Multiple movement dependencies and parasitic gaps

ISAAC GOULD

Ewha Womans University

gould@ewha.ac.kr

1. INTRODUCTION

Nissenbaum (2000) and Heck and Himmelreich (2017) (henceforth HH) are two prominent works that are noteworthy for focusing on the interaction of multiple movement dependencies (MMDs) and parasitic gaps (PGs).[1] However, these two works consider disjoint and contrasting types of data paradigms. In this squib, I take both types of paradigms into consideration. In doing so, there are three inter-related questions that arise for our understanding of PGs (and in particular for PGs that involve certain MMDs). The first question asks what the correct descriptive generalizations about PG paradigms are. Going beyond such a generalization, we can also ask how to properly account for the data (and whether a unified analysis is possible). Finally, and more specifically, we can ask how to account for the compositional semantics of PGs, especially those involving MMDs.

Nissenbaum presents a generalization based on English data regarding (a) the PGs in a PG-containing *v*P-adjunct, and (b) the XPs that move to the edge of a *v*P immediately dominating such an adjunct, crucially attaching above the adjunct. For convenience, I will refer to these XPs as edge-XPs. As we will see, Nissenbaum's generalization is that in these specific circumstances, each edge-XP must associate with a PG in the adjunct (i.e., be interpreted as the filler for the gap of a PG). However, HH introduce a data paradigm from German that challenges this generalization.[2] I also present a tweaked version of the generalization based on Nissenbaum's analysis that is again challenged by the data. In particular, we will see the possibility of multiple edge-XPs where only one of them associates

[1]The following abbreviations are used: HNPS: Heavy NP Shift; MMD: multiple movement dependency; PG: parasitic gap; PM: Predicate Modification.

[2]Note this is a paradigm that properly includes supplemental data from Heck and Himmelreich (p.c.).

with a PG. This disparity raises the question of what the correct descriptive generalization of the PG paradigms should be.

When we consider the theoretical proposals of Nissenbaum and HH in sections 2 and 3, we will see that neither can account for the full set of attested data. Nissenbaum cannot fully account for the German data, and conversely, HH cannot fully account for Nissenbaum's data. We will see that it is not clear how to properly account for the data, and whether a unified analysis of the data is possible.

Finally, there is the related question of what the compositional semantics of the PG construction is. In contrast to HH, Nissenbaum focuses on the semantics of PGs, and indeed, he attempts to derive the generalization he proposes in a principled way via the nature of semantic composition. We will see that within the framework of these proposals, Nissenbaum's semantic treatment of the construction is not able to account for Heck and Himmelreich's data. We are then left with the broad question of how we can account for the compositional semantics of PGs, especially those involving MMDs.

Before continuing, I note two things in relation to the literature on German. First, PG data involving putative MMDs has appeared earlier in the literature (e.g., Fanselow 1993: 34, Müller 1995: 261–264, and Kathol 2001: 329). However, HH are noteworthy in developing a data set that allows for a clearer focus on the interaction of multiple A'-dependencies. HH's data, as far as I can tell, help us to see a potential confound (see section 3 and note 8) in the data that have been presented in all previous works when it comes to considering the application of Nissenbaum's generalization to German. In particular, HH's data set points us toward supplemental Heck and Himmelreich data (p.c.) that do not suffer from this potential confound. One reason to therefore focus on the current paradigm is its more comprehensive scope regarding A'-dependencies and PGs. Their proposal also distinguishes itself from what is found in these earlier works in having a detailed and restrictive theory for PGs involving MMDs (see Fanselow 1993: 35 for a sketch of a proposal). Looking at the various HH examples allows us to see clearly how their proposal works (and I note that applying HH's proposal to earlier German data is less than straightforward because of differences in reported judgments). Second, it should be mentioned that Kathol (2001: 329) suggests in passing that there is a difference between English and German regarding the disparity between the number of edge-XPs and PGs. Importantly, though, Kathol (along with other earlier works) does not discuss the theoretical and empirical significance of such an observation, which is the primary contribution of this squib.[3]

In what follows, I first review in section 2 some of Nissenbaum's data and discuss how the data support/are accounted for by Nissenbaum's generalization. Then, in section 3, I review the alternative paradigm raised by HH, discuss how it

---

[3]Kathol (2001) assumes that German lacks PGs altogether (see Haider and Rosengren 2003). The core of this claim rests on differences between German data on the one hand, and what are considered canonical PG data found in other languages on the other; but such a claim is undermined by the kinds of cross-linguistic differences reviewed in Culicover and Postal (2001). See also Assmann (2010) for a critical review of Kathol's position.

does not conform to Nissenbaum's generalization, and review how HH's proposal can account for the new data (at least from a syntactic perspective). Finally, in section 4, I discuss how neither proposal can account for the full data set, and I highlight the challenge that this poses for a compositional account of PGs.

