

Configurationality Parameter in Korean and its Computational Implication - An HPSG Approach

Minhaeng Lee

Cheju National University

1. Introduction

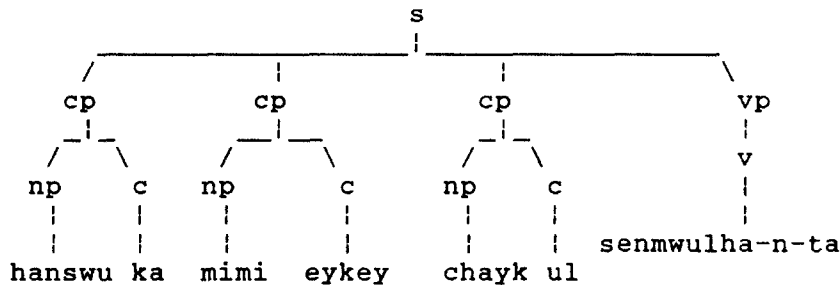
This paper has the following two purposes. First, I will show that the configurational hypothesis provides a good basis for the adequate description of the so-called "case conversion" phenomena involved in nominalization and causativization in Korean. With respect to the case conversions, I will discuss the argument structure inheritance and present a precise formulation of it within the HPSG framework. Second, I will show that the configurational hypothesis can have a computational implication. I would like to describe an experimental system of machine translation from Korean to English which assumes the configurational structure as a universal characteristic of Korean and English.

2. Nominalization in Korean and Case Conversion

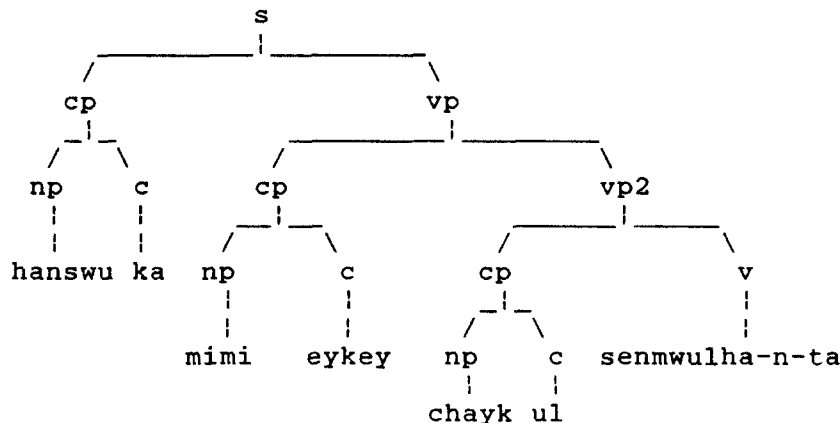
According to the configurationalist, Korean syntax does not follow a flat structure but a hierarchical configurational structure. In line with the position, the structure of the sentence (1a) would be presented not as in (1b) but as in (1c).¹

- (1) a. hanswu-ka mimi-eykey chayk-ul senmwulha-n-ta.
 hanswu N mimi D book A give Tense Mood
 Hanswu gives Mimi a book.

b.



c.



The configurationality hypothesis seems to be at first supported by the case conversion phenomenon in nominalization. Morphologically, the nominalization construction in Korean is formed by adding the nominalizers "-(u)m" and "-ki" to the verb stem. Syntactically, the nominalizer inherits the unsaturated arguments from the verbal expression and assigns the genitive case to the inherited arguments.² An outstanding characteristic of nominalization in Korean is that it takes place continuously. Consider the following examples.³

- (2) hanswu-ka mimi-eykey chayk-ul senmwulha-n-ta.
 hanswu N mimi D book A give- Tense Mood
 Hanswu gives mimi a book.
- (3) a. hanswu-uy mimi-eykey-uy chayk-uy senmwulha -m -i
 Nm N
 (kuney-lul kippu-key hay-ss-ta.)
 her delight-caus Tense-Mood
 It delighted Mimi, that Hanswu gave her a book.
- b. hanswu-uy mimi-eykey-uy chyak-ul senmwulha -m- i
 A
 (kuney-lul kippu-key hay-ss-ta.)
 her delight-caus Tense-Mood
 It delighted Mimi, that Hans gave her a book.
- c. hanswu-uy mimi-eykey chayk-ul senmwulha -m-i
 D A
 (kuney-lul kippu-key hay-ss-ta.)
 her delight-caus Tense-Mood
 It delighted Mimi, that Hanswu gave her a book.
- d. hanswu-ka mimi-eykey chayk-ul senmwulha -m -i
 N D A
 (kuney-lul kippu-key hay-ss-ta.)
 her delight-caus Tense-Mood
 It delighted Mimi, that Hanswu gave her a book.

