Structural Ambiguities in the Light Verb Constructions: Lexical Relatedness and Divergence*

Jong-Bok Kim Kyung-Sup Lim Jaehyung Yang

June 5, 2007

Abstract

One of the most widely used constructions in Korean is the so-called light verb construction (LVC) involving an active-denoting verbal noun (VN) together with the light verb *ha-ta* 'do'. This paper first discusses the argument composition of the LVC, mixed properties of VNs both of which have provided a challenge to syntactic analyses with a strict version of X-bar theory. The paper shows the mechanism of multiple classification of category types with systematic inheritance can provide an effective way of capturing these mixed properties. In particular, it assumes that VNs have both [N +] and [V +] features to reflect their dual properties. The paper also addresses the issue of relatedness and divergence between the VNs with an accusative argument and those without it. An implementation of the analysis within the LKB (Linguistics Knowledge Building) system also proves its feasibility and efficiency.

Key words: light verb, verbal noun, argument composition, HPSG, implementation

1 Issues

The first main theoretical and computational issue we encounter in the analysis of the LVC is the status of the light verb and argument composition. One of the main properties the light verb *ha* 'do' carries is that it does not affect the argument structure of the VN (verbal noun) it combines with:[1]

*We thank three anonymous reviewers of this journal for questions and criticisms. This research was supported by the Korea Research Foundation (KRF-2006-A00577)

[1]The abbreviations for the glosses and attributes used in this paper are ACC (ACCUSATIVE), ARG (ARGUMENT), C-CONT (CONSTRUCTIONAL CONTENT), DAT (DATIVE), DECL (DECLARATIVE), LBL (LABEL), LOC (LOCATIVE), LTOP (LOCAL TOP), NOM (NOMINATIVE), PL (PLURAL), PRE (PREDICATE), PST (PAST), IND (INDEX), RELS (RELATIONS), TOP (TOPIC), etc.

(1) a. John-i Mary-eykey cenhwa(-lul hayessta)
John-NOM Mary-DAT phone-ACC did
'John phoned Mary.'

b. John-i Mary-lul myengtan-ey chwuka(-lul hayessta)
John-NOM Mary-ACC list-LOC addition-ACC did
'John added Mary to the list.'

c. John-i ku chayk-ul Mary-lopwuthe manwon-ey kwuip(-ul hayessta)
John-NOM the book-ACC Mary-from 1,000 won-LOC buy-ACC did
'John bought the book from Mary at 1,000 won.'

As observed here, it is the type of VN (*cenhwa, chwuka, kwuip*) that decides the types of arguments in the given sentence: the light verb *ha-* does not influence the needed arguments. This fact has led the literature to view that the light verb has no argument structure on its own but inherits the argument structure of the theta-transparent VN.

We can also observe that like auxiliary verbs, the light verb itself does not assign a particular theta role to the subject as noted in (2):

(2) a. John-i ton-ul unhayng-ey yekum-ul hayessta
John-NOM money-ACC bank-LOC deposit-ACC did
'John deposited the money in the bank.'

b. hwasal-i kwanyek-ey myengcwung-ul hayessta
arrow-NOM target-ACC mark-GOAL did
'The arrow marked the target.'

The NP *John* here acts as an agent, whereas *hwasal* is a theme, implying that the light verb is thematically underspecified with its subject role. This is what we find in the auxiliary construction too.

In addition, there exist further arguments to support the view that the VN forms a complex predicate with the following light verb, inducing monoclausal properties. For example, the LVC also undergoes the passivization process, which is a canonical monoclausal property:

(3) tampay-ka mikwuk-ulopwute swuip-i toyessta
cigarette-NOM America-from import-NOM became
'The cigarette was imported from America.'

The NPI phenomenon also indicates that the VN and the light verb behave like one unit.

(4) a. ku hwoysa-nun mikwuk-ulopwute amwukesto [swuip ha-ci anhassta]
the company-TOP America-from anything import do-COMP not
'The company imported nothing from America.'

b. *John-un ku hwoysa-ka [mikwuk-ulopwute amwukesto swuipha-tolok]
John-TOP the company-TOP America-from anything import-COMP
seltukha-ci anhassta
persuade-COMP not
'John did not persuade the company to import anything.'

Unlike the VP selecting predicate *seltukha-* 'persuade' as in (4)b, the NPI object *amwukesto* in (4)a is licensed in the LVC. The contrast can be captured if we take the VN and the light verb and the negative auxiliary all to form one complex predicate.

The second main issue concerns the grammatical status of VNs. It is well-observed that in terms of the internal properties, VNs behave like verbs, whereas in terms of external syntax, they act like nouns. For example, as observed in (1), VNs select their own arguments and assign verbal cases such as ACC, regardless of the light verb's presence. Adverbial modification also supports the verbal properties of VNs: the VN can be modified by an adverb but not by an adjectival element.

(5) catongcha-ul mikwuk-ey **elyepkey**/*elyewun swuchwul(-ul hayessta)
car-ACC America-LOC hard/difficult export-ACC did
'(They) exported cars to America with difficulty.'

