

A Computational Treatment of Korean Serial Verb Constructions*

Jong-Bok Kim^a, Jaehyung Yang^b, and Sanghoun Song^c

^a School of English, Kyung Hee University
jongbok@khu.ac.kr

^b Dept. of Computer Science, Kangnam University
jaehyung@kangnam.ac.kr

^c Dept. of Linguistics, Univ. of Washington
sanghoun@gmail.com

Abstract. The so-called serial verb construction (SVC) is a complex predicate structure consisting of two or more verbal heads but denotes one single event. This paper first discusses the grammatical properties of Korean SVCs and provides a lexicalist, construction-based analysis couched upon a typed-feature structure grammar. We also show the results of implementing the grammar in the LKB (Linguistics Knowledge Building) system couched upon the existing the KRG (Korean Resource Grammar) which has been developed since 2003. The implementation results provides us with a feasible direction of expanding the analysis to cover a wider range of relevant data.

Keywords: serial verb construction, Korean Resource Grammar, LKB (Linguistic Knowledge Building)

1 Introduction

Korean is one of the languages that employ the so-called serial verb construction (SVC) in which more than one verb occurs without any specific coordination or subordination markings:

- (1) a. Mia-ka hakkyo-ey kel-e ka-ss-ta
Mia-NOM school-to walk-COMP go-PAST-DECL
'Mia walked to school.'
- b. Mia-ka cwul-ul cap-a tangki-ess-ta
Mia-NOM rope-ACC draw-COMP pull-PAST-DECL
'Mia pulled a rope, drawing it.'

Both sentences here, though including two serial verbs with their own predicate relations, semantically represent only a single event. These sentences display the canonical grammatical properties of SVCs in that the two successive verbs behave like a complex predicate, sharing one tense and subject value, and even the object value as in (1b).

The generation of Korean SVCs is quite flexible as attested by the corpus examples of the verb *mek-ta* 'eat':

- (2) *nanwu-e mek-ta* 'divide and eat', *kkulhi-e mek-ta* 'boil and eat', *mandul-e mek-ta* 'make and eat', *cap-a mek-ta* 'catch and eat', *cip-e mek-ta* 'pick up and eat', *ssip-e mek-ta* 'chew and eat', *kwu-e mek-ta* 'broil and eat', *ppal-a mek-ta* 'suck and eat', etc.

The examples, extracted from the Sejong POS-tagged Corpus, show us that the activity verb *mek-ta* 'eat' can combine with various types of verbs, forming an SVC.¹

* Many thanks go to anonymous reviewers for helpful comments. Main theoretical aspects sketched here follow those in Kim (2010).

¹ The Sejong Corpus, released by the National Institute of the Korean Language, is a balanced corpus consisting of about 12 million words with 311,048 sentences.

In this paper, we try to explore the syntactic and semantic properties of the Korean SVCs and build a computationally feasible analysis that can parse the desired set of examples. We then test the analysis with implementing it in the LKB (Linguistic Knowledge Building) System within the preexisting computational grammar, KRG (Korean Resource Grammar).

2 Grammatical Properties of the SVCs

2.1 Syntactic Headedness and Types of the SVC

As noted, one main property of the Korean SVCs is that each SVC sentence has only one tense, aspect, and mood value realized on the final verb. The appearance of this value on the first verb is thus not licensed:

- (3) a. Mia-ka sayngsen-ul kwu-(*ess)-e mek-ess-ta
 Mia-NOM fish-ACC roast-PAST-COMP eat-PAST-DECL
 ‘Mia roasted fish and ate it.’
 b. Mia-ka ttwi-(*ess)-e ka-ss-ta
 Mia-NOM run-PAST-DECL go-PAST-DECL
 ‘Mia left, running.’

Honorification information, which canonically surfaces in the subject and verb together as an agreement in the language, also appears on the final verb:

- (4) sensayng-nim-kkeyse John-ul tolli-(*si)-e ponay-si-ess-ta
 teacher-HON-NOM Chelswu-ACC turn-HON-COMP send-HON-PAST-DECL
 ‘The teacher sent Chelswu back.’

