A Unified Approach to VP-ellipsis and VP-anaphora

Yong-hun Lee

Department of English Language and Literature, Chungnam National University
99 Daehak-ro, Yuseong-gu, Daejeon 305-764, Korea
ylee@cnu.ac.kr

Summary. It is known that VP-ellipsis and VP-anaphora are typologically different phenomena. English has VP-ellipses whereas Korean has VP-anaphora. The goals of this paper are to develop a unified algorithm which can analyze these two different phenomena and to explain them using the algorithm. In order to analyze these phenomena, this paper incorporates Jäger(2010)'s anaphora resolution mechanism into the typed feature structure formalism of Head-driven Phrase Structure Grammar (HPSG). In this paper, VP-ellipsis and VP-anaphora are analyzed as follows. First, English *do* and Korean *kuleha-ta* are introduced with the Geach value, and this value is changed with a *slash-elimination* rule. Then, one constituent combines with another by ordinary syntactic rules, while the information on the target predicate is percolated up. When a potential source appears, a *slash-introduction* rule is applied. Then, the source predicate activates the *VP-resolution* rule, and the target predicate is connected with the source in the semantic representations.

Keywords: VP-ellipsis, VP-anaphora, Geach, slash rules

1 Introduction

Ellipsis is one of the interesting topics in syntax and semantics, since syntactically elided parts have to be recovered in the semantic interpretation (Johnson, 2010). It is also an interesting area in computational linguistics where the syntax and semantics of words and sentences are computationally implemented.

As the sentences in (1) illustrates, English has VP-ellipsis phenomena.

(1) a. John came, and Mary [$_{VP}$ came], too.
b. John came, and Mary did [$_{VP}$ ~~come~~], too.

The VP parts of two conjuncts are identical in (1a), and the VP of the second conjunct is elided in (1b) while the dummy auxiliary *do* takes the past tense in the second conjunct.

Whereas English has VP-ellipsis, Korean has VP-anaphora phenomena. Let's see the example sentences in (2).

(2) a. *Chelsoo-ka* *o-ass-ko,* *Younghee-to* [$_{VP}$ *o*]-*ass-ta*.
Chelsoo.NOM come.PAST.and Younghee.too come.PAST.DECL
'Chelsoo came, and Younghee came, too.'
b.* *Chelsoo-ka* *o-ass-ko,* *Younghee-to* [$_{VP}$ ~~*o*~~]-*ass-ta*.
Chelsoo.NOM come.PAST.and Younghee.too come.PAST.DECL
'Chelsoo came, and Younghee came, too.'
c.* *Chelsoo-ka* *o-ass-ko,* *Younghee-to* [$_{VP}$ *ha*]-*ass-ta*.
Chelsoo.NOM come.PAST.and Younghee.too do.PAST.DECL
'Chelsoo came, and Younghee came, too.'

d. *Chelsoo-ka o-ass-ko, Younghee-to* [$_{VP}$ *kuleha*]-*yss-ta*.
Chelsoo.NOM come.PAST.and Younghee.too do-so.PAST.DECL
'Chelsoo came, and Younghee came, too.'

Since the VP parts of two conjuncts in (2a) are identical, the VP of the second conjunct is elided in (2b) and it results in an ungrammatical sentence. In (2c), we have a *ha*-support, which is similar to *do*-support in the English sentence in (1b), to take a past tense morpheme *-ess*. However, the *ha*-support does not save the sentence. On the other hand, in (2d), the pro-form *kuleha-ta* is inserted into the elided VP position, and it makes the sentence grammatical. In this sentence, the pro-form *kuleha-ta* refers to the verb *o-ass-ta* 'come' in the first conjunct. The example sentences in (1) and (2) demonstrate that English has VP-ellipsis phenomena but that Korean has VP-anaphora.

The goals of this paper are (i) to develop a unified algorithm which can analyze these two different phenomena and (ii) to explain them using the developed algorithm. In order to analyze both VP-ellipsis and VP-anaphora phenomena in HPSG (Pollard and Sag, 1994; Sag et al., 2003; Kim and Sells, 2008), this paper borrows basic ideas from Categorial Grammar (CG) and provides resolution algorithms for these two different phenomena.