2.   NISSENBAUM (2000)

Nissenbaum presents a variety of data types which support his generalization that each edge-XP must associate with a PG in the adjunct. These include examples involving single or multiple edge-XPs, as well as edge-XPs that move both overtly and covertly. Further, the examples involve edge-XPs undergoing a range of movement dependencies, such as *wh*-movement, clefting, relative clause formation, and Heavy NP Shift (HNPS). As the focus here is on MMDs, I will use one set of examples involving two edge-XPs as an illustration of the generalization. See Nissenbaum (pp. 98, 112, and 117) and also Williams (1990: 277) for further illustrative sets of examples with two edge-XPs (moved overtly or covertly) involved in different types of movement dependencies.

      Let us now consider how the examples in (1) from Nissenbaum (p. 113–114) can be taken to support Nissenbaum's generalization. These examples involve leftward, cyclic *wh*-movement to the edge of vP and rightward HNPS movement to the edge of vP.[4] Setting aside the height of the adjunct for the moment, first notice the following. In (1a), which is largely acceptable, two XPs move to the edge of vP, and the adjunct *after putting a copy of next to* correspondingly contains two PGs. In contrast, the adjuncts in both (1b) and (1c) contain only one PG, which associates with just one of the XPs, and these examples are ungrammatical.

(1)   a.   ? Which book$_1$ did Smith find $\_1$ on top of $\_2$,
            after putting a copy of $\_{PG1}$ next to $\_{PG2}$, the table in the corner$_2$?

      b.   * Which book$_1$ did Smith find $\_1$ on top of $\_2$, after reading a review of $\_{PG1}$,
            the table in the corner$_2$?

      c.   * Which book$_1$ did Smith find $\_1$ on top of $\_2$, after wiping $\_{PG2}$ with a sponge, the
            table in the corner$_2$?

Second, note that Nissenbaum proposes that the relevant XPs are indeed edge-XPs in that they are higher than the adjunct. That is, each adjunct in (1) is in a position below the two edge-XPs and above the vP-internal position of the subject *Smith*. Given the order of rightward attachment, with the Heavy NP following the adjunct, we can assume that the adjunct is indeed below the *table*-Heavy NP. As for the relative position of the *wh*-phrase and the adjunct, Nissenbaum assumes that the adjunct must be below the *wh*-phrase because the Heavy NP must have merged below the *wh*-phrase. This latter assumption is derived via two auxiliary assumptions. The first is that as the *wh*-phrase is closer to v than the Heavy NP, it will move to the vP edge first (see

---

[4]See Nissenbaum (pp. 113–114) for arguments that these examples do indeed contain HNPS movement (and not right node raising).

Richards [1997]). Second, Nissenbaum (p. 101) proposes the constraint in (2), which is modeled on a similar constraint in Richards ([1997]).

(2) Tucking-in condition: Movement does not extend the tree if an alternative [to extending the tree] exists (it must tuck in *below the outermost segment* whenever possible).

Nissenbaum does not fully flesh out when it is not possible to tuck in, although his discussion does indicate several conditions where tucking does not occur. Relevant here is that tucking in is impossible if the tucking in intervenes between a thematic argument and the predicate selecting that argument. Thus in the derivation of (1), we first have movement of *which book*, which can extend the tree by merging with *v*P in a position above the external argument *Smith*, which is a thematic argument of *v* (see (3) below).[5] Subsequently, we have movement of the Heavy NP to the *v*P edge, where it must tuck in below the *wh*-phrase, and above the subject *Smith*.[6]

The relevant hierarchy above *v* in (1a) is summarized in the annotated tree in (3), which I discuss further below. Given the hierarchical position of the adjunct and the observation in (1) that two PGs are necessary, we can see how the MMD data involving two edge-XPs in (1) support Nissenbaum's generalization that each edge-XP associates with a PG.