These facts support the idea that the final verb functions as the syntactic head.

The following question is then what kind of verbs can be combined in the SVC. For this purpose, we have performed a corpus search and extracted VV sequence verbs from the Sejong POS-tagged Corpus. Depending on the transitivity of the two verbs (intransitive, transitive, and ditransitive), we classified the extracted VV (called V1 and V2) sequence instances as following:²

(5) Frequency of VV Sequences by Transitivity

V1	V2	# of type	# of token	percentage	examples
intran	intran	3,566	14,658	32.07%	kel-e kata ‘go on foot’
intran	tran	1,794	5,217	11.41%	ttwi-e nem-ta ‘jump over’
intran	ditran	86	180	0.39%	nayli-e pat-ta ‘download’
tran	intran	2,501	9,651	21.11%	cip-e ka-ta ‘pick up and go’
tran	tran	3,902	14,499	31.72%	cap-a tangki-ta ‘catch and pull’
tran	ditran	142	359	0.79%	cip-e cwu-ta ‘pick up and give’
ditran	intran	82	350	0.77%	ponay-e o-ta ‘send to me’
ditran	tran	127	756	1.65%	pat-a mekta ‘receive’
ditran	ditran	6	43	0.09%	pat-a kalochay-ta ‘usurp’
sum		12,206	45,713	100.00%	

The table shows us that the most frequent (or natural) combinations are between transitive and intransitive verbs, for example, intransitive + intransitive/transitive or transitive + transitive/intransitive verb combinations, covering almost 95%. The corpus search has shown that since the SVCs describe an event or process, rather than a state, we observe that stative (or adjectival)

² There are more than three verb types in terms of argumenthood, but for simplicity, we introduce only these three types.

verbs do not appear in the constructions. After checking the grammatical properties of these sequences, we have first build an analysis that can capture these and then implemented it in the LKB system.³

2.2 Argument Sharing and Composition

One main characteristic of the SVCs is that the succession of multiple verbs behaves like a complex predicate with mono-clausal properties. Korean SVCs are not exception. For example, we can observe that only one subject or one object is required though there are more than one verb:

- (6) a. Mia-ka (*Chelswu-ka) hakkyo-ey kel-e ka-ss-ta
Mia-NOM Chelswu-NOM school-to walk-COMP go-PAST-DECL
'(int.) Mia walked to school.'
- b. Mia-ka (*koki-lul) sayngsen-ul kwuw-e mek-ess-ta
Mia-NOM meat-ACC fish-ACC roast-COMP eat-PAST-DECL
'(int) Mia roasted fish and ate it.'

When both verbs are transitive, they share the object as illustrated in (7a). The object in the SVC can also be linked only to one of the two verbs. In (7b), the object Mia is the argument of the first verb *chac-a* only, whereas the locative argument *hakkyo-ey* is selected only by the final verb 'go'.

- (7) a. Mia-ka lopu-lul kkul-e tangki-ess-ta
Mia-NOM rope-ACC draw-COMP pull-PAST-DECL
'Mia pulled a rope, drawing it.'
- b. Chelswu-ka hakkyo-ey Mia-lul chac-a ka-ass-ta.
Chelswu-NOM school-to Mia-ACC look.for-COMP go-Past-Decl
'Chelswu went to school to look for Mia.'

Unlike (7b), the object can be selected by the final verb too:

- (8) a. Mia-ka kang-ul hyeemchi-e kenne-ess-ta
Mia-NOM river-ACC swim-COMP cross-PAST-DECL
'Mia crossed the river, swimming.'
- b. Mia-ka Chelswu-lul toli-e ponay-ess-ta.
Mia-NOM Chelswu-ACC turn-COMP send-PAST-DECL
'Mia sent Chelswu back.'

