

Capturing and Parsing the Mixed Properties of Light Verb Constructions in a Typed Feature Structure Grammar

Jong-Bok Kim
Kyung Hee University
School of English
jongbok@khu.ac.kr

Jaehyung Yang
Kangnam University
School of Computer Engineering
jhyang@kangnam.ac.kr

Incheol Choi
Kyung Hee University
Language Institute
incheol@khu.ac.kr

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 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. The paper then restates the argument composition properties of the LVC and reinforces them with a constraint-based analysis. This paper also offers answers to the the puzzling syntactic variations in the LVC. Following these empirical and theoretical discussions is a short report on the implementation of the analysis within the LKB (Linguistics Knowledge Building) system.

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* carries is that it does not affect the argument structure of the VN it combines with.

- (1) a. John-i Mary-eykey cenhwa(-lul hayessta)
John-NOM Mary-DAT phone-ACC did
'John phoned Mary.'
- b. John-i Mary-lul myengt'an-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 that decides the types of arguments in the given sentence. This has led the literature to view that the light verb has no argument structure on its own and 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-ACC did
'The arrow marked the target.'

The data imply that the light verb is thematically underspecified with the subject role which in turn is thematically identical with that of the light verb.

The second main issue we have in the analysis of the LVC is 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 are just like nouns. It is not difficult to find out that VNs exhibit verbal properties. 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 support the verbal properties of VNs: the VN can be modified by an adverb but not by an adjectival element.

- (3) ku ceypwum-ul mikwuk-ey **elyepkey**/*elyewun swuchwul(-ul hayessta)
 the item-ACC America-LOC hard/difficult export-ACC did
 '(They) exported the item 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 (2). In addition, the VN can assign the nominal case GEN to its argument(s):

- (4) 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:

- (5) 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'

Another main issue in the LVC concerns 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 pronominalized:

- (6) 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, we observe no such frozen effects. All the syntactic processes are allowed:

- (7) 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:

- (8) 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

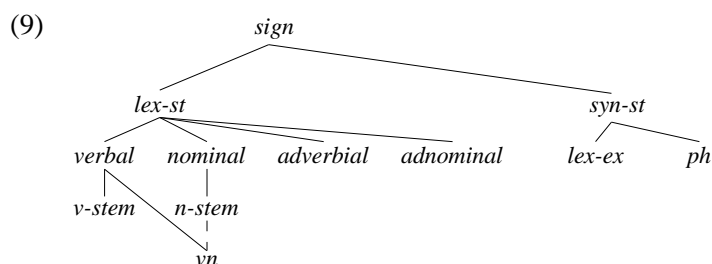
These clear differences raise the questions of ‘are these VNs or the light verbs different?’ Should we differentiate the VNs with the accusative NP from those without it? Or should we differentiate the two instances of the light verb? Most of the literature, except a few, have accepted the view that sentences like (7) 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 (8)a is different from the one without it in (8)b. In addition, it appears also nonintuitive to assume that the dative argument ‘Bill-eykey’ here is different in each of these sentences.

There have been various attempts to account for the properties of argument composition, mixed properties, and syntactic variations, but not at a satisfactory level.¹ In what follows, we lay out a more viable analysis that invokes the mechanism of multiple inheritance system that enables us to capture the mixed properties as well as other related ones in a much more streamlined manner.

2 A Type-Feature Structure Grammar: KPSG

2.1 Mixed Properties within a Multiple Inheritance System

Within the KPSG (Korean Phrase Structure Grammar) we develop here to build a computationally feasible grammar with a comprehensive coverage of Korean, all the linguistic expressions are types of *sign* which in turn has *lex-st* (lexical structure) and *syn-st* (syntactic structure) as its subtypes. Following the traditional wisdom, the KPSG takes the basic lexical categories of the grammar (*lex-st*) 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:²



The key point of capturing the mixed properties of VNs here comes from the cross-classification and multiple inheritance mechanism. As we can notice here in the hierarchy, the type *vn* is declared to be the subtype of both *verbal* and *n-stem*. This in turn means that the type *vn* will inherit all the constraints of these supertypes some of which are given in the following:

- (10) a. *nominal*: [NOMINAL +]
 b. *verbal*: $\left[\begin{array}{l} \text{VERBAL +} \\ \text{ARG-ST} \langle [\dots] \rangle \end{array} \right]$

¹See Ahn (1989), Chae (1996), Grimshaw and Mester (1988), Lapointe (1993), Manning (1993), Sells (1995), Choi and Wechsler (2001), and references cited therein.