(3)
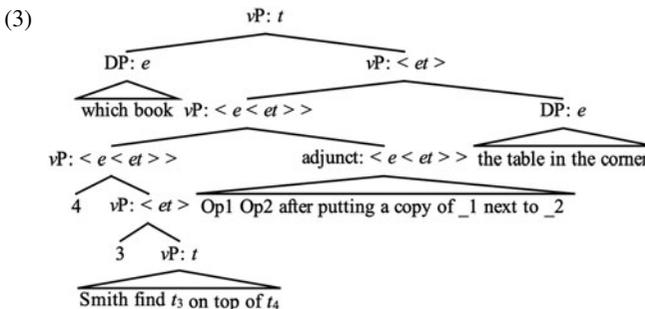
Nissenbaum accounts for the data above and his generalization by relying on principles of semantic composition, given the internal structure of the *v*P and the adjunct. As for the structure of the PG-containing adjunct, Nissenbaum assumes it is formed via movement of semantically vacuous null operators (Ops) (see Chomsky [1986]). This operator movement turns the adjunct into a derived predicate, and as there are

---

[5]Note that the assumption here is that subject, when raising out of the *v*P, does not move and merge again with the maximal projection of *v*P: as an external argument, its base position is already high enough to escape the *v*P phase without merging again with *v*P.

[6]Nissenbaum is somewhat agnostic about when the adjunct is inserted. It must be inserted after movement of the *wh*-phrase. If it were not, then both the *wh*-phrase and Heavy NP would need to tuck in below the adjunct. This would result in the following problems: (a) the *wh*-phrase and/or the Heavy NP could not bind the appropriate PG in the adjunct in (1) (see (4) below); and (b) the Heavy NP would presumably not appear to the right of the adjunct as it does in (1). However, the adjunct can be inserted below the Heavy NP before or after the Heavy NP moves. Either way, such HNPS movement will satisfy (2) by tucking in below the *wh*-phrase.

two instances of operator movement, the result is a two-place predicate over individuals, as indicated in (3).[7] Such an open predicate cannot semantically compose with the thematically saturated lowermost vP in (3). However, this vP itself becomes a derived predicate via movement of the edge-XPs. This movement to the edge of vP is not semantically vacuous here, and accordingly will leave a lambda-binder, resulting in a derived vP-predicate (see Heim and Kratzer 1998). And as there are two instances of edge-XPs moving, the resulting vP in (3) is also a two-place predicate over individuals.

With this structure in hand, we can now see the crux of Nissenbaum's proposal. Because there is predicate-forming edge-XP movement, the adjunct can now compose as an intersective modifier with vP via Predicate Modification (see Heim and Kratzer 1998 for a definition of Predicate Modification), the result of which is given schematically in (4) for the vP in (1a).

(4) which book [the table in the corner $\lambda x.\lambda y$ [Smith find y on top of x
                              $\wedge$ after putting a copy of y next to x]]

Crucially, the deviance of (1b) and (1c) now follows from mismatches of semantic types. In those examples there is only one PG in the adjunct, meaning that the adjunct (with only one instance of operator movement) will be a one-place predicate over individuals. Nevertheless, there are still two edge-XPs moving to positions that are higher than the adjunct, resulting in a two-place vP predicate that must combine with the adjunct. Predicate Modification can no longer apply between the two mismatching derived predicates, and the adjunct cannot be semantically composed, resulting in ungrammaticality. The consequence is that each edge-XP must associate with its own PG.

Note that the logic here generalizes to examples with different numbers of edge-XPs (although for the sake of brevity, I do not discuss such examples here). Because (non-semantically vacuous) movement of each edge-XP increases the valency of the derived vP predicate, it forces there to be a corresponding valency-increasing movement of a null operator linked to a PG within the adjunct. In this way Nissenbaum captures his generalization regarding the correspondence between the number of edge-XPs and their associated PGs.

