Syntax and Semantics of Numeral Classifier Constructions in Korean*

June 8, 2006

Abstract

The syntactic and semantic complexity of the so-called numeral classifier (Num-Cl) constructions in Asian languages challenges theoretical as well as computational linguists. We provide a constraint-based analysis of these constructions within the framework of HPSG and MRS (Minimal Recursion Semantics). The present analysis enables us to capture the intriguing syntactic properties of these constructions as well as provide appropriate semantic representations for the constructions in question.

1 Basic Properties

One of the most salient features of languages like Korean is the complex behavior of numeral classifiers linked to an NP it classifies. There exist at least three different environments where numeral classifiers (Num-CL) can appear:

(1) a. Genitive-Case (GC) Type:
sey myeng-uy haksayng-i o-ass-ta
three CL-GEN student-NOM come-PST-DECL
'Three students came.'

b. Noun Initial (NI) Type:
haksayng sey myeng(-i) o-ass-ta
student three CL-NOM come-PST-DECL

*I thank Chan Chung, Incheol Choi, Jaewoong Choe, Jaehyung Yang for comments and criticisms. My thanks also go to two anonymous reviewers of this journal. This work was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2005-042-A00056).

c. Noun-Case (NC) Type:
haksayng-i sey myeng-i o-ass-ta
student-NOM three CL-NOM come-PST-DECL

In the GC type, the Num-CL appears with the genitive case marking, preceding the modifying NP. In the NI, the Num-CL sequence follows a caseless N, whereas in the NC type the noun is case marked.

There exist overt syntactic and semantic differences among these three types. As indicated by the name of each type, the possible case value in each type is different. In particular, even though the NI type can host almost any semantic case marker, the NC type canonically allows only NOM or ACC (cf. Choi 1988, Choi 2001):

(2) a. Kim-un phyenci-lul haksayng tases myeng-eykey ponayessta.
Kim-TOP letter-ACC student five CL-DAT sent
'Kim sent letters to five students.'
b. *Kim-un phyenci-lul haksayng-eykey tases myeng-(eykey) ponay-ss-ta

In terms of syntactic structures, the three types also display clear differences. As noted in the following contrast, only the NC type allows the Num-CL to float away from the NP it quantifies:[1]

(3) a. *sey myeng-uy cengmal pemin-i te iss-ta
three CL-GEN really criminal-NOM more exist-DECL
'There are three more criminals.'
b. ??*pemin cengmal sey myeng-i te iss-ta
criminal really three CL-NOM more exist-DECL
c. pemin-i cengmal sey myeng-i te iss-ta
criminal-NOM really three CL-NOM more exist-DECL

The cleft constructions also show that in the NI and GC type, the N + Num-CL sequence forms a syntactic unit whereas in the NC, it does not (cf. Choi 2001).

(4) a. ku sensayngnim-ul mos ka-key ha-n kes-un
that teacher-ACC not go-COMP do thing-PNE
sey myeng-uy haksayng-i-essta.
three-CL-GEN student-COP-PAST
'What made the teacher not leave were five students.'

[1] (2b) can become better if we place a long pause after *pemin* 'criminal'. Such an example can be interpreted as a NC-type example.

b. ku sensayng-nim-ul moskakey han kes-un haksayng sey myeng-i-ess-ta.
c. *ku sensayng-nim-ul moskakey han kes-un haksayng-i sey myeng-i-ess-ta.

Coordination also shows another difference: the GC and the NI type can participate in the coordinate constructions, but the NC type cannot:

(5) a. Kim-un [[sey kwen-uy kongchayk]-kwa [twu calwu-uy
Kim-TOP three-CL notebook-CONJ two-CL-GEN
yenphil]]-ul sassta.
pencil-ACC bought
'Kim bought three notes and two pencils.'
b. Kim-un [[kongchayk sey kwen]-kwa [yenphil twu calwu-lul]] sassta
c. *Kim-un [[kongchayk-ul sey kwen]-kwa [yenpil-ul twu calwu]]-lul sassta.

Beside such syntactic differences, there is also a subtle semantic difference. Consider the entailment relationship in the following sentences.

(6) a. Seoul-lo tomangka-n tases myeng-uy haksayng-i
Seoul-GOAL run-away-PNE five CL-GEN student-NOM
tolawassta
returned
'The five students who ran away for Seoul returned.'
b. Seoul-lo tomangkan haksayng tases myeng-i tolawassta.
'The five students who ran away for Seoul returned.'
c. Seoul-lo tomangkan haksayng-i tases-myeng-(i) tolawassta.
'Of those who ran away for Seoul, just five returned.'

As can be noticed from the English translations, the example (6a) and (6b) is true in the situation where there are five students who left for Seoul, and they all came back. Meanwhile, the preferred reading of (6c) is such that there are more than five students who left for Seoul and of them just five returned.

2 Forming a Numeral-Classifier Sequence

Syntactically, numeral classifiers are a subclass of nouns (just like Japanese, cf. Bond and Paik 2000). However, unlike common nouns, they cannot stand

alone and must combine with a numeral or a limited set of determiner:[2]

(7) a. *(twu) kay 'two CL' (Numeral)
b. *(yeleo/myech) kay 'several CL' (Quantifier)
c. *(myech) kay 'how many' (Interrogative)

Semantically, classifiers both classify and quantify the referent of the NP they collocate with, but can modify nouns whose semantic features match with their own. For example, we can roughly group classifiers depending on the semantic groups as following (cf. Bond and Paik 2000):

(8)

CL Type	Referents classified	Examples
kay	general objects	sakwa han kay 'apple'
pen	events	nolay han pen 'song'
myeng	person	haksang han myeng 'student'
pangwul	liquid	nwunmwul han pangwul 'tear' '
cang	flat objects	congi han cang 'paper'
tay	machinary	cacenke han tay 'bike'
ken	incidents	kyeyak han ken 'contract'
mali	animals	saca han mali 'lion'
ca	length	oskam han ca 'cloth'

These semantic criteria tell us why classifiers like *kay* cannot modify liquid nouns like **nwumwul han kae* 'tear one-CL' since it just classifies a concrete object.[3]

Reflecting these syntactic and semantic properties, we can assign the following lexical information to numerals and classifiers within the feature structure system of HPSG and MRS.[4]

[2]A limited set of common nouns such as *salam* 'person', *kulus* 'vessel', *can* 'cup', *khep* 'cup', *thong* 'bucket' can also function as classifiers.