In the examples (8) here, it is the final verb that selects the object *kang-ul* 'river-ACC' and *Chelswu-lul* 'Chelswu-ACC', respectively. It is also possible that the complement(s) in the SVC is selected only by one of the verbs:

- (9) a. Mia-ka hakkyo-ey kel-e ka-ass-ta.
Mia-NOM school-to walk-COMP go-PAST-DECL
'Mia went to school, walking.'
- b. Mia-ka Chelswu-lul ccoch-a ka-ass-ta.
Mia-NOM Chelswu-ACC follow-COMP go-PAST-DECL
'Mia went (somewhere), chasing Chelswu.'

In (9a), the locative complement is selected by the second one whereas in (9b) the accusative object is selected by the first one.

³ For this experimental study, we have made a random selection of 100 sentences whose number of words are less than 10 and tried to see how the present analysis can parse these. Of these total 45,713 tokens, we have first made a random selection of 10,000 and removed non-serial verb instances and again extracted 100 sentences for the test.

3 A Computational Treatment: Typed-feature Structure Approach

The Korean Resource Grammar (KRG) has been developed since 2003, aiming at building a computationally feasible grammar with a comprehensive coverage (see Kim and Yang 2004). 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 KRG 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 *verbal* category includes verbs and adjectives.

As for the combination of the sequence of such lexical expressions to form a bigger constituent, the KRG 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 ($XP \rightarrow \text{Mod}, XP$) as given in the following:

- (10) a. Head-Subject Rule:
 $XP[hd\text{-}subj\text{-}cx] \rightarrow \underline{1}, \mathbf{H}[\text{SUBJ } \langle \underline{1} \rangle]$
- b. Head-Complement Rule:
 $XP[hd\text{-}comp\text{-}cx] \rightarrow \underline{1}, \mathbf{H}[\text{COMPS } \langle \dots, \underline{1}, \dots \rangle]$
- c. Head-Modifier Rule:
 $XP[hd\text{-}mod\text{-}cx] \rightarrow [\text{MOD } \langle \underline{1} \rangle], \underline{1}\mathbf{H}$

These simple rules can license major phrasal constructions in the language. The Head-Subject Rule, generating a *hd-subj-cx*, 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-cx*. 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-cx*.⁴ In addition to these basic grammar rules, the KRG assume the following grammar rule to license the combination of two lexical expressions:

- (11) Head-Lexical-Cx:
 $[hd\text{-}lex\text{-}cx] \rightarrow [\text{LEX } +], \mathbf{H}[\text{LEX } +]$

This grammar rule basically licenses the combination of two verbs as in the following:

- (12) a. [[cal] [mek-ta]] (Adv + V)
 well eat-DECL('eat well')
- b. [[mek-ko] [siph-ta]] (Main V + Aux)
 eat-COMP would.like-DECL('would like to eat')
- c. [[cap-a] [tangki-ta]] (V + V)
 catch-COMP pull-DECL('catch and pull')

In (12a), the lexical adverb occurring only with a verb (not a VP) combines with the main verb whereas in (12b), the main verb forms a complex predicate with the auxiliary verb. (12c) is a SVC where two lexical verbs are combined.

In dealing with the generation of SVCs, another thing to note is that the verbs participating in the SVC are non-stative (activity) verbs. This constraint applies to both verbs in the SVC:

- (13) a. *cap-a ppalu-ta/*ppalu-a mek-ess-ta
 catch-COMP fast-DECL/fast-COMP eat-PAST-DECL
- b. *kel-e apu-ta/*himitul-e ka-ass-ta
 walk-COMP sick-decl/difficult-COMP go-PAST-DECL

⁴ 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 2004 for details.

