-linguistic variation. In *Proceedings of the 39th Meeting of the North East Linguistic Society*.
- Sudo, Yasutada. 2009. Invisible degree nominals in Japanese clausal comparatives. In *Proceedings of the 5th Workshop on Altaic Formal Linguistics*, ed. Reiko Vermeulen and Ryosuke Shibagaki. MITWPL, Cambridge, Mass.

Submitted on: March 29, 2010

Accepted on: April 23, 2010