[3]There can be several ways of grouping classifiers. For example, classifiers can be divided into five types, depending on their properties: 'sortal' classifying the kind of the NP they quantify, 'event' quantifying events, mensuaral measuring the amount of some property, group referring to a collection of members, taxonomic denoting generic kind. See Bond and Paik 2000.

[4]MRS (Minimal Recursion Semantics), developed by Copestake et al. (2003), is a framework of computational semantics designed to enable semantic composition using only the unification of type feature structures. See Copestake et al. 2003 and Bender et al. 2002. The feature abbreviations used in this paper are as follows: ARG (ARGUMENT), LBL (LABEL), LTOP (LOCAL TOP), ORTH (ORTHOGRAPHY), SYN (SYNTAX), SEM (SEMANTICS), RELS (RELATIONS), PRED (PREDICATE), etc.

(9) a.
$$\begin{bmatrix} \textit{num-det} \\ \text{ORTH } \langle \text{sey `'} \rangle \\ \text{SYN} \mid \text{HEAD} \begin{bmatrix} \text{POS } \textit{det} \\ \text{NUM } + \end{bmatrix} \\ \text{SEM} \begin{bmatrix} \text{HOOK} \begin{bmatrix} \text{INDEX } i \\ \text{LTOP } \textit{h2} \end{bmatrix} \\ \text{RELS} \left\langle \begin{bmatrix} \text{PRED } \textit{card\_rel} \\ \text{LBL } \textit{h2} \\ \text{ARG0 } i \\ \text{CARG 3} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

b.
$$\begin{bmatrix} \textit{cl-n} \\ \text{ORTH } \langle \text{myeng `'} \rangle \\ \text{SYN} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{POS } \textit{noun} \\ \text{CLTYPE } + \end{bmatrix} \\ \text{VAL} \mid \text{SPR } \langle \begin{bmatrix} \textit{num-det} \\ \text{INDEX } i \end{bmatrix} \rangle \end{bmatrix} \\ \text{SEM} \begin{bmatrix} \text{HOOK} \begin{bmatrix} \text{INDEX } i \\ \text{LTOP } \textit{h1} \end{bmatrix} \\ \text{RELS} \left\langle \begin{bmatrix} \text{PRED } \textit{person\_rel} \\ \text{LBL } \textit{h1} \\ \text{ARG0 } i \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

The feature structure in (9a) represents that there exists an individual x whose CARG (constant argument) value is "3".[5] The feature NUM is assigned to the numerals as well as to determiners like *yelye* 'several' and *myech* 'some' that combine with classifiers. Meanwhile, (9b) indicates that syntactically a classifier selects a specifier, whereas semantically it belongs to the ontological category *person*.

[5]The feature CARG takes as its value a string representing the name of the named entity.

The feature CLTYPE differentiates classifiers from common nouns.[6] Assuming that only [NUM +] elements can combine with the [CLTYPE +], we can rule out unwanted forms such as **ku myeng* 'the CL'. In addition, unlike quantifier determiners *motun* 'all' as in *ku motun haksayng* 'the all student', nothing can intervene between the NUM and CL. Our grammar capture these constraints by assuming that the Num-CL sequence is a multiword (*mw*) expression as given in the following rule:[7]

(10) Num-CL Rule:

$$\left[\textit{num-cl-mw}\right] \rightarrow \begin{bmatrix}\textit{num-det} \\ \text{NUM} +\end{bmatrix}, \text{[CLTYPE +]}$$

This rule means that when the head classifier combines with its SPR element (a numeral), it will form a multiword expression *num-cl-mw* that still can look for a determiner so that it can predict the contrast between *ku sey myeng* 'the three CL' and **sey ku myeng* 'three the CL'. The present system will then generate the following structure:

---

[6] The value of LBL is a token to a given EP (elementary predicate). The feature HOOK includes externally visible attributes of the atomic predications in RELS (RELATIONS). The value of LTOP is the local top handle, the handle of the relations with the widest scope within the constituent. See Copestake et al. 2003 for the exact functions of each attribute.

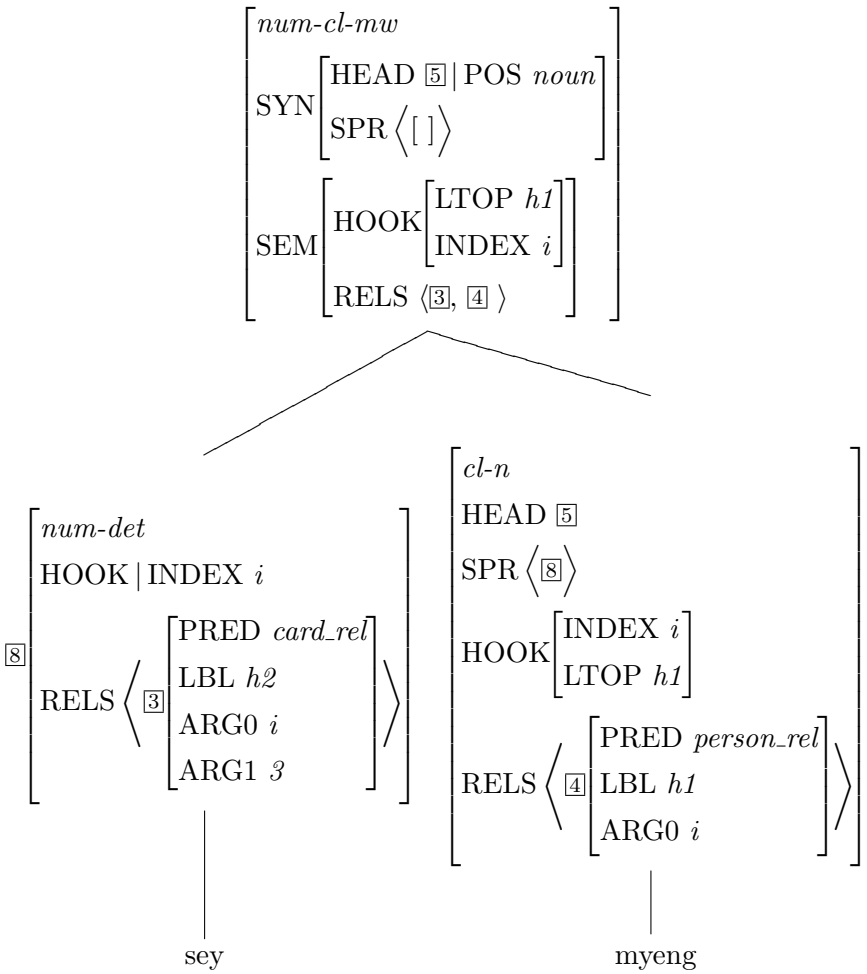
[7] The expression *num-cl-mw* is a subtype of *hd-spr-ph* generated by the Head-SPR (specifier) Rule:

(i) Head-SPR Rule:

$$[\textit{hd-spr-ph}] \rightarrow \boxed{1}, \mathbf{H}[\text{SPR} \left\langle \boxed{1} \right\rangle]$$

This rule allows a head to combine with its specifier. We thus do not treat pure determiners like *i* 'this' and *ku* 'the' ad a modifier, but a determiner selected by a head noun. This will block us from generating examples like **i ku haksayng* 'this the student'. However, other adnominal elements like *motun* 'all', *enu* 'some', genitive phrases like *na-uy* 'I-GEN' can function as modifiers, to generate examples like (13).

