Case Markers and Word Order in Multi-{} Combinatory Categorial Grammar

Keun Young Shin
(Hanyang University)

Shin, Keun Young. 2011. Case Markers and Word Order in Multi-{} Combinatory Categorial Grammar. *The Linguistic Association of Korea Journal.* 19(2). 57-78. The traditional Combinatory Categorial Grammar account of the scrambling phenomena of languages with case markers operates under the assumption that case markers play a computational role in word order by licensing grammatical relations between NPs and predicates. This paper shows that case markers do not always have such a powerful function and that various word order cues are interactive. An alternative analysis will be proposed in the framework of Multi-set Combinatory Categorial Grammar, which enables us to formalize the effects of case markers and other word order cues in the computation of sentences. The explanation offered here depends on complementarily using a case feature provided by a verb and assigning different modalities for different lexical items.

Key Words: case markers, word order, combinatory categorial grammar (CCG), multi-set CCG, multi-modal CCG

1. Introduction

Korean and Japanese are classified as free word order languages, which allow local and long distance scrambling. For instance, Korean allows two permutations of a simple transitive sentence, as shown in (1), and NPs can be scrambled out of their clausal boundaries as in (2).

(1) a. Cheli-ka Mina-lul cohaha-n-ta
Cheli-Nom Mina-Acc like-Pres-Dec

'Cheli likes Mina.'

b. Mina-lul Cheli-ka cohaha-n-ta
Mina-Acc Cheli-Nom like-Pres-Dec
'Cheli likes Mina.'

(2) phica-lul Cheli-ka [Mina-ka mek-ess-ta-ko] malhay-ss-ta
pizza-Acc Cheli-Nom [Mina-Nom eat-Pst-Dec-that] say-Pst-Dec
'Cheli said that Mina ate the pizza.'

In Combinatory Categorial Grammar (CCC), scrambling in languages with case markers has been generally dealt with by assigning a computational role to case markers (Karttunen 1989; O'Grady 1991, 1998; J Yoon 1998; Bozsahin 1998). In this computational case approach, case markers are treated as combinatory functors that take NPs and allow them to combine with predicates under the assumption that grammatical relations are expressed by case.

(3) Computational Case Rules (O'Grady 1991)
(i) The nominative case marks an NP that combines with an intransitive verbal category (IV)
(ii) The accusative case marks an NP that combines with a transitive verbal category (TV)
(iii) The genitive case marks an NP that combines with a noun category

In other words, in the computational case approach, scrambling is allowed because case markers determine grammatical relations between argument NPs and predicates.

However, this paper will show that case markers do not always specify a particular grammatical relation between an NP and a predicate and that other factors can control word order in Korean. An alternative analysis will be proposed in the framework of Multi-set Combinatory Categorial Grammar (Multi-{} CCG) in which word order variations are captured under the assumption that there is no rigid word order for argument NPs. The explanation offered here depends on complemetarily using a case feature provided by a verb and assigning different modalities for different lexical items.

This paper is organized as follows. Section 2 presents the computational case approach and its shortcomings. As an alternative approach to the computational case approach, section 3 discusses Hoffman's (1995a,b) Multi-{} CCG. By modifying the current multi-set approach using modalities, section 4 proposes a Korean Multi-{} CCG and shows how this approach can handle various word order cues. Section 5 summarizes the conclusions of this paper.

2. The Computational Case Approach

In the computational case approach, case markers are analyzed as functors that allow NPs to combine with predicates (Karttunen, 1989; O'Grady, 1991, 1998; Yoon, 1998; Bozsahin, 1998). This computational function of case markers is made explicit by treating them as type-raisers in CCG. That is, a nominative case triggers type-raising and changes an NP into a function over an intransitive verb (i.e., S/(S\NP)), whereas an accusative-case-marked NP is type-raised to become a functor that combines with a transitive verb (i.e., (S\NP)/((S\NP)\NP)).[1)]

As illustrated in (5) and (6), this approach permits both SOV and OSV readings with a transitive verb.

(4) mek- 'eat' := (S\NP1)\NP2: $\lambda x \lambda y$ eat'$(x)(y)$

(5)

Cheli-ka	Mina-lul	cohaha-n-ta
Cheli-Nom	Mina-Acc	like-Pres-Dec
NP1	NP2	(S\NP1)\NP2
----------> T	----------->T	
S/(S\NP1)	(S\NP1)/((S\NP1)\NP2)	
----------> B		
S/((S\NP1)\NP2)		
---------->		
S		

1) Throughout this paper, I assume familiarity with Combinatory Categorial Grammar.

