

Abstract

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.

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 (cf. Kim and Sells 2010). 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 in forming comparatives (cf. Jhang 2001, Choe 2008, Kim and Sell 2009):

- (1) a. *tongsayng-i hyeng-pota chayk-ul (te) manhi ilkessta*
 young.bro-NOM old.bro-than book-ACC more many read
 ‘The younger brother read more books than his older brother.’
- b. *tongsayng-i [[hyeng-i __ ilk-un] kes-pota] (te) manhi*
 young.bro-NOM old.bro-NOM read-MOD kes-than more many
 ilkessta
 read
 ‘The younger brother read more than his older brother did.’

Phrasal comparatives (PC) in (1a) involve two compared nominals whereas clausal comparatives (CC) in (1b) have core clausal properties. With the strong motivation for capturing the truth conditionally identical meaning between phrasal and clausal comparatives, it is commonly assumed that phrasal comparatives are derived from clausal sources through deletion rules (cf. Bresnan 1973, Pancheva 2006, Bhatt and Takahashi 2007).

To see if all Korean comparatives can be grouped into these two clausal and phrasal types, we extracted comparative sentences from the sample examples in the verbal (*vv*) and adjectival (*va*) lexical entries of the Sejong Electronic Dictionary (compiled on the basis of the 100 million words of the Sejong Corpus):

[†]This work was supported by the Korea Research Foundation Grant (KRF-2009-A00065).

(2) Comparative Sentences from the Sejong Electronic Dictionary

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

As indicated here, from the total 67,797 sample sentences in the adjectival (*va*) and verbal (*vv*) lexical entries, we extracted total 486 comparative sentences including an NP-*pota* ‘than’ expression and 41 sentences where *pota* is used as a comparative marker (CM). We analyzed these 486 sentences and could identify 9 additional types that cannot be identified either as PC or CC examples, including the following two types:

- (3) a. John-un seykey kkilok-pota ppalli talliessta
John-TOP world.record-than fast ran
‘John ran faster than the world record.’
- b. ku-uy ima-ka na-pota te pantulkel-yess-ta
he-GEN forehead-NOM I-than more shiny-PAST-DECL
‘(lit.) His forehead is more shiny than me.’

The presumed source sentence for (3a) ‘the world record runs’ does not make any sense. Examples like (3b) are also peculiar since the friend’s forehead is syntactically compared with not my forehead but ‘me’, which is not possible in English. Such an empirical investigation tells us that we cannot reduce all phrasal comparatives to corresponding clausal comparatives as often assumed in the transformational framework.

In this paper, we provide a surface-based, lexicalist analysis that can parse the complex Korean comparative constructions as well as a context-dependent semantic analysis. We then sketch the results of implementing our analysis within the LKB system.

2 Parsing the Structure

2.1 Clausal Comparatives

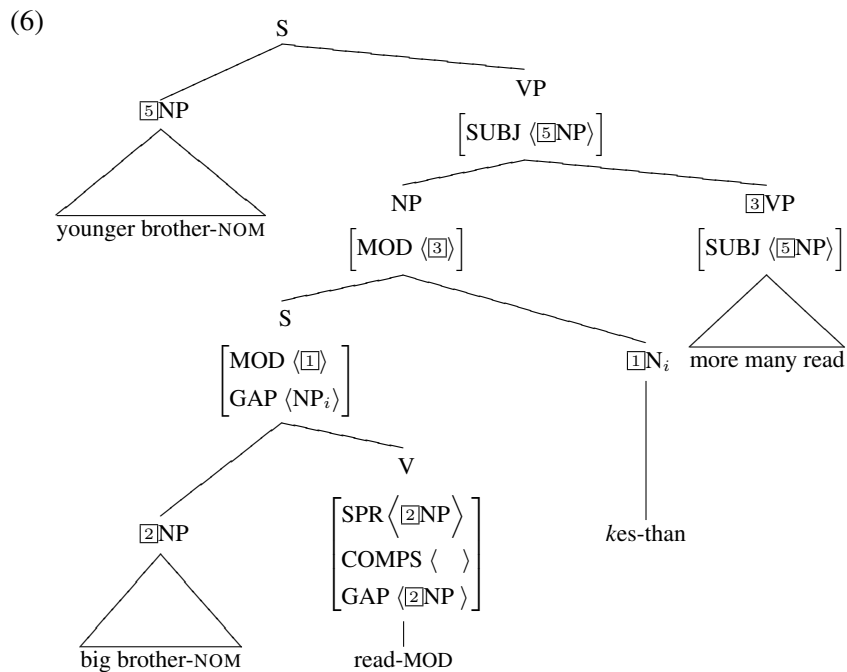
A rich set of empirical data indicates that the clause-like complement in CC is in fact a free relative NP headed by *kes*. Previous literature has assumed that the noun *kes* is a complementizer introducing a CP (e.g., Lee 2002, Park 2009). However, rich evidence undermines this assumption. For example, the complement clause of *pota* can occur only in the NP position, and *kes* in clause-like comparatives can be replaced by a common noun and even be preceded by a determiner:

- (4) John-un [Tom-i sa-n (ku) kes/chayksang]-pota pissan
 John-TOP Tom-NOM buy-MOD the kes/desk-than expensive
 chayksang-ul sa-ass-ta
 desk-ACC buy-PAST-DECL
 ‘John bought a more expensive desk than what Tom bought.’