The nominalizer "-m" is contained in the examples (3a)-(3d). The verb stem *senmwulha-* is nominalized in (3a) and therefore the genitive case is assigned to all the nominal phrases. I will refer to the derivation as VØ-nominalization. In (3b) the verb phrase *chayk-ul senmwulha-* is nominalized and therefore the genitive case is assigned to nominal phrases *hanswu* and *mimi*. I will refer to the derivation as Vmax-2 -nominalization. In (3c) the verb phrase *swumi-eykey chayk-ul senmwulha-* is nominalized and therefore the genitive case is assigned to the nominal phrase *hanswu*. I will refer to the derivation as Vmax-1-nominalization. The whole sentence is nominalized in (3d) and therefore no genitive case appears in the sentence. I will refer to the derivation as Vmax-nominalization. This continuation property of nominalization convincingly explains the case conversions involved in it and supports the assumption that Korean syntax has not a flat structure but a configurational structure, as mentioned above.

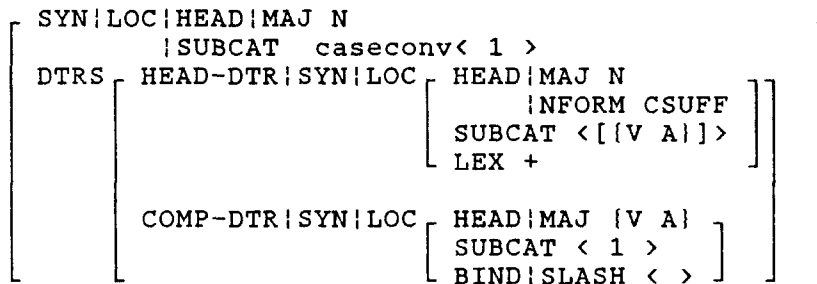
The ungrammaticality of the following examples can be explained by the fact that the nominalized expression is not able to assign a nominative case to its inherited arguments.

- (4) a. *hanswu-ka mimi-eykey chayk-uy senmwulha-m-i sasil...
 hanswu N mimi D book G give Nm N fact ...
 It is a fact, that Hanswu gives Mimi a book.
- b. *hanswu-ka mimi-eykey-uy chayk-uy senmwulha-m-i sasil...
 hanswu N mimi D book G give Nm N fact ...
 It is a fact, that Hanswu gives Mimi a book.
- c. *hanswu-ka mimi-eykey-uy chayk-ul senmwulha-m-i sasil...
 hanswu N mimi D book A give Nm N fact ...
 It is a fact, that Hanswu gives Mimi a book.

Now, turn to the formulation within the framework of HPSG of the above discussed nominalization process. We can posit the rule for the nominalization as in (5).

(5) Syntactic Rule for Korean Nominalization (SRK)

SRK1

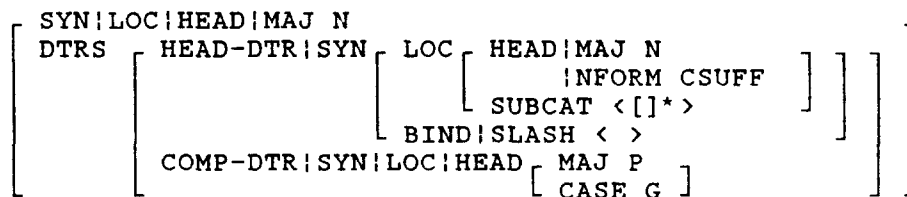


In the rule SRK1 above, the argument inheritance and case conversion are taken into consideration. The operation *caseconv* takes the subcat-frame of the verbal expressions as input and gives a variant of the subcat-frame as output, where case conversion takes place. Nominative/Genitive and Accusative/Genitive have occurred. The rule also takes into consideration that a scrambled verbal expression may not be partially nominalized, as discussed in Han (1987) but can be wholly nominalized. Observe the following examples.⁴

- (6) a. *mimi-uy hanswu-ka coaha-m-i sasil-i-ta
 mimi A hanswu G love Nom N fact be Mood
 It is a fact, that Hanswu loves Mimi.
- b. mimi-lul hanswu-ka coaha-m-i sasil-i-ta
 mimi A hanswu N love Nom N fact be Mood
 It is a fact, that Hanswu loves Mimi.