(6) phica-lul Cheli-ka mek-ess-ta
pizza-Acc Cheli-Nom eat-Pst-Dec
NP2 NP1 (S\NP1)\NP2
---------->T ---------->T
(S\NP1)/((S\NP1)\NP2) S/(S\NP1)
---------- <Bx
S/((S\NP1)\NP2)
---------->
S

In (5) and (6), the case-marked subjects and objects are converted to functions over predicates, and they can combine by the function composition rules, yielding a functor seeking a transitive verb on the right side regardless of their relative order.

Since Saito 1985, it has been claimed that free word order languages such as Korean and Japanese are configurational languages that have a rigid word order. Kim (1996) claims that the configurationality of Korean can be supported by the fact that a word order is fixed when argument NPs appear without overt case markers: a simple transitive sentence is only interpreted as SOV if the subject and the object are not case-marked, as illustrated in (7).

(7) Cheli Mina cohaha-n-ta
Cheli Mina like-Pres-Dec
'Cheli likes Mina.'

The computational case approach can explain this word order freezing phenomenon because it assumes that Korean has a basic or default word order. The default word order is given in the verbal category by ordering its arguments as in (4). An NP without a case marker cannot be type-raised and only combines with a verb by the function application rule, as in (8).

(8) Cheli Mina cohahanta
Cheli Mina like
NP NP (S\NP1)\NP2
---------- >
S\NP1
---------- >
S

The non-case-marked NPs in (8) are interpreted following the word order specified in the verbal category; hence, we can obtain only an SOV reading for (8): 'Cheli likes Mina'.

However, the computational case approach faces empirical problems. One problem concerns the assumption that the nominative case marker always seeks an intransitive verb category. Contrary to the computational case approach's speculation, there is no one-to-one correspondence between a grammatical relation and a case. In Korean, some transitive verbs take nominative-case-marked objects. In (9) and (10), the objects are marked not with the accusative case but with the nominative case.

(9) Cheli-nun saca-ka/*lul mwusep-ta / kekcengi-ta
Cheli-Top lion-Nom/*Acc be afraid-Dec/ worry-Dec
'Cheli is afraid of/is worried about the lion.'

(10) Cheli-nun kohayng-i/*ul kulip-ta / coh-ta
Cheli-Top hometown-Nom/*Acc miss-Dec / like-Dec
'Cheli misses/ likes the hometown.'

Under the computational case approach, a nominative-case-marked NP would be interpreted as the subject of the verb in the preverbal position because it undergoes type-raising and becomes a functor that combines with an intransitive verb, as exemplified in (11). Example (11) would be wrongly interpreted to mean that the hometown missed Cheli.

(11) Cheli-ka kohayng-i kulipta
Cheli-Nom hometown-Nom miss
NP NP $(S\backslash NP_1)\backslash NP_2$
---------- >T ---------- >T
S/(S\NP) S/(S\NP)
---------- >Bx
$S\backslash NP_2$: λx_2miss′(x)(hometown′)
---------- <
S: miss′(Cheli′)(hometown′)

A nominative-case-marked object is considered an idiosyncratic property of certain classes of predicates that express psychological or emotional states in Korean and Japanese (Kuno, 1973; Kang, 1986; Kim, 1990; Ura, 1999).[2] Thus, the lexical information of a verb plays a significant role in the interpretation of the nominative-case-marked NP. The computational case approach has difficulty in dealing with nominative-case-marked objects because a verb does not have case information in the lexicon.

Note that the computational case approach cannot rule out sentences where NPs are marked by wrong cases. As in derivation (11), it allows two arguments of a transitive verb to be marked with the nominative case; hence, the approach cannot predict that a typical transitive verb cannot take a nominative-case-marked object, as in (12).

(12) *Cheli-ka phica-ka mek-ess-ta
Cheli-Nom pizza-Nom eat-Pst-Dec
'Cheli ate the pizza.'

In addition, non-case-marked NPs can be scrambled in certain circumstances, contrary to the clam in Kim (1996). Case markers tend to be dropped when the grammatical role of an NP can be easily identified by semantic or pragmatic factors (Lee and Thompson, 1989; Lee, 2002; Lee, 2006). Consider example (13), where the NP arguments are not case-marked but carry semantic information that specifies their grammatical roles in the sentence. We can obtain only an

2) Kang (1986) and Kim (1990) propose that ACC is assigned by a [-stative] verb, whereas NOM is assigned by default to caseless NPs.

OSV reading for (13).

(13) phica ce cohahay-yo
pizza I like-Pol
'I like pizza.'

This shows that non-case-marked NPs can undergo scrambling and that case marking is not the only necessary condition for scrambling. Semantic features associated with NPs can function as another powerful word order cue in resolving parsing ambiguities when case markers are not available. However, to date, the computational case approach has largely ignored word order cues other than case markers and has not provided a satisfying account for word order variations in Korean.