(11)



The classifier *myeng* and the numeral *sey* have their own semantic (RELS) values which will be accumulated to the top node. The index value of the resulting expression is identical to that of the numeral and of the classifier.

3 Genitive Case Type

As we have seen, the NUM-CL can have the GEN case marking.

(12) a. sey myeng-uy haksayng 'three CL-GEN student'
b. sey mali-uy kangaci 'three CL-GEN pet'

One thing to note is that just like a Det(P), a genitive NP precedes a definite NP or can occur after a determiner:

(13) a. John-uy ku chinkwu-tul
John-GEN the friend-PL
b. ku John-uy chinkwu-tul
the John-GEN friend

As noted here, even though the head noun combines with its GEN-marked NP, the resulting phrase still can combine with a determiner. This means the GEN-marked NP does not saturate the selectional requirement of its head noun but functions as a modifier; the combination of these two expressions are licensed by the Head-MOD Rule:[8]

(14) Head-MOD Rule:
$[\textit{hd-mod-ph}] \rightarrow [\text{HEAD} \mid \text{MOD} \langle \boxed{1} \rangle]$, $\mathbf{H}\boxed{1}$

Following Kim (2004) and Kim and Yang (2004), we assume that nominal particles are treated as suffixes attached to the nominal stems in the lexicon, adding appropriate information.[9] For example, the attachment of GEN case particle *-uy* will add the information on the GCASE (grammatical case) value as well as the additional specification on the MOD feature:

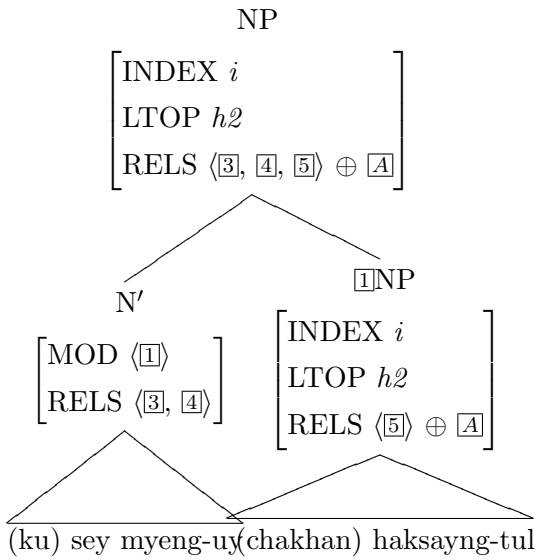
(15)
$$\begin{bmatrix} \textit{num-cl-mw} \\ \text{ORTH} \langle \text{sey myeng-uy} \rangle \\ \text{SYN} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{POS } \textit{noun} \\ \text{CASE} \mid \text{GCASE } \textit{gen} \\ \text{MOD} \langle \boxed{1}\text{NP}_i \rangle \end{bmatrix} \end{bmatrix} \\ \text{SEM} \begin{bmatrix} \text{HOOK} \begin{bmatrix} \text{INDEX } i \\ \text{LTOP } \textit{h1} \end{bmatrix} \\ \text{RELS} \left\langle \boxed{3} \begin{bmatrix} \text{PRED } \textit{card\_rel} \\ \text{LBL } \textit{h2} \\ \text{ARG0 } i \\ \text{ARG1 } \textit{3} \end{bmatrix}, \boxed{4} \begin{bmatrix} \text{PRED } \textit{person\_rel} \\ \text{LBL } \textit{h1} \\ \text{ARG0 } i \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

[8]To be more precise, we distinguish *hd-mod-ph* in accordance with the type of head: *hd-n-mod-ph* and *hd-v-mod-ph*. One additional constraint on the *hd-n-nom-ph* is that the modifier can get the GEN case value.

[9]See Kim (2004) for the formation of nominal elements in detail.

Notice the difference between the information for *sey myeng* in (11) and *sey myeng-uy* here: The GEN-marked *num-cl-gen* places a restriction on what it can modify. The modified element must be a nominal expression whose index value is identical with its own index value. The present grammar assigns the following structure:[10]

(16)



The head noun *haksayng-tul* 'student-PL' or the NP *ku haksayng-tul* combines with *sey meyng-uy* in accordance with the Head-MOD rule. The resulting NP will collect all the RELS values from the daughters, but its INDEX value is identical with that of its head. The final MRS representation will then be something like the following:

(17)
$$\begin{bmatrix} \text{HOOK} \begin{bmatrix} \text{LTOP } h3 \\ \text{INDEX } i \end{bmatrix} \\ \text{RELS} \left\langle \begin{array}{l} \boxed{5} \begin{bmatrix} \text{PRED } \textit{\_student\_n\_rel} \\ \text{LBL } h3 \\ \text{ARG0 } i \end{bmatrix}, \boxed{4} \begin{bmatrix} \text{PRED } \textit{person\_rel} \\ \text{LBL } h1 \\ \text{ARG0 } i \end{bmatrix}, \\ \boxed{3} \begin{bmatrix} \text{PRED } \textit{card\_rel} \\ \text{LBL } h2 \\ \text{ARG0 } i \\ \text{ARG1 } 3 \end{bmatrix} \end{array} \right\rangle \end{bmatrix}$$

[10]The variable A here indicates additional PRED values that the DetP may have.

As given here, the final NP means that there are three or more than 'x' which belong to the semantic category 'person'. This 'x' is anchored to the set of students.

