Weak Connectivity in (Un)bounded Dependency Constructions[∗]

Yong-Beom Kim

Kwangwoon University, Seoul , Korea

ybkim@kw.ac.kr

Abstract. This paper argues that various kinds of displaced structures in English should be licensed by a more explicitly formulated type of rule schema in order to deal with what is called weak connectivity in English. This paper claims that the filler and the gap site cannot maintain the total identity of features but a partial overlap since the two positions need to obey the structural forces that come from occupying respective positions. One such case is the missing object construction where the subject fillers and the object gaps are to observe requirements that are imposed on the respective positions. Others include passive constructions and topicalized structures. In this paper, it is argued that the feature discrepancy comes from the different syntactic positions in which the fillers are assumed to be located before and after displacement. In order to capture this type of mismatch, syntactically relevant features are handled separately from the semantically motivated features in order to deal with the syntactically imposed requirements.

Keywords: Weak connectivity, feature mismatch, missing object construction, passives, topicalized construction, filler, gap, unbounded dependency.

1. Introduction

This paper attempts to deal with filler-gap mismatches found in what Pollard and Sag (1994) calls the weak unbounded dependency construction. One of the key issues in the construction is that there is a feature or category mismatch between the filler and gap, as shown in (1).

(1) a. That John will help us is hard for us to rely on ___.
     b. That Chomsky might be wrong is hard to think of _____.

Prepositions like *on* or *of* would normally be subcategorized to take a nominal complement, but in (1) the filler site for the corresponding gap is occupied by a *that* clause, which would pose a puzzle to a constraint-based analysis as well as to a minimalist approach. This kind of mismatch is not peculiar to the missing object construction (MOC, hereafter). The same kind of mismatch can also be found in a topicalized construction as shown in (2).

(2) a. That some passives lack active counterparts, no theory can capture ____.
     b. *No theory can capture that some passives lack active counterparts.
     c. That some passives lack active counterparts, few teachers were aware of ___.

The verb *capture* does not take as its complement a *that* clause in an ordinary construction as shown in (2b), but in a topicalized construction like (2a), a *that* clause can initiate the construction with the same verb governing the gap site.

   This kind of weak connectivity is not restricted to the above two constructions. It can be found in passives as well.

(3) a. That some passives lack active counterparts can be captured by no theory.
     b. That Chomsky might be wrong has never been thought of ___.

---

cf. No one has ever thought (*of) that Chomsky might be wrong.

  c. That John will help us may not be relied on.

As shown in (3a) and (3b), a *that* clause can appear as the subject in these passives, but their active counterparts cannot take the clause-type constituent as their complement. There is also a less deviant and less pernicious kind of problem, as in the ordinary missing object construction, as shown in (4)

  (4) a. He/*Him is easy to please ___.

       b. They/*Them will take only five minutes to boil ____.

There arises a case clash between the subject and the corresponding gap in (4) but this issue has been more or less elegantly handled in the framework of HPSG. It is taken care of by requiring not the total identity but a partial identity of feature specifications. However, as we will see in section 2.1, such an account is not still satisfactory.

This paper will argue that the filler-gap relations in these constructions can in general be semantically accounted for, since the categorical mismatch is sometimes inevitable in syntax and should be allowed so far as other components of grammar require such mismatch.

2. Proposals

In this section we will claim that the filler-gap dependency cannot be a relation of total identity between the two elements but one of partial overlap, considering that the features may differ where other grammatical requirements or other structural forces dictate them to be non-identical. One of the proposals in this line of conception has already been presented in PSG framework (See Hukari and Levine (1991)). We will consider missing object constructions, first.

2.1 Missing Object Constructions (MOCs)

Feature mismatches in MOCs can be plainly seen in the examples in (1) through (4) where the 'displaced' main clause subject should be somehow related to the object position of the embedded clause. This construction is relatively easy to deal with since there is a licensing lexical item and its related lexical rule can mediate the feature mismatch. For instance, Hukari and Levine (1991) formulated a rule that ignores [CASE] feature in this construction, and Levine and Hukari (2006, p 355) also suggested a rule as shown in (5)

(5) partial representation of SYNSEM value for *tough* lexical rule

$$
\left[\text{LOCAL} \mid \text{CAT} \left[\begin{array}{ll} \text{HEAD} & adjective \\ \text{SUBJ} & <[1]_i>, \\ \text{COMPS} & < ..., \text{VP} \left[\begin{array}{l} \text{INHER|SLASH } \{[1]\text{NP}_i\} \\ \text{TO-BIND|SLASH } \{[1]_i\} \end{array} \right] > \end{array} \right] \right]
$$

This rule may take care of case clashes between the filler and the gap, since its identity requirement is limited to INDEX value. However, this rule wrongly demands that the filler and the gap should be simultaneously an NP. So the syntactic category mismatch shown in (1a) and (1b) is not taken care of by Levine and Hukari.