²The dot line here means the existence of other types between the two types. For example, there exist types such as *n-p-stem* and *n-xdel-stem*, each of which have particles attached to. For a more elaborated type hierarchy system, see Kim and Yang (2003).

- c. *v-stem*: [POS *verb*]
- d. *n-stem*: [POS *noun*]

Our grammar, differing from standard approaches, thus introduces three HEAD features POS, VERBAL, and NOMINAL. All instances of the type *verbal* will have the feature [VERBAL +] and have non-empty ARG-ST values. In addition, the grammar assigns [POS *verb*] to the type *v-stem* and [POS *noun*] to the type *n-stem*. Within these constraint specifications, the inheritance mechanism will eventually ensure the instances of *vn* to have at least the following information inherited:

$$(11) \quad vn \rightarrow \left[\begin{array}{l} \text{SYN | HEAD} \left[\begin{array}{l} \text{POS } \mathit{noun} \\ \text{NOMINAL } + \\ \text{VERBAL } + \end{array} \right] \\ \text{ARG-ST } \langle [] , \dots \rangle \\ \text{SEM } \dots \end{array} \right]$$

For lexical entries we need to specify only the minimally required information on the VNs, as exemplified in (12):

$$(12) \quad \text{a. } \left[\begin{array}{l} \mathit{vn} \\ \text{PHON } \langle \mathit{kongpwu} \rangle \\ \text{ARG-ST } \langle \text{NP}, \text{NP} \rangle \\ \text{SEM } \mathit{study-rel} \end{array} \right] \quad \text{b. } \left[\begin{array}{l} \mathit{vn} \\ \text{PHON } \langle \mathit{swuchwul} \rangle \\ \text{ARG-ST } \langle \text{NP}, \text{NP}, \text{NP} \rangle \\ \text{SEM } \mathit{export-rel} \end{array} \right]$$

Such simple lexical information encoded in the lexicon will then be enriched as in (13) when the type *vn* inherits all the relevant constraints from its supertypes:

$$(13) \quad \text{a. } \left[\begin{array}{l} \mathit{vn} \\ \text{PHON } \langle \mathit{kongpwu} \rangle \\ \text{SYN} \left[\text{HEAD} \left[\begin{array}{l} \text{POS } \mathit{noun} \\ \text{VERBAL } + \\ \text{NOMINAL } + \end{array} \right] \right] \\ \text{ARG-ST } \langle \text{NP}_i, \text{NP}_j \rangle \\ \text{MODE } \mathit{prop} \\ \text{INDEX } \mathit{s1} \\ \text{SEM} \left[\text{RESTR} \left\langle \left[\begin{array}{l} \text{RELN } \mathit{study-rel} \\ \text{ARG0 } \mathit{s1} \\ \text{ARG1 } \mathit{i} \\ \text{ARG2 } \mathit{j} \end{array} \right] \right\rangle \right] \end{array} \right] \quad \text{b. } \left[\begin{array}{l} \mathit{vn} \\ \text{PHON } \langle \mathit{swuchwul} \rangle \\ \text{SYN} \left[\text{HEAD} \left[\begin{array}{l} \text{POS } \mathit{noun} \\ \text{VERBAL } + \\ \text{NOMINAL } + \end{array} \right] \right] \\ \text{ARG-ST } \langle \text{NP}_i, \text{NP}_j, \text{NP}_k \rangle \\ \text{MODE } \mathit{prop} \\ \text{INDEX } \mathit{s1} \\ \text{SEM} \left[\text{RESTR} \left\langle \left[\begin{array}{l} \text{RELN } \mathit{export-rel} \\ \text{ARG0 } \mathit{s1} \\ \text{ARG1 } \mathit{i} \\ \text{ARG2 } \mathit{j} \\ \text{ARG3 } \mathit{k} \end{array} \right] \right\rangle \right] \end{array} \right]$$

We can easily notice that this system clearly represents why VNs are in part nominal ([NOMINAL +]) and are in part verbal ([VERBAL +]) 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 VERBAL element. This would then predict the adverb modification in the LVC we discussed earlier. In addition, since the type *vn* as a subtype of *n-stem* bears [NOMINAL +] and [POS *noun*], we naturally predict 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*] element.³

³Note that a phrase projected from a gerundive verb cannot serve as the head of a relative phrase since it is still *verb* though it both has [NOMINAL +] and [VERBAL +]. See Kim and Yang (2003, 2004).