Meanwhile, in terms of the external properties, VNs act like nominals. For example, the grammatical ACC case can optionally be attached to the VN, as observed in (1). In addition, the VN can assign the nominal case GEN to its argument(s):

(6) a. cek-uy mwuchapyelcekin tosi-uy kongkyok
enemy-GEN merciless city-GEN attack
'the enemy's merciless attack on the city'

b. John-uy Mary-wa-uy kyelhwon
John-GEN Mary-with-GEN marriage
'John's marriage with Mary'

A further nominal property can be observed from the fact that the phrase projected from a VN (such as *coseng* 'establishment') can function as the head of a relative clause construction:

(7) haksayngtul-ul topki-wihan [hakkwa-uy canghakkum-uy coseng]
student-ACC help-PUR department-GEN scholarship-GEN establishing
'the department's establishment of the scholarship to help students'

Though VNs display the mixed properties of nominals and verbals, this does not mean that they have the full distribution of canonical NPs or Ss. For example, we could observe that the phrases projected from a VN cannot be coordinated either with a canonical sentence or with an NP:

(8) a. $_{VNP}$[ku tanchey-ka pepan-ul ceyan] kuliko $_{VNP}$[kwukhoye-ka i-lul simuy]
organization-NOM the bill suggest and assembly it-ACC review
'The organization's suggesting the bill and the assembly's reviewing it'

b. *$_{VNP}$[ku tanchey-ka pepan-ul ceyan] kuliko $_{NP}$[kwukhoye-ka i-uy simuy]
organization-NOM the bill suggest and assembly it-GEN review

c. *$_{VNP}$[ku tanchey-ka pepan-ul ceyan] kuliko $_{S}$[kwukhoye-ka i-lul molassta]
organization-NOM the bill suggest and assembly it-ACC not.know
'The organization's suggesting the bill and the assembly didn't know this.'

Another main issue in the LVC comes from syntactic variations. It is well-observed that the VN in the true LVC has frozen effects: it does not undergo relativization, scrambling, clefting, and topicalization. The VN further cannot be wh-questioned or pronominlizaed:

(9) a. John-i Bill-eykey tocaki-lul **senmwul-ul** hayssta
John-NOM Bill-DAT china-ACC present-ACC did
'John gave a china to Bill as a present.'

b. *John-i Bill-eykey tocaki-lul han **senmwul** (relativization)

c. *John-i **senmwul-ul** Bill-eykey tocaki-lul hayssta. (scrambling)

d. *John-i Bill-eykey han kes-un **senmwul**-i-ta (clefting)

e. *John-i Bill-eykey **ku kes-ul** hayssni? (pronominalization)

f. *John-i Bill-eykey **mwues-ul** hayssni? (wh-question)

Intriguing facts emerge when the VN does not appear with the accusative object. In such cases, the frozen effects disappear: all these syntactic processes in (18) are possible as shown in the following:

(10) a. John-i Bill-eykey **senmwul-ul** hayssta
John-NOM Bill-DAT present-ACC did
'John gave a present to Bill.'

b. John-i Bill-eykey han **senmwul** (relativization)

c. John-i **senmwul-ul** Bill-eykey hayssta. (scrambling)

d. John-i Bill-eykey han kes-un **senmwul**-i-ta (clefting)

e. John-i Bill-eykey **ku kes-ul** hayssni? (pronominalization)

f. John-i Bill-eykey **mwues-ul** hayssni? (question)

The difference can be further observed with the usage of adverb or adjectival modification. With no ACC argument, the VN can be modified by an adjective:

(11) a. John-i Bill-eykey tocaki-lul *__caymiissnun__ senmwul-ul hayssta
John-NOM Bill-DAT chinaware-ACC interesting present-ACC did

b. John-i Bill-eykey **caymiissnun** senmwul-ul hayssta
John-NOM Bill-DAT interesting/interestingly present-ACC did
'John gave an interesting present to Bill.'

These clear differences raise the questions of 'are these two VNs and light verbs in (9) and (10) different from each other?' Should we differentiate the VNs with the accusative NP from those without it? Or should we differentiate the two instances of the light verb? Most of the literature, except a few, have accepted the view that sentences like (10) are not the LVC but the MVC (main verb construction). However, it is rather hard to claim that the *senmwul* with the ACC object in (11)a is different from the one without it in (11)b. In addition, it appears also nonintuitive to assume that the dative argument 'Bill-eykey' in (9) and (10) is different in each of these sentences.

There have been various attempts to account for these aforementioned properties of LVC constructions.[2] In what follows, we lay out a constraint-based analysis adopting the mechanism of multiple inheritance hierarchies that enables us to capture the mixed properties as well as other related ones in a much more streamlined manner.

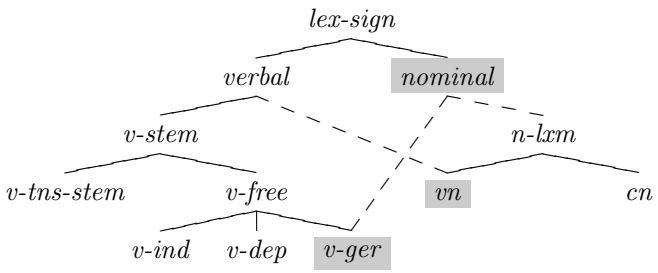
[2]See Ahn (1989), Chae (1996), Grimshaw and Mester (1988), Lapointe (1993), Manning (1993), Sells (1995), Choi and Wechlser (2001), and references cited therein.