In order to combine the nominalized expression and its arguments, we need another grammatical rule as follows.

(7) **SRK2**



Now, let us see an analysis of a nominalized sentence.

(8) a. hanswu-uy mimi-uy coaha-m-i sasil-i-ta
 hanswu G mimi G love Nm N fact be Mood
 It is a fact, that Hanswu loves Mimi.

b. hanswu-uy mimi-uy coaha -m (-i)
 [MAJ P] [MAJ P] [V] [N]
 [CASE G] [CASE G] [SUBCAT<[P],[P]>] [Nsuff]
 \ [A] [N] /
 \ / <-- SRK1
 [N]
 [Nsuff]
 [SUBCAT<[P],[P]>]
 [G],[G]
 \ / <-- SRK2
 [N]
 [Nsuff]
 [SUBCAT<[P]>]
 [G]
 \ / <-- SRK2
 [N]
 [Nsuff]
 [SUBCAT< >]

3. Causativization and Case Conversion

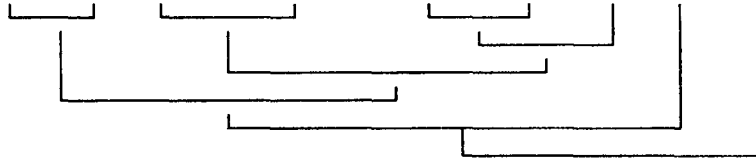
Korean has two types of causative constructions, i.e. lexical and affixal causative. The affixal causativization, with which we are only concerned in this section, exhibits interesting features, as for the case conversion. Let us consider the following sentences.

- (9) a. emeni -ka ai -lul/-eykey ca -key ha -n -ta
 mother N baby A/D sleep Caus Aux Tense Mood
 The mother let the baby sleep.
 b. emeni -ka ai -ka ca -key ha -n -ta
 mother N baby N sleep Caus Aux Tense Mood
 The mother let the baby sleep.
 c. ai -ka ca -n -ta
 baby N sleep- Tense Mood.
 The baby sleeps.
- (10) a. emeni -ka ai-lul/-eykey cec -ul mek -key ha -n -ta
 mother N baby A/D milk A drink Caus Aux Tense Mood
 The mother let her baby drink milk.
 b. emeni -ka ai-ka cec -ul mek -key ha -n -ta
 mother N baby N milk A drink Caus Aux Tense Mood
 The mother let her baby drink milk.
 c. ai -ka cec -ul mek -nun -ta
 baby N milk A drink- Tense Mood
 The baby drinks milk.

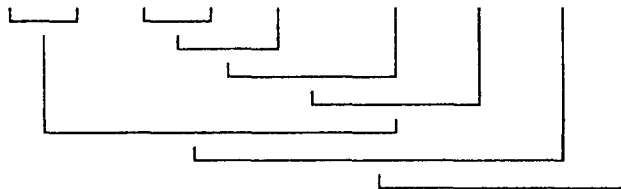
I will assume, that the difference between (9a) and (9b) and between (10a) and (10b) can be attributed to the difference of the syntactic derivations. In (9a), the verb phrase of the Vmax-1-level, ca- is causativized. In (9b), the whole embedded sentence ai-ka ca- is causativized. The same explanation can be given to (10a) and (10b). In (10a), the Vmax-1-level verb phrase

cec-ul mek is causativized. In (10b), the whole sentence, Vmax-level verb phrase *ai-ka cec-ul mek* is causativized. If we describe informally the derivational difference, we get the following analysis trees.

- (11) a. *emeni-ka ai -lul/-eykey ca -key ha -n -ta*
 mother N baby A D sleep- Caus Aux Tense Mood



- b. *emeni -ka ai -ka ca - key ha -n -ta*
 mother N baby N sleep- Caus Aux Tense Mood



An argument for the differentiation of two types of affixal causativization is provided by the fact that in (12b) the adverb "ppalli" modifies solely the embedded verb *ca-* but in (12a) the adverb modifies the auxiliary *-key ha-* as well as the embedded verb *ca-*.

