Problems with Complement Extraction

Sae-Youn Cho
(Honam University)

Cho, Sae-Youn. 1997. Problems with Complement Extraction. *Linguistics,* 5-2, 21-34. This paper is mainly concerned with complement extraction in English. Recently, BMS (1997) have proposed the so-called Dependent Realization Constraint to allow a lexical head to be realized as either a local dependent or a nonlocal dependent. Though this constraint might seem to attain a theoretical achievement in that it enables us to provide a unified account of extraction dependencies such as subject, complement and adverb extraction, it appears to be empirically not fully tested. Hence, this paper shows what data can be problematic for this constraint and suggests a solution to them. **(Honam University)**

1. Introduction

It is a well-known fact that there are various analyses of extraction phenomena in UDC (Unbounded Dependency Constructions) depending on syntactic theories. Regardless of the variety of the analyses on the phenomena, they share a common property of assuming the existence of traces. Hence, the analyses having such property is called the trace approach in this paper.[1] Since traces have seemed to be empirically motivated, the previous analyses on extraction in UDC have focused on the explanation for how traces are related to appropriate fillers in the sentences involving extraction. The traditional Government and Binding (GB) approach heavily relys on movement while Generalized Phrase Structure Grammar (GPSG (1985)) and Head-Driven Phrase Structure Grammar (HPSG (1987)) employ feature percolation in accounting for the "filler-gap" dependencies as in (1).

1 I would like to thank Ivan Sag, Jong-Bok Kim, Kwang-Ho Lee, and all members of the HPSG circle of the Linguistic Association of Korea for their comments and suggestions.

(1) a. I wonder who John likes ___ .
b. Apples, I like ___ .

Bouma, Malouf and Sag (BMS (1997)), however, have proposed a new traceless analysis of UDC within the frame work of current HPSG on the basis of the fact that the argument for the existence of traces is untenable or neutral. The key mechanism of their analysis is the Dependent Realization Constraint (DRC) to allow a lexical head to be realized as either a local dependent or a nonlocal dependent. Though this constraint might seem to provide a unified account of extraction dependencies such as subject, complement and adverb extraction, it needs to be fully tested for empirical support.

In this paper, it is claimed that the DRC is empirically motivated but the head-filler phrase assumed in current HPSG must be revised in order for the traceless analysis to succeed. For simplicity, I limit myself to mainly exploring complement extraction in UDC.

To argue for this claim, section 2 reviews how complement extraction can be treated under the trace approach including GB (1981), GPSG (1985) and HPSG (1987). Section 3 introduces the traceless approach proposed by BMS (1997) after providing the evidence that the argument for the existence of traces is untenable or neutral. Section 4 examines some complement extraction data which seem to be problematic for the DRC and suggests that the difficulty of explaining them is due not to the DRC but to the head-filler phrase. In conclusion, some solutions and remaining problems are provided.

2. Trace Approach

To see how a theory assuming the existence of traces differs from a theory without them, it is essential to understand how complement extraction phenomena are analyzed under the trace approach. In the traditional GB analysis, sentences like (2) can be represented as (3). Various principles such as ECP (Empty Category Principle), Subjacency and Theta theory must be satisfied for the dependency between the

filler *who* and the trace e_i in (3) to be legitimate.

(2) Who did Johnny hit?
(3) $_{IP}$[Johnny hit who]? -> $_{CP}$[Who$_i$ e $_{IP}$[Johnny hit e_i]]?

The analysis on UDC by GPSG (1985) employs various mechanisms such as Slash Termination Metarule 1 (STM1) in (4), an Immediate Dominance rule in (5), and the informal definition of Foot Feature Principle (FFP) in (6).

(4) **Slash Termination Metarule 1 (STM 1)**

$$X \rightarrow W,\ X^2$$
$$\downarrow$$
$$X \rightarrow W,\ X^2[+NULL]$$

(5) $S \rightarrow X^2,\ H/X^2$

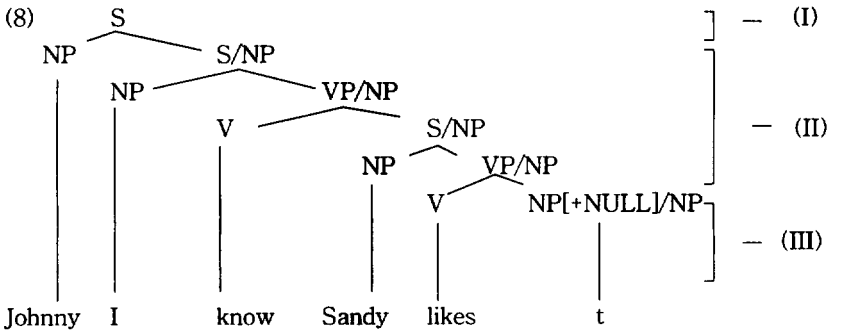
(6) **Foot Feature Principle (FFP)**

The FOOT feature specifications that are instantiated on a mother category in a tree must be identical to the unification of the instantiated FOOT feature specifications in all of its daughter categories.