3. The Multi-{} Approach

An alternative way to deal with scrambling is to relax the strict ordering of the arguments of a verb and allow a flexible category that can project multiple word orders directly in combinations. This is the general idea behind Multi-set Combinatory Categorial Grammar (Multi-{} CCG), originally proposed by Hoffman (1995a, b) in order to deal with the free word order of Turkish.

In Multi-{} CCG, a single category can contain a multi-set of arguments that are not ordered. For example, the transitive verb *cohahata* 'like' is assigned the lexical category in (14): it is defined as a functor looking for a set of arguments.

(14) *cohahata* 'like':= S: like(y, x)|{NPnom: y, NPacc: x}

A transitive verb is allowed to combine with its argument in any order by applying the application rule in (15).

(15) Multi-set CCG Composition Rules (Hoffman 1995b)

a. $X|(Args \uplus \{/Y\})\quad Y \Rightarrow X|\ Args \quad (>)$

$Y\quad X|(Args \uplus \{\backslash Y\}) \Rightarrow X|Args \quad (<)$

b. $X|(ArgsX \uplus \{/Y\})\quad Y|ArgsY \Rightarrow X|(ArgsX \uplus ArgsY) \quad (>B)$

$Y|ArgsY\quad X|(ArgsX \uplus \{\backslash Y\}) \Rightarrow X|(ArgsX \uplus ArgsY) \quad (<B)$

As exemplified in (16), case features of the verb force a nominative-case-marked NP and an accusative-case-marked NP to be respectively interpreted as the subject and the object regardless of their surface order.[3)]

(16) a. Cheli-ka Mina-lul cohahanta

Cheli-Nom Mina-Acc like

NPnom: Cheli′ NPacc: Mina′ S:like′(y, x)|{NPnom:y, NPacc:x}

---<

S: like′(y, Mina′)| {NPnom: y}

--- <

S: like′ (Cheli′, Mina′)

b. Mina-lul Cheli-ka cohahanta

Mina-Acc Cheli-Nom like

NPacc: Mina′ NPnom: Cheli′ S: like′(y, x)|{NPnom:y, NPacc:x}

---<

S: like′(Cheli′, x)| {NPacc: x}

--- <

S: like′(Cheli′, Mina′)

This mechanism allows us to describe nominative-case-marked objects since a case-marker on the NP alone simply does not determine its grammatical role, and a verb bears its own case features in the lexicon. It also has the potential to capture the scrambling of non-case-marked NPs because it makes no assumptions about a canonical word order.

Multi-{} approach can solve some of the problems with the computational case approach, but it also faces it own problems. In order to handle restrictions

3) Note that the CCG combination rules are modified to be sensitive to multisets, as defined in (15), where Args is a variable for a set of categories. Adopting Baldridge's (2002) notation, I use multiset union $\uplus$ and slashes instead of regular set union $\cup$ and an arrow above the argument for a direction feature.

in scrambling, Hoffman (1995a, b) suggests that the Pure Multi-{} CCG should be extended to a Prioritized Multi-{} CCG. In the Prioritized Multi-{} CCG, multisets of arguments are prioritized in a certain linear order in a lexicon. For example, if a simple transitive verb can take a non-case-marked object, as in (17), the lexical information of the verb should specify the order of the two multisets of arguments, as in (18).

(17) Cheli-ka phica mek-ess-ta
Cheli-Nom pizza eat-Pst-Dec
'Cheli ate the pizza.'

(18) a. cohahata 'like' := S | {NPnom, NPacc}
b. cohahata 'like' := S | {NPnom}| {\NP}

The category (18b) forces the verb to combine first with an unmarked object on the immediate left before combining with a nominative-case-marked subject. In this way, it can capture the fact that the non-case-marked NP in (17) is interpreted as the object. Multi-{} CCG lexicalizes syntactic restrictions in word order, but this solution is unattractive because it essentially denies the function of case markers in scrambling, which is not sensitive to the quirks of individual verbs. Furthermore, since all local word orders cannot be specified with a single category in the system, this approach will increase the complexity of the grammar.

In addition, Multi-{} CCG, as well as the standard CCG, causes the grammar to overgenerate. It is well known that CCG's composition and type-raising rules are crucial in handling coordination and scrambling without transformational movement, but they cause overgeneration problems and often need to be restricted in an ad hoc way (Baldridge & Kruijff, 2003).[4] For example, in Korean, the word order is not completely free. Adjectives such as *say* 'new' cannot be separated from NPs, as in (19); nevertheless, Multi-{} CCG cannot prevent an NP modifier from combining with a verb, as shown in derivation (19b). In order to avoid derivation (19b), we need to restrict the (backward) composition rules: Y≠NP in (15).