3. HECK AND HIMMELREICH (2017)

Heck and Himmelreich introduce a type of paradigm from German PGs containing MMDs involving *wh*-movement and scrambling that challenges Nissenbaum's generalization. The relevant data are in (5): the examples in (5a), (5c), and (5d) come from HH (pp. 52, 53), and are supplemented by (5b) from Heck and Himmelreich (p.c.).

---

[7]The tree in (3) is slightly different from Nissenbaum's presentation in that it uses indices. This is done for ease of exposition, and, as far as I can tell, this does not result in any differences in predictions as regards the PG data. Note also that I follow Nissenbaum in abstracting away from additional semantic details of the vP, such as event variables, and for simplicity treat the saturated vP as type t.

(5) a. * Wem$_1$ hat   der Fritz das Buch   [anstatt __$_{PG1}$ zu helfen]
         who.DAT has the Fritz the book.ACC   instead   to help
         weggenommen?
         away.taken
         'Who did Fritz take the book from instead of helping him?'

   b. Wem   hat   der Fritz das Buch$_1$ [ohne __$_{PG1}$ selbst lesen zu wollen]
      who.DAT has the Fritz the book.ACC without   self   read to   want
      weggenommen?
      away.taken
      'Who did Fritz take the book from without wanting to read it himself?'

   c. Was$_1$ hat   der Fritz der Maria   [anstatt __$_{PG1}$ wegzuwerfen] zu
      what.ACC has the Fritz the Maria.DAT instead   away.to.throw to
      essen angeboten?
      eat offered
      'What did Fritz offer Maria to eat instead of throwing it away?'

   d. Wem$_1$ hat   der Fritz [anstatt __$_{PG1}$ zu helfen] das Buch
      who.DAT has the Fritz   instead   to help   the book.ACC
      weggenommen?
      away.taken
      'Who did Fritz take the book from instead of helping him?'

In (5a), the dative *wh*-phrase and the scrambled accusative DP are the edge-XPs that move from post-adjunct positions to pre-adjunct positions. The *wh*-phrase moves cyclically to the edge of *v*P on its way to the CP level, while the scrambled accusative moves only as high as the *v*P edge. HH observe that in this configuration, the accusative blocks the dative from associating with (and licensing) the adjunct's lone PG. In contrast, with the same configuration in (5b) (modulo adjustment of the adjunct's predicate for pragmatic reasons), the accusative can associate with the single PG in the adjunct. The *wh*-phrase can associate with the single PG given the dative and accusative internal arguments in the these examples, but only if either (a) the *wh*-phrase is accusative, as in (5c), which parallels (5b) in that it is the accusative DP that associates with the PG; or (b) the accusative does not scramble past the adjunct (5d).

   Given this description of (5), (5b) and (5c) appear *prima facie* to be counterexamples to Nissenbaum's generalization. Both involve two edge-XPs, but only one of them associates with a PG in the adjunct. It should be noted, though, that unlike the MMDs that Nissenbaum considers, the data in (5) involve a scrambling dependency. A question that arises now is whether a scrambled edge-XP is actually relevant for Nissenbaum's generalization, or a tweaked version of it. In other words, perhaps (5b) and (5c) are not counterexamples, and (a tweaked version of) Nissenbaum's generalization can be maintained. I will now discuss in what sense a scrambled edge-XP could in principle be irrelevant for the generalization. But then we will see how the data in (5) – (5b) in particular – force us to abandon such a view, meaning that at the very least (5b) appears to be a genuine counterexample to Nissenbaum's generalization.

Consider again Nissenbaum's generalization and his proposal to account for it. First, the generalization is about edge-XPs, which crucially must have moved to the edge of *v*P. One approach to reconciling the data in (5) with the generalization, then, would be to suppose that the scrambled DP is not in fact an edge-XP because it does not always move; rather it can be base-generated at the *v*P edge (see Bayer and Kornfilt 1994). The issue above regarding having two edge-XPs but only one PG might then disappear because there would be only one edge-XP, namely the *wh*-phrase. I return to this possibility below. Alternatively, recall that in Nissenbaum's analysis it is crucial that the movement of the edge-XP is predicate deriving, and that this derived predicate composes at LF with the adjunct. Accordingly, we could tweak the generalization such that the edge-XPs must be understood as undergoing non-semantically vacuous movement to the edge of *v*P, thereby introducing a lambda-binder at the *v*P level. With this in mind, we might attempt to reconcile the data in (5) with this tweaked generalization by supposing that a scrambled DP does not always leave a lambda-binder at the *v*P level at LF. This is conceivable if the scrambling in (5) is purely PF movement (and thus not predicate deriving). It could also be possible if the scrambling occurs in the syntax and introduces a lambda-binder, but if at LF this scrambling is effectively undone, with the lambda-binder being deleted and the scrambled DP being interpreted in its base position (see reconstruction in Hornstein 1995). Either of these latter scenarios again might in principle allow for the issue above to disappear; this time it is because the data would involve only a one-place derived predicate at LF via edge-XP movement (namely, movement of the *wh*-phrase) and one PG.