- c. *ccocho-a apu-ta/*apu-a nay-ss-ta
 chase-COMP sick-DECL/sick-COMP take.out-PAST-DECL

The properties of the Korean SVCs we have seen, we assume, are reflections of the constructional properties (cf. Ginzberg and Sag 2001, Sag et al. 2003). That is, the Korean grammar introduces the SVC construction whose general constraint is given in the following (cf. Chung and Kim 2008):⁵

(14) Head-SVC:

$$\left[\begin{array}{c} hd-svc \\ C-CONT \left[\begin{array}{c} IND e0 \\ RELS \left\langle \begin{array}{c} PRED svc-rel \\ ARG1 e1 \\ ARG2 e2 \end{array} \right\rangle \end{array} \right] \end{array} \right] \rightarrow \left[\begin{array}{c} nonstative-v \\ VFORM (a/e)se \\ IND e1 \end{array} \right], H \left[\begin{array}{c} nonstative-v \\ IND e2 \end{array} \right]$$

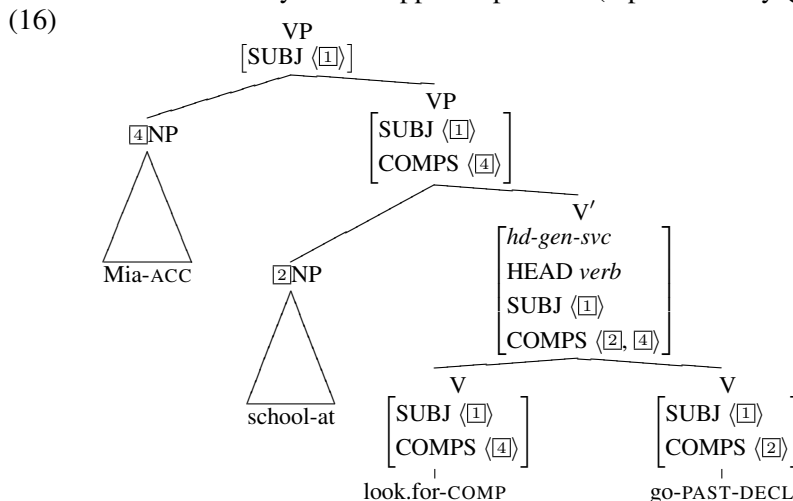
The constructional declaration on the head serial verb construction (*hd-svc*) specifies that two nonstative verbs are combined with the first carrying the *a/e* VFORM value. Each of these two denotes its own event *e1* and *e2*, and these two events are in the semantic relation *svc-rel* which includes the semantic relations such as a temporally-precedence (for SSVC) or overlap, manner or direction-relation. At this stage, the rule does not tease out all these three different ways of meaning composition in a formal way: we just assume that the semantic relation in the SVCs is constructionally-related as represented with the C-CONT (constructional meaning).⁶

As we have noted so far, the Korean SVCs allow the argument composition: the subject and object are shared while all the remaining arguments are composed together. To formalize this argument composition, we differentiate object-sharing cases from the other general SVCs cases, assuming two different SVCs.⁷ Consider the general cases with no object sharing first:

(15) Head-Gen-SVC:

$$\left[\begin{array}{c} hd-gen-svc \\ COMPS [A] \oplus [B] \end{array} \right] \rightarrow \left[\begin{array}{c} nonstative-v \\ COMPS [A] \end{array} \right], H \left[\begin{array}{c} nonstative-v \\ COMPS [B] \end{array} \right]$$

This constructional declaration means that a nonstative verb will combine with a preceding nonstative main verb, forming a *head-gen-svc*. This resulting construction will compose the COMPS value of these two verbs by the list append operation (represented by \oplus).



⁵ The generation of idiomatic SVCs is different since they are generated in the lexicon. That is, given the two nonstative verbs as input, the lexicon generates a serial verb compound with an idiomatic meaning. See Hashimoto and Bond (2005) for similar Japanese examples.

⁶ Representing this *svc-rel* in a finer and more precise way is thus our future project.

⁷ The main reason for positing these two subconstructions has to do with computational ones: we have found there is no clear formal way of representing object sharing cases together.