2.2 Argument Composition and the Syntax of the LVC

2.2.1 Three Types of Head-Complement Combination

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-Subj Rule ($XP \rightarrow ZP, X'$), Head-Comp Rule ($X' \rightarrow YP^* X$), and Head-MOD Rule, similar to the traditional X' system. Of these, the Head-Comp Rule can have three different types:

- (14) a. *chayk-ul sangca-ey nehessta*
 book-ACC box-LOC put
 'placed the books in the box'
 b. *sakwa-lul mek-ko sipessta*
 apple-ACC eat-COMP would.like
 'wanted to eat apples.'
 c. *mwillihak-ul kongpwu-lul hayessta*
 physics-ACC study-ACC did
 'studied physics'

The main verb *nehessta* 'put' in (14)a takes three arguments which are in syntax realized as SUBJ and COMPS, respectively, according to the Argument Realization Constraint:⁴

- (15)
$$\left[\begin{array}{l} \text{HEAD} \mid \text{POS verb} \\ \text{VAL} \left[\begin{array}{l} \text{SUBJ} \langle \boxed{1} \rangle \\ \text{COMPS} \langle \boxed{2}, \boxed{3} \rangle \end{array} \right] \\ \text{ARG-ST} \langle \boxed{1}, \boxed{2}, \boxed{3} \rangle \end{array} \right]$$

The Head-Comp Rule given in (16), generating a *hd-comp-ph*, licenses the combination of the head verb *nehessta* 'put' with one of its complements:⁵

- (16) Head-Comp Rule:

$$hd\text{-}comp\text{-}ph \Rightarrow \boxed{1}[\text{LITE } -], \mathbf{H}[\text{COMPS} \langle \dots, \boxed{1}, \dots \rangle]$$

The rule simply says that when a head combines with one of its complements, the resulting phrase is a well-formed phrase.⁶

The situation in the auxiliary verb construction (AVC) in (14)b is different. Unlike canonical cases (14)a, in the AVC the main verb and the following auxiliary form a complex predicate, and the two further display a tight syntactic cohesion:

- (17) *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-Lex Rule given in (18):

- (18) Head-Lex Rule:

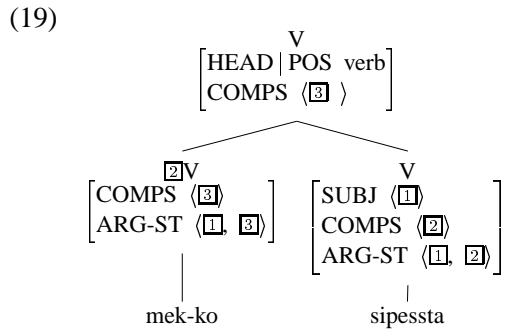
$$\left[\begin{array}{l} hd\text{-}lex\text{-}ex \\ \text{COMPS} \langle \boxed{A} \rangle \end{array} \right] \rightarrow \boxed{1} \left[\begin{array}{l} \text{LEX} + \\ \text{COMPS} \langle \boxed{A} \rangle \end{array} \right], \mathbf{H} \left[\begin{array}{l} \text{AUX} + \\ \text{COMPS} \langle \boxed{1} \rangle \end{array} \right]$$

⁴See Kim (2004).

⁵See the following section for the role of the feature LITE which is assigned to the VNs.

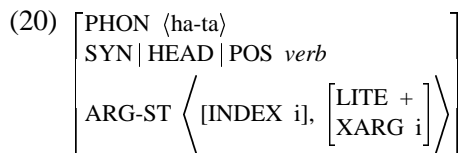
⁶Note that this rule places no restriction on the SUBJ value: this allows the head to combine with the subject before combining with a complement. One great advantage of this is to allow sentential internal scrambling with no further operation or mechanism. Kim and Yang (2003).