2 A Typed Feature Structure Grammar: KPSG

2.1 Mixed Properties within a Multiple Inheritance System

Our grammar KPSG (Korean Phrase Structure Grammar), based on the framework of HPSG (head-driven phrase structure grammar), aims at building a computationally feasible Korean grammar with a comprehensive coverage. In the grammar, all the linguistic expressions are types of *sign* which in turn has *lex-sign* (lexical sign) and *syn-sign* (syntactic sign) as its subtypes. Following traditional Korean grammar, the KPSG takes the basic lexical categories of the grammar (*lex-sign*) to include *verbal*, *nominal*, *adverbial*, and *adnominal* as its subtypes which again are subclassified according to their properties. The following is a simplified hierarchy, representing the relevant part:[3]

(12)



The key point of capturing the mixed properties of VNs lies in the cross-classification and multiple inheritance mechanism.[4] As noticed in the hierarchy, the type *vn* is declared to be the subtype of both *verbal* and *n-lxm*, implying that it will inherit all the constraints of these supertypes. The type *verbal* is declared to have the value [V +] with a non-empty ARG-ST value, whereas *n-lxm* has the value [POS *noun*]. The inheritance mechanism will then ensure that the type *vn* has at least the following information:

(13)
$$\begin{bmatrix} vn & \\ \text{SYN} \mid \text{HEAD} & \begin{bmatrix} \text{POS } noun \\ \text{N } + \\ \text{V } + \end{bmatrix} \\ \text{ARG-ST } \langle [\,], \ldots \rangle & \\ \text{SEM } \ldots & \end{bmatrix}$$

[3]The dot line here means the existence of other types between the two types. The type glosses mean *v-ind(ependent), v-dep(endent), v-ger(undive).*

[4]The type *v-ger* is for gerundive verbs like *ilk-ess-um* 'read-PST-NMLZ' which also display mixed properties. See Kim and Yang (2004).

This lexical information will then be enriched when each lexical instance inherits all the relevant constraints from its supertypes:[5]

(14)

a.
$$\begin{bmatrix} \textit{vn-tr} \\ \text{PHON } \langle \text{kongpwu} \rangle \\ \text{SYN} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{POS } \textit{noun} \\ \text{V } + \\ \text{N } + \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST } \langle \text{NP}_i, \text{NP}_j \rangle \\ \text{SEM} \begin{bmatrix} \text{INDEX } \textit{s1} \\ \text{RELS} \left\langle \begin{bmatrix} \text{PRED } \textit{study-rel} \\ \text{ARG0 } \textit{s1} \\ \text{ARG1 } i \\ \text{ARG2 } j \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

b.
$$\begin{bmatrix} \textit{vn-ditr} \\ \text{PHON } \langle \text{swuchwul} \rangle \\ \text{SYN} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{POS } \textit{noun} \\ \text{V } + \\ \text{N } + \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST } \langle \text{NP}_i, \text{NP}_j, \text{NP}_k \rangle \\ \text{SEM} \begin{bmatrix} \text{INDEX } \textit{s1} \\ \text{RELS} \left\langle \begin{bmatrix} \text{PRED } \textit{export-rel} \\ \text{ARG0 } \textit{s1} \\ \text{ARG1 } i \\ \text{ARG2 } j \\ \text{ARG3 } k \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

As observed here, the system explicitly represents why VNs are in part nominal ([N +]) and are in part verbal ([V +]) though in terms of POS, they are more like nouns. In addition, by referring to a proper feature value, the grammar can be flexible enough to capture other related properties. For example, the KPSG allows an adverb to modify a [V +] element. This would then predict the adverb modification in the LVC we discussed in (2). In addition, since the type *vn* as a subtype of *n-stem* bears [N +] and [POS *noun*], it is expected that the VNs will act like other nominal elements: the VNs can have case markings attached to them, have the GEN grammatical case, and can serve as the head of a relative clause construction like the other [POS *noun*] elements.

2.2 Argument Composition and the Syntax of the LVC

Once we understand the basic properties of VN and the light verb, the next issue is the syntactic structure of the LVC: what allows the combination of the VN and the light verb, what are the results of the combination, and what kind of constraints exist in the combination? The KPSG we developed here posits a small set of well-formed syntactic combination rules such as Head-Subject Rule (XP $\rightarrow$ ZP X′), Head-Complement Rule (X $\rightarrow$ YP$*$ X), and Head-Modifier Rule (XP $\rightarrow$ Mod, XP$*$) as given in the following:[6]

[5]The semantics we represent here is a simplified version of a flat semantic formalism MRS (minimal recursion semantics). See Bender et al. 2002 and Copestake et al. 2003 for details.

[6]In the current version of the KPSG, the grammar rules include restrictions on the case values (i.e., nominative and accusative). The space does not allow us to explicate the discussion of case phenomena in the language. See

(15) a. Head-Subject Rule:
XP[*hd-subj-ph*] → 1, **H**[SUBJ ⟨1⟩]

b. Head-Complement Rule:
XP[*hd-comp-ph*] → 1, **H**[COMPS ⟨..., 1, ...⟩]

c. Head-Modifier Rule:
XP[*hd-mod-ph*] → [MOD ⟨1⟩], 1**H**

These simple rules can license major phrases in the language. The Head-Subject Rule, generating a *hd-subj-ph*, allows a VP to combine with its subject. The Head-Complement Rule ensures a head to combine with one of its COMPS(COMPLEMENTS) elements, forming a *hd-comp-ph*. The Head-Modifier Rule allows a head to form a well-formed phrase with an adverbial element that modifies the head, resulting in *hd-mod-ph*.[7]

To see how the system works, let us consider one simple sentence:

(16) John-i chayk-ul ilk-ess-ta
John-NOM book-ACC read-PST-DECL
'John read a book.'