Each mechanism above plays an important role in explaining UDC like (7). In GPSG (1985), (7) can be represented as in (8).

(7) Johnny, I know Sandy likes ____ .

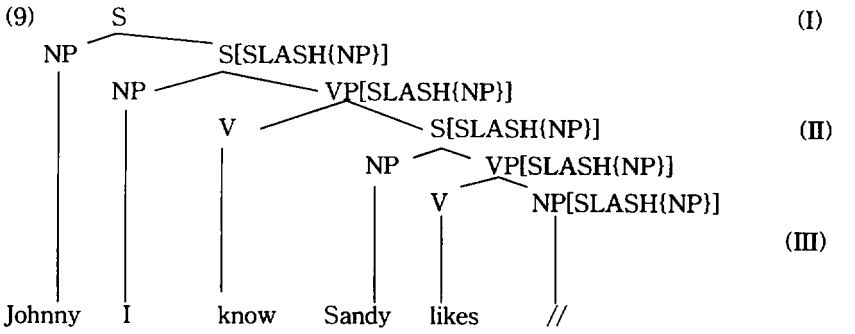
(8)



In the tree representation of (8), the trace of the bottom part (III) is guaranteed by the STM 1 in (4). The instantiated SLASH feature, which is one of FOOT features, is percolated through the middle part (II) in terms of the FFP in (6). Finally, the top part (I) is licensed by the ID rule in (5). In this local tree, the SLASH value in S node must be identical with the filler NP by the definition of the CAP (Control Agreement Principle) in GPSG (1985: *pp* 89).

Similarly, HPSG (1987) also assumes the existence of traces in accounting for filler-gap dependencies. The sentence (7) can be represented as in (9).

(9)



There seems to be no big difference between the analysis of earlier HPSG and that of GPSG (1985) in explaining the UDC. It is, however, notable that traces specified as // in the bottom part (III) also are assumed in this earlier HPSG. Even HPSG (1994) posits a trace in the lexicon as follows:

(10) Trace in the lexicon (simplified version):

$$\begin{bmatrix} \text{PHONOLOGY} < > \\ \text{SYNSEM} \begin{bmatrix} \text{LOCAL [1]} \\ \text{NONLOCAL} \begin{bmatrix} \text{SLASH } \{[1]\} \\ \text{QUE } \{\ \} \\ \text{REL } \{\ \} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(HPSG (1994; *pp* 161)

As demonstrated above, traces are, after all, key categories to account for filler-dependencies for this approach, though, depending on the theories, various different mechanisms are employed. The question to be answered is why the phonologically unrealized element, namely trace, should be regarded as a syntactic category. It is needless to say that a theory with empty categories is preferable to that without them if other conditions are equal. In addition to this theoretical advantage, a theory without empty categories is also supported by the easiness of explaining human language acquisition and the efficiency of parsing sentences in Computational Linguistics. To defend their position, the trace approach provides several syntactic arguments based on empirical data.

3. Traceless Approach

Sag (1996) has proposed a new traceless analysis, while arguing that the arguments for the existence of traces are untenable or neutral. There is a three-part syntactic argument for the trace approach: *wanna* contraction, auxiliary contraction and adverb position.[2] First of all, *want* and *to* can contract to *wanna* in (11a) whereas they cannot in (11b).

(11)a. Who$_j$ does Johnny$_i$ want PRO$_i$ to(wanna)go to the store with $\_\_{}_j$?
b. Who$_j$ does Johnny$_i$ want $\_\_\_{}_j$ to (*wanna) go to the store?

To explain the distributional behavior, the trace approach claims that WH-trace is visible to the *wanna* contraction rule but PRO is invisible. The bottom line of this claim is that traces are indispensible to account for this contraction phenomenon. This explanation, however, does not work for (12) where there is a WH-trace between *think* and *is*.

(12) Who$_i$ does Johnny think $\_\_\_\_{}_i$ is (think's) smart?

2 The claim that the argument for three syntactic evidence is untenable is originally made by Sag (1996).

According to this explanation, *think* and *is* must not contract to *think's*, but it is not the case. So this argument is untenable.