This assumption will basically license the combination of two nonstative verbs with different argument structure values as illustrated in the above structure for the sentence (8). In the SVC sentence, meaning ‘(Chelswu) went to school to look for Mia’, the two verbs ‘look-for’ and ‘go’ will combine first.⁸ This resulting complex predicate will combine with the two arguments *Mia-lul* and *hakkyo-ey*. This is possible from the argument composition in accordance with the Head-Gen-SVC rule. That is, the rule allows us to compose the argument ‘school-at’ selected by the second verb ‘go’ with the object ‘Mia’ selected by the first verb ‘look.for’. Note that this argument process is not a lexical one, but licensed by the constructional constraint.⁹

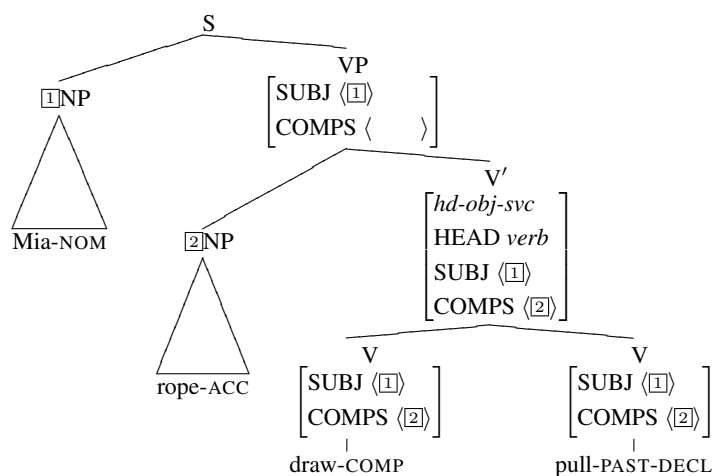
In addition to this general case, as we have seen, there are cases where the two verbs in the SVC share their object. We separate this kind of SVC case from the general cases with the following constructional constraint:¹⁰

(17) Head-Obj-SVC:

$$\left[\begin{array}{l} hd-obj-svc \\ COMPS \langle [2] \oplus [A] \oplus [B] \rangle \end{array} \right] \rightarrow [1] \left[\begin{array}{l} COMPS \langle [2][GCASE \textit{acc}] \oplus [A] \rangle \\ H \left[\begin{array}{l} nonstative-v \\ COMPS \langle [2][GCASE \textit{acc}] \oplus [B] \rangle \end{array} \right] \end{array} \right]$$

This rule refers to the GCASE (grammatical case) whose value is *acc*. That is, when both verbs in the SVC select an object whose structural case value is *acc*, the combination of these two verbs will then share this object ([2]) while the remaining complements are composed. This will then allow a structure like the following for (1b):

(18)



In the sentence, meaning ‘Mia pulled a rope, drawing it’, the head serial verb *tangki-ta* ‘draw’ now shares its object with the preceding verb *cap-a* ‘catch’. This is made possible by the constructional constraint of the Head-Obj-SVC. Once again, this argument sharing happens in syntax, not in the lexicon.

4 Evaluation and Conclusion

The analysis we have presented so far has been incorporated in the typed-feature structure grammar HPSG for the KRG aiming at working with real-world data (Copestake and Flickinger 2000

⁸ Due to the lexical process, it now selects the main verb ‘look for’ as its additional complement.

⁹ Another difference, as seen from the structure, is that the auxiliary verb selects its main verb, but the two serial verbs are not in a selectional relation.

¹⁰ In the KRG, the feature GCASE represents grammatical cases such as nominative or accusative whereas SCASE means semantic cases such as locative or instrument.

for English, Kim and Yang 2004 for Korean.) To check the computational feasibility of the analysis, we have implemented the analysis into the LKB (Linguistic Knowledge Building) system.¹¹ As the first step we selected 100 test suite sentences. The test results give us the proper syntactic as well as semantic structures for Korean serial verb constructions. For example, the following two figures are sample results of parsing the two main SVC sentences.