2 Previous Categorial Approaches to VP-Ellipsis and VP-Anaphora

Through a series of papers (Jacobson, 1996, 1999, 2000, 2001), Pauline Jacobson has developed an alternative categorial approach to pronominal anaphora resolution and applied it to a wide range of empirical phenomena. She introduced a third slash connective that is responsible for anaphoric dependencies, and she used the notation A^B for signs of category A that needs an antecedent of category B. On the other hand, Jäger (2001) used other notation $A|B$ to stress the similarity with the other slashes. In their analyses of anaphora, a pronominal *him* has a category *NP*|*NP*, and it translates into $\lambda x.x$. Then, the dummy auxiliary *do* in English and the pro-form *kuleha-ta* in Korean may have a category (*S**NP*)|(*S**NP*) whose translation is $\lambda P.P$. As for category combinatorics for the anaphora, Jacobson adopted Geach rules while Jäger (2010) used |-elimination and |-introduction rules in the analyses.

3 VP-Ellipsis and VP-Anaphora in HPSG

3.1 Basic Ideas

For the purpose of analyzing both VP-ellipsis and VP-anaphora phenomena in HPSG, this paper borrows Jäger's anaphora resolution algorithms. In this paper, English VP-ellipsis and Korean VP-anaphora are analyzed as follows. First, the English auxiliary *do* and the Korean pro-form *kuleha-ta* are introduced with the Geach value, and this value is changed with a *slash-elimination* rule. Then, one constituent combines with another by ordinary syntactic rules in HPSG, while the information on the target predicate is percolated up. When the target predicate meets a potential source predicate, a *slash-introduction* rule is applied and the Geach value was changed again. Then the potential source predicate activates the *VP-resolution* rule, and the target predicate is connected with the source in the semantic representations.

3.2 Type Hierarchy and Attribute-Value Matrix

In order to provide a unified analysis to VP-ellipsis and VP-anaphora phenomena, this paper incorporates Jager's ideas into the typed feature structure formalism of HPSG and modifies type hierarchy and feature structures as follows. In the Lexicon, a new type *ellip-ana-aux-v-lxm* is introduced into the type hierarchy as in Figure 1, and English *do* and Korean *kuleha-ta* are instances of *ellip-aux-v-lxm* and *ana-aux-v-lxm* respectively. The Attribute-Value Matrix (AVM) for the type *ellip- ana-aux-v-lxm* is shown in Figure 2.

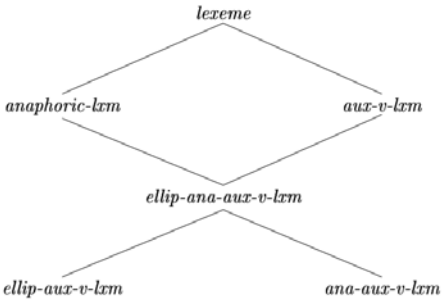

Figure 1: Type Hierarchy for the Type *ellip- ana-aux-v-lxm*

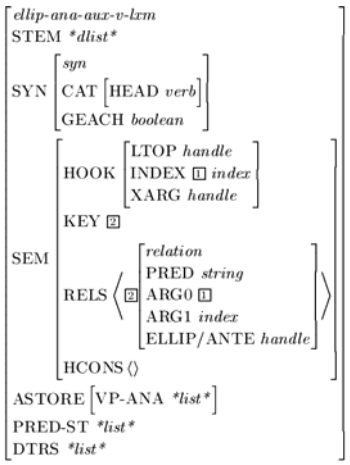

Figure 2: AVM for the Type *ellip-ana-aux-v-lxm*

Four attributes/features are introduced into the typed feature formalism: GEACH, ELLIP/ANTE, ASTORE, and PRED-ST. The first one encodes whether a Geach rule is applied or not. If a Geach rule is applied, its value becomes +. If the VP-ellipsis/VP-anaphora resolution algorithms are activated, its value becomes -. For the second attribute, if the given auxiliary is an instantiation of *ellip-aux-v-lxm*, the auxiliary has ELLIP and it refers to the label of the elided VP. If the given auxiliary is an instantiation of *ana-aux-v-lxm*, the auxiliary has ANTE instead and it refers to the label of the antecedent VP. The third attribute PRED-ST contains the predicates of the sentences. The fourth attribute ASTORE (anaphoric expression store) contains the HCONS values in A-HCONS (anaphoric expression HCONS), that encodes which source predicate refers to which target predicate.

3.3 Slash Rules

Three slash rules are introduced into the type hierarchy to analyze VP-ellipsis and VP-anaphora phenomena in HPSG. They are *slash-elimination* rule (|E), *slash introduction rule* (|I), and *VP-resolution rule* (VP-Resol). These rules are organized in the type hierarchy as follows.