The second argument is related to the observation that the contraction of auxiliaries immediately being followed by WH-trace as in (13) is not allowed. This evidence seems to be strong enough to show why trace is needed as an independent category.

(13) How tall did Johnny think she is (*s) _____ ?

But Selkirk (1984) shows that this behavior follows from some principles of metrical phonology. It can, then, be treated within the phonological domain, rather than the syntactic component. Hence, this argument cannot be tenable, either.

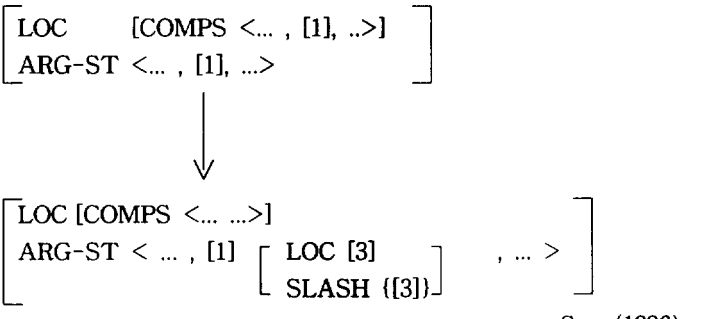
The last argument is based on the distributional bahavior of the adverb *all* in (14). The adverb *all* in (14a) can occur between *they* and *were* whereas all in (14b), where it immediately precedes trace, cannot. In order to differentiate one from the other in grammaticality, it appears to be unavoidable to assume the existence of traces.

(14) a. How satisfied do you think they all were ____ ?
b. *How satisfied do you think they were all ____ ?

For this, Brodie (1983) proposed that *all* as a floating quantifier can be regarded as base-generated adjoined modifiers. The importance of this proposal is that traces do not play any role in accounting for the behavior. So this argument is also neutral. If all the arguments for the trace approach are untenable, the only thing to demonstrate is how well this traceless approach can analyze UDC in order to show that the traceless approach is more preferable than the trace one.

The key mechnism of the extraction analysis by Sag (1996) is a lexical rule that reduces the valence of an input word by eliminating an element from its COMP list, while encoding information to the corresponding member in ARG-ST. This lexical rule is as follows:

(15) Complement Extraction Lexical Rule (CELR):



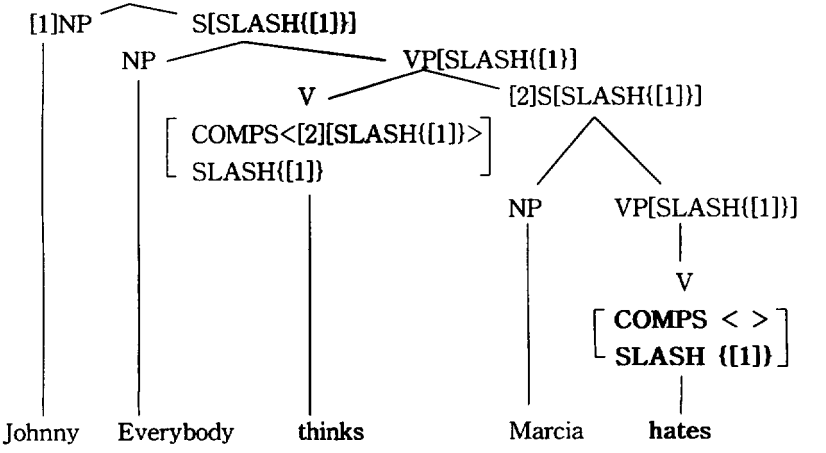
Sag (1996)

In conjunction with CELR and SLASH Inheritance Constraint (SLASH IC) in (16), sentences like (17) can be represented as in (18).

(16) SLASH Inheritance Constraint (SLASH IC)
The SLASH value of a head-valence-phrase is indential to the SLASH value of the phrase's head daughter.

(17) Johnny, Everybody thinks Marcia hates ____ .

(18)



As can be seen in (18), the value for SLASH is percolated up to the highest local tree by the definition of SLASH IC in (16). The bold faced Attribute Value Matrix (AVM) is licensed by CELR and thus, traces are completely removed successfully in this way. To ensure that the SLASH value in the head daughter of the top part is identical to the filler, the following schema is required:

(19) Head-Filler Schema:

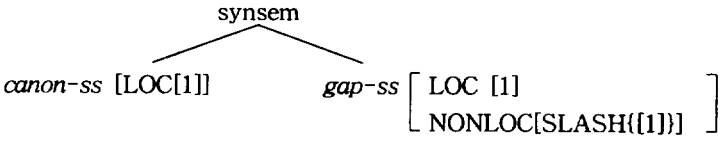
S	->	Filler-Dtr,	Head-Dtr Phrase
[SLASH { }]		[LOC[1]]	[SLASH {[1]}]

Though this traceless analysis seems to achieve a theoretical improvement, it still needs at least two more extraction lexical rules since there are three kinds of extraction, namely subject, complement, and adverb extraction. To overcome this difficulty, BMS (1997) has proposed the DRC in (20) to cover all extraction phenomena, after providing a new sort, *gap-synsem*, in (21).