This is because the feature specification in the *tough* lexical rule explicitly mentions NP as the SLASH value. To correct this problem, we may loosen the lexical rule to such a degree that the filler and the gap may be minimally different from each other, and thus the SLASH value could be simply a phrase (i.e., XP) instead of an NP. The possible revised lexical rule would contain (6) as the VP specification as the value of COMPS.

(6) VP $\begin{bmatrix} \text{INHER|SLASH} \ \{ \ [1]_i \ \} \\ \text{TO-BIND|SLASH} \ \{[1]_i \ \} \end{bmatrix}$

With this revision, we could displace any type of category from the complement position in the MOC, and the sentences like (1a) and (1b) may be licensed by the grammar regardless of whether the preposition licensing the gap in the infinitival phrase can take S as its complement or not[1].

2.2 Passives

The types of passive sentences shown in (3) are not accounted for within the current `constraint-based` phrase structure grammar. The problems with these examples are twofold: one involves a reanalysis assumption; the other relates to category mismatch as in MOCs. These examples show that passives are not simply a process of demoting and promoting some arguments. If the passive sentences like (3b) and (3c) are to be generated, there should be a process that concatenates *relied* and *on*, in addition to the ordinary passive lexical rule since *rely on* or *think of* cannot be an input the Passive Lexical Rule (PLR) of Sag and Wasow (1999, p235) as it is. Even though the *verb-preposition* sequence can somehow be fed into the Passive Lexical Rule, the least oblique complement cannot be S-bar or S, since it would be governed by a preposition, not by the verb in question, in the active counterpart.

Furthermore, the PRL demotes the first element of ARG-STR value and promotes the least oblique argument in the remainder so that the category next to the main verb becomes the subject in ordinary cases. But this general rule would license the following ill-formed cases shown in (7a), (7b) and (7c). Furthermore, the well-formed ones in (7a'), (7b'), and (7c') are not generated by the PLR.

(7) a. *That John is a liar is not thought.
     a'. It is not thought that John is a liar
     b. *That John is a liar is said.
     b'. It is said that John is a liar.
     c. ?*That John is a liar has never been thought
     c'. That John is a liar has never been thought of.

I assume there is a slight difference in what are presupposed by (7a') and (7c'): (7a') does not carry as much factive presupposition as (7c') does. In order to take care of this situation, some parochial passive rules need to be posited. What is clear is that the Sag and Wasow's (1999) PLR cannot deal with the examples shown above and that it has to be flanked with subsidiary rules. One instance of such rules that can deal with (7c') would look like the one proposed in (8) and this would take care of the two problems: need for reanalysis and categorical mismatch.

In fact, the rule proposed by Sag and Wasow does not take care of the examples in (7a') and (7b'). This cannot be dealt with by a kind of extraposition operation, either, since the (7a) and (7b) are not well-formed sentences. Therefore, the passive forms without prepositions, like *thought* and *said* are to be dealt with separately from the ones with prepositions.

---

[1] This point may be debatable according to the theoretical assumptions within HPSG grammar because this statement is based on a particular theory-internal assumption that SLASH termination at the bottom of the UDC is performed by no special rule but is taken care of by such mechanisms like Argument Realization Principle (Sag and Wasow 1999:402). Then there seems to be no apparatus within this version that can incorporate this generalized category (XP, for instance) into specific rules because a gap is a product of lexicon-internal operation regarding argument structure, but not a rule-to-rule operation. The point is that the lexicon-internal operation cannot deal with generalized categories like XP but full-fledged syntactic categories regarding *part-of-speech* features.

(8) Passive Lexical Rule for some propositional verbs

$$
\langle [1],\ \begin{bmatrix} \textit{tv-lxm} \\ \text{AGR-ST} < NP_k,\ PP \begin{bmatrix} \text{PFORM } \alpha \\ \text{P-OBJ } NP_i \end{bmatrix} > \end{bmatrix} > \quad \Rightarrow
$$

$$
\langle\ F_{PSP}([1]),\ \begin{bmatrix} \textit{Word} \\ \text{SUB } < [\ \ ]_i > \\ \text{SYN } [\text{HEAD } [\text{FORM pass}]] \\ \text{AGR-ST} < P[\text{PFORM } \alpha],\ PP \begin{bmatrix} \text{PFORM } \textit{by} \\ \text{P-OBJ } NP_k \end{bmatrix} \end{bmatrix} >
$$

The rule in (8), first, takes a verb-preposition sequence and changes it to a passive form and concatenates the preposition to the passive form. Second, the prepositional object is promoted to a subject position but the promoted one only carries an index identical to INDEX value of the original object.