If *kes* in comparatives were simply a complementizer, such a behavior would not be expected. In addition, the noun *kes* cannot refer to a person. This restriction also holds in comparative constructions, indicating its nominal status:

- (5) John-un [Tom-i manna-n *kes/salam]-pota chakha-n
 John-TOP Tom-NOM meet-MOD kes/person-than honest-MOD
 salam-ul mannassta
 man-ACC met
 ‘John met a more honest man than Tom met.’

Based on these observations, we assume that clausal-like comparatives basically involve a relative clause headed by the noun *kes* as represented in the following structure for (1b):



As given in the structure here, the comparative marker *pota* is attached to the noun *kes*, heading the complex NP consisting of *kes* and an S with a missing element. Like a relative clause, the gapped object of *ilk-un* ‘read-MOD’ in the modifier S is coindexed with the head noun *kes*. The complex NP functioning as standard expression also modifies the gradable predicate *te manhi ilkessta* ‘more many read’. The structure thus assumes that clausal comparatives are in fact NP-phrasal comparatives.

There are also cases where *kes* clauses with no syntactic gap as in (7). Within the relative clause analysis we adopt here, such gapless examples are expected when considering Korean also has amount relative clauses. In fact, all the clause-like comparatives with no overt gap can be reinterpreted as amount or degree relative clause with the replacement by a noun like *cengto* ‘degree’, *sokto* ‘speed’, or *kil* ‘way’:

- (7) a. John-un [Mary-ka talli-n kes/degree]-pota te ppali
 John-TOP Mary-NOM run-MOD kes/degree-than more fast
 kel-ess-ta
 walk-PAST-DECL
 ‘John 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
 kes/pangpep-pota] phyenha-ta
 kes/way-than convenient-DECL
 ‘For us to go is a more convenient way than for students to come.’

2.2 Phrasal Comparatives

The standard marker *-pota* can be attached to a nominal element, allowing only an NP-*pota* phrase. This NP-*pota* phrase has rather flexible distributional possibilities. For example the standard expression NP-*pota* can either precede or following the associate NP. However, when the standard phrase is semantic-case marked, the possibility of scrambling the NP-*pota* disappears:

- (8) *chaykpang-eyse tosekwan(-eyse)-pota kongpwu-ka te cal
 bookstore-at library-at-than study-NOM more well
 toynta
 become
 ‘It is better to study at a bookstore than at a library.’

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

- (9) 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

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, in addition suggesting 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. Based on the observations that the simple NP-*pota* prefers to combine with the associate NP when it is immediately followed, as illustrated in the following for (9a):

- (10) [_{NP} English-than [_{NP} Chinese-than [_{NP} Korean-NOM]]] difficult-DECL

As indicated here, the standard expression combines with the associate NP, forming a bigger NP.¹ This analysis, assuming the existence of base-generated phrasal comparatives, thus treats the ‘standard’ and compared phrase as a kind of NP modifying structure.

The ordering patterns we observe from our 486 samples also provides support for this kind of analysis:

- (11)

Ordering Patterns in the Sejong Electronic Dictionary

Word Ordering Patterns	Frequency
Pattern 1: NP- <i>pota</i> + NP-associate	136
Pattern 2: NP- <i>pota</i> YP NP-associate	240
Pattern 3: NP-associate + NP- <i>pota</i>	5
Pattern 4: NP-associate YP NP- <i>pota</i>	105
Total	486

The NP-*pota* standard expression can immediately precede its associate (Pattern 1) but there is no example where it immediately follows the associate (Pattern 3) though the standard expression can follow it when there is an intervening expression (Pattern 4). We interpret the rare instances of Pattern 3 as supporting evidence for the postulation of the NP-modifying structure, allowing the two NPs in Pattern 1 to combine first, but not those two NPs in Pattern 3. We believe that this NP modifying structure can support the preference to have an coordination-like NP structures for Korean as for English (cf. Napoli 1983).