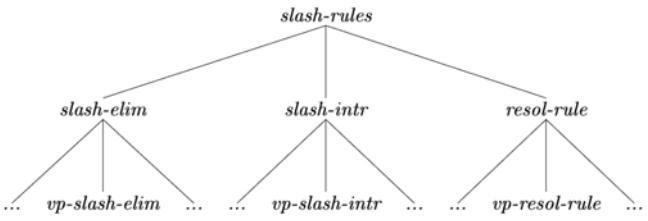

Figure 3: Type Hierarchy for Slash Rules

A *slash-elimination* rule changes the value of GEACH from – to +. Along with this change, a hook for the target predicate has to be stored in other parts of the AVM. A *slash-introduction* rule is triggered when the target predicate meets a potential source predicate, and this rule changes the value of GEACH from + to –. A *VP-resolution* rule finds out the source predicate and connects the target predicate with its source predicate.

4 An Analysis of VP-Ellipsis in English

Based on the AVM of the type *ellip-aux-v-lxm* and slashes rules in Figure 3, the overall analysis processes of English VP-ellipsis are as follows.

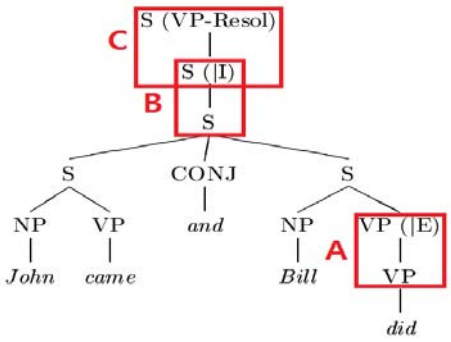

Figure 4: An Example Analysis of VP-Ellipsis

Here, the important operations are marked with Step A, Step B, and Step C.

In the Step A, the English *do* introduced into syntax with the feature [GEACH -]. Then, when there is an *ellip-aux-v-lxm* with [GEACH -], a *slash-elimination* rule (|E) is applied and the feature structure of *do* are changed as shown in Figure 5.

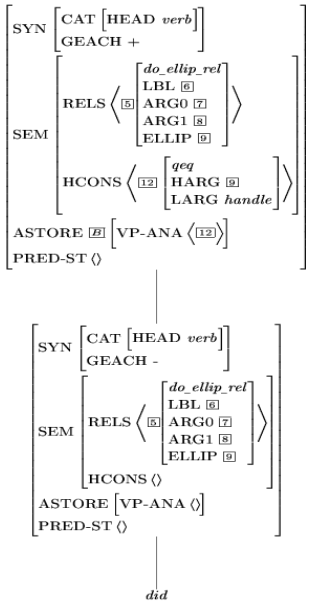

Figure 5: Applying a *slash-elimination* Rule

After the *slash-elimination* is applied, the GEACH value is changed from - to +, and HCONS includes a new *qeq* whose HARG value is equal to the ELLIP value of English *do*. Here, LARG

will refer to the handle of the source predicate in the final step of the algorithm. This HCONS value is stored in A-HCONS of VP-ANA.

Then, the top part of feature structure in Figure 5 is percolated up until PRED-ST contains a potential source predicate. In the English sentence (1), when the first conjunct *John came* combines with *Mary did*, since PRED-ST contains a potential predicate (*came*), a *slash-introduction* (|I) is applied in Step B and the AVM of Figure 5 is changed into that of Figure 6.

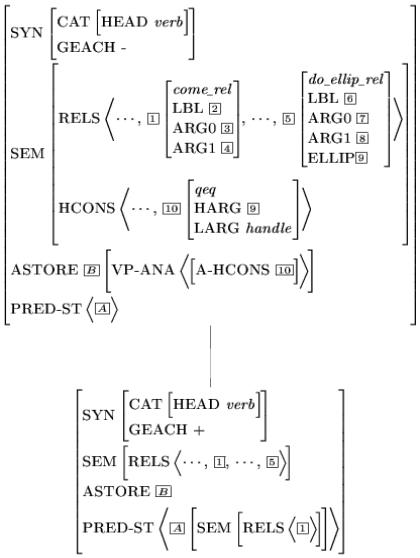

Figure 6: Applying a *slash-introduction* Rule

This rule changes the value of GEACH from + to -, which implies that there is a potential source predicate for the VP-ellipsis phenomena. This potential source predicate will activate the *VP-resolution* rule.

In Step C, the *VP-resolution* rule (VP-Resol) is applied when (i) the value of GEACH is - and (ii) VP-ANA is not empty. Then, the AVM of Figure 6 is changed into that of Figure 7.