4 NI (Noun-Initial) Type

As we have seen in section 1, cleft and coordination constructions indicate that unlike in the NC type, in the NI type the N forms a strong syntactic unit with a following Num-CL. There exist various examples indicating that the NI type behaves like a synthetic compound. For example, the N and the following Num-CL sequence cannot be separated at all:

(18) a. haksayng sey myeng-i o-ass-ta
student three CL-NOM come-PST-DECL
'Three students came.
b. *haksayng ku sey myeng-i o-ass-ta

The situation is not different with the object:[11]

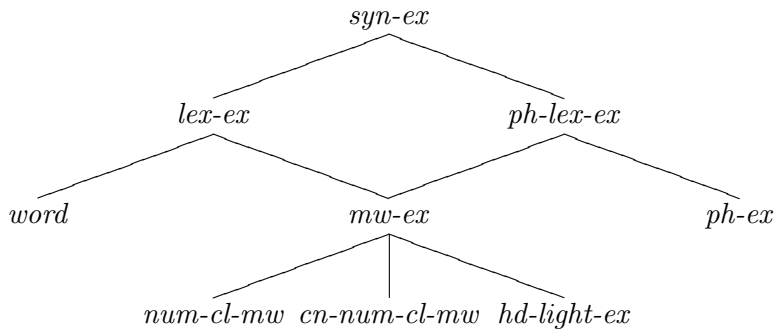
(19) a. maykcwu-lul John-i sey pyeng masi-ess-ta
beer-ACC John-NOM three CL drink-PST-DE CL
'(He) drank three bottles of beer.'
b. *maykcwu John-i sey pyeng masi-essta
beer-ACC John-NOM three bottle drink-PST-DECL

As noted here, no element can appear between the N and the following Num-CL sequence. Considering the flexible distributions of Det(P) in the language, such a tight syntactic cohesion is unexpected. These facts support the idea that the NI sequence is another synthetic compound as noted in the following hierarchy:[12]

[11] Again, (18b) becomes better if we put a long pause after *maykcwu*.

[12] One main difference between *lex-ex* and *ph-lex-ex* is that *lex-ex* behaves like a lexical element.

(20)



The hierarchy indicates that both *num-cl-mw* and *cn-num-cl-mw* are an instance of *mw-ex* which behaves like synthetic compounds.

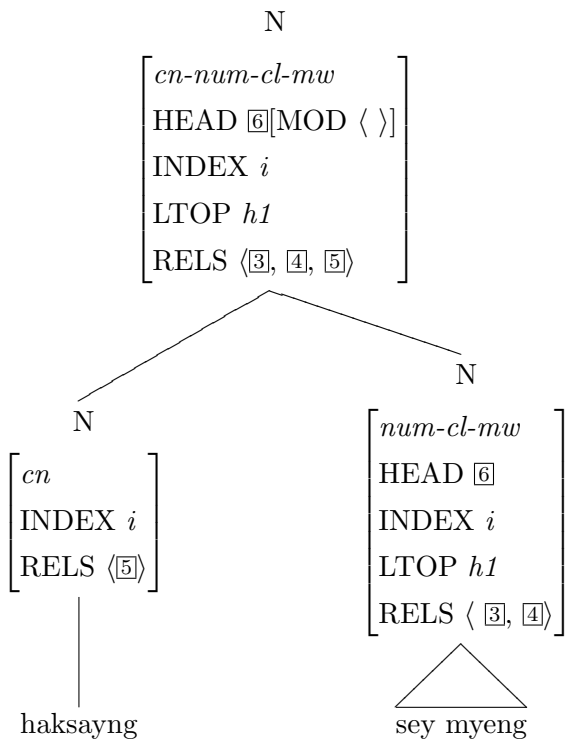
(21)

NI Compound Formation Rule:

$$\begin{bmatrix} \textit{cn-num-cl-mw} \\ \text{HEAD} \mid \text{MOD} \ \langle\ \rangle \end{bmatrix} \rightarrow \begin{bmatrix} \textit{cn} \\ \text{INDEX}\ i \end{bmatrix},\ H \begin{bmatrix} \textit{num-cl-mw} \\ \text{CLTYPE}\ + \\ \text{INDEX}\ i \end{bmatrix}$$

As represented in the rule, a common noun *cn* and *num-cl-mw* forms a well-formed expression *cn-num-cl-mw*. Unlike *num-cl-mw*, the resulting expression has an empty MOD value, indicating that it has no restriction in its grammatical function. This formation rule will eventually license the following structure:

(22)

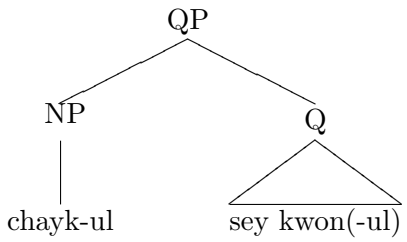


As represented in the structure, the common noun *haksayng* 'student' combines with the *sey myeng* in accordance with the formation rule in (21). They both have the same index value with their own semantic contributions as given in the RELS values. The final MRS of this compound will be the same as the GC type *sey myeng-uy haksayng* 'three CL-GEN student' in (17).

4.1 Noun-Case Type

The NC type allows NOM or ACC to the NP followed by the case-marked CL-NUM (FQ) which can float away from the NP. In terms of its structure, one could assume that the Num + CL sequence is the head of a quantifier phrase (Terada 1990, Lee 1989) as given in (23):

(23)



Such a constituent suffers from many problems. As noted earlier, coordination and cleft constructions tell us that the QP here cannot form a constituent at all.[13]

There exist several supporting phenomena indicating that the FQ modifies the following verbal expression.[14] One phenomenon is the substitution by *kulay-ss-ta* 'do so-PST-DECL'. As noted in (24), unlike the NI type, only the NC type allows an FQ and a following main verb together to be substituted by the proverb *kulay-ss-ta*:

(24) a. namca-ka [sey myeng o-ass-ko], yeca-to kulay-ss-ta
man-NOM three CL come-PST-CONJ woman-also do-PST-DECL.
'As for man, three came, and as for woman, the same number came.'
b. *[namca sey myeng-i] o-ass-ko, yeca-to [kulay-ss-ta]

This means that the FQ in the NC type functions as a VP constituent, unlike the Num-CL in the NI type. Coordination data also support a VP modifier analysis:

(25) nam haksayng-tul-kwa kuliko yehaksayng-i sey myeng-i oassta
boy student-PL-and and girl student-NOM three CL-NOM came
'Total three boys and girls came.'

If the grammar assume a QP structure in (24), we expect a reading in which the number of girls are three. However, this sentence has only one reading such that the number of boys and girls is in total three.

[13]The only case where the QP here seems to form a unit is gapping constructions.