The rule specifies that the auxiliary head combines with a [LEX +] complement⁷, and that to the resulting combination the COMPS value of this lexical complement is passed up.⁸ This rule, interacting with appropriate lexical entries for auxiliary verbs, will allow the following structure:

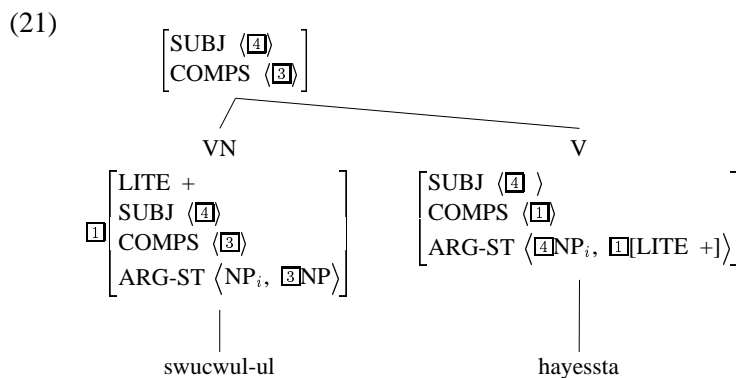


The auxiliary verb *sipessta* ‘would-like’ takes two arguments: one realized as subject and the other as a complement. When the auxiliary combines with the main verb, the result forms a *hd-lex-ex* and inherits the main verb’s COMPS value in accordance to the rule in (18).

In Section 1, we discussed the argument composition properties between the VN and the following light verb. This can lead us to take the light verb to form a complex predicate with the VN with the following lexical entry for the former:⁹



According to this lexical information, just like an auxiliary verb, the light verb is syntactically transitive, selecting a subject argument and a VN expression (lexical or phrasal) with the positive feature LITE whose main functions we will discuss in due course. 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-Lex Rule in (18) will then generate the following:



⁷The feature LEX is assigned to all words and a complex predicate.

⁸This 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.

⁹The 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) [ARG-ST ⟨NP, NP[INDEX i], VP[XARG i]⟩]

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.

However, issues arise from the fact that there also exist crucial differences from the AVC. That is, though there is a tight syntactic cohesion between the main verb and the auxiliary verb, the VN and the light verb are rather loose in terms of syntax. The LVC, even though forming a complex predicate with the preceding predicate, does have no such tight syntactic coherence:

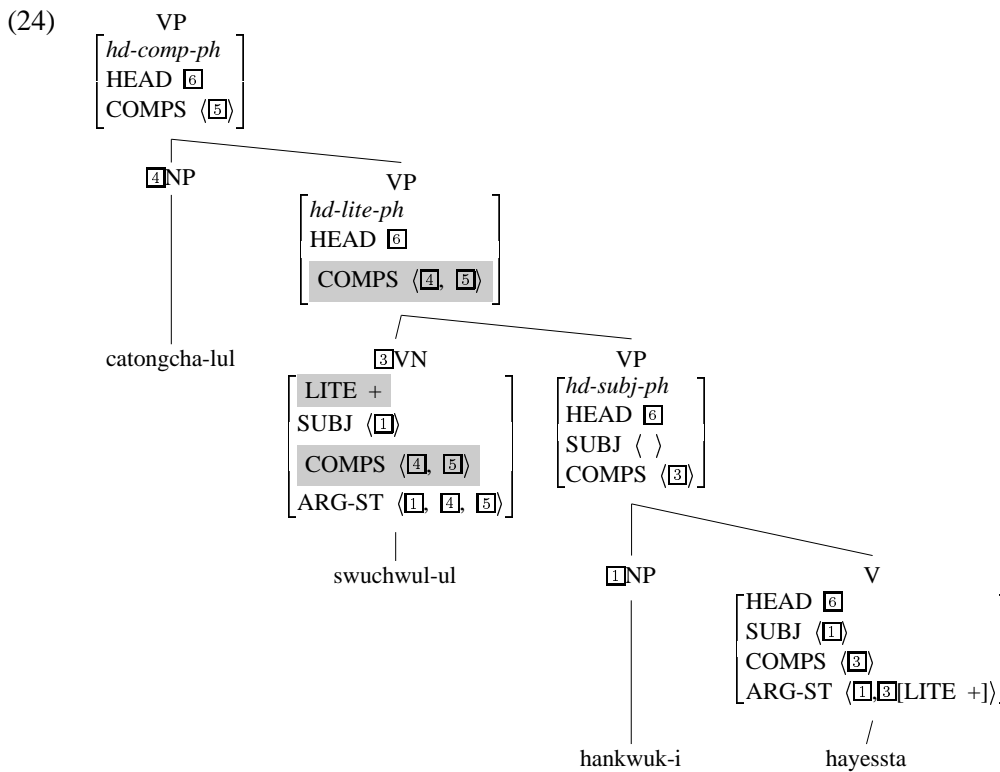
- (22) a. catongcha-lul mikwuk-ey swuchwul-ul **hankwuk-i** hayessta
cars-ACC America-LOC export-ACC Korea did
‘Korea exported cars to America.’
b. hankwuk-i catongcha-lul mikwuk-ey swuchwul-ul **cengmal manhi** hayessta
Korea-NOM cars-ACC America-LOC export-ACC really many did
‘Korea exported really a lot of cars to America.’