- (12) a. *emeni-ka ai-ka ppalli ca-key ha-n-ta.*
 soon
 The mother let her baby her baby sleep soon.
 b. *emeni-ka ai-lul ppalli ca-key ha-n-ta.*
 soon
 The mother let her baby soon sleep.

Another argument for the line can be brought in connection with scrambling. Consider the following examples.

- (13) a. **ai-ka emeni-ka ca-key ha-n-ta*
 b. *ai-lul emeni-ka ca-key ha-n-ta*

Under the assumption that the scrambling in Korean is clause-bounded, the ungrammaticality of the sentence (13a) is easily explainable, because scrambling takes place across the clause in (13a).

Now, let us turn to the formulation of the causativization in the HPSG framework. The lexical entries for the causative suffix *key*, auxiliary *ha-* are assumed as follows.

(14)

[PHON	key]	
	SYN LOC	HEAD		[MAJ V VFORM VSUFF VFNCT CAUS ...]]
		SUBCAT		< [MAJ V VFORM BSE ...] >
		LEX +]
SEM CONT]		

(15)

[PHON	ha]	
	SYN LOC	HEAD		[MAJ V AUX CAUSATIVE MODAL - ...]]
		SUBCAT		< [MAJ V VFORM VSUFF] [MAJ N] > [VFNCT CAUS] [CASE N]
		LEX +]
SEM CONT]		

The syntactic rule for the combining of the V_{max-1}-level verb phrase or the V_{max}-level verb phrase with the causative suffix *-key* can be formulated as in (16).

(16) **SRK3**

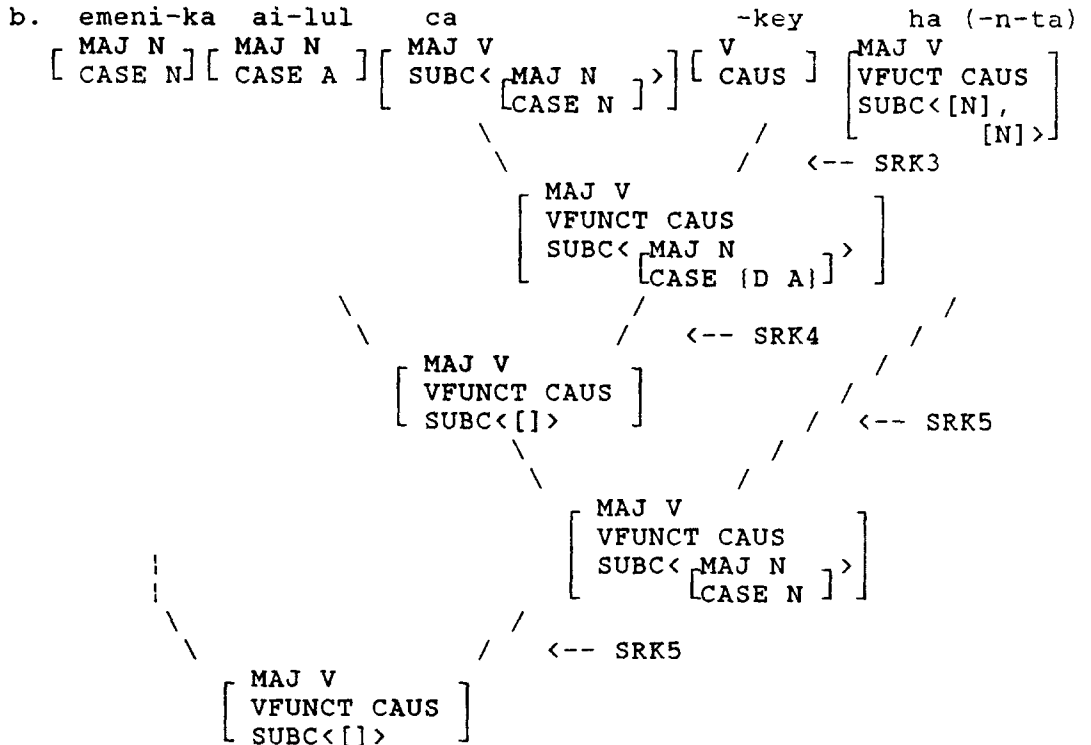
[SYN LOC HEAD MAJ N]
	SUBCAT caseconv2< 1 >	
DTRS	HEAD-DTR SYN LOC HEAD MAJ V]
	VFORM VSUFF VFNCT CAUS SUBCAT < [V] > LEX +	
[COMP-DTR SYN LOC HEAD MAJ V]
	SUBCAT < 1 >	
	BIND SLASH < >	