How well do these three ideas (base-generation, PF movement, lambda-binder deletion) actually fare with (5b) and (5c)? Although (5c) is in principle amenable to any of these three suggestions, (5b) is not in any clear way compatible with them. This is because it is the scrambled DP that associates with the PG. It is widely held that it is A'-movement of a PG's associate that licenses the PG. If we assume such movement to the edge of *v*P in (5b), then doing so rules out base-generation of the scrambled DP. Further, a PG's associate semantically binds the gap (HH: 69). This rules out the latter two suggestions. For this variable binding to be possible given movement of the scrambled DP, the scrambled DP must introduce a lambda-binder via movement to its scrambled position, and this lambda-binder must remain at LF.

In sum, in (5b) at the very least, remains as a valid counterexample to Nissenbaum's generalization. We must treat the scrambled accusative as a relevant edge-XP. Further, (5b) also contains a *wh*-phrase edge-XP that, via its cyclic A'-movement through the edge of *v*P, also leaves a lambda-binder at the *v*P level. Consequently, in (5b) there are two edge-XPs in the relevant sense, but only one of them (the scrambled one) is associated with a PG.[8] This now raises the first of

---

[8]The German data from the literature noted in section 1 that are potential counterexamples to Nissenbaum's generalization parallel (5c). In much of these data we see a scrambled full DP nominal in the *Mittelfeld*, as in (5c), that does not associate with a PG. These data could be treated along the lines of the suggestions in the text, and this would mean there is a potential

the questions from the introduction, namely the question of what the correct gener-
alization is regarding a restriction between edge-XPs and the number of PGs.
I will not attempt to address this question, my goal here being simply to draw atten-
tion to the question itself. In section 4, I discuss a consequence that emerges from the
discussion here, but in the remainder of this section I first review HH's proposal for
how to account for the type of paradigm in (5).

The core of HH's proposal is as follows. They assume that before scrambling or
wh-movement occurs, both of which are triggered via attraction by *v*, the adjunct (a)
adjoins to the left of a *v*P phrase marker, as per German word order; and (b) within the
VP, the dative DP is structurally higher than the accusative (6a) (see discussion of the
feature subscripts in (6) immediately following (6)). Along with Nissenbaum, HH
assume that the adjunct is formed via null operator movement to its edge. In contrast
to Nissenbaum, though, HH attempt to account for the data without any counter-cyclic
movement, meaning that tucking in is not a viable theoretical option. Instead HH assume
that when the two DPs move to the *v*P edge (as in (5a–c)), their relative order will be
preserved at the *v*P level by stacking them in a first-in, last-out buffer system. As the
dative DP is structurally higher, it will be attracted by *v* first, and will be stacked first
into a buffer hosting the constituents that *v* attracts. The accusative DP is attracted
second, and is placed on the top of the stack in the buffer. As *v* is now done attracting
constituents, the objects in the buffer will now merge with *v*P in the reverse order in
which they were stacked. As the accusative DP is on the top of the stack, it will
merge first (6b), followed by the dative on the bottom of the stack (6c). Later, when
C is merged it will attract the *wh*-phrase, giving us the word orders in (5a-c).