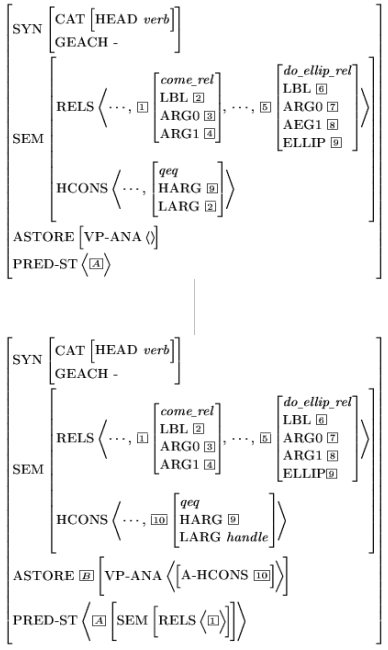

Figure 7: Applying a *VP-resolution* Rule

Since the PRED-ST value refers to the AVM of the verb *came* in the first conjunct, it also contains the RELS value of *came* in the MRS. Then, the *VP-resolution* rule searches for the LBL value of the source predicate *came* and it connects the value to the LARG value of A-HCONS (in VP-ANA). Then, after the LARG of A-HCONS gets its value, VP-ANA becomes empty. This implies that the handle of the ELLIP value of *do_ellip_rel* is identical to the LBL value of *come_rel*, which in turn means that the head of the elided VP is *come*.

5 An Analysis of VP-Anaphora in Korean

On the other hand, based on the AVM of the type *ana-aux-v-lxm* and slashes rules in Figure 3, the overall analysis processes for Korean VP-anaphora are as follows.

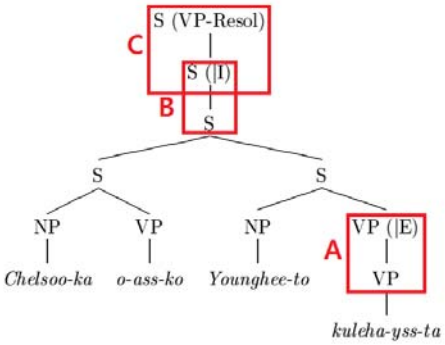

Figure 8: An Example Analysis of VP-Anaphora

As in English example, the important operations are marked with Step A, Step B, and Step C.

In the Step A, the Korean pro-form *kuleha-ta* introduced into syntax with the feature [GEACH -]. Then, since there is an *ana-aux-v-lxm* with [GEACH -], a *slash-elimination* rule (|E) is applied and the AVM of *kuleha-ta* are changed as shown in Figure 9.

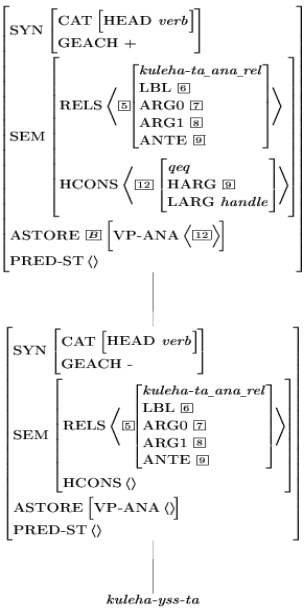

Figure 9: Applying a *slash-introduction* Rule

After the *slash-elimination* is applied, the GEACH value is changed from - to +, and HCONS includes a new *qeq* whose HARG value is equal to the ANTE value of the Korean pro-form

kuleha-ta. Here, LARG will refer to the handle of source predicate in the final step of the algorithm. This HCONS value is stored in A-HCONS of VP-ANA.

Then, the top part of feature structure in Figure 9 is percolated up until PRED-ST contains a potential source. In the sentence (2), when the first conjunct *Chelsoo-ka o-ass-ko* combines with *Younghee-to kuleha-yss-ta*, since PRED-ST contains a potential predicate (*o-ass-ko*), a *slash-introduction* is applied and the AVM of Figure 9 is changed into that of Figure 10.

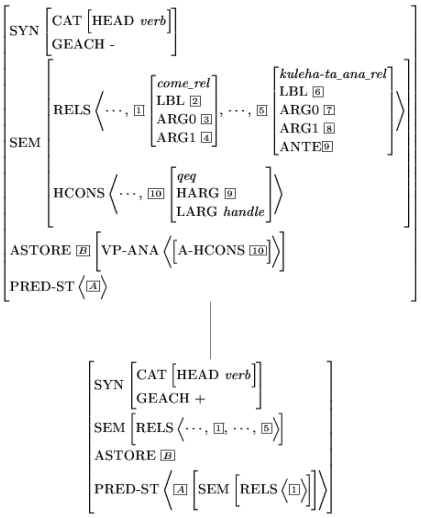

Figure 10: Applying a *slash-introduction* Rule

This rule changes the value of GEACH from + to -, which implies that there is a potential source predicate for the VP-anaphora phenomena. This potential source predicate will activate the *VP-resolution* rule.