As Pattern 2 and 4 orderings, they are many contexts where the NP-*pota* and its associate are not adjacent with no precedence constraint. In order to capture such flexible, distributional possibilities of the standard of comparison NP-*pota* expression in a surface-oriented grammar, we assume that in addition to the coordination-like structure, the NP-*pota* ‘than’ can also syntactically modify a verbal element. For example, (3a) will have the following VP modifying structure:

- (12) [_S John-TOP [_{VP} world.record-than [_{VP} fast ran]]]

¹The 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).

In this structure, the NP-*pota* (world.record-than) modifies the verbal predicate *fast ran*, forming a modifier structure. An issue can arise from assuming two different functions of the NP-*pota*, one modifying the following associate NP and the other modifying a verbal predicate. This may be a burden to the grammar, but seems to be inevitable when considering the distributional possibilities and preferences of the NP-*pota* as well as its semantic interactions.

3 Contextual Dependent Interpretation

In terms of semantics, phrasal comparatives appear to be similar to clausal comparatives. For example, the PC in (1a) and the CC in (1b) will have the identical LF structure:

- (13) [[MORE]] (λd the younger brother is *d*-much tall) (λd the older brother is *d*-much tall).

As noted earlier, the rational move to capture this kind of systematic meaning relationships between phrasal and clausal comparatives seems to posit a clausal source and then compute the semantics in a compositional way. In a compositional analysis as given in (13), 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 in Korean as hinted earlier. The first issue is the status of the functor ‘MORE’ that selects two propositional arguments. In languages like Korean, the comparative marker is not present in syntax always: that is, unlike *more* in English, its counterpart *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. 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 (3). Our 486 examples include dozens of examples where the standard expression NP-*pota* is not the expression that is really compared:

- (14) a. *nay yenge sillyek-un Chelswu-pota nasta*
 my English ability-TOP Chelswu-than better
 ‘(lit.) My English is better than Chelswu.’
 b. *i ccok-eyse tangki-nun him-i ce ccok-pota nemwu yakhay*
 this side pulling-MOD power-NOM that side-than more weak
 ‘(lit.) The pulling power in this side is much weaker than that side.’

In such examples, the NP complement of *pota* does not express the head element which is compared with the associate NP. For example in (14b), the compared elements are not this side and that side: they are the power in both sides. The standard expression thus just sets the context which will help us to conjecture the target of comparison. Such examples strong support the assumption that comparison highly depends on context.

As Beck et al. (2004) and Oda (2008) suggest, there are many cases in languages like Japanese where the interpretation of comparatives also hinge on context. In such a context-dependent analysis, the standard expression denotes just a set of individuals, 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’. Given these kinds of paraphrase, the truth conditions of comparatives can be something like the following:

$$(15) \quad \max(\lambda d \text{ Mary wrote } d\text{-many papers}) > c$$

c = the number made salient by the utterance context

$:$ – the number of papers John wrote

The variable c is a contextually provided degree whose value is provided by the complement of *pota* ‘than’. This means the value of c is inferred from the set of individuals denoted by the standard NP-*pota* expression. This context-dependent analysis, providing contextual information for the value of a free variable c , means that there is no degree movement in the matrix clause.²

Adopting this contextual dependent analysis, we treat all the NP-*pota* as a modifier whose semantic argument is just the standard expression. In addition, the NP-*pota* introduces the contextual background relation *contextual-comparison*, reflecting its context-setting function. We can represent this as lexical information:

$$(16) \quad \left[\begin{array}{l} n\text{-than-mod} \\ \text{SYN} \left[\begin{array}{l} \text{HEAD} \mid \text{POS } \textit{noun} \\ \text{MOD} \left\langle \text{XP}[\text{IND } \boxed{1}] \right\rangle \end{array} \right] \\ \text{SEM} \left[\begin{array}{l} \text{IND } i \\ \text{RELS} \left\langle \left[\begin{array}{l} \text{PRED } \textit{pota_rel} \\ \text{ARG1 } i \end{array} \right] \right\rangle \end{array} \right] \\ \text{CNXT} \mid \text{BKGR} \left[\begin{array}{l} \text{PRED } \textit{contextual-comparison} \\ \text{ARG1 } \boxed{1} \\ \text{ARG2 } i \end{array} \right] \end{array} \right]$$

²An 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.