So, the passive subject need not match all the feature specifications of the (original) prepositional object. I assume, as Sag and Wasow do, verbal categories (or S) can carry exactly the same type of INDEX specifications as nouns, especially if the nouns are propositional nouns like *fact, proposal, assumption, conception*, etc.

Without this kind of rule, it would be impossible to generate appropriate structures for the above passive tokens[2]. This proposal will eventually deal with what has been put aside for the so-called reanalysis of a verb phrase.

2.3 Topicalized Constituents

Topicalized constituents carry some special discourse functions that make the constituents prominent and such discourse properties seem to be shared by MOCs and Passives as well. Investigating into this discourse property in full length would be beyond the scope of this paper, but I will talk briefly about this property, assuming that the topic phrase, passive subject and MOC's subject act as a *link* between the preceding utterance and the following. Consider the following sentence.

(9)   A: What happened to John?
       B: He was hit by a truck.
       Cf. A truck hit him.

(10) A; I hate these math problems.
       B: Yeah, those are very difficult to solve.
       Cf. Yeah, it is very difficult to solve those.

(11) A: We should reexamine various kinds of premises before going any further.
       B: Did you know that John is Brazilian?
       A: John is Brazilian, I have never thought of ___.
       Cf. I have never thought that John is a Brazilian.

What is common among these constructions is that the constituents in question are discourse oriented and, thus, their grammatical functions are somewhat 'diluted' in the sense that the

---

[2] There is a possibility of over-generation of ill-formed structures with this revision and a possible solution to this problem will be mentioned at the end of section 2.3.

phrases in question are not adjacent to their lexical governor. In some cases, the meanings are not identical. For example, the third utterance in (11) is not identical to the compared pair in meaning and it may be the more appropriate form of utterance than the other in this context.

So the category mismatch is somewhat related to non-adjacency to the lexical governor and this is common to many discourse-oriented constructions. So the solution to the mismatch should reflect this property. So the requirement imposed on this construction will be whether it is interpretable on a discourse level as well as on a semantic level. In the same way as we dealted with passives, the rule schema for the topicalized construction should be loosened as shown in (12):

(12) Head-Filler Rule (Revised)

$$
\begin{bmatrix} phrase \\ \text{GAP} < > \end{bmatrix} \rightarrow \begin{bmatrix} phrase \\ \text{SYN} \mid \text{GAP} < > \\ \text{SEM} \mid \text{INDEX} \quad i \end{bmatrix} \quad \text{H} \quad \begin{bmatrix} phrase \\ \text{FORM} \quad \text{fin} \\ \text{SYN} \quad \text{SPR} \quad < > \\ \text{GAP} \quad <[1]_i > \\ \quad \text{STOP-GAP} <[1]_i > \end{bmatrix}
$$

What this rule does is to allow the mismatch between the filler and gap as long as they share a certain semantic property, that is, the same index. On a discourse level, some propositions can be referred to by nouns like *the fact*, *the proposal*, *the assumption*, and so on. This type of anaphoric reference may be reflected partially in the rule in (12). This formulation of head-filler rule may over-generate some ill-formed sentences like the ones in (13a) and (13c).

(13) a. *That John is Canadian, we put the book on [$_{NP}$ e].
     b. #We put the book on the fact that John is Canadian.
     c. *The table, we put the book [$_{PP}$ e].

In this paper, the solution to this over-generation will be sought based on the (im)possibility of a kind of reference or processing. In (13a), the awkwardness comes from semantic mismatch since the location role imposed by semantics of the verb *put* cannot be satisfied by the noun *fact*. This is why (13b) is also ill-formed. In other word, (13a) is not acceptable for the same reason that (13b) is unacceptable. As for (13c), we would say that there is no known (co-)reference relation between NP and PP, unlike the case of NP and S-bar[3]. Let us compare (13c) with (14):

(14) a. [That John is Canadian]$_i$, we have never thought of [$_{NP}$ e] $_i$.
     b. We have never thought of [the fact] $_i$ [that John is Canadian] $_i$.

The kind of relation that is frequently found between NP and S-bar within or across sentences as shown in (14) cannot be found between NP and PP. We think this is why (13c) is unacceptable.