As noted here, it is acceptable to have the subject or an adverb between the VN and the light verb. This asks us to allow argument composition at a phrasal level. To accommodate this main difference, we introduce a similar grammar rule as in the following:¹⁰

(23) Head-Lite Rule:

$$\left[\begin{array}{l} hd\text{-}lite\text{-}ph \\ COMPS \langle A \rangle \end{array} \right] \rightarrow \left[\begin{array}{l} LITE + \\ COMPS \langle A_{nelist} \rangle \end{array} \right], H \left[COMPS \langle \square \rangle \right]$$

The Head-Lite Rule, generating a well-formed *hd-lite-ph*, specifies that a head (lexical or phrasal) element combines with a complement that bears the head feature LITE. This feature in a sense differentiates the VNs in the true LVC from common nouns that do not undergo argument composition. Also, just like the Head-Lex Rule, the COMPS value of this lexical light complement is passed up to the resulting mother, reflecting the argument composition properties. Notice that the grammar will then allow structures like the following for (22)a where the phrase projected from the VN is combined at a later level.



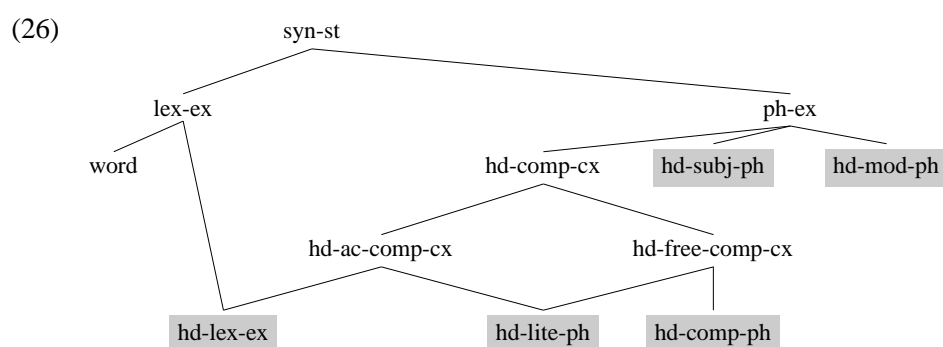
¹⁰The type *nelist* means it is not an empty list.

The light verb here combines with its subject first and forms a phrasal level. This phrase then combines with the VN. Since the Head-Lite Rule allows a head to combine with a LITE element and then inherit its COMPS value to the resulting value, this structure is a well-formed structure.¹¹

The rule in (23) requires its nonhead daughter to have a nonempty COMPS list. This basically rules out the possibility of combining the VN with its COMPS argument before combining with the verb. Thus for example, for the sentence (14c) we will have only the structure (25)b, but not (25)a in which the VN has an empty COMPS value.

- (25) a. [[mwullihak-ul kongpwu-lul] hayessta].
 b. [mwullihak-ul [kongpwu-lul hayessta]].