4) For a detailed discussion, see Baldridge (2002), Baldridge and Kruijff (2003), and Steedman & Baldridge (2002).

(19) a. [say chayk-i] nawa-ss-ta
[new book-Nom] come out-Pst-Dec
'A new book was published.'

b. *chayk-i nawassta say
book-Nom came out new
NP S|{NP} NP|{NP}
---------- >B
S|{NP}
---------- <
S

However, different languages require different restrictions on combinatory rules. English grammar should prohibit the permutation-inducing rule >Bx (Steedman, 2000); otherwise, the embedded subject would be scrambled into the higher clause: **I Bill think that met John.* In Dutch, two restricted versions of each of the rules >B, >Bx, and <Bx are needed (Steedman, 2000). As pointed out by Baldridge and Kruijff (2003), those restrictions on certain combinatory rules are ad hoc, and they are not cross-linguistically motivated, which may lead to lose the appeal of the purely type-driven nature of Categorial Grammar.

Despite these unappealing aspects, Multi-{} CCG enables us to capture the fact that case markers do not always play a computational role in licensing a combinatorial dependency. This paper proposes a formal analysis for Korean word order variations by modifying this Multi-{} approach.[5)]

---

5) In fact, an attempt to explain Korean word order using Multi-{} CCG exists. Adopting Hoffman's approach, Cha et al. (2002) develop Korean multi-{} CCG, but they still assume that case-marked NPs obligatorily undergo type-raising as in (i) and (ii). Therefore, it is not very different from the computational case approach.

(i) NP+case-marker ⇒ S/(S\NP[case-feature])

(ii) ku-ka sakwa-lul meknunda
he-Nom apple-Acc eat
S/(S\NP[nom]) S/(S\NP[acc]) S\{NP[nom], NP[acc]}
---------- <B
S\{NP[nom]}
---------- <
S

4. The Proposal: Multi-{} CCG with Modalities

In order to avoid overgeneration problems without stipulating ad hoc constraints on the CCG rules, Baldridge and Kruijff (2003) and Baldrige (2002) brought modalities into CCG as a device to control the applicability of the combinatory rules. Multi-Modal Combinatory Categorial Grammar (Multi-Modal CCG) is a hybrid categorial grammar (Baldridge, 2002; Baldridge & Kruijff, 2003; Steedman & Baldridge, 2003), incorporating the CCG computational approach into the resource-sensitivity of Categorial Type Logic (Morill, 1994; Moortgat, 1997). I will modify Hoffman's (1995a,b) Multi-{} CCG rules adopting this multi-modal system. This alternative approach can overcome the problems discussed in the preceding sections.

In Multi-Modal CCG, rules and function categories are modalized. The four basic modalities are as follows:

Table 1. Modals in Multi-Modal CCG

	Non-permutative	Permutative
Non-associative	★	x
Associative	◇	•

Each combinatory rule is sensitive to a particular modality, and thus applies only to input categories with the appropriate type slashes. The followings are the modified application and composition rules using these modals:

(20) a. $X/_{\star}\{Y,\cdots\}$ Y $\Rightarrow X/_{\star}\{\cdots\}$ (>)

Y $X\backslash_{\star}\{Y\cdots\}$ $\Rightarrow X\backslash_{\star}\{\cdots\}$ (<)

b. $X/_{\diamond}\{Y, \cdots_1\}$ $Y/_{\diamond}\{\cdots_2\}$ $\Rightarrow X/_{\diamond}\{\cdots_1, \cdots_2\}$ (>B)

$Y\backslash_{\diamond}\{\cdots_2\}$ $X\backslash_{\diamond}\{Y, \cdots_1\}$ $\Rightarrow X\backslash_{\diamond}\{\cdots_1, \cdots_2\}$ (<B)

c. $X/_{\times}\{Y, \cdots_1\}$ $Y\backslash_{\times}\{\cdots_2\}$ $\Rightarrow X/_{\times}\{\cdots_1\}\backslash_{\times}\{\cdots_2\}$ (>Bx)

$Y/_{\times}\{\cdots_2\}$ $X\backslash_{\times}\{Y, \cdots_1\}$ $\Rightarrow X\backslash_{\times}\{\cdots_1\}/_{\times}\{\cdots_2\}$ (<Bx)

d. $X \Rightarrow T/_i(T\backslash_i X)$ (<T)

$X \Rightarrow T\backslash_i((T/_i X)$ (>T)

The ★ modality is the most restricted since it is non-permutative and

non-associative; hence, any categories can combine by the application rules. The ◇ and × modalities are used to distinguish between the harmonic composition rules and the permutative composition rules.[6] Note that the type-raising rules in (20d) use a variable modality *i*, and they are applied only to primitive argument categories in the system. The resource-sensitivity of Categorial Type Logic allows the CCG to have a universal grammar. In other words, the combinatory rules in (20) are universal, and all typological variations are placed in the lexicon.