(6) a. $[_{vP} [\text{Op}_{1[\_F]} \ldots \underline{\quad}_{PG1} \ldots] [_{vP} \ldots [_{VP} \text{DP}_{DAT[xF]} \ldots [_{VP} \text{DP}_{ACC[yF]} \ldots]]]]$

     b. $[_{vP} \text{DP}_{ACC[yF]} [_{vP} \text{Op}_{1[yF]} \ldots \underline{\quad}_{PG1} \ldots] [_{vP} \ldots [_{VP} \text{DP}_{DAT[xF]} \underline{\quad}_{ACC}]]]]$

     c. $[_{vP} \text{DP}_{DAT[xF]} [_{vP} \text{DP}_{ACC[yF]} [_{vP} \text{Op}_{1[yF]} \ldots \underline{\quad}_{PG1} \ldots] [_{vP} \underline{\quad}_{DAT} \underline{\quad}_{ACC}]]]]$

Crucially, HH assume that the PG is licensed via feature valuation under Agree as
soon as possible by a c-commanding probe (see Chomsky 2001). In contrast to
Nissenbaum, HH assume that the null operator has an unvalued feature [_F] that
must be valued via Agree by a c-commanding DP probe with a valued feature
[xF], and it is agreement with this DP that determines which DP the PG will associate
with via binding (see HH for further details on these features). All DPs are such
probes, and they need not Agree with a goal for the derivation to converge, but as
soon as a DP can Agree with a goal, it must, after which it can no longer probe.

---

confound in treating such data as genuine counterexamples. In the remainder of these data from
the literature, instead of a full DP nominal, there is a weak pronoun that does not associate with
a PG. However, such pronominals are also in principle amenable to the suggestions above. As
best I can tell, HH's (5a) contrasts with all earlier data from the literature in clearly suggesting
the possibility of the two edge-XPs in a crucial data point such as the supplemental (5b). This is
done by showing how a scrambled DP can interact with a clear case of A'-movement
(*wh*-movement), suggesting the possibility of two A'-movement dependencies (and thus two
edge-XPs) in (5b). The juxtaposition of (5b), then, with examples of type (5c), which poten-
tially have only one A'-dependency/edge-XP, helps make the potential confound clear.

Thus, in (6b) the accusative DP Agrees with Op and values its feature. This valuation results in Op no longer being an active goal for subsequent DP probes, such as the dative DP in (6c), and fixes the interpretation of the PG as being necessarily associated with the accusative DP.

The consequence is that DP arguments that are merged lower in the VP will prevent DP arguments that are merged higher in the VP from associating with the lone PG in the structure, because the former will merge with *v*P and Agree with Op before the latter can. This is what we see in (5a–c). In these examples, the accusative DP, which is lower in the VP, Agrees with Op first. Doing so (i) prevents the dative DP from Agreeing with Op and associating with the PG in (5a), and (ii) allows the accusative DP to associate with the PG in (5b) and (5c). As Agree involving the dative DP is blocked in (5a), then, the association indicated by co-indexation in the example is illicit, and HH predict correctly the ungrammaticality of this example. And in the examples in (5b) and (5c), once the accusative DP Agrees with Op, the dative DP is free to not Agree, and thus HH correctly predict these examples to be grammatical. Finally, we can note that as only the *wh*-phrase is attracted by *v* in (5d), it is able to Agree with Op and associate with the PG.[9]

In sharp contrast to Nissenbaum, HH thus propose shifting much of the explanation of their data to the syntax, crucially relying on the mechanics of Agree and buffering. And as will be discussed further in the following section, there is nothing in these mechanics that will regularly force each edge-XP to associate with a PG, and this again is a point of contrast with Nissenbaum's proposal. Thus we saw in (5b)/(6c) that the dative DP does not need to Agree with any Op, meaning that it is possible for this edge-XP to not associate with a PG.

4. FURTHER CHALLENGES OF THE DATA

I begin this section by considering how the proposals in both Nissenbaum and HH struggle to account for all the data discussed so far. Thus I first focus on the grammaticality predictions that the two proposals make based on what the authors discuss themselves. For Nissenbaum, this will involve the compositional semantics, but for HH this will not. Later in the section, I consider the broader semantic picture for the data and proposals at hand.

First, Nissenbaum's proposal struggles with the German data that do not conform to his generalization(s) discussed above. For the sake of simplicity, in the discussion here I will treat scrambling as always leaving a lambda-binder at LF, but nothing crucial hinges on this, and the full data set from section 3 remains problematic even if this lambda-binder remains at LF only in (5b). Recall that for Nissenbaum, each operator that moves in the PG-containing adjunct is predicate

---

[9]HH observe that external arguments never block internal arguments from associating with a PG. They assume that when internal arguments are attracted by *v*, they internally merge with *v*P before the external argument first-merges with *v*P, thereby allowing the internal arguments to Agree with Op before the external argument ever has a chance to probe and do so.

deriving, and that this adjunct must combine via Predicate Modification (PM) with a *v*P of the same type.