Even though we posit three different rules for the combination of a head and a complement, this does not mean that they are unrelated or we need to specify these each, missing generalizations. The close membership of these three rules is once again captured by the mechanism of multiple inheritance hierarchy:¹²



For example, within the hierarchy, the types *word* and *hd-lex-ex* are identical in the respect that they all bear the feature [LEX +]; the type *hd-lex-ex* and *hd-lite-ph* share the constraints on argument composition. Filtering out the common properties, the grammar then just requires to specify the minimal constraint on each terminal subtype. This modular of grammar allows us to tease out the differences between the three different types of syntactic phrases, while capturing the similarities in a precise way.

2.3 Dissolving Puzzles in the Syntactic Variations

The grammar we set forth in the previous section can provide a simple way of capturing the similarities and differences we find in the syntactic variations we noted in Section 1.

The first step we take is to accept that the VNs with no ACC argument are not VNs but common nouns (are not subtypes of *verbal*). For example, the VN-like nouns in (27) are different from the argument-taking VNs even though they combine with the light verb.¹³

- (27) a. John-i **kongpwu-ul** hayessta
 John-NOM study-ACC did
 ‘John studied.’

¹¹This analysis predicts that it is most natural to have the subject in this position. To most speakers having elements other than the subject here is rather unnatural, as expected:

- (i) a. ??/*swucwul-ul hankwuk-i catongca-lul mikwuk-ey hayessta.
 b. ??/*hankwuk-i swucwul-ul catongca-lul mikwuk-ey hayessta.
 c. ??/*hankwuk-i catongca-lul swucwul-ul mikwuk-ey hayessta.

A stress on the element in the post light verb position improves the acceptability.

¹²This system can minimize the burden of specifying redundant constraints. For example, the type *hd-lite-ph* generated by the Head-Lite Rule just needs the constraint that the nonhead is [LITE +]. All the other constraints are inherited from its supertypes.

¹³All the VNs are selecting a subject and an argument which are realized as NOM and ACC.

- b. John-i Bill-eykey **sokay-lul** hayssta
 John-NOM Bill-DAT introduction-ACC did
 ‘John did an introduction to Bill.’

These VNs are common nouns with the feature [NOMINAL +, VERBAL –], unlike the true VNs [NOMINAL +, VERBAL +]. However, this doesn’t mean that they are unrelated. They are closely linked by the feature LITE at least and are further derived from VNs from the following lexical rule:

(28) VN-to-CN Lexical Rule:

$$\left[\begin{array}{l} vn-tr \\ LITE + \\ ARG-ST \langle []_1, [] \rangle \oplus [A] \end{array} \right] \rightarrow \left[\begin{array}{l} cn-nonst-lite \\ LITE + \\ ARG-ST [A] \end{array} \right]$$

This lexical rule turns any transitive VNs selecting two or more arguments into CNs with no change in the meaning. This in turn means that as for the classification of nominal elements, we assume that there are at least two types of common nouns those with the LITE and those without the LITE feature:

- (29) a. *cn*: *cn-gen*, *cn-lite*
 b. *cn-lite*: *cn-st-lite*, *cn-nonst-lite*
- (30) a. *cn-st-lite*: kwumyon ‘diligence’ kenkang ‘healthy’ chimchak ‘calmness’,.....
 b. *cn-nonst-lite*: tayhwa ‘talk’, sokay ‘introduction’, senmwul ‘present’, koychoi ‘held’,...

The type *cn-lite* is the nouns that can combine with the light verbs.¹⁴ Though the stative VNs are classified as common nouns from the beginning, all the nonstative VNs are generated from the lexical rule.¹⁵ The following illustrates an example of this lexical process:

$$(31) \left[\begin{array}{l} vn-tr \\ PHON \langle sokay \rangle \\ SYN \left[\begin{array}{l} HEAD \left[\begin{array}{l} POS \textit{ noun} \\ VERBAL + \\ NOMINAL + \\ LITE + \end{array} \right] \end{array} \right] \\ ARG-ST \langle NP_i, NP_j, NP_k \rangle \\ SEM \left[\begin{array}{l} MODE \textit{ prop} \\ INDEX \textit{ s1} \\ XARG \textit{ i} \end{array} \right] \end{array} \right] \rightarrow \left[\begin{array}{l} cn \\ PHON \langle sokay \rangle \\ SYN \left[\begin{array}{l} HEAD \left[\begin{array}{l} POS \textit{ noun} \\ VERBAL - \\ NOMINAL + \\ LITE + \end{array} \right] \end{array} \right] \\ ARG-ST \langle NP_k \rangle \\ SEM \left[\begin{array}{l} MODE \textit{ prop} \\ INDEX \textit{ s1} \\ XARG \textit{ i} \end{array} \right] \end{array} \right]$$