The main verb *ilk-ess-ta* 'read-PST-DECL' takes two arguments which are in syntax realized as SUBJ and COMPS, respectively:[8]

(17) $\begin{bmatrix} \text{HEAD} \mid \text{POS } \textit{verb} \\ \text{VAL} \begin{bmatrix} \text{SUBJ } \langle 1 \rangle \\ \text{COMPS } \langle 2 \rangle \end{bmatrix} \\ \text{ARG-ST } \langle 1, 2 \rangle \end{bmatrix}$

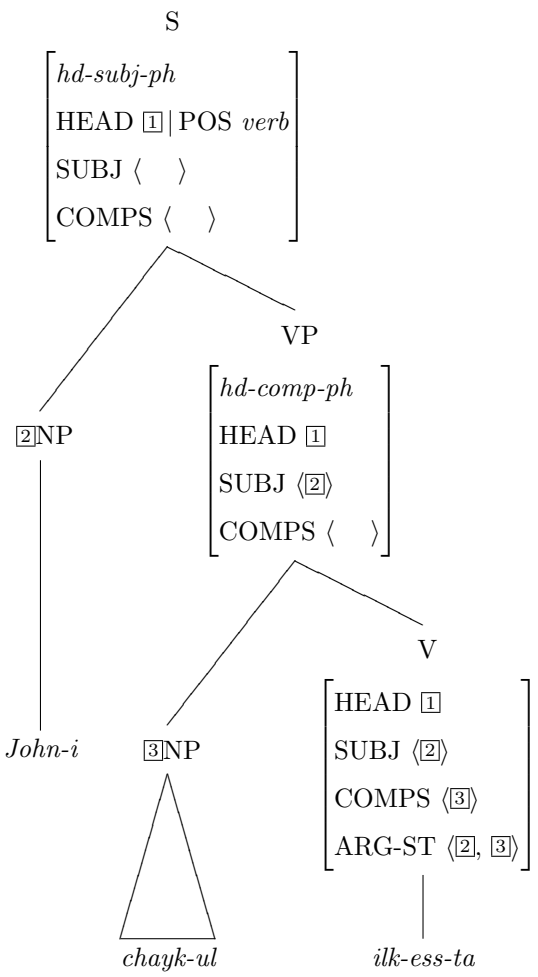
It is not difficult to see that the grammar rules can eventually generate a sentence like the following, projected from this lexical realization:

Kim (2004) for the analysis of Korean case phenomena.

[7]Note that the grammar rules here place no restriction on the SUBJ value: this allows the head to combine with the subject before combining with a complement. One great advantage of this is to allow sentential internal scrambling with no further operation or mechanism. See Kim and Yang (2003) for details.

[8]See Kim and Yang (2003).

(18)



The verb *ilk-ess-ta* 'read-PST-DECL' selects two arguments, each of which is realized as SUBJ and COMPS according to the Argument Realization Constraint that ensures the first argument be realized as SUBJ while the remaining ones as COMPS element (see Kim 2002 and Kim and Yang 2003). The head verb then combines with its COMPS *chayk-ul*, forming a well-formed *hd-comp-ph* in accordance with the Head-Complement Rule. The resulting VP then combines with the subject *John-i*, forming a *hd-subj-ph* licensed by the Head-Subject Rule.

The situation in the auxiliary verb construction (AVC) is different. Unlike canonical cases like (16), in the AVC, the main verb and the following auxiliary form a complex predicate, and the two further display a tight syntactic cohesion:

(19) John-i sakwa-ka/lul mek-ko (*cengmal) siph-ess-ta
John-NOM apple-NOM/ACC eat-COMP really would.like
'John would really like to eat apples.'

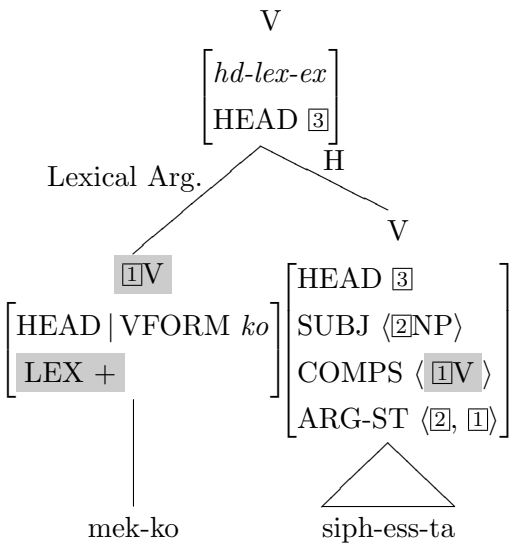
As argued and shown by Kim and Yang (2003), one effective way of capturing such complex predicate-like properties of the AVC is to introduce the Head-Lexical Rule given in (20):

(20) Head-Lexical Rule:

$$\begin{bmatrix} \textit{hd-lex-ex} \\ \text{COMPS } \boxed{A} \end{bmatrix} \rightarrow \boxed{1} \begin{bmatrix} \text{LEX } + \\ \text{COMPS } \boxed{A} \end{bmatrix}, \text{H} \begin{bmatrix} \text{AUX } + \\ \text{COMPS } \left\langle \boxed{1} \right\rangle \end{bmatrix}$$

The rule specifies that the auxiliary head combines with a lexical complement ($\boxed{1}$), and that to the resulting combination the COMPS value ($\boxed{A}$) of this lexical complement is composed.[9] This system, interacting with appropriate lexical entries for auxiliary verbs, will allow the following structure:

(21)



The auxiliary verb *siphessta* 'would-like' takes two arguments: one realized as subject and the other as a complement. When the auxiliary combines with the main verb, the result forms a *hd-lex-ph* and inherits the main verb's COMPS value in accordance to the rule in (20).

The LVC is not different from this AVC as we have seen: the light verb forms a complex predicate with the VN as in the following lexical entry:[10]

[9] This kind of argument composition is different from the previous analyses (cf. Bratt 1996, Chung 1998, Kim 2002), mainly in that the composition happens in syntax rather than in the lexicon.