4.1. Word Order Variations: Case Markers

By adopting this modified version of Multi-{} CCG, we can capture word order variations in Korean. Unlike Hoffman's original approach, a directional slash is used in the system for Korean, a head-final language: the forward and backward slashes indicate whether a given category is an adjunct or a head. The backward slash (\) in (21) indicates that a verb takes its arguments from the left.

(21) a. ketta 'walk' := S\.{NP[nom]}
b. mekta 'eat' := S\.{NP[acc], NP[nom]}
c. kulipta 'miss' := S\.{NP[nom], NP[nom]}

It should be stressed that case features in (21) are not as strong as those in Hoffman's (1995a,b) Multi-{} CCG. Case-features such as [acc] and [nom] in a verbal category do not require argument NPs to be case-marked. Their function is to prevent a case-marked NP from being construed as a wrong argument of the verb and to enable parsing a sentence with case-marked NPs quickly, as shown in (22).

---

6) For instance, given that the ◇ modality is associative and non-permutative, the forward composition rule actually represents the following four instantiations:

X/◇{Y, ⋯1} Y/◇{⋯2} => X/◇{⋯1,⋯2}
X/◇{Y, ⋯1} Y/.{⋯2} => X/.{⋯1,⋯2}
X/.{Y,⋯1} Y/◇{⋯2} => X/◇{⋯1,⋯2}
X/.{Y,⋯1} Y/.{⋯2} => X/.{⋯1,⋯2}

(22) a. NP[nom] = Nom-marked or non-case-marked NP
b. NP[acc] = Acc-marked or non-case-marked NP

For example, NP[nom] in (22) means that this NP argument cannot be saturated by an NP which is marked with a case other than the nominative case; NP[nom] may be saturated with a nominative-case-marked NP or a non-case-marked NP. This complementary use of the case feature can capture the case-dropping phenomenon and the scrambling of non-case-marked NPs without lexicalizing all possible word orders.

Examples (23a) and (23b) demonstrate how a simple transitive sentence can be computed when NPs are case-marked. Even though the nominative-case-marked NP occurs in front of the transitive verb in (23b), it cannot be interpreted as the object because NPacc blocks it.

(23) a. Cheli-ka phica-lul mek-ess-ta
Cheli-Nom pizza-Acc eat-Pst-Dec
NPnom: Cheli′ NPacc: pizza′ S:eat′(y,x)\.{NP[acc]:x, NP[nom]:y}
---<
S: eat′(y, pizza′)\.{NP[nom]: y}
---<
S: eat′(Cheli′, pizza′)

b. phica-lul Cheli-ka mek-ess-ta
pizza-Acc Cheli-Nom eat-Pst-Dec
NPacc: pizza′ NPnom:Cheli′ S:eat′(y,x)\.{NP[acc]:x, NP[nom]:y}
---<
S: eat′(Cheli′, x)\.{NP[acc]:y}
---<
S: eat′(Cheli′, pizza′)

Long-distance scrambling can be easily handled by applying the backward harmonic composition rule, as in (25).

(24) malhata 'say':= S: say′(x, p)\.{Sko: p, NP[nom]: x}

(25) phica-lul Cheli-ka [Mini-ka mekessta-ko] malhay-ss-ta
pizza-Acc Cheli-Nom [Mini-Nom ate-ko('that')] say-Pst-Dec
NPacc NPnom NPnom Sko\.{NPacc, NPnom} S\.{Sko, NP[nom]}
----------<
Sko\.{NPacc}
----------<B
S\.{NPnom, NPacc}
----------<
S\.{NPacc}
----------<
S

As discussed above, the computational case approach has difficulty in dealing with nominative-case-marked objects in Korean because it assumes that a case marker specifies a particular syntactic relation between an NP and a predicate. In the proposed Multi-{} approach, a verb category has case information in the lexicon, and hence, it can express the fact that a stative verb takes a nominative-case-marked NP as its object argument, as was seen in (10) and repeated in (27) below. Sentence (27) can be derived using lexical entry (26).