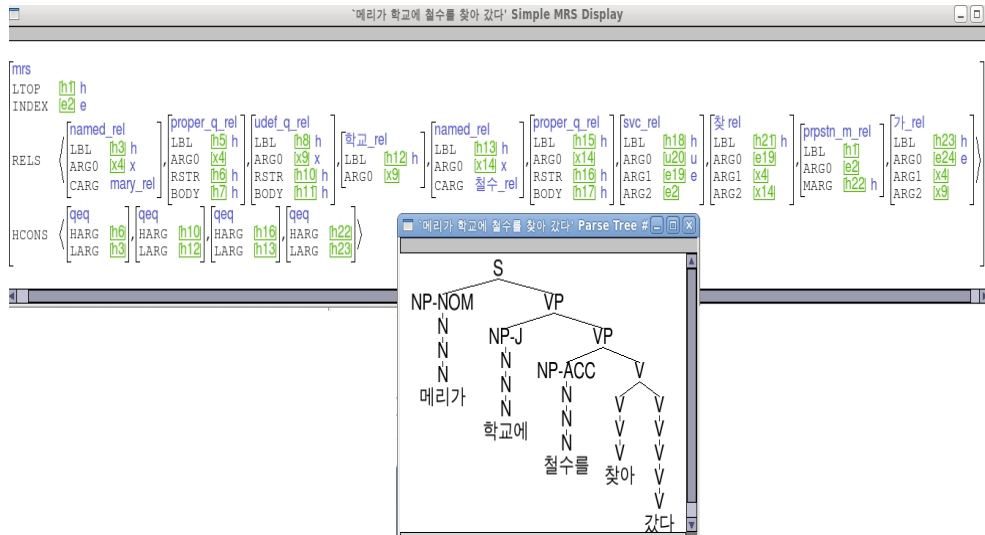


Figure 1: Parsed Tree and MRS for the phrasal comparative *Mary went to school, looking for Chelswu.*

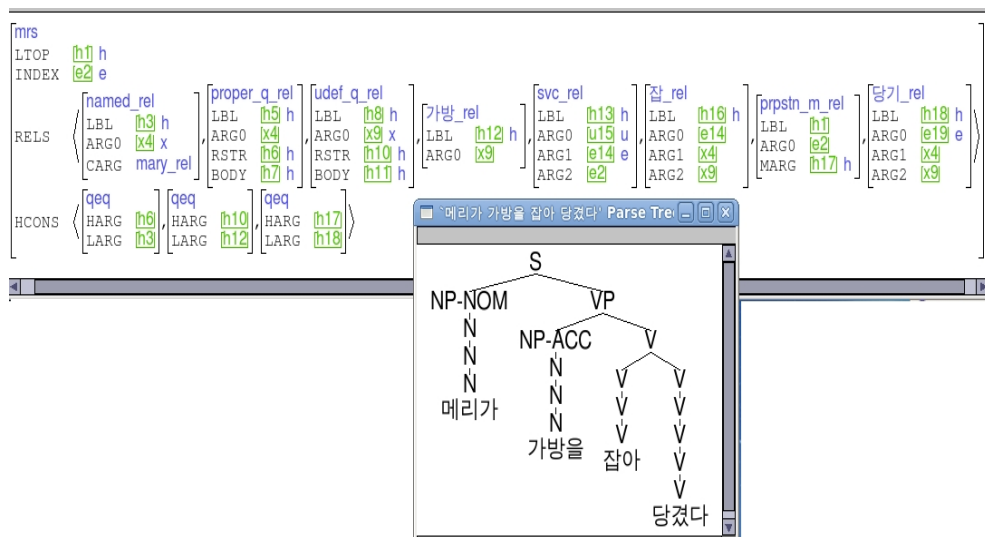


Figure 2: Parsed Tree and MRS for the phrasal comparative *Mary caught the bag and pull it.*