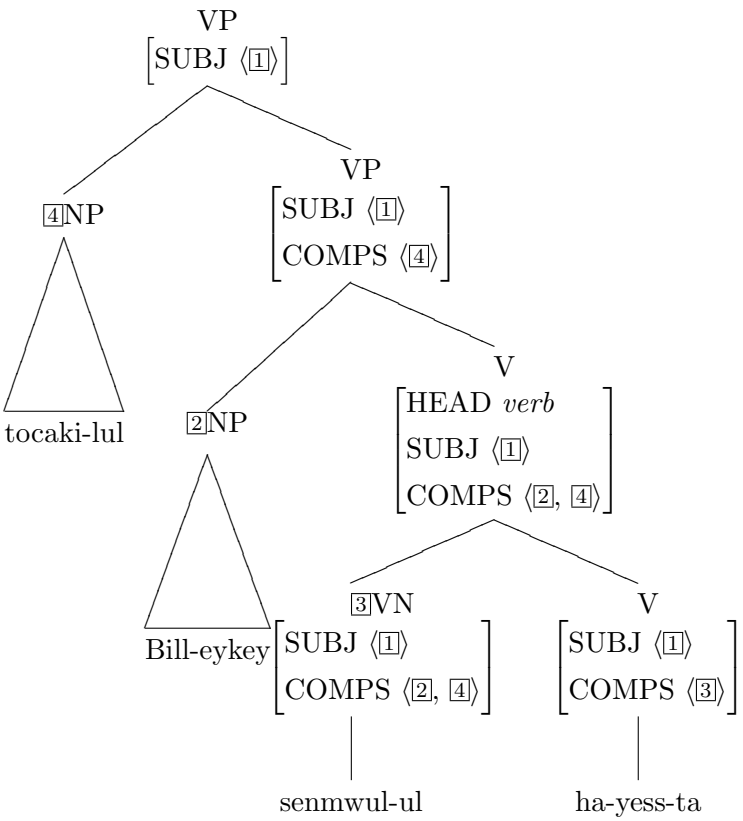
[10] The semantic attribute XARG relevant for equi and raising phenomena, identifies the semantic index of a phrase's external argument, usually the subject of a verb phrase. For example, the following would be the lexical entry for *seltukha-* 'persuade':

(i) $\left[\text{ARG-ST } \langle \text{NP, NP[INDEX } i\text{], VP[XARG } i\text{]} \rangle\right]$

(22) $\begin{bmatrix} \text{PHON } \langle \text{ha-ta} \rangle \text{ 'do'} \\ \text{SYN} \mid \text{HEAD} \mid \text{POS } \textit{verb} \\ \text{ARG-ST} \left\langle [\text{INDEX } i], \begin{bmatrix} \text{LEX } + \\ \text{XARG } i \end{bmatrix} \right\rangle \end{bmatrix}$

According to this lexical information, just like an auxiliary verb, the light verb is syntactically transitive, selecting a subject argument and a VN expression with the positive feature LEX. Since the external argument of the light verb is identical with the first argument, it in turn means the subject of the LVC is determined by the VN. The Head-Lexical Rule in (20) and the Head-Complement Rule in (15b) combined will then generate the following.

(23)



The VN *senmwul-ul* 'present' combines with the light verb *ha-yess-ta* in accordance with the Head-Lexical Rule. The resulting expression *senmwul-ul ha-yess-ta*, inheriting the COMPS value of

As noted, the XARG of the VP complement is identical with the object NP's index value. This attribute is visible for control of subject-unsaturated complements in the process of the semantic composition.

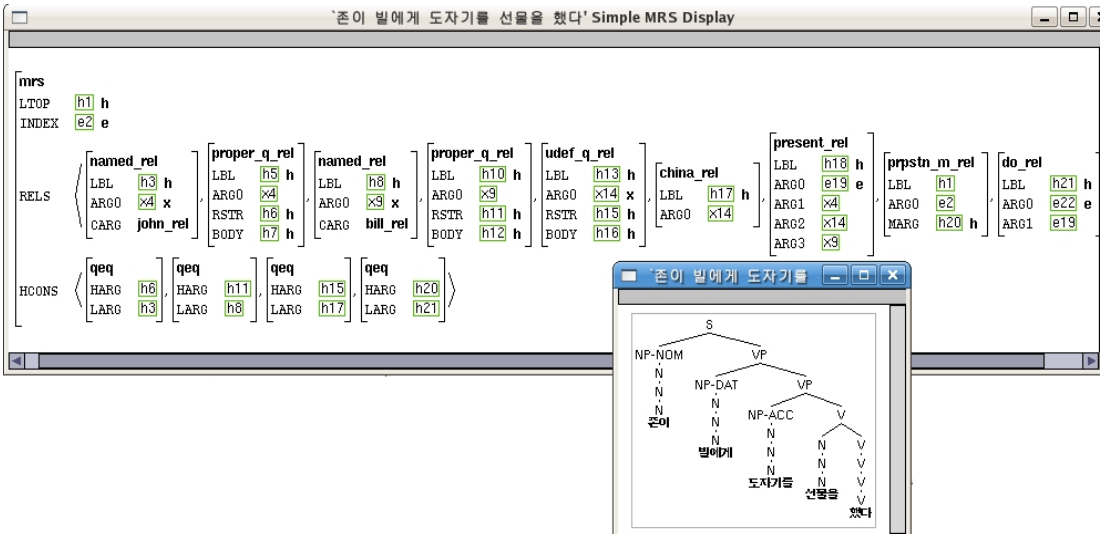


Figure 1: Parsed Tree and MRS for (9a)

the VN, then combines with the complement *Bill-eykey*. The Head-COMP rule then allows this resulting VP to again combine with the last complement *tocaki-lul*.