(17) **SRK4**

[SYN LOC HEAD MAJ V]
	DTRS	
[HEAD-DTR SYN]
	LOC	
[HEAD]
	MAJ V VFORM VSUFF VFNCT CAUS SUBCAT < []* >	
[BIND SLASH ...]
	COMP-DTR SYN LOC HEAD	
[MAJ N]
	CASE {D A}]	

The operation `caseconv2` in the rule takes care that the nominative argument of the `Vmax-1`-level verb phrase, i.e. the causee is inherited to the causativized mother node and its case is converted to dative or accusative. Finally, observe an analysis example:⁴

(18)a. `emeni-ka ai -lul/-eykey ca -key ha -n -ta (= (11a))`
 mother-N baby-A/-D sleep-V[caus] V[caus]-tense-mood
 The mother let her baby sleep



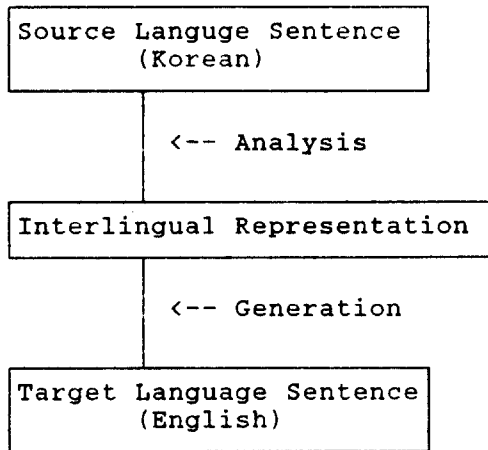
4. Configurational Structure and Computational Implication

In this section, I discuss the computational implication of the configurational structure. I will describe a Korean-English experimental machine translation system under development.

The discussions above provide us with a good basis for the configurational structure of the Korean syntax. As generally accepted, the configurational structure is indispensable for English syntax. Here I propose to adopt the configurational structure as a universal linguistic structure for the interlingual representation of machine translation system based on the interlingual model. In the interlingual approach, analysis produces a representation which contains semantic primitives rather than words of a particular language. This representation is used directly by the generator to produce a target language text.

A Korean-English machine translation system based on the interlingual model is presented, as follows:

(19)



A left-corner-parser takes over the task of analysing the Korean sentences.⁵ It takes a list as input and gives an interlingual representation in the form of the functor-argument-structure as output. As for the interlingual representation, I assume that not only English but also Korean has tense-, mood- and case phrase, where the cases are understood as semantic cases such as agent, patient and addressee, which respectively corresponds to nominative, objective and dative in Fillmore (1968). A definite-clause-grammar generator, which is unification-based and augmented with feature structure, is integrated into the system for the generation of English.⁶ The feature structure can be used for the description of subject-verb agreement in English. The system is able to solve the problems related to the different coding mechanisms of cases and tenses. In English, cases are configurationally determined and tenses are synthetically coded by verbs, while not only cases but also tenses are analytically coded by the relevant morphemes in Korean. Now, I would like to describe the translation process with an example.

(20) a. ?- mtkes([tonio,ka,yeonkuk,ul,coaha,n,ta]).

b. ?- parse([tonio,ka,yeonkuk,ul,coaha,n,ta],[,IL1),
IL1 = ..[_,IL].