(20) Dependent Realization Constraint

word => [SUBJ [1]
COMPS [2]list(*canon-ss*)
DEPS [1] + ([2] 0 list(*gap-ss*)]

(21) canonical-synsem (canon-ss) & gap-synsem (gap-ss)



The DRC, a constraint for the sort *word*, provides an answer to the question whether a dependent of a lexical head is to be realized as a local dependent (i.e. on one of the valence features SUBJ or COMPS) or a nonlocal dependent. For example, if there is the feature structure of

the verb *hates* in (22), it can be realized as in (23) or (24) in terms of the DRC.

(22)
$$\begin{bmatrix} \text{I-FORM} & \textit{hates} \\ \text{SUBJ} & \langle [\ ] \rangle \\ \text{DEPS} & \langle \begin{bmatrix} \text{LOC NP} \\ \text{SLASH [1]} \end{bmatrix}, \begin{bmatrix} \text{LOC NP}[\textit{acc}] \\ \text{SLASH [2]} \end{bmatrix} \rangle \\ \text{SLASH} & [1] \cup [2] \end{bmatrix}$$

(23)
$$\begin{bmatrix} \text{I-FORM} & \textit{hates} \\ \text{SUBJ} & \langle [3] \begin{bmatrix} \text{LOC} & \text{NP} \\ \text{SLASH} & [1] \end{bmatrix} \rangle \\ \text{COMPS} & \langle [4] \begin{bmatrix} \text{LOC} & \text{NP}[\textit{acc}] \\ \text{SLASH} & [2] \end{bmatrix} \rangle \\ \text{DEPS} & \langle [3],[4] \rangle \\ \text{SLASH} & [1] \cup [2] \end{bmatrix}$$

(24)
$$\begin{bmatrix} \text{I-FORM} & \textit{hates} \\ \text{SUBJ} & \langle [3] \begin{bmatrix} \text{LOC} & \text{NP} \\ \text{SLASH} & [1] \end{bmatrix} \rangle \\ \text{COMPS} & \langle\ \rangle \\ \text{DEPS} & \langle [3], \begin{bmatrix} \text{gap-ss} \\ \text{LOC [2]} \\ \text{SLASH } \{[2]\} \end{bmatrix} \rangle \\ \text{SLASH} & [1] \cup \{[2]\text{NP}\{\textit{acc}\}\} \end{bmatrix}$$

BMS (1997)

So, a two-place predicate like *hates* can be realized as a verb with a subject and a complement or a verb with a local dependent and a gap. In principle, the DRC works not only for subject and complement extraction but for adverb extraction. Thus, the DRC by BMS (1997) must be a corner stone in order for the traceless approach to provide a unified account of all types of extraction phenomena.

4. Problems with DRC or Head-Filler Phrase

It is by now obvious that the DRC theoretically plays an important role in the sense that it enables us to account for various extraction phenomena without postulating any additional rule or constraint. But it still does not appear to be empirically fully tested, though BMS (1997) showed how their analysis works for complement and adverb extraction.

The data which seem to be problematic for the DRC is UDCs headed by a ditransitive verb in (25). The distributional behavior in (25) shows that when there are two different values of SLASH, the sentences are ungrammatical.

(25) a. *Johnny a book, I gave ___ ___ .
b. *A book Johnny, I gave ___ ___ .

This is problematic for the DRC since, by the definition, it must allow the feature structure for *gave* to be realized as in (26), where it has two complement gaps in DEPS.

(26) gave-4 (simplified version)

I-FORM	*gave*
SUBJ	<[1]>
COMPS	< >
DEPS	<[1], [2][LOC[3], SLASH{[3]}], [4][LOC[5], SLASH{[5]}]>

So the easiest way to solve this problem might be to constrain the DRC somehow so that the feature structures having two SLASH values cannot be realized. This strategy, however, does not work since counter examples like (27) are found.

(27) a. I wondered to which students this vase would be a good present

to give ___ ___ .

b. That famous old professor in Chicago, I can't remember which papers I sent copies of ___ to ___ .