To check the feasibility of our grammar, we implemented this grammar in the LKB (Linguistic Knowledge Building) System (cf. Copestake 2002). The LKB system is a grammar and lexicon development environment for use with constraint-based linguistic formalisms such as HPSG.[11] Figure 1 is the parsed tree and semantic representation of sentences like (9a). The tree structure in the small box indicates that the light verb *hayssta* 'did' here combines with its VN complement *senmwul* 'present', forming a well-formed *hd-lex-ex*. This resulting combination also inherits the COMPS value of the VN in accordance with the Head-Lexical Rule in (20). This will then combines with the argument *tocaki* 'china' whose resulting VP again combines with the dative argument *Bill-eykey*.

The bigger window in Figure 1 represents the semantics of the sentence in the MRS (Minimal Recursion Semantics), developed by Copestake et al. (2003). The MRS is a framework of computational semantics designed to enable semantic composition using only the unification of type feature structures. We can observe that the parsed MRS provides enriched information of the sentence. The value of LTOP is the local top handle, the handle of the relation with the widest scope within the sentence. The INDEX value here is identified with the ARG0 value of the *prpstn_m_rel* (propositional message). The attribute RELS is basically a bag of elementary predications (EP)

[11]The LKB is freely available with open source (http://lingo.stanford.edu).

each of whose value is a *relation*.[12] Each of the types *relation* has at least three features LBL, PRED (represented here as a type), and ARG0. We can notice that the MRS correctly represents the propositional meaning such that John did the action of giving a china as a present to Bill. Observe that the EP *present_rel* in the RELS: it denotes an event e19 in which ARG1 (x4), ARG2 (x14), and ARG3 (x9) participate: x4 is linked to John, x14 to china, and x9 to Bill. The EP *do_rel* selects two arguments: Bill and the event *present_rel*. This indicates that Bill is involved in the event in which Bill is presenting a china to Bill.

2.3 VN as Common Noun Usages

VNs can also be used as common nouns when they take no ACC arguments. For example, the VN-like nouns in (24) are different from the argument-taking VNs even though they combine with the light verb.[13]

(24) a. John-i **kongpwu-ul** hayessta
John-NOM study-ACC did
'John studied.'

b. John-i Bill-eykey **senmwul-ul** hayssta
John-NOM Bill-DAT present-ACC did
'John did an introduction to Bill.'

Unlike the true VNs with the feature [N +, V +], these VNs are common nouns with the feature [N +, V −]. As noted in (10), they also can be modified by an adjectival element and they do not have frozen effects as VNs. In addition, even though they do not select an ACC argument, they still keep the dative argument *Bill-eykey*.

As we have seen in section 1, note that the verb *ha* here is different from the verb *ha* in the cases where the VN occur with its ACC object. Unlike the accusative example, all syntactic processes are possible, whose data repeated here again:

(25) a. John-i Bill-eykey **senmwul-ul** hayssta
John-NOM Bill-DAT present-ACC did
'John gave a present to Bill.'

b. John-i Bill-eykey han **senmwul** (relativization)

c. John-i **senmwul-ul** Bill-eykey hayssta. (scrambling)

[12]The attribute HCONS is to represent quantificational information. See Bender et al. 2002.
[13]All the VNs are selecting a subject and an argument which are realized as NOM and ACC.

d. John-i Bill-eykey han kes-un **senmwul**-i-ta (clefting)

e. John-i Bill-eykey **ku kes-ul** hayssni? (pronominalization)

f. John-i Bill-eykey **mwues-ul** hayssni? (question)

The VN in such cases can be modified by an adjective, whose data we repeat here:

(26) John-i Bill-eykey caymiissnun senmwul-ul hayssta
John-NOM Bill-DAT interesting/interestingly present-ACC did

A similar case can be found with VNs like *hapsek*:

(27) a. John-i Bill-kwa **hapsek-ul** hayssta
John-NOM Bill-with sitting-ACC did
'John sat with Bill.'

b. John-i Bill-kwa han **hapsek** (relativization)

c. John-i **hapsek-ul** Bill-kwa hayssta. (scrambling)

d. John-i Bill-kwa han kes-un **hapsek**-i-ta (clefting)

e. John-i Bill-kwa **mwues-ul** hayssni? (question)

Such examples give us reason to treat the verb *ha* here as a main verb and the VN as a canonical noun but not a verbal noun. If the verb *ha* is a main verb, the issue is then the number of its arguments. Does this verb select a dative argument like *Bill-eykey* which is obviously linked to the VN-like noun *senmwul*? In this paper we assume that the verb *ha* in this context selects two arguments as in the following example:

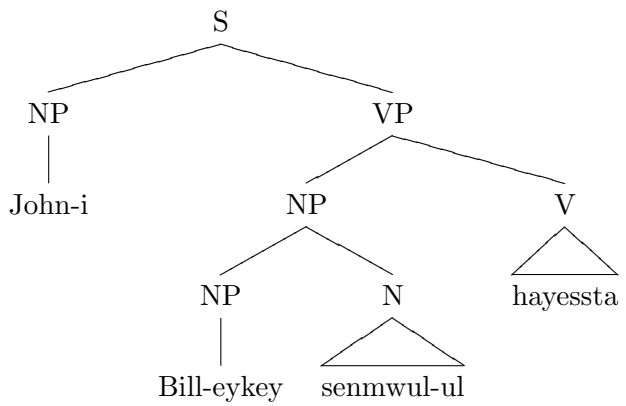
(28) John-i kongpwu-lul hayssta
John-NOM study-ACC did
'John did the action of study.'