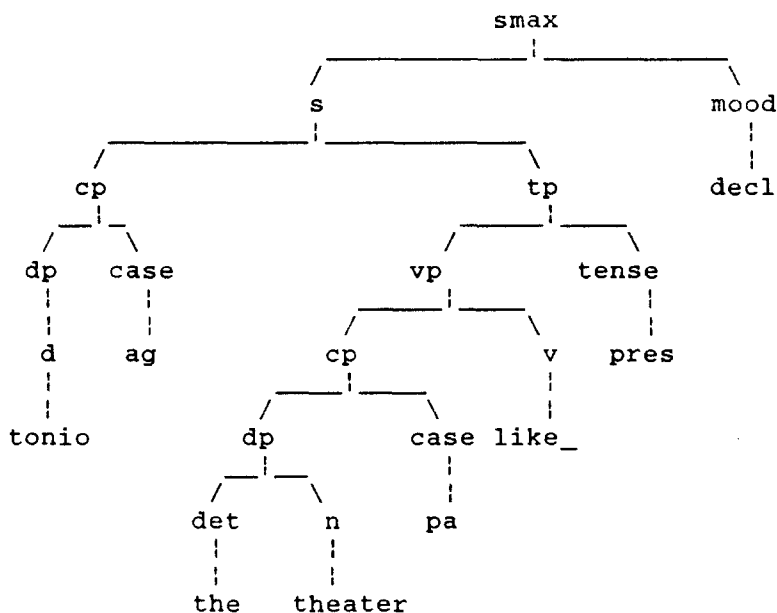
c. IL1 = smax(smax(s(cp(dp(d(tonio)),case(ag)), !!!
tp(vp(cp(dp(det(d_),n(theater)),case(pa)),
v(like_)),
tense(pres))),
mood(decl))
)

=>> Interlingual Representation in F-A-Structure

```

IL = smax(s(cp(dp(d(tonio)),case(ag)),
tp(vp(cp(dp(det(d_),n(theater)),case(pa)),
v(like_)),
tense(pres))),
mood(decl))
  
```


d. =>> Interlingual Representation in C-Structure



e. `mtkes(KS) :- parse(KS,[],L), L =.. [_ ,L1],
 esmax(S,[],L1,F),nl, write(S),nl,nl,
 write('>>> Information Structure '),
 nl,nl, print_fstr(F).`

(21) a. `?- esmax(S,[],
 smax(s(cp(dp(d(tonio)),case(ag)),
 tp(vp(cp(dp(det(the),n(theater)),case(pa)),
 v(like_)),
 tense(pres))),
 mood(decl)) ,
 F).`

b. >>> K-E-Translation:

S = [tonio,likes,the,theater,.]

c. F = [phon:likes,cat:v,
 subc:[first:[agr:[num:sg,per:3!_],
 case:nom,phon:tonio,cat:det!_],
 second:[case:acc,phon:the,cat:det,
 comp:[phon:theater,cat:n
 !_!_!_],
 mood:[phon:.,cat:mood!_!_]

d. >>> Information Structure

```
I--->phon:likes
I--->cat:v
I--->subc
I   I--->first
I   I   I--->agr
I   I   I   I--->num:sg
I   I   I   I--->per:3
I   I   I--->case:nom
I   I   I--->phon:tonio
I   I   I--->cat:det
I   I--->second
I       I--->case:acc
I       I--->phon:the
I       I--->cat:det
I       I--->comp
I           I--->phon:theater
I           I--->cat:n
I--->mood
    I--->phon:.
    I--->cat:mood
```

The predicate which is responsible for the whole process of the machine translation is *mtkes/1* which is stated in (20a). *parse/3* of the left-corner-parser is the predicate which is relevant to the transfer of the Korean sentence to the interlingual representation. In (20b) *parse/3* takes the list [tonio,ka,yeonkuk,ul,coaha,n,ta] as input and gives the configurational structure IL1 in (20c) as output. If we delete the functor *smax* of the structure IL1, we get the interlingual representation IL as in (20c), which can be converted into an analysis tree by the predicate *print_cstr/1* (Lehner 1990: 168-173), as in (20d). The predicate *emax/3* in (21a) which is responsible for the generation takes the interlingual representation as input and gives the list of the strings of the target language, English, i.e. [tonio,likes,the,theater,.] as in (21b) and the corresponding feature structure F as in (21c), which can be transformed to an analysis tree as in (21d) by the predicate *print_fstr/1*.

5. Concluding Remarks

This study is motivated by the need to find a convincing evidence for the existence of the configurational structure in Korean. I have shown that case conversions involved in nominalization and causativization in Korean can be adequately described by the HPSG framework based on the configurationality hypothesis. I have given also a description of an experimental machine translation system from Korean to English which adopts configurational structure as interlingual representation, in order to support the configurationality hypothesis from the computational viewpoint.