Examples (5a) and (5d), then, are predicted correctly to be ungrammatical and grammatical respectively. In (5a), the adjunct is a one-place predicate, but the *v*P it needs to combine with, which is below the two edge-XPs, is a two-place derived predicate as a result of the two edge-XPs moving. These semantic types do not allow for composition via PM, hence the ungrammaticality. But in (5d), a comparable adjunct can compose via PM with the *v*P, which is below the edge-XP, because this *v*P is similarly a one-place predicate as a result of only one edge-XP moving.

However, (5b) and (5c) are predicted to be ungrammatical, contra the judgments reported above. In both these examples we have a one-place predicate adjunct combining with a *v*P below two edge-XPs. But because there are these two edge-XPs, it means that this *v*P will be a two-place predicate. As in (5a), then, composition with the adjunct via PM should be impossible, and according to the Nissenbaum's proposal, examples (5b) and (5c) should be ungrammatical just as (5a) is. That they are not ungrammatical, then, remains a puzzle for Nissenbaum.[10]

Similarly, some of Nissenbaum's English data from section 2 is challenging for HH's proposal. Consider first example (1a), which has two edge-XPs and two PGs in the adjunct, which is merged below these edge-XPs. Such a construction is in principle compatible with HH's proposal, which would conform with the acceptability given to (1a). In (1a), the hierarchically lower of the two edge-XPs (the Heavy NP), when it is merged first with *v*P, could Agree with one of the operators in the adjunct, and then the second edge-XP (the *wh*-phrase) could agree with the other operator in the adjunct. The requirements of the Agreement dependencies having

---

[10]A reviewer suggests an approach to (5) that would make the data consistent with Nissenbaum's generalization, but it is not clear how this approach would be tenable. The approach attempts to capitalize on the structure of ditransitives in assuming an ApplP phase (presumably) below a *v*P phase. Crucially, the approach assumes that an adjunct with a PG licensed by an associating accusative DP attaches to ApplP, whereas adjuncts with a PG attach to *v*P when that PG is licensed by a dative DP. If we assume that dative DPs are selected arguments by the Appl-head, and that accusative DPs originate lower in the VP, then it is technically true that Nissenbaum's analysis could be applied to capture the data in (5). Space limitations preclude detailed discussion of such an account, but I believe two compelling complications can be succinctly mentioned here. First, this approach depends on different attachment sites of the adjunct, sites that depend on the case of the PG's associate. Short of some ad hoc stipulation (perhaps specific to PG constructions), it is not clear what would enforce these particular attachment restrictions. After all, if an adjunct containing a PG can sometimes attach to, for example, *v*P, why would it not always be able to do so? Second, this approach faces a significant difficulty with grammatical German examples involving adjuncts that contain two PGs (again, not discussed here for space reasons; see HH, pp. 54, 73), one with an accusative associate, and one with a dative associate. It is not clear how such an adjunct could possibly be merged into the structure, as it would seemingly run afoul of the competing attachment restrictions from above: given the contrasting cases of the associates, there are seemingly contradictory requirements on the adjunct to attach to both *v*P and ApplP.

been thus satisfied, the construction is predicted correctly to be acceptable. At least one of the examples in (1b) and (1c), however, is also predicted to be grammatical, contra the judgments reported above. These examples parallel examples (5b) and (5c) in the relevant respects: there is an adjunct containing a single PG, which merges below two edge-XPs. Recall that when there are two probes (the two edge-XPs) and only one goal (the operator in the adjunct), after one probe Agrees with the goal, the remaining probe cannot Agree with the goal, and the grammar allows that remaining probe to not undergo Agree. Thus just as we saw only one probe Agree in (5b) and (5c), we expect the same to be possible in either (1b) or (1c) without this leading to ungrammaticality. That both (1b) and (1c) are in fact ungrammatical constitutes a puzzle now for HH.