The lexical entry for *ha-*, functioning as a main verb, will then look like the following:

(29) $\begin{bmatrix} \text{PHON } \langle \text{ha-ta} \rangle \text{ 'do'} \\ \text{SYN} \mid \text{HEAD} \mid \text{POS } \textit{verb} \\ \text{ARG-ST} \left\langle \text{NP, NP[COMPS } \langle \quad \rangle] \right\rangle \end{bmatrix}$

Notice that the second argument NP is a fully saturated NP with the empty COMPS value. This is to allow the non-ACC argument to combine with the VN, forming a full NP as represented in the following tree structure:

(30)



As given in the parsed tree, the N *senmwul* first combines with its DAT argument *Bill-eykey*. This happens because *senmwul* is no longer a [LEX +] expression. As we have noted before the noun *senmwul* here does not have a verbal property, but functions as a common noun, generated from the following lexical process:

(31) VN-to-CN Lexical Rule:

$$\begin{bmatrix} \textit{vn-ditr} \\ \text{ARG-ST } \langle \boxed{1}, [\quad] \rangle \oplus \boxed{A} \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} \textit{cn-vn} \\ \text{HEAD} \mid \text{V} - \\ \text{ARG-ST } \langle \boxed{1} \rangle \oplus \boxed{A} \end{bmatrix}$$

This lexical rule turns any di-transitive VNs selecting two or more arguments (including an ACC argument) into a canonical noun with the negative LEX value. In addition, the output has no verbal properties any more as indicated from the [V –] value. This lexical process will allow the following:

(32)

$$\begin{bmatrix} \textit{vn-tr} \\ \text{PHON } \langle \text{senmwul} \rangle \\ \text{SYN} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{POS } \textit{noun} \\ \text{V} + \\ \text{N} + \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST } \langle \text{NP}_i, \text{NP}_j, \text{NP}_k \rangle \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} \textit{cn-vn} \\ \text{PHON } \langle \text{senmwul} \rangle \\ \text{SYN} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{POS } \textit{noun} \\ \text{V} - \\ \text{N} + \end{bmatrix} \\ \text{LEX} - \end{bmatrix} \\ \text{ARG-ST } \langle \text{NP}_i, \text{NP}_k \rangle \end{bmatrix}$$

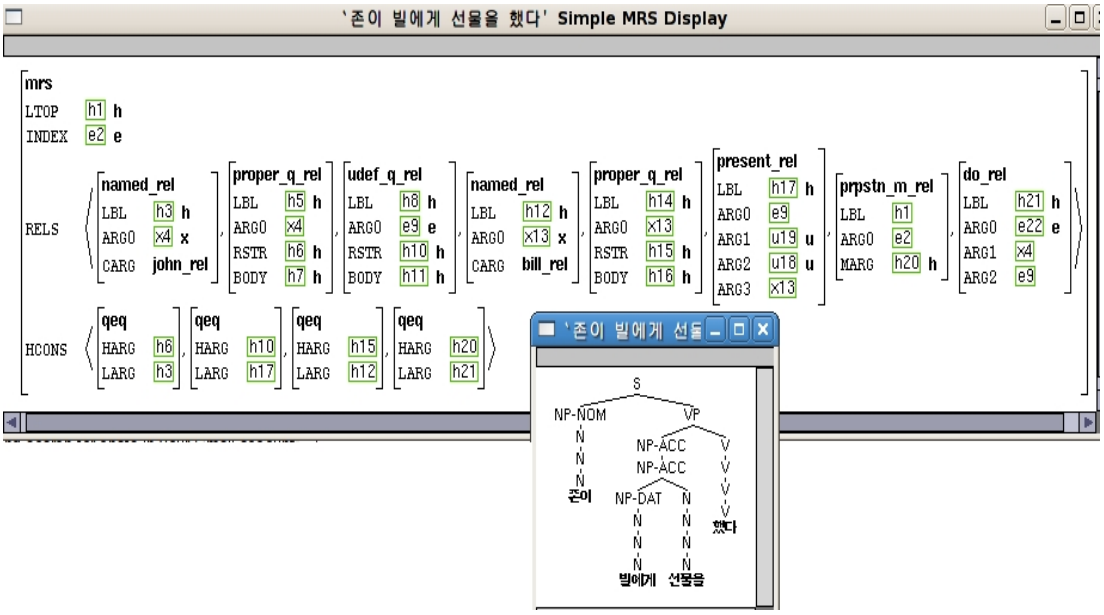


Figure 2: Parsed Tree and MRS for (25a)

As noted here, the output *vn-cn* has lost the verbal property and become [V –]. The output noun also has just two arguments, unlike the input verbal noun.

We also implemented this system in the LKB system, and produced Fig 2 as the parsing structures and meaning representation for the sentence (25a). As given here, the parsing results show us that the system generates correct tree structures with the proper meaning representations. Figure 2 represents that the meaning of this sentence is similar to that of (9a) given in Figure 1. The only difference is that the theme argument (referring to a china) is unbounded.