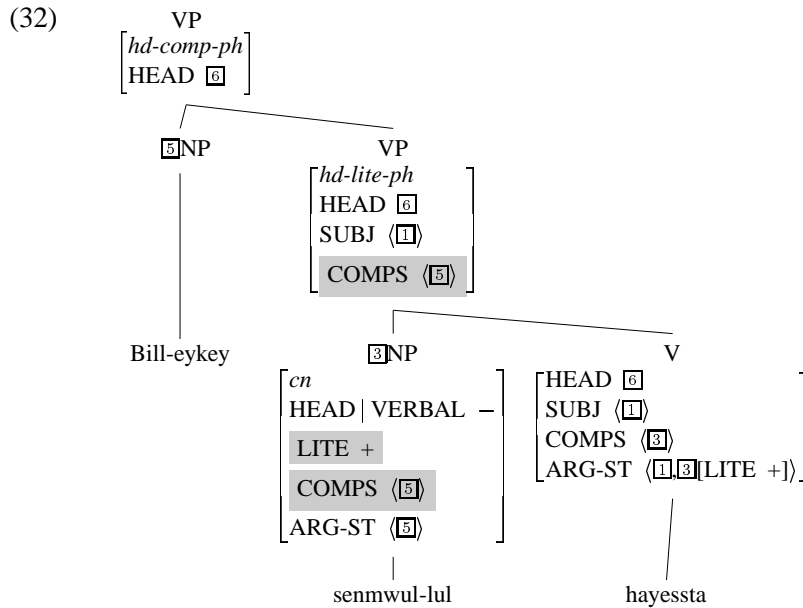
As noted, the verbal noun is lexicalized as a common noun with the oblique argument, still keeping the LITE feature. One thing to note here is that even though the output is a common noun, there is no semantic type change. This in a sense means that the output functions as a kind of predicative noun.¹⁶

Given this, the differences in the syntactic processes we observed in Section 1 are simply the matter of the general assumption that the syntactic processes are true of just for the common nouns with the features [NOMINAL +, VERBAL –]. VNs are [VERBAL +], and thus the phrases projected from VNs cannot undergo any of these syntactic processes. In addition, this simple explication, the analysis, allows argument composition for the common nouns with the feature LITE. Consider the structure of (7) our analysis generates:

¹⁴For example, nouns like *chayksang* ‘desk’ is a member of *cn-gen*, but not *cn-lite*.

¹⁵In general the argument elements of all the nominal elements will be mapped onto COMPS. We could make this as a constraint on the n-word.

¹⁶Even though the XARG value of the output will remain unanchored, the restriction placed on the light verb will ensure the subject of the light verb is identical to the noun’s.



Even though the noun *sokay* ‘introduction’ here is a common noun, it has [LITE +] feature and a dative argument. In accordance with the Head-Lite Rule, when the light verb combines with *sokay*, its COMPS value is inherited to the resulting phrase here.

One additional welcoming consequence is that there is no need to posit another lexical entry for the light verb *ha-ta*. Its lexical entry in (20) is enough for its combination with a common noun NP.

- (33) a. John-i [kongpwu-ul [hayessta]] ‘John studied’
 b. John-i [kenkang-un [hayessta]] ‘John is healthy’

As noted in (20), the verb *hata* requires its second argument just to be [LITE +]. The nouns *kongpwu* and *kenkang* here are all [LITE +]. The only difference is that the combination of this is just generated as a *hd-lex-ph*, but not as a *hd-lite-ph* since the VN here has an empty COMPS list.¹⁷

As set forth here so far, the grammar does not posit two lexical entries for the LITE verbal nouns and LITE common nouns. They are related by the lexical rule. One further consequence is that there is no need to posit two lexical entries for the light verb *hata* even if it combines with a common noun, given that the type *cn-lite* nouns capture LITE.