In sum, Nissenbaum's proposal cannot account for all the data in section 3, and HH's proposal cannot account for all the data from section 2. This raises the question of what a unified analysis of all the data considered here might look like, or whether it is possible.

One might now consider the analytical possibility that there is cross-linguistic variation, such that in some grammars (e.g., that of English), an approach based on Nissenbaum's could be used to account for PGs, while in other grammars (e.g., that of German), an approach based on HH's could be used. Although this idea is not exactly in the spirit of what the authors above have in mind – for example, HH (pp. 88–91) explicitly argue against tucking-in, and we have seen how tucking-in plays a role in Nissenbaum's account of his generalization – something along these lines may indeed prove to be a fruitful line of investigation.

Nevertheless, such an idea overlooks the important question of how the compositional semantics of PGs works. In order to fully account for the data we have seen, we must have some understanding of how these PG constructions can converge at LF compositionally. As we have seen, Nissenbaum attempts give such a compositional semantics, whereas HH do not. In light of this, we could suppose that a Nissenbaum-style proposal could be maintained for the English data to account for LF convergence compositionally. Now, something along the lines of HH's proposal for Agree and PGs could very well be correct, but applying such a proposal to the German data cannot be sufficient, because it does not take LF compositionality into account.

Moreover, we can now say that an important consequence of the counterexample (5b) is to complicate the question of compositionality. If we assume that Nissenbaum and HH are correct in that the PG-containing adjuncts involve operator movement, and if we follow the standard assumption (e.g., Heim and Kratzer 1998) that movement is predicate deriving, we are left with the challenge discussed in this section with respect to Nissenbaum as to how (5b) can be interpreted. As we have seen, HH have nothing to say about this, and it is not clear what they could, given the assumptions laid out above. In light of this discussion, then, the counterexample in (5b) presents us with a significant challenge to understanding the semantics of PGs.

This squib has thus raised several inter-related questions that revolve around the MMDs in the German example (5b), and I will leave these as open questions to stimulate future research. There is the question of what the correct generalization is

regarding the number of PGs and edge-XPs. There is the question of how to account for the English and German data discussed here. And perhaps most significantly, there is the question of how to understand the compositional semantics of the PG construction.

REFERENCES

Assmann, Anke. 2010. Parasitic gaps in derivational grammar. Master's thesis, Universität Leipzig.

Bayer, Josef, and Jaklin Kornfilt. 1994. Against scrambling as an instance of Move-alpha. In *Studies on scrambling: Movement and non-movement approaches to free word-order phenomena*, ed. Norbert Corver and Henk van Riemsdijk, 17–60. Berlin: Mouton de Gruyter.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Culicover, Peter W., and Paul M. Postal, eds. 2001. *Parasitic gaps*. Cambridge, MA: MIT Press.

Fanselow, Gisbert. 1993. Die Rückkehr der Basisgenerierer. *Groninger Arbeiten zur germanistischen Linguistik* [The return of base-generation. Groningen Working Papers in Germanic Linguistics] 36: 1–74.

Haider, Hubert, and Inger Rosengren. 2003. Scrambling: Nontriggered chain formation in OV languages. *Journal of Germanic Linguistics* 15(3): 203–267.

Heck, Fabian, and Anke Himmelreich. 2017. Opaque intervention. *Linguistic Inquiry* 48(1): 47–97.

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in generative grammar*. Oxford: Blackwell.

Hornstein, Norbert. 1995. *Logical Form: From GB to Minimalism*. Oxford: Wiley-Blackwell.

Kathol, Andreas. 2001. On the nonexistence of true parasitic gaps in Standard German. In *Parasitic gaps*, ed. Peter W. Culicover and Paul M. Postal, 315–338. Cambridge, MA: MIT Press.

Müller, Gereon. 1995. *A-bar syntax: A study in movement types*. Berlin: Mouton de Gruyter.

Nissenbaum, Jonathan. 2000. Investigations of covert phrase movement. Doctoral dissertation, Massachusetts Institute of Technology.

Richards, Norvin. 1997. What moves where when in which language. Doctoral dissertation, Massachusetts Institute of Technology.

Williams, Edwin. 1990. The ATB theory of parasitic gaps. *The Linguistic Review* 6(3): 265–279.