(i) namca-ka sey myeng kuliko yeca-ka twu myeng o-ass-ta
man-NOM three CL-NOM and woman-NOM two CL come-PST-CONJ
'As for men, two and as for women three came.'

Treating such examples to be gapping, we need not posit the QP analysis, either.

[14]Dowty and Brodie (1984) claim that English floating quantifiers like *all* are better treated as an adverb.

Notice that these do not mean that the FQ is just like an adverb. For example, unlike a pure adverb, the FQ cannot be topicalized:

(26) a. Ecey-nun Kim-i sinna-key nolassta.
yesterday-TOP Kim-NOM pleasantly played
'As for yesterday, Kim played.'
b. *Sey kay-nun Kim-i sakwa-lul mekessta.
three CL-TOP Kim-NOM apple-ACC ate

In addition, we can observe that unlike a locative adverbial element, it cannot function as the head of a relative phrase:

(27) a. Kim-i __ sakwa-lul mek-ess-ten kos
Kim-NOM apple-ACC eat-PST-PNE place
'the place where Kim played'
b.*Kim-i sakaw-lul __ mek-ess-ten sey kay
Kim-NOM apple-ACC eat-PST-PNE three CL

Given these facts, we assume that syntactically, the FQ functions as a modifier to a verbal expression, whereas semantically it is linked to a preceding antecedent, functioning as an argument of the predicate. Thus, unlike the GEN-marked *num-cl-mw*, the NOM or ACC-marked *num-cl-mw* will modify a verbal expression.

The remaining issue is then how to connect the FQ with an appropriate antecedent. There exist several constraints in identifying the antecedents. When the floating quantifier is case-marked, it is linked to the subject or object in the same case marking (see Choi 2001, O'Grady 1982, Gerdts 1987).

(28) a. haksayng-tul-i sey myeng-i sakwa-lul cengmal mek-ess-ta
student-PL-NOM three CL-NOM apple-ACC really eat-PST-DECL
'As for the students, three really ate apples.'
b. haksayng-tul-i sakwa-lul sey myeng-i cengmal mek-ess-ta
c. haksayng-tul-i sakwa-lul sey myeng-i cengmal mek-ess-ta

(29) a. haksayng-tul-i sakwa-lul sey kay-lul cengmal mek-ess-ta
student-PL-NOM apple-ACC three CL-ACC really eat-PST-DECL
'As for the apples, three really ate three.'
b. sakwa-lul haksayng-tul-i sey kay-lul cengmal mek-ess-ta
c. haksayng-tul-i sakwa-lul cengmal sey kay-lul mek-ess-ta

Regardless of its location, we can observe that the NOM FQ is linked to the subject whereas the ACC FQ is linked to the object. As a way of capturing such constraints, we can assume that the case-marked *num-cl-n* functions as a modifier to a verbal expression, but quantifies over an argument with the same case value. That is, examples like *sey myeng-i* 'three CL-NOM' and *sey myeng-ul* 'three CL-ACC' will have the following lexical entries:

(30) a.
$$\begin{bmatrix} \textit{num-cl-mw} \\ \text{ORTH } \langle \text{sey myeng-i} \rangle \\ \text{HEAD} \begin{bmatrix} \text{POS } \textit{noun} \\ \text{CASE} \mid \text{GCASE } \textit{nom} \\ \text{MOD} \left\langle \begin{bmatrix} \text{POS } \textit{verb} \\ \text{SUBJ} \left\langle \left[\text{INDEX } i \right] \right\rangle \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{SEM} \mid \text{HOOK} \mid \text{INDEX } i \end{bmatrix}$$

b.
$$\begin{bmatrix} \textit{num-cl-mw} \\ \text{ORTH } \langle \text{sey myeng-ul} \rangle \\ \text{HEAD} \begin{bmatrix} \text{POS } \textit{noun} \\ \text{CASE} \mid \text{GCASE } \textit{acc} \\ \text{MOD} \left\langle \begin{bmatrix} \text{POS } \textit{verb} \\ \text{COMPS} \left\langle ..., \begin{bmatrix} \text{GCASE } \textit{acc} \\ \text{INDEX } i \end{bmatrix}, ... \right\rangle \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{SEM} \mid \text{HOOK} \mid \text{INDEX i} \end{bmatrix}$$

We can notice here that *num-cl-n-nom* modifies a verbal element whose SUBJ has the same index value, whereas *num-cl-n-acc* modifies a verbal element which has one unsaturated COMPS element whose INDEX value is identical with its own INDEX value. What this means is that the NOM or ACC marked *num-cl-n* is semantically linked to the SUBJ or COMPS element through the INDEX value.[15]

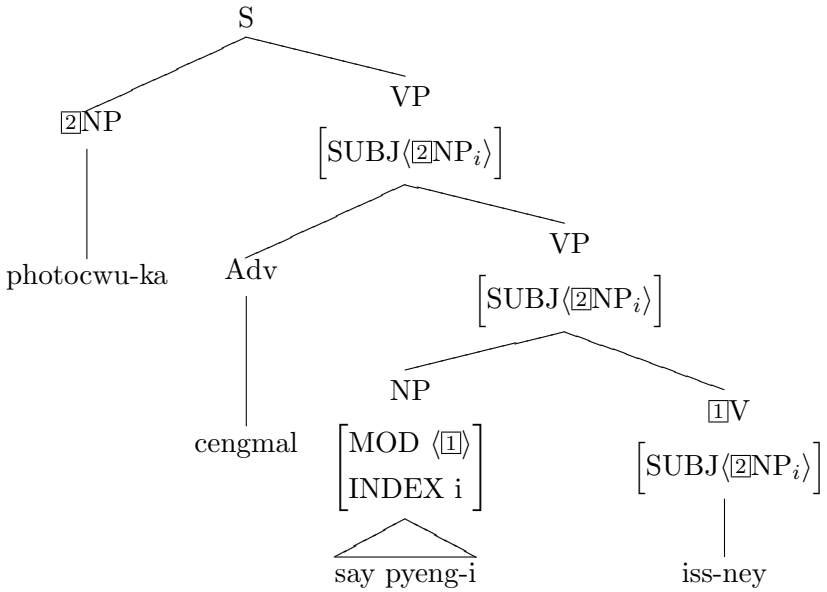
[15]The indirect object with DAT case cannot be the antecedent of an FQ:

(i) a. haksayng-tul-i ku pyonci-ul sensayngnim ney-pwun-eykey ponayessta.
student-PL-NOM that letter-ACC teacher four-CL-DAT sent
'Students sent the letter to four teachers.'
b. *haksayng-tul-i pyonci-lul sensayngnim-eykey neypwun-eykey ponayessta.
c. *haksayng-tul-i pyonci-lul sensayngnim-eykey cinancwumal-ey neypwun-eykey ponayessta.