(26) kulipta 'miss':= S: miss'(y, x)\.{NP[nom]: x, NP[nom]: y}

(27) Cheli-ka kohayng-i kulip-ta
Cheli-Nom hometown-Nom miss-Dec
NPnom NPnom S:miss'(y,x)\.{NP[nom]:x, NP[nom]:y}
----------<
S: miss'(y, hometown')\.{NP[nom]: y}
----------<
S: miss' (Cheli', hometown')

Since the two NPs are marked with the same case, we can expect that the nominative-case-marked NP immediately preceding the stative verb will be construed as either the object or the subject. In (27), an OSV reading is not acceptable for a semantic reason: the verb 'miss' requires its subject to be an animate entity.[7]

7) Note that it is still predictable that the object is not marked by the nominative case in a

It is also possible to describe the fact that only a limited number of ditransitive verbs can take two accusative-marked arguments in Korean: for example, *cwuta* 'give', *kaluchita* 'teach', and *chipwulhata* 'pay' (Jung & Miyagawa, 2004). In a typical ditransitive sentence, the indirect object and the direct object should be marked by the dative case and the accusative case respectively. Compare (28) and (29).

(28) Mina-ka sakwa-lul Cheli-eykey/lul cwu-ess-ta
Mina-Nom apple-Acc Cheli-Dat/Acc give-Pst-Dec
'Mina gave an apple to Cheli.'

(29) Mina-ka pyenci-lul Cheli-eykey/*lul ponay-ss-ta
Mina-Nom letter-Acc Cheli-Dat/*Acc send-Pst-Dec
'Mina sent a letter to Cheli.'

The proposed Multi-{} CCG can lexicalize the difference between the two ditransitive verbs in (28) and (29), as shown in (30) and (31).

(30) cwuta:= S: give'(z, y, x)\.{NP[acc]:x, NP[acc/dat]:y, NP[nom]:z}
(31) ponayta:=S: send'(z, y, x)\.{NP[acc]:x, NP[dat]:y, NP[nom]:z}

Since *ponayta* 'send' can take only one accusative-case-marked NP, sentence (33) is ruled out unlike (32).

---

typical transitive sentence due to case features specified in a verbal category. For example, derivation (i) is blocked because NP[acc] in the verbal category prevents the nominative-case-marked NP from filling the subject position.

(i) * Cheli-ka phica-ka mek-ess-ta
Cheli-Nom pizza-Nom eat-Pst-Dec
NP[nom] NP[nom] S: eat'(y,x)\.{NP[acc]:x, NP[nom]:y}
----------<
S: eat'(pizza', x)\.NP[acc]: x
----------XXXXXXX----------

(32) Mina-ka sakwa-lul Cheli-lul cwu-ess-ta
Mina-Nom apple-Acc Cheli-Acc give-Pst-Dec
NPnom NPacc NPacc S\.{NP[acc], NP[acc], NP[nom]}
----------<
S\.{NP[acc], NP[nom] }
----------<
S\.{NP[nom] }
---------- <
S

(33) *Mina-ka pyenci-lul Cheli-lul ponay-ss-ta
Mina-Nom letter-Acc Cheli-Acc send-Pst-Dec
NPnom NPacc NPacc S\.{NP[acc], NP[dat], NP[nom]}
----------<
S\.{NP[dat], NP[nom] }
----------XXXXX----------

4.2. NP Modifiers

In Korean, the NP modifier *say* 'new' cannot be separated from an NP. This NP attachment property can be described by assigning a restricted modality to the NP modifier (Shin 2008), as in (34). Note that the forward slash (/) in (34) indicates that an NP must occur to the right of the modifier.

(34) say 'new':= NP/★{NP}

The adjective *say* 'new' can form a constituent with the NP it modifies, as shown in (35).

(35) a. say chayk
new book
NP/★NP NP
---------- <
NP

However, the ★ modality on the slash of the category for *say* 'new' is

incompatible via the forward and backward crossed composition rules. Adjectives of this type therefore cannot be separated, ruling out the sentences in (36) and (37).[8]

(36) *chayk-i nawassta say
book-Nom came out new
NPnom S\•{NP[nom]} NP/⋆{NP}
----------XXXXX----------

(37) *chayk-i say nawassta
book-Nom new came out
NPnom NP/⋆{NP} S\•{NP[nom]}
----------XXXXX----------

The proposed analysis has an advantage over the original Multi-{} CCG in that it can explain that Korean word order is not completely free without restricting compositional rules in an ad hoc manner.

4.3. Case-dropping: Semantic Features and Discourse Factors

It has been stated without argument in the literature that non-case-marked NPs cannot be scrambled in Korean: a sentence is interpreted as the basic SOV word order without case marking. Recent studies have shown that case-dropping occurs only in particular circumstances (Ko, 2000; Matsuda, 1996; Lee, 2002; Lee, 2006, among others). Lee (2006) examines a Korean corpus of telephone conversations and finds that case markers tend to be dropped when semantic factors such as person, animacy, and specificity can disambiguate the grammatical roles of NPs: NPs are left unmarked when subjects are high in person, animacy, and definiteness, and objects are low in those dimensions. In other words, case markers are eliminated when the grammatical role of an NP is

8) The genitive case marks an NP as modifying another NP. The genitive case can be analyzed as a marker shifting an NP category type:

(i) NP: y+*uy* := NP: R(y, x) /⋆{NP: x} where R indicates a relationship between the two NPs.

easily identified without them. The scrambling of non-case-marked NPs is possible when subjects are not likely to be confused with objects, as was seen in (13), and repeated in (38).