Figure 1 represents a general SVC where all the arguments are composed whereas Figure 2 is the object sharing case. The small boxes in both the figures indicate parsed tree structures whereas the big boxes denote the semantic representations. In both small boxes, the grammar licenses binary structures in which two serial verbs are combined and then the resulting phrase combines one argument in turn. The big boxes represent semantics. We can notice here that the MRS the

¹¹ The current Korean Resource Grammar, version 2.0, as of June 2010, has 659 lexical types and 114 phrasal types, 99 grammar rules, 304 inflectional rules, 39,688 lexical entries, and 1198 test-suite sentences, and 77% successful parsing rates.

grammar generates provides enriched information of the phrase.¹² The value of LTOP is the local top handle, the handle of the relation with the widest scope within the constituent. The attribute RELS is basically a bag of elementary predications (EP) each of whose value is a *relation*. The attribute HCONS is to represent quantificational information (see ?). Each of the types *relation* has at least three features LBL, PRED (represented here as a type), and ARG0. For the proper noun English and Korean, each has two related EPs: *named_rel* and *proper_q_rel*.

As noted earlier, the meaning of the SVC construction is added as a construction constraint, represented as *svc_rel*. The constructional-added relation, as noted, is inferred from the world knowledge or lexical properties. In terms of computational implementation, there still are more issues for our analysis to be resolved. However, we can observe that the grammar implemented in the LKB system appears to be feasible enough to extend to more complex data in a process of building a comprehensive KRG.

References

- Bender, Emily M. and Dan Flickinger and Stephan Oepen. 2002. The Grammar Matrix: An Open-Source Starter-Kit for the Rapid Development of Cross-Linguistically Consistent Broad-Coverage Precision Grammars. In Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics, Taipei, Taiwan.
- Chung, Taegoo. 1994. Argument structure and serial verbs in Korean. Doctoral. dissertation, University of Texas at Austin.
- Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. Stanford, CA: CSLI Publications.
- Copestake, Ann and Dan Flickinger. 2000. An Open-Source Grammar Development Environment and Broad-Coverage English Grammar using HPSG. In Proceedings of the 2nd conference on Language Resources and Evaluation. Athens, Greece.
- Copestake, Ann. and Dan Flickinger Carl Pollard Ivan A. Sag. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language & Computation* 3(4): pp. 281–332.
- Hashimoto, C. and F. Bond. 2005. A Computational Treatment of V-V Compounds in Japanese. Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, pp. 143-156.
- Kim, Jong-Bok. 2004. Korean Phrase Structure Grammar. (In Korean). Hankuk Publishing.
- Kim, Jong-Bok. 2005. Korean Passive Constructions: A Nonderivational, Constraint-Based Perspective (In Korean). *Korean Linguistics*, 26: 67-92.
- Kim, Jong-Bok. 2010. Argument composition in Korean Serial Verb Constructions. To appear in *Studies in Modern Grammar*.
- Kim, Jongbok. and Jaehyung. Yang. 2004. A Constraint-based Approach to Korean Auxiliary Constructions and its Computational Implementation. *Language Research*, 40(1): 195-226.
- Lee, Youngjoo. 2002. V-Complementation in serial verb constructions. *CLS*, 38-1.
- Sag, Ivan A., Thomas Wasow, and Emily M. Bender. 2003. *Syntactic Theory: A Formal Introduction*. Stanford, CA: CSLI Publications.
- Zubizarreta, Maria and Eunjeong Oh. 2007. On the Syntactic Composition of Manner and Motion. *Linguistic Inquiry Monograph* 48. Cambridge, Massachusetts: The MIT Press.

¹² MRS (Minimal Recursion Semantics), developed by Copestake et al. 2005 is a framework of computational semantics designed to enable semantic composition using only the unification of type feature structures. See ? and ? for its implementation in English and Korean, respectively. The value of the attribute SEM(ANTICS) in our system represents a simplified MRS. ARG0 canonically refers to the index value of the EP (elementary predicate) itself whereas ARG1 or ARG2 refers to the predicate's semantic arguments.