In order to see how this system works, let us then consider a simple example in which the FQ is not adjacent to its antecedent NP:

(31) a. photocwu-ka cengmal sey pyeng-i iss-ney
wine-NOM really three CL-NOM exist-DECL
'There are really three bottles of wine.'

b.



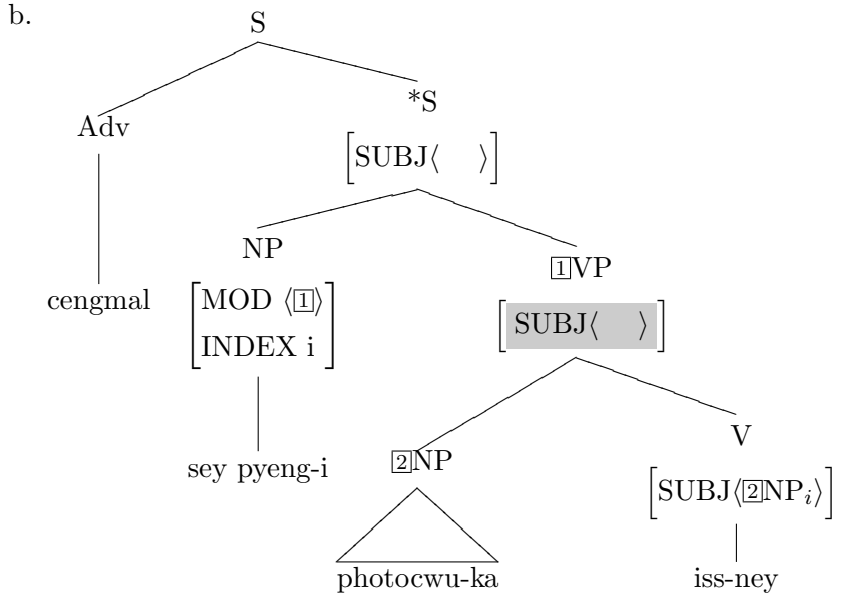
Notice here that it is crucial that the Num-CL refers to the unsatruated SUBJ value of the matrix verb. If the subject is saturated, then the Num-CL does not have any NP to be coindexed. This explains why (32a) is odd as give in its structure (32b):

(32) a. *cengmal sey pyeng-i photocwu-ka iss-ney
really three CL-NOM wine-NOM exist-DECL

d. ?haksayng-tul-i pyonci-lul sensayngnim-ul cinancwumal-ey neypwun-ul pon-ayessta.

However, when the goal argument has ACC only as in (id), the argument can serve as the FQ's antecedent.

b.



Here, even though *sey pyeng-i* modifies a verbal expression, it has no SUBJ whose index is identical with its own INDEX value.

This system explains one clear difference between a pure adverb and a floating quantifier: an adverb can occur almost in any place within a clause, but an FQ cannot. That is, even an ACC marked FQ cannot precedes its host NP as given in (33a):

(33) a. *Kim-un sey-kay-lul sakwa-lul mekessta.
Kim-TOP three-CL apple-ACC three-CL ate
'Kim ate three apples.'
b. *sey kay-lul Kim-un sakwa-lul mekessta.

The ungrammaticality of (33b) simply follows from the current system: The FQ *sey kay-lul* has no SUBJ or COMPS element whose INDEX value is identical with its own INDEX value. This in turn means that when the verb or its verbal projection has unsaturated valence features (SUBJ and COMPS), it makes the indices of its arguments 'visible' to any modifiers attaching to it.

5 Case Variations

More complicated phenomena occur in raising, causatives, and topicalization where the FQ and its antecedent have different case values. In raising

constructions, the ACC-marked raising NP argument can take either a NOM or ACC-marked floating quantifier as its modifiee. When the embedded subject is raised as the matrix object in (34b), both NOM and ACC-marked FQ can be the antecedent of the raised object (cf. Gerdts 1987):

(34) a. John-i haksayng-i sey-myeng-i/*ul
John-NOM student-NOM three-CL-NOM/*ACC

chencay-i-lako mit-ess-ta.
genius-COP-COMP believed
'John believed that three students are genius.'

b. John-i haksayng-ul sey-myeng-i/ul chencaylako mit-ess-ta.

In the present analysis in which the case-marked FQ is linked to either the SUBJ or a COMPS element, we can expect these variations. Let us consider the lexical entry for the raising verb:

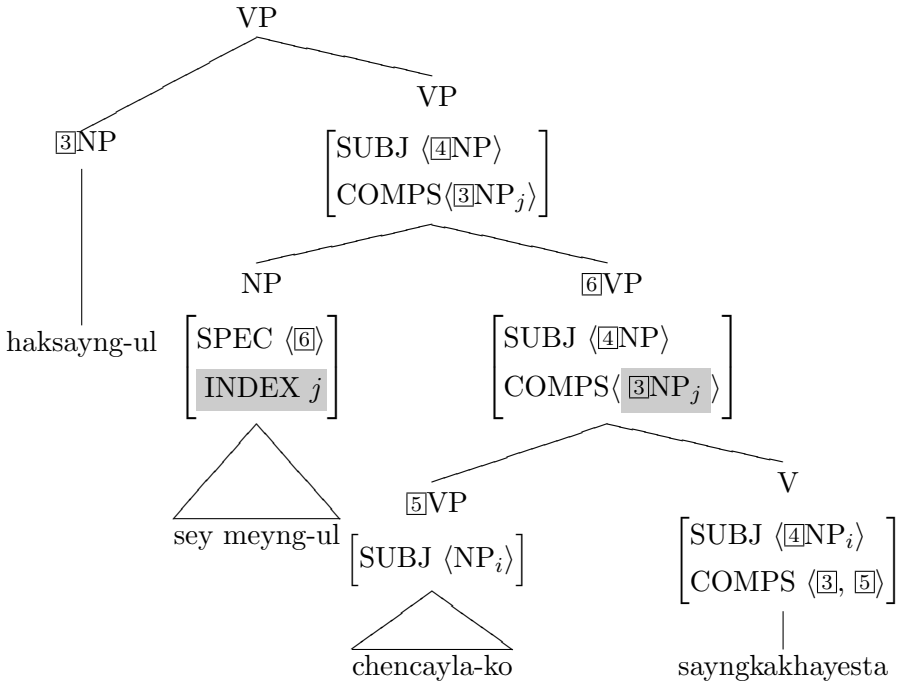
(35) a. $\begin{bmatrix} \text{HEAD} \mid \text{POS } \textit{verb} \\ \text{SUBJ } \langle \boxed{1}\text{NP} \rangle \\ \text{COMPS } \langle \boxed{2}\text{S} \rangle \\ \text{ARG-ST } \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix}$

b. $\begin{bmatrix} \text{HEAD} \mid \text{POS } \textit{verb} \\ \text{SUBJ } \langle \boxed{1}\text{NP} \rangle \\ \text{COMPS } \langle \boxed{2}\text{NP}_i, \boxed{3}\text{VP}[\text{SUBJ } \langle \text{NP}_i \rangle] \rangle \\ \text{ARG-ST } \langle \boxed{1}, \boxed{2}, \boxed{3} \rangle \end{bmatrix}$