2.4 Sum

The grammar we have developed here have teased out the differences in the combination between head and complement, and then introduced the feature LITE to allow the argument composition at a phrasal level. The grammar we set up here in a sense can say that the light verb is a verb that is semantically light in the sense that it bears no significant semantics, but also light in terms of syntax in the sense that it combines with a light VN element and forms a complex predicate. This allowed us to build up a simpler grammar, while capturing rather unpredictable differences and similarities in the constructions with verbal nouns.

3 A Computational Implementation of the Grammar

3.1 Building Up a Test Suite from Corpora

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

¹⁷The verb *hata* inherits the agentivity property of *kenkang* here and thus lose the ability of assigning any case value here, just like the copula verb *i-ta*.

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*.¹⁸ 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.

3.2 Implementing the Grammar in the LKB

With these base test sentences, we further formalized the analysis of the LVC sketched in the previous section and implemented it into the LKB system. As noted earlier, we have developed the KPSG (Korean Phrase Structure Grammar) in a larger context, aiming for an open source grammar of Korean.¹⁹ The KPSG, whose basic grammatical framework is HPSG (Head-driven Phrase Structure Grammar), adopts the mechanism of type hierarchy in which every linguistic sign is typed with appropriate constraints and hierarchically organized. The characteristic of such typed feature structure formalisms facilitates the extension of grammar in a systematic and efficient way, resulting in linguistically precise and theoretically motivated descriptions of languages including Korean. Reflecting this, the KPSG thus consists of grammar rules, inflection rules, lexical rules, type definitions, and lexicon. All the linguistic information is represented in terms of signs whose main subtypes include *lex-st* and *syn-st* as we have seen before. It has a broad syntactic coverage and adopts a flat semantic formalism Minimal Recursion Semantics (MRS) in representing semantics (Copestake et al. 2001).

As the basic tool for writing, testing, and processing the KPSG, we adopt the LKB (Linguistic Knowledge Building) system (Copestake 2002). The LKB system is a grammar and lexicon development environment for use with constraint-based linguistic formalisms such as HPSG.²⁰

3.3 Results

The following shows the resulting of parsing our test suits:

Corpus Types	# of S	# of Parsed S	# of Test LVC Ss	Parsed LVC Ss
(34) 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 to the wider range of authentic corpus data that display more complex properties of the language, the parsing results indicate that the current system is rather robust in terms of the possibility of deep processing.

¹⁸The Sejong Corpus thus does not distinguish general nouns from verbal nouns.

¹⁹See Siegel and Bender (2002) for the Japanese grammar developed in the same environment.

²⁰The LKB is freely available with open source (<http://ling.stanford.edu>).

4 Conclusion

The grammar we developed here is based on the interactions of declarative constraints on types of signs. In capturing linguistic generalizations in a precise and concise manner, linguistic types are arranged into a multiple-inheritance hierarchy. The mechanism of multiple inheritance hierarchy has provided a succinct way of encoding generalizations in mixed constructions. The paper has further provided a clean, streamlined way of capturing the mixed properties of Korean LVC. The feasibility of the grammar developed has been checked with its implementation into the LKB system. The result of an existing test suite and self-constructed experimental data is quite promising though there still remains an issue of testing it with a large scale of corpora.

References

- Ahn, Hee-Don. 1989. On Light Verb Constructions in Korean and Japanese. *Japanese/Korean Linguistics* 1:221-238.
- Bratt, Elizabeth. 1996. *Argument composition and the lexicon: Lexical and periphrastic causatives in Korean*. Ph.D. dissertation, Stanford University.
- 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.
- Grimshaw, Jane, and Mester, Armin. 1988. Light Verbs and (Theta)-Marking. *Linguistic Inquiry* 19:205-232.
- Kim, Jong-Bok. 2004. The Korean Case System: A Unified, Constraint-based Approach. *Language Research*, 40.4.
- 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.
- Manning, Christopher D. 1993. Analyzing the Verbal Noun: Internal and External Constraints. *Japanese/Korean Linguistics* 3:236-253.
- 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 Hangeul and Korean Language Information Processing*.