3 An Implementation and Its Results

In testing the performance and feasibility of the grammar, we first built up our test sets from (1) the SERI Test Suites '97, (2) the Sejong Project Basic Corpus, and (3) self-constructed examples adopted from the literature. The SERI Test Suites (Sung and Jang 1997), designed to evaluate the performance of Korean syntactic parsers, consists of total 472 sentences (292 test sentences representing the core phenomena of the language and 180 sentences representing different types of predicate). Meanwhile, the Sejong Corpus have about 2,061,977 word instances with 179,082 sentences. Of these, we found total 95,570 instances of the combination of a noun (tagged as NNG)

with the light verb *ha-ta*.[14] Some of the nouns with the higher frequency are given here:

5111 말/NNG+하/XSV 'speak' 3021 생각/NNG+하/XSV 'think'
1730 시작/NNG+하/XSV 'begin' 897 필요/NNG+하/XS 'need'
834 중요/XR+하/XSA 'important' 619 사용/NNG+하/XSV 'use'
543 주장/NNG+하/XSV 'claim' 528 시작/NNG+되/XSV 'begin'

Based on the frequency list, we first extracted the most frequently used 100 VNs, and from these VNs we selected 100 simple sentences (one from each VN type) that could show us at least the basic patterns of the LVC.

The following shows the results of parsing our test suits:

Corpus Types	# of S	# of Parsed S	# of LVC Ss	Parsed LVC Ss
SERI Test Suite	472	443 (93.7%)	12	12 (100 %)
Self-designed Test Suite	350	330 (94.2%)	100	94 (94 %)
Ss from the Sejong Corpus	179, 082		100	87 (87 %)
Total LVC Ss			212	190 (89%)

As the table shows, our system correctly parsed about 93 percent of the total 472 Seri Test Suite sentences which include those sentences that theoretical literature have often discussed. The system also parsed about 94% of the self-designed test sentences most of which are also collected from the major literature on the LVC. As for the Sejong corpus, the system parsed about 87% of the simple sentences from the Sejong corpus. Though there is need for extending this current grammar to the wider range of authentic corpus data that display more complex properties of the langauge, the parsing results indicate that the current grammatical system is feasible enough to capture the mixed properties and gives us the possibility of deep processing for such phenomena.

References

Ahn, Hee-Don. 1989. On Light Verb Constructions in Korean and Japanese. *Japanese/Korean Linguistics* 1:221-238.

Bender, E.M., Flickinger, D.P., Oepen, S. 2002. The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In Carroll, J., Oostdijk, N., Sutcliffe, R., eds.: Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics, Taipei, Taiwan. 8–14

[14]The Sejong Corpus thus does not distinguish general nouns from verbal nouns.

Bratt, Elizabeth. 1996. *Argument composition and the lexicon: Lexical and periphrastic causatives in Korean.* Ph.D. dissertation, Stanford University.

Cattell, Ray. 1984. *Composite Predicates in English.* North Ryde, New South Wales: Academic Press Australia.

Chae, Hee-Rahk. 1996. Verbal Nouns and Light Verbs in Korean. *Language Research*, 33: 581–600.

Choi, Incheol and Steve Weschler. 2001. Verbal Nouns and Light Verbs in Korean and Japanese. Paper Presented in HPSG-2001.

Chung, Chan. 1998. Argument Composition and Long Distance Scrambling in Korean: An Extension of the Complex Predicate Analysis. In Hinriches et al. (eds.), *Complex Predicates in Nonderivational Syntax*, 159–220. New York: Academic Press.

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars.* CSLI Publications.

Copestake, A., Flickenger, D., Sag, I., Pollard, C. 2003. Minimal recursion semantics: An introduction. Manuscript.

Dubinsky, Stanley. 1997. Syntactic underspecification and light phenomena in Japanese. *Linguistics* 35, 627–672.

Grimshaw, Jane, and Mester, Armin. 1988. Light Verbs and (Theta)-Marking. *Linguistic Inquiry* 19:205-232.

Kim, Jong-Bok. 2004. Korean Phrase Structure Grammar. Hankwuk Publishing, Seoul

Kim, Jong-Bok and Jaehyung Yang. 2003. Korean Phrase Structure Grammar and Implementing it into the LKB System (In Korean). *Korean Linguistics* 21: 1–41.

Kim, Jong-Bok and Jaehyung Yang. 2004. Projections from Morphology to Syntax in the Korean Resource Grammar: Implementing Typed Feature Structures. In *Lecture Notes in Computer Science* Vol 2945, pp 13-24, Springer-Verlag.

Lapointe, Steven. 1993. Dual Lexical Categories and the Syntax of Mixed Category Phrases. In A. Kathol & M. Bernstein (eds.), *Proceedings of ESCOL*, 199-210.

Malouf, Robert. 1998. *Mixed Categories in the Hierarchical Lexicon.* Stanford: CSLI Publications.

Matsumoto, Yo. 1996. A Syntactic Account of Light Verb Phenomena in Japanese. *Journal of East Asian Linguistics.* 5.2: 107–149.

Manning, Christopher D. 1993. Analyzing the Verbal Noun: Internal and External Constraints. *Japanese/Korean Linguistics* 3:236-253.

Miyagawa, Shigeru. 1989. Light Verbs and the Ergative Hypothesis. *Linguistic Inquiry* 20:659-668.

Park, Kabyong. 1995. Verbal Nouns and Do-Insertion. Studies in Generative Grammar 5.1:319-358.

Siegel, Melanie and Emily M. Bender. 2002. Efficient Deep Processing of Japanese. In *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization.* Coling 2002 Post-Conference Workshop. Taipei, Taiwan.

Sells, Peter. 1995. The category and case marking properties of Verbal Nouns in Korean. *Harvard Studies in Korean Linguistics VI*, 370-386.

Sung, Won-Kyung and Myung-Gil Jang. 1997. SERI Test Suites '95. In *Proceedings of the Conference on Hanguel and Korean Language Information Processing.*