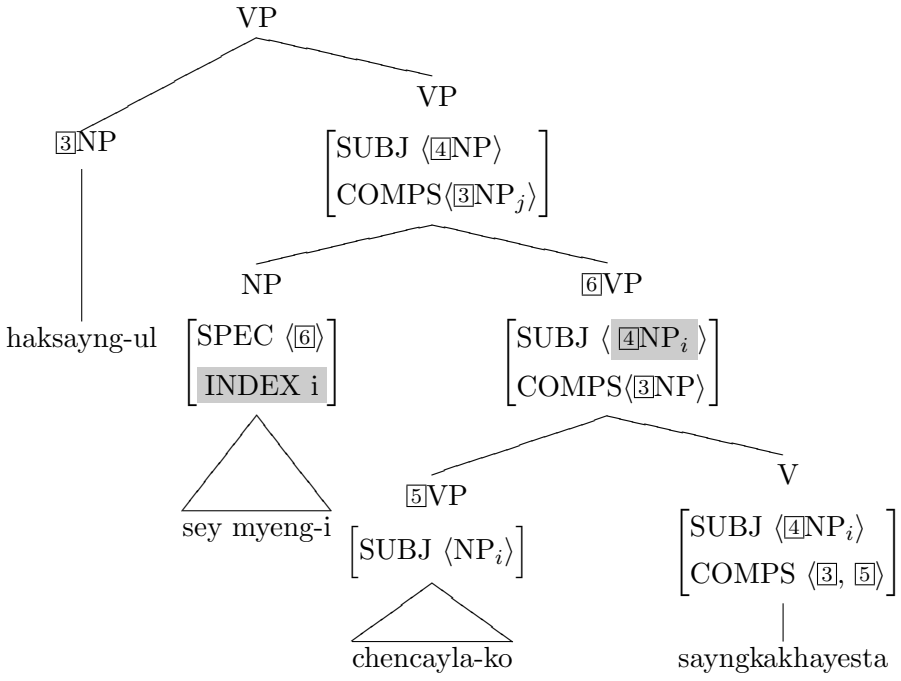
(35a) represents the lexical entry for *mit-ess-ta* in (34a) which selects a sentential complement whereas (35b) represents the raising verb *mit-ess-ta* in (34b) in which *haksayng* 'student' functions as its object.

Given the present analysis, the lexical entry (35b) will project at least three three possible structures related to the antecedent of the FQ:

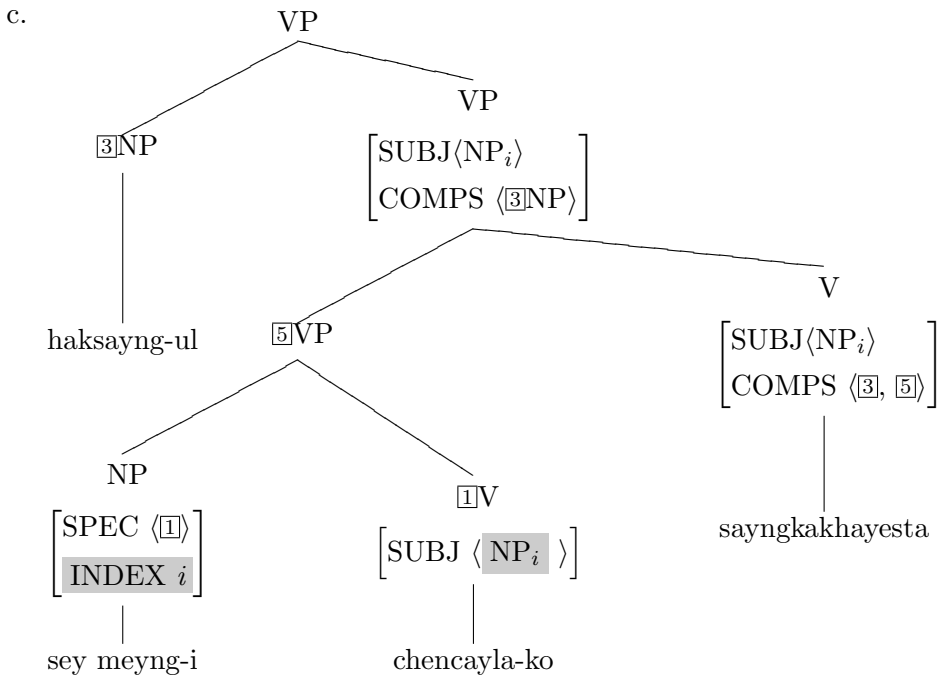
(36) a.


VP
VP
3NP
[SUBJ ⟨4NP⟩ COMPS⟨3NP$_j$⟩]
NP
6VP
haksayng-ul
[SPEC ⟨6⟩ INDEX j]
[SUBJ ⟨4NP⟩ COMPS⟨ 3NP$_j$ ⟩]
5VP
V
sey meyng-ul
[SUBJ ⟨NP$_i$⟩]
[SUBJ ⟨4NP$_i$⟩ COMPS ⟨3, 5⟩]
chencayla-ko
sayngkakhayesta

b.


VP
VP
3NP
[SUBJ ⟨4NP⟩ COMPS⟨3NP$_j$⟩]
NP
6VP
haksayng-ul
[SPEC ⟨6⟩ INDEX i]
[SUBJ ⟨ 4NP$_i$ ⟩ COMPS⟨3NP⟩]
5VP
V
sey myeng-i
[SUBJ ⟨NP$_i$⟩]
[SUBJ ⟨4NP$_i$⟩ COMPS ⟨3, 5⟩]
chencayla-ko
sayngkakhayesta

c.