(38) phica ce cohahay-yo
pizza I like-Pol
'I like pizza.'

Multi-{} CCG enables us to capture the fact that semantic features, as well as case features, can play a role in word order in Korean. Under the proposed analysis, non-case-marked NPs in a sentence are not strictly ordered, and hence SOV and OSV readings are possible for (38). The SOV reading can be ruled out by the semantic mismatch because the verb 'eat' does not take an inanimate entity as its subject argument as in (39); it would be acceptable if it had a figurative meaning, as it would be in the sentence 'the pizza ate Cheli.'

(39) S: eat'(y, x)\.{NP[acc]: x, NP[nom, +animate]: y}

The question then arises as to what determines word order if case features and semantic features cannot resolve parsing ambiguities. Consider the following sentence where the unmarked subject and object are both animate.

(40) Cheli Mina choahay-yo.
Cheli Mina like-Pol

At first glance, it seems that (40) can be interpreted only as SOV (Kim, 1996; Cho & Choe, 2001), but a close examination shows that both SOV and OSV readings are available for (40). (40) is interpreted as OSV if *Cheli* represents old information and serves as a topic. We can see this more clearly when the subject and object are marked with focus or topic-markers.

(41) Cheli-nun Mina-to cohahay-yo
Cheli-Top Mina-delim 'also' like-Pres-Dec
'Mina also likes Cheli.' 'Cheli likes Mina as well.'

As discussed above, case-dropping is common and natural in informal speech. According to Lee and Thompson (1989), the accusative case is more likely to drop in contexts where it is less necessary to specify the grammatical roles of NPs because there is a greater amount of "shardeness between communicators," that is, shared experience, context, and cultural background. The grammatical role of an unmarked or discourse-marked NP is determined by a given context, and hence, the ultimate word order of a sentence is affected by pragmatics and interactional contexts. The parser should allow both SOV and OSV readings for (40) and (41). Unlike the computational case approach, the proposed Multi-{} CCG has a potential to capture the scrambling of non-case-marked NPs because no rigid word order is assumed.

Recall that we can get the OSV reading for (40) when the subject is the topic, a link to the previous context. It has been claimed that different word orders encode different information structures in Korean (Choi, 1999; Jo, 2004). In the traditional CCG approach, different interpretations of information structure and discourse focus are obtained by different derivational surface structures corresponding to different intonation structures (Steedman, 2000). Hoffman (1995a) shows that it is also possible to incorporate information structure into Multi-{} CCG, suggesting that a verbal category is associated with an ordering category which serves as a template for information structure. This is one option that can be adopted for accounting for the effect of scrambling on information structure. In the present moment, finding a method to integrate Korean Multi-{} CCG with information structure will be left as the object of further research.[9]

9) It is possible to explain the strong preference of the SOV reading for (40) by redefining a default word order in terms of frequency. It has been observed that the actual frequency of scrambled sentences is very low in Japanese (Kuno, 1973; Yamashita & Suzuki, 1995), and the same might be said of Korean. One of the advantages of CCG in computational linguistics is that it can integrate with the statistical approach and resolve parsing ambiguities through probability (Cha et al., 2002). The preferred reading of (40) can be taken as the result of resolving the ambiguity of the thematic roles of non-case-marked NPs in favor of the most probable analysis.

5. Conclusions

This paper has shown that case markers do not always play a computational role by determining grammatical relations between NPs and predicates and that various word order cues interact with one another in resolving parsing ambiguities. Multi-{} CCG provides an effective framework to display the fact that case marking, verb types, parts of speech, semantic features and discourse factors can influence word order in Korean. In the proposed Multi-{} CCG, a verb has its own case features in the lexicon, which are used complementarily. Not only does this capture that different cases may be assigned to NPs depending on the semantics of the verb, it also allows us to explain that non-case-marked NPs can undergo scrambling in Korean. Even though Multi-{} CCG assumes that there is no rigid word order for arguments, word order restrictions are accounted for by assigning different modalities for different lexical items.

References

Baldridge, J. (2002). *Lexically Specified Derivational Control in Combinatory Categorial Grammar*. Unpublished doctoral dissertation. University of Edinburgh.

Baldridge, J & Kruijff, G-J M. (2003). Multi-Modal Combinatory Categorial Grammar. *Proceedings of the 10th EACL, 1,* 211-218.