NOTES

- * The original title was "Nominalization, Causativization, and Configurationality Parameter in Korean --- An HPSG Approach". I would like to thank to Prof. Kiyong Lee, Prof. Soo-Song Shin and Prof. Byung-Soo Park for their valuable comments on earlier version of this paper.

- 1 The category "cp" stands for case phrase.
- 2 If we assume that the nominalizer is the head and the verbal expression is the complement, it is also a peculiarity that the nominalizer inherits the arguments from its complement, i.e., the verbal expression.
- 3 Han (1987) doesn't accept the Accusative/Genitive conversion in nominalization, based on his intuition on the unacceptability of the sentence (3a). But the sentence is acceptable, according to my intuition.
- 4 The rules SRK5 and SRK6 take care of the combining of the constituents in Korean. In the presented analysis the rule SRK5 is involved, which combines a head constituent with a complement constituent..

a. **SRK5**

$$\left[\begin{array}{l} \text{SYN|LOC|SUBCAT} \langle \rangle \\ \text{DTRS} \left[\begin{array}{l} \text{HEAD-DTR|SYN|LOC|LEX} + \\ \text{COMP-DTRS} \langle [] \rangle \end{array} \right] \end{array} \right]$$

b. **SRK6**

$$\left[\begin{array}{l} \text{SYN|LOC|HEAD} \left[\langle 1 \rangle \right] \\ \text{DTRS} \left[\begin{array}{l} \text{HEAD-DTR|SYN|LOC|HEAD} \langle 1 \rangle \\ \text{ADJ-DTR|SYN|LOC|LEX} + \end{array} \right] \end{array} \right]$$

- 5 The left-corner-parser is originated from Shieber/Pereira (1987).
- 6 A definite-clause-grammar parser augmented with the feature-structure is presented in Gazdar/Mellish (1989).

REFERENCES

- Gazdar, G. and C. Mellish (1990) *Natural Language Processing in PROLOG - An Introduction to Computational Linguistics*, Wokingham.
- Han, H.-S. (1987) *The Configurational Structure of the Korean Language*, Seoul.
- Johnson, D.-E. and H. Watanabe (1991) "Relational-Grammar-Based Generation in the JETS Japanese-English Machine Translation System." *Machine Translation* 6. 1-20.
- Lee, K. (1992) "An Analysis of Some Fragments of Korean for Computational Semantics." ms, Seoul.
- Lee, M. (1992) *Kontrastive Syntax und maschinelle Sprachanalyse im Rahmen einer Unifikationsgrammatik: Untersuchungen zum Deutschen und Koreanischen*, Frankfurt a. M.
- Lehner, Ch. (1990) *Prolog und Linguistik*, Muenchen.
- Pereira, F.C.N. and S.M. Shieber (1987) *Prolog and Natural Language Analysis*. CSLI Lecture Notes 10.
- Pollard, C. and Ivan A. Sag (1987) *Information-Based Syntax and Semantics*, Vol.I: Fundamental Notions, CSLI Lecture Notes 13.
- Shibatani, M. (1973) "Lexical versus periphrastic causatives in Korean." *Journal of Linguistics*, Vol.9, 281-297.
- Shin, S.-S. (1988) "On Long Distance Movement in German." *Linguistische Berichte* 115. 229-254.

APPENDIX

: *Grammatical Rules & Lexical Entries Formats*

/* Analysis */

```
smax(smax(S,mood(M))) ---> [s(S), mood(M)].  
s(s(CP,T)) --->  
  [cp(CP,ag), tp(T,TNS)].      /* ag == Agent */  
word(coaha,v2(v(like_))).
```

/* Generation */

```
esmax(smax(S,mood(M)),Fsmax) --> es(S,Fs), emood(M,Fmood),  
  {Fsmax === Fs, Fsmax:mood === Fmood}.  
es(s(CP,T),Fs) --> ecp(CP,ag,Fcp), etp(T,Ftp),  
  {Fs === Ftp, Fs:subc:first === Fcp}.  
ev2(v(like_),pres,Fv) --> [likes], {Fv:phon === likes,  
  Fv:cat === v,  
  Fv:subc:first:agr:num === sg,  
  Fv:subc:first:agr:per === 3,  
  Fv:subc:first:case === nom,  
  Fv:subc:second:case === acc}.
```