In (27), both sentences are perfectly grammatical though they contain two independently motivated gaps. Hence, constraining the DRC is not the right way to exclude the ungrammatical sentences in (25).

The next alternative suggested by Ivan Sag is that the head-filler rule like (19) should not be applied recursively.[3] If this alternative is accepted, the difference in grammaticality between (25) and (27) follows. Specifically, sentences in (25) are ungrammatical because (19) is recursively applied, whereas those in (27) are grammatical because (19) is not applied recursively.

(19) Head-Filler Schema:

S	->	Filler-Dtr,	Head-Dtr Phrase
[SLASH { }]		[LOC[1]]	[SLASH {[1]}]

But the constraint of this alternative is too strong to explain complement extraction in considering that there are grammatical sentences, in which they contain two independently motivated gaps and the head-filler rule is recursively applied to them, as in (28)

(28) a. To Johnny a book, I gave ___ ___ .

b. A book to Johnny, I gave ___ ___ .

There seems to be a clue to avoid this difficulty, since the fillers in (25) are all NPs but those in (28) are an NP and a PP. On the basis of this observation, the head-filler rule must be revised as follows:

(29) When the head-filler rule is recursively applied in a sentence, the fillers' CATEGORY (Part of Speech) values should be different.[4]

---

3 This suggestion was made by I. Sag through personal communication.

Though the grammatical difference between (25) and (28) simply follows from the generalization on complement extraction in (29), it looks hard to extend it to account for the optional complement extraction phenomena in (30) and (31). The ungrammaticality of (30) can be correctly predicted by the generalization in (29), because both fillers have the same category value, namely PP. However, (29) wrongly predicts that sentences in (31) are all grammatical, since both fillers are two different category values.

(30) a. *Johnny the book, I talked about ___ to ___ .
b. *To Johnny about the book, I talked ___ ___ .
(31) a. ?To Johnny the book, I talked about ___ ___ .
b.*About the book Johnny, I talked to ___ ___ .

It is worthwhile noting that the grammatical degree in (31) is similar to those in (32). This grammatical similarity might suggest that optional complement extraction behave like adjunct extraction, rather than real complement extraction.

(32) a. (?)The apples yesterday, I ate.
b. *Yesterday the apples, I ate.

It is unclear at this point whether the shortcoming of my analysis is due to the insufficient constraints to the head-filler rule or the different distributional behavior in extraction between real complement and optional complement. This issue will remain for a future study.

5. Conclusion

It has been claimed that the DRC is empirically motivated but the

4 The implementation of this generalization into current HPSG format must be trivial, so it is not provided here.

head-filler phrase assumed in current HPSG must be revised in order for the traceless approach to succeed. To show that my claim is tenable, various seemingly problematic data for the DRC have been provided and then argued that they are counterexamples to the head-filler phrase, rather than the DRC. Furthermore, a generalization in (29) has been made on the basis of the observations in various types of complement extraction data. This generalization still has some challenges to explain some optional complement extraction, though it sufficiently accounts for the typical extraction cases. Nonetheless, my analysis might claim to be on the right tract since there is a similarity of distributional behavior in extraction between optional complement and adjunct. It will be finally decided when all extraction phenomena are completely analyzed. It must be, then, a long term project, and thus, the issue at stake remains for further research. Though my analysis of extraction phenomena is far from complete, it seems, however, to be obvious that this paper will contribute to constructing a unified traceless account of extraction phenomena in HPSG.

References

Bouma, Gosse, Rob Malouf, and Ivan Sag. 1997. Satisfying Constraints on Extraction and Adjunction. Ms. Groningen University and Stanford University.

Brodie, Belinda. 1983. *English Adverb Placement in Generalized Phrase Structure Grammar*. M.A. Thesis, Ohio State University.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan Sag. 1985. *Generalized Phrase Structure Grammar*. Cambridge, MA: Harvard University Press.

Pollard, Carl and Ivan Sag. 1987. *Information-based Syntax and Semantics*. CSLI.

Pollard, Carl and Ivan Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press, and Stanford: CSLI.

Sag, Ivan. 1996. Constraint-Based Extraction (Without a Trace). Ms. Stanford University.

Sag, Ivan. 1996. Notes on Phrases. Ms. Stanford University.
Selkirk, Elisabeth. 1984. *Phonology and Syntax*. Cambrige, MA: MIT Press.

Department of English
Honam University
59-1, Seobong-dong, Kwangsan-gu
Kwangju 506-090, Korea
E-mail: sycho@honam.honam.ac.kr
Fax: +82-62-940-5116