The lexical entry 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 we introduce the relation *contextual-comparison* 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. Each of these two subtypes will have the following lexical specifications:

$$(17) \quad \begin{array}{cc} \text{a.} & \left[\begin{array}{l} \textit{n-than-nmod} \\ \text{SYN} \left[\begin{array}{l} \text{HEAD} \mid \text{POS } \textit{noun} \\ \text{MOD} \left\langle \begin{array}{l} \text{DEG } + \\ \text{POS } \textit{nominal} \\ \text{IND } j \end{array} \right\rangle \end{array} \right] \\ \text{SEM} \mid \text{RELS} \left\langle \begin{array}{l} \text{PRED } \textit{than_rel} \\ \text{ARG1 } i \\ \text{ARG2 } j \end{array} \right\rangle \end{array} \right] \\ \text{b.} & \left[\begin{array}{l} \textit{n-than-vmod} \\ \text{SYN} \left[\begin{array}{l} \text{HEAD} \mid \text{POS } \textit{noun} \\ \text{MOD} \left\langle \begin{array}{l} \text{DEG } + \\ \text{POS } \textit{verbal} \\ \text{IND } eI \end{array} \right\rangle \end{array} \right] \\ \text{SEM} \mid \text{RELS} \left\langle \begin{array}{l} \text{PRED } \textit{than_rel} \\ \text{ARG1 } i \end{array} \right\rangle \end{array} \right] \end{array}$$

The NP-*pota* projected from (17a) 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. Meanwhile, the NP-*pota* projected from (17b) modifies a gradable predicate. 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 standard of comparison is 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 deleted 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, Kim 2004 for Korean.) To check the computational feasibility of the analysis, we have implemented the analysis into the LKB (Linguistic Knowledge Building) system.³

³The 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.

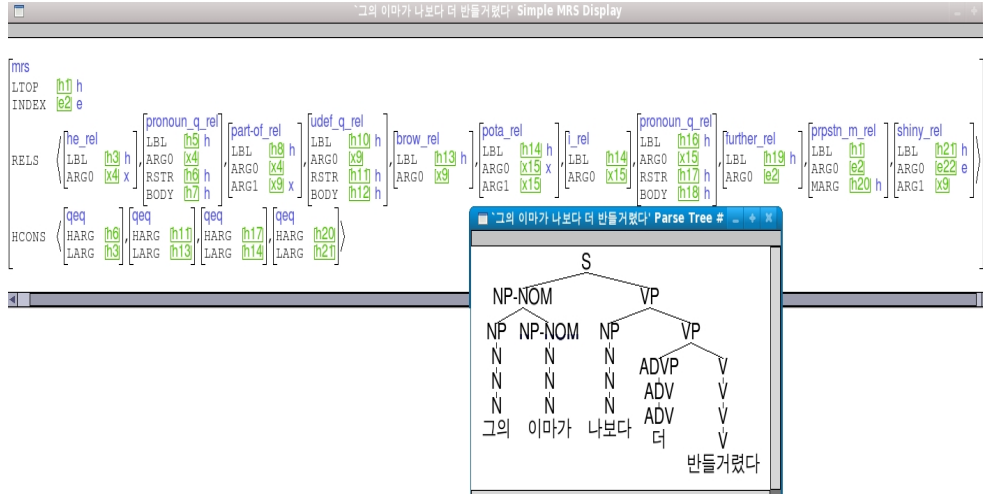


Figure 1: Parsed Tree and MRS for the gapless clausal comparative *His forehead is more shiny than mine*

‘krg/comparative/research/sj/comparative’ Coverage Profile							
Aggregate	total items #	positive items #	word string ϕ	lexical items ϕ	distinct analyses ϕ	total results #	overall coverage %
$5 \leq i\text{-length} < 10$	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
Total	100	100	5.03	31.32	75.31	94	94.0

Figure 2: Profile of the Two Test Suites

As the first step we selected 100 test suite sentences from our 486 sample sentences as well as literature. Figure 1 is one sample syntactic and semantic structure that our implementation produces as the parsing results for the sentence (3b). The small box in Figure 1 indicates parsed tree structures whereas the big box denotes the MRS representations. In terms of the syntactic structure, we can observe the grammar thus generates the structure in which the standard phrase NP-*pota* modifying the predicate. We can notice here that the MRS, though not clearly visible, also provides a proper *pota* ‘than’ semantic relation. The contextual comparison is given in the contextual information.

In addition, as a way of evaluating the computational feasibility of the analysis, we also established two `[incr tsdb()]` test suites; the ‘baseline’ to be parsed with the existing KRG (Korean Resource Grammar) and the ‘comparative’ to be parsed with the new grammar. Figure 2 is the resulting profile we obtained: As shown in Figure 2, the overall coverage of ‘comparative’ is 94% as shown below, which is the same as that of ‘baseline’, but the resulting readings of ‘comparative’ (6,043) are almost twice as many as those of ‘baseline’ (3,083), which means our revised grammar yields the promising parsing results as well as the same results

that the previous one does.⁴

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.

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.
- 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.
- 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.
- Jhang, Sea-Eun. 2001. Comparative Constructions in Korean. In *Comparative Korean Studies*, 9: 1–18.
- 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 and Jaehyung Yang. 2004. Projections from Morphology to Syntax in the Korean Resource Grammar: Implementing Typed Feature Structures, *Lecture Notes in Computer Science*, 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*

⁴The 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.

tics 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.