In (36a), the FQ links to the COMPS value *haksayng-ul*, whereas (36b), it links to the SUBJ value of the main verb *sayngkakhayessta*. In (36c), the INDEX value of the FQ is linked to the SUBJ value of *chencayla-ko*.

Causative constructions also show a similar behavior.

(37) a. sensayngnim-i haksayng-i sey-myeng-i/*ul/*eykey ttena-key
teacher-NOM student-NOM three-CL-NOM/*ACC/*DAT leave-COMP
hayessta
did
'The teacher made three students to leave.'

b. sensayngnim-i haksayng-tul-eykey sey-myeng-i/*eykey ttena-key hayessta

c. sensayngnim-i haksayngtul-ul sey-myeng-i/ul/*eykey ttena-key hayessta

The causee can be Nom, DAT, or ACC in Korean causative constructions. The data tell us at least three facts: Regardless of the case of the causee, the DAT causee in (37b) can be the antecedent of the FQ, and the FQ's case need not agree with its modifying NP (causee), and finally the FQ should be either NOM or ACC.

The present analysis provides a clean analysis for these. The NP *haksayng-i* in (37)a is only the SUBJ of the verb *ttena-key*. However, *haksayng-eykey*

functions as the SUBJ of *ttena-key*. This is why the FQ can get NOM. The same logic applies to *haksayngtul-ul* in (38): This was the subject of *ttena-key* but at the same time functions as the COMPS element of *hayessta*.[16]

6 Conclusion

Numeral classifiers display complex properties in terms of both syntax and semantics. In particular, syntax approaches have met difficulties in linking a floated quantifier and its antecedent located remotely. The present analysis, allowing interactions among the syntactic valence features SUBJ and COMPS, ARG-ST information, SPEC information, and semantic information, can capture these complexities without resorting to movement operations.

In addition, we have seen that MRS (Minimal Recursion Semantics) can provides a straightforward way of composing semantics in a proper manner. It is needless to say that a precision grammar requires compositional semantics so that it can be suitable for applications requiring natural language understanding such as machine translation and automated email response. This paper shows us the possibility of building such a precision grammar.

References

Bender, Emily M., Dan Flickinger, 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 Evolution*, Coling 2002, pages 8-14, Taipei.

[16]It has been assumed that the caseless FQ quantifying a subject cannot be separated from the subject or appear in the postverbal position. However, as noted in Kang (2001), when quantification information is important, the caseless FQ can be acceptable even if it is not adjacent to its antecedent (cf. Downing 1993):

(38) a. A: haksayng-i myech-myeng-ina ku chayk-ul sassni?
student-NOM how many the book-ACC bought
'How many students bought the book?
b. B: haksayngtul-i ku chayk-ul ama twu myeng sassulkeya
student-NOM the book-ACC probably two CL bought
'Probably two students bought the book.'

This in turn then means that the grammar cannot rule out the examples where the FQ appears with no case, which our grammar expects t happen.

Bender, Emily M. and Melanie Siegel. 2004. Implementing the Syntax of Japanese Numeral Classifiers. *Proceedings of IJCNLP-04*, Hainan Island, China.

Bond, Francis, Daniel Kurz, and Satoshi Shirai. 1998. Anchoring Floating Quantifiers in Japanese-to-English Machine Translation. In *COLING-ACL98*.

Bond, Francis and Kyoung-Hee Paik. 2000. Reusing an ontology to generate numeral classifiers. In *Coling 2000*, Saarbrucken, Germany.

Choi, Kiyong. 2001 The Structure and Interpretation of Non-genitive Numeral Classifier Constructions in Korean. *Language Research* 37(3), 445-480. [written in Korean]

Choi, Young-Seok. 1988. A Reanalysis of Quantifier Floating in Korean. *Korean Journal of Linguistics* 13.

Copestake, Ann, Daniel P. Flickinger, Ivan A. Sag, and Carl Pollard. 2003. Minimal Recursion Semantics. An introduction. Under review.

Dowty, David. and B. Brodie. 1984. The semantics of "floated" quantifiers in transformationless grammar. *WCCFL* 4. 75-90.

Downing, Pamela. 1993. Pragmatic and semantic constraints on numeral quantifier position in Japanese. *Linguistics* 29, 65-93.

Gerdts, Donna B. 1987. Surface case and grammatical relations in Korean: The evidence from quantifier float. *Studies in Language II*, 181-197.

Gunji, Takao and Koiti Hasida. 1998. Measurement and Quantification. In T. Gunji and K. Hasida (eds.) *Topics in Constraint-Based Grammar of Japanese.* Kluwer Academic Publishers, Dordrecht, pp. 39-79.

Fukushima, K. 1991. Phrase Structure Grammar, Montague Semantics, and Floating Quantifiers in Japanese. *Linguistics and Philosophy* 14, 581-628.

Gunji, Takao. 2005. Measurement and Quantification Revisited. *Theoretical and Applied Linguistics at Kobe Shoin* 8:21–36

Haig, John H. 1980. Some observations on quantifier floating in Japanese. *Linguistics* 18, 1065-1083.

Kang, Beom-mo. 2002. Categories and meanings of Korean floating quantifiers-with some reference to Japanese. *Journal of East Asian Linguistics* 11, 375-398.

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

Kim, Jong-Bok, 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.

Matsumoto, Yo. 1993. Japanese numeral classifiers: A study of semantic categories and lexical organization. *Linguistics*, 31:667-713.

Miyagawa, Shigeru. 1989. *Structure and Case Marking in Japanese*. Academic Press, New York.

O'Grady, William. 1982. The Syntax and semantics of quantifier placement. *Linguistics* 20, 519-539.

Paik, Kyounghee and Francis Bond. 2002. Spatial representation and shape classifiers in Japanese and Korean. In David I. Beaver, Luis D. Casillas Martinez, Brady Z. Clark, and Stefan Kaufmann, editors, *The Construction of Meaning*, pages 163-180. CSLI Publications, Stanford CA.

Shibatani, Masayoshi. 1977. Grammatical Relations and Surface Cases. *Language* 53.4, 789-609.

Shi, Chung-Kon. 2000. The Syntactic structure of Quantifier Phrase in Korean. *Korean Journal of Linguisitcs*, 25-1, 73-101.

Siegel, Melanie and Emily M. Bender. 2002. Efficient deep processing of Japanese. In Proceedings of the 3rd Workshop on Asian Language Resources and Standardization, Coling 2002, Taipei.