The loosened head-filler rule in (12), however, may over-generate the ill-formed ones in (15b) and (15c):

(15) a. Him, I very much doubt anyone likes _____.
     b. *Him, I very much doubt nayone likes _____.
     c. *He, I very much doubt anyone likes _____.
     d. He, I very much doubt _____ likes anyone.
        (Levine and Hukari, 2006: 23)

---

[3] The same type of over-generation is expected with passives and MOCs, but we can provide the same kind of solution that is derivable from the filler-gap construction, as we can see regarding the examples in (13).

Since the freer head –filler rule places no restriction on the possibility of diversity in features between filler and the gap, it in fact does generate all the sentences in (15) and it should be restricted or tightened. The dilemma is that we need a rule tighter than the one in (12) while we don't want a rule that is as tight as shown in (16):

(16)
$$\begin{bmatrix} phrase \\ \text{GAP} < \ > \end{bmatrix} \rightarrow [1] \begin{bmatrix} phrase \\ \text{GAP} < \ > \end{bmatrix} \text{H} \begin{bmatrix} phrase \\ \text{FORM} \quad fin \\ \text{SYN} \quad \text{SPR} \quad < \ > \\ \text{GAP} \quad <[1] > \\ \quad \text{STOP-GAP} <[1]> \end{bmatrix}$$

A solution to this dilemma may be found by revising the rule in (12) so that the CASE feature should be relevant in this rule. Of course, only nominal (and pronominal) categories will always bear the CASE feature and clauses will never carry any CASE feature. Therefore, we will need a non-defeasible constraint like (17):

(17) (1) No clause-type categories carry the CASE feature.
    (2) Every (pro)nominal categories carry the CASE feature.

Given the above constraints which are needed independently, we can devise a separate schema as shown in (18):

(18) Head-Filler Rule (proposed)[4]

$$\begin{bmatrix} clause \\ \text{GAP} < \ > \end{bmatrix} \rightarrow \begin{bmatrix} phrase \\ \text{GAP} < \ > \\ \text{CASE} [2][\ \beta\ ]/\text{NIL} \\ \quad \text{SEM} |\text{INDEX} \quad i \end{bmatrix} \text{H} \begin{bmatrix} phrase \\ \text{FORM} \quad fin \\ \text{SPR} \quad < \ > \\ \text{GAP} <[1][\ \text{CASE} [2]]> \\ \quad \text{STOP-GAP} <[1]> \\ \text{SEM} |\text{INDEX} \qquad i \end{bmatrix}$$

By introducing the feature CASE and allowing its value to alternate between a certain value and NIL, the proposed Head-Filler Rule can achieve the effect of collapsing two different rules in conjunction with the constraint stated in (17). One rule will conatain the CASE feature in the filler position and it will generate a clause in which (pro)nominal phrases are extracted; and the other will not carry the CASE feature and it will allow for cases where clauses or non-nominal phrases are displaced.

3. Conclusions

This paper has proposed that categorical mismatch can be accounted for by revising the filler-gap rule or setting up some parochial rules within an HPSG framework. There seem to be a lot of cases that require a detailed refinement or re-examination in the grammar. We also need to tighten up the semantic and pragmatic portion of grammar in order to prevent (over)generation of ill-formed sentences. Especially, more research is also need in the discourse functions of these constructions, which seems to call for a larger paper.

References

Ginzburg, J. and I. A. .Sag 2001. *Interrogative Investigations*. Stanford: Center for the Language and Information.

---

[4] The slash notation in this rule renders one of the the values around it present in the structure.

Hukari, T. and Levine, R. D. 1987. Rethinking Connectivity in Unbounded Dependency Constructions. In Megan Crowhurst ed. *Proceedings of the Sixth West Coast Confernece on Formal Linguistics*, Stanford: Stanford Linguistics Association.

Hukari, T. and Levine, R. D. 1991. On the Disunity of Unbounded Dependency constructions, *Natural Language and Linguistic Theory* 9:97-144.

Jacobson, P. 1984. Connectivity in Phrase Structure Grammar, *Natural Language and Linguistic Theory* 1: 535-581

Levine, R. D. and T. Hukari. 2006. *The Unity of Unbounded Dependency Constructions*, CSLI Lecture Notes #126, Stanford, California: CLSI Publications.

Pollard, C. and I. A. Sag. 1994. *Head-Driven Phrase Structure Grammar*, Center for the Study of Language and Information, Stanford. California: University of Chicago Press.

Sag, I. A. and T. Wasow. 1999. *Syntactic Theory*, *A formal Introduction*, Stanford University: CSLI Publications.