In Step C, the *VP-resolution* rule (VP-Resol) is applied when (i) the value of GEACH is - and (ii) VP-ANA is not empty. Then, the AVM of Figure 10 is changed into that of Figure 11.

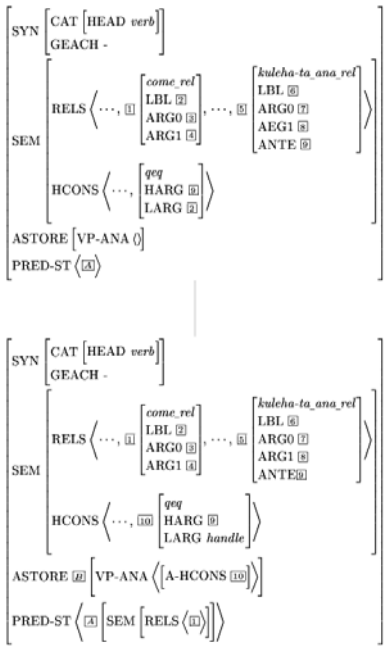

Figure 11: Applying a *VP-resolution* Rule

Since the PRED-ST value refers to the AVM of the verb *o-ass-ko* of the first conjunct, it also contains the RELS value of *come_rel* in the MRS. Then, the *VP-resolution* rule searches for the LBL value of source predicate *come_rel* and it connects the value to the LARG value of A-HCONS (in VP-ANA). Then, after LARG of A-HCONS gets its value, VP-ANA becomes empty. This implies that the handle of the ANTE value of *kuleha-ta_ana_rel* is identical to the LBL value of *come_rel*, which in turn means that *kuleha-yss-ta* refers to *o-ass-ko*.

6 Conclusion

In this paper, a unified resolution algorithm was developed which can account for both VP-ellipsis and VP-anaphora in HPSG. In order to analyze these two phenomena, this paper incorporated Jäger's anaphora resolution mechanism into the typed feature structure formalism of HPSG, and these two typologically phenomena were explained using the unified resolution algorithm.

In this paper, English VP-ellipsis and Korean VP-anaphora were analyzed as follows. First, the English auxiliary *do* and the Korean pro-form *kuleha-ta* were introduced with the Geach value, and this value was changed with a *slash-elimination* rule. Then, one constituent combined with another by ordinary syntactic rules in HPSG, while the information on the target predicate was percolated up. When the target predicate met a potential source predicate, a *slash-introduction* rule is applied and the Geach value was changed again. Then, the source predicate activates the *VP-resolution* rule, and the target predicate is connected with the source in the semantic representations.

Through the analysis, we observed that both VP-ellipsis and VP-anaphora could be analyzed with a unified resolution algorithm. It was possible by incorporating the type *ellip-ana-aux-v-lxm* and three kinds of slash rules.

References

Jacobson, P. 1996. The Syntax/Semantics Interface in Categorial Grammar. In S. Lappin, ed., *The Handbook of Contemporary Semantic Theory*, pp. 89-116. Oxford: Blackwell.

Jacobson, P. 1999. Toward a Variable-free Semantics. *Linguistics and Philosophy*, 22(2), 117-184.

Jacobson, P. 2000. Paycheck Pronouns, Bach-Peter Sentences, and Variable-free Semantics. *Natural Language Semantics*, 8(2), 77-155.

Jacobson, P. 2001. Binding without Pronouns (and Pronouns without Binding). Manuscript. Brown University.

Jäger, G. 2010. *Anaphora and Type Logical Grammar*. Dordrecht: Springer.

Johnson, K. 2010. *Topics in Ellipsis*. Cambridge: Cambridge University Press.

Kim, J. and P. Sells. 2008. *English Syntax*: *An Introduction*. Stanford, CA: CSLI.

Pollard, C. and I. A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Stanford: CSLI Publications and Chicago: The University of Chicago Press.

Sag, I., T. Wasow, and E. Bender. 2003. *Syntactic Theory*: *A Formal Introduction*. Second Edition. Stanford, CA: CSLI.