Bozsahin, C. (1998). Deriving Predicate-Argument Structure for a Free Word Order Language. *Proceedings of the 36th EACL and 17th COLING,* 167-173.

Cha, J, Lee, G., & Lee, J. (2002). Korean Combinatory Categorial Grammar and Statistical Parsing. *Computers and the Humanities* 36, 431-453.

Cho, S.-Y., & Choe, J. (2001). Nominal Markers and Word Order in Korean. *Proceedings of the 15th PAC,* 247-258.

Choi, H.-W. (1999). *Optimizing Structure in Context*. Stanford, CA: CSLI.

Hoffman, B. (1995a.) Integrating "Free" Word Order Syntax and Information Structure. *Proceedings of the 1995 EACL,* 245-252.

Hoffman, B. (1995b). *Computational Analysis of the Syntax and Interpretation of "Free" Word-Order in Turkish.* Doctoral dissertation. University of Pennsylvania.

Jo, J.-M. (2004). *Grammatical effects of Topic and Focus Information.* Doctoral dissertation. University of Illinois at Urbana-Champaign.

Jung, Y.-J., & Miyagawa, S. (2004). Decomposing Ditransitive Verbs. *Proceedings of the 6th SICOGG,* 101-120.

Kang, Y.-S. (1986). *Korean Syntax and Universal Grammar.* Doctoral dissertation. Harvard University.

Karttunen, L. (1989). Radical Lexicalism. In Baltin, M., & Kroch, A. (Eds.), *Alternative Conceptions of Phrase Structure.* Chicago: Chicago University Press.

Kim, Y.-H. (1996). Overt Case and Covert Case in Korean. *Studies in Generative Grammar 8*(1), 177-236.

Kim, Y.-J. (1990). *The Syntax and Semantics of Korean Case: the Interaction between Lexical and Syntactic Levels of Representation.* Doctoral dissertation. Harvard University.

Ko, E.-S. (2000). A Discourse Analysis of the Realization of Objects in Korean. *Japanese/Korean Linguistics* 9, 195-208.

Kuno, S. (1973). *The Structure of the Japanese Language.* Cambridge, Mass: MIT Press.

Lee, D.-Y. (2002). The Function of the Zero Particle with Special Reference to Spoken Japanese. *Journal of Pragmatics* 34, 645 – 682.

Lee, H. (2006). Parallel Optimization in Case Systems: Evidence from Case Ellipsis in Korean. *Journal of East Asian Linguistics* 15, 69-96.

Lee, H. S., & Thompson, S. A. (1989). A Discourse Account of the Korean Accusative Marker. *Studies in Language* 13, 105 – 128.

Matsuda, K. (1996). *Variable Zero-Marking of (o) in Tokyo Japanese,* Doctoral dissertation. University of Pennsylvania.

Moortgat, M. (1997). Categorial type logics. In Johan, B., & Meulen, A. (Eds.), *Handbook of Logic and Language.* Cambridge, MA: The MIT Press.

Morill, G. V. (1994). *Type Logical Grammar: Categorial Logic of Signs.* London: Kluwer Academic Publishers.

O'Grady, W. (1991). *Categories and Case: The Sentence Structure of Korean.* Philadelphia: John Benjamins.

O'Grady, W. (1998). Korean Case: a Computational Approach. In Park, B.-S., & Yoon, J. (Eds.), *Papers from the 11th International Conference on Korean*

Linguistics, Seoul: Hankwuk mwunhwasa.
O'Grady, W. (1999). *Computing Korean*. Unpublished manuscript.
Saito, M. (1985). *Some Asymmetries in Japanese and their Theoretical Implications*. Doctoral dissertation. MIT.
Shin, K. Y. (2008). Quantified Noun Phrases in a Head-Final Language. *Proceedings of the 37th NELS*, 197-208.
Steedman, M. (2000). *The Syntactic Process*. Cambridge, MA: The MIT Press.
Steedman, M., & Baldridge, J. (2003). Combinatory Categorial Grammar. Unpublished manuscript.
Ura, H. (1999). Checking Theory and Dative Subject Construction in Japanese and Korean. *Journal of East Asian Linguistics* 8, 223-254.
Yamashita, H. & Suzuki, Y. (1995). Frequency of Scrambled Sentences in Japanese Texts. Unpublished manuscript.
Yoon, J. H.-S. (1998). RE: Korean Case: A Computational Approach (O'Grady 1998). Unpublished manuscript.

Keun Young Shin
Department of English Language and Culture
Hanyang University
1271 Sa 3-dong, Sangnok-gu
Ansan-si, Gyeonggi-do 426-791, Korea
Phone: 82-31-400-5344
Email: kyshin@hanyang.ac.kr

Received on 12 March, 2011
Revised on 21 May, 2011
Accepted on 21 May, 2011