

Isaac Gould*

Parasitic gap patterns and hierarchy preservation in German

<https://doi.org/10.1515/psicl-2022-1021>

Received June 2, 2022; accepted May 8, 2023

Abstract: This paper discusses how German parasitic gap data from various earlier publications illustrate two patterns of systematic grammatical variation in the language, which have not been previously identified as such in the literature. I show how Heck and Himmelreich's (Heck, Fabian & Anke Himmelreich. 2017. Opaque intervention. *Linguistic Inquiry* 48. 47–97) analysis for one pattern, although not able to currently capture both patterns, can be extended by allowing for variation in the positions targeted by scrambling along with the phrase markers that constitute domains for linearization. The resulting unifying analysis highlights how different grammatical mechanisms can in various ways (both local and global) have the effect of preserving the hierarchical relations involved in multiple movement dependencies.

Keywords: cyclic linearization; grammatical variation; movement stacks; scrambling; shape conservation

1 Introduction

This paper focuses on parasitic gap (PG) data from German that involve scrambling of two non-pronominal internal argument DPs. The core configurations are schematized in (1c, d), but by way of introduction, consider first (1a, b), with scrambling of just one DP. In (1a, b), a single internal argument – an accusative direct object (DO) in (1a) (from Heck and Himmelreich 2017: 53) or a dative indirect object (IO) in (1b) (from Assmann 2010: 114) – is the antecedent of a gap (the PG) contained within a vP-level adjunct that it has scrambled past. Such adjuncts are typically islands for extraction, but nevertheless the scrambled DP can serve as the antecedent for the gap inside the adjunct. I follow the common analysis that such a gap inside an island crucially is licensed by (i.e. is parasitic on) A'-movement (here, scrambling) of its antecedent from a position below the adjunct to a position above it (cf., for example, Chomsky 1982; Engdahl 1983). Indeed, this line of analysis has been taken

*Corresponding author: Isaac Gould, Ewha Womans University, Seoul, South Korea,
E-mail: gould@ewha.ac.kr

in numerous works that discuss comparable German examples (see, e.g., Felix 1985; Mahajan 1990; but for alternative views of scrambling and PGs in German, see, e.g., Haider and Rosengren 2003; Kathol 2001).¹ The core empirical concern of this paper, now, is which DP can function as the PG's antecedent when both an accusative DO and dative IO scramble past the adjunct. Given the different word orders possible from scrambling, a simple representation of this can be as in (1c) or (1d).

- (1) a. Hans hat Maria₂ [ohne PG₂ anzuschauen] 2 geküsst.
 Hans has Maria_{ACC} without at.to.look kissed
 'Hans has kissed Maria without looking at her.'
- b. wenn mander Maria₂ [anstatt PG₂ zu helfen] 2 das Buch wegnimmt
 if one the Maria_{DAT} instead to help the book_{ACC} away.takes
 'if one takes away the book from Maria instead of helping her'
- c. ... IO₂ DO₃ [Adjunct ... **PG**_{???} ...] ... 2 3
 d. ... DO₃ IO₂ [Adjunct ... **PG**_{???} ...] ... 2 3

As we have seen in (1a, b), in the case of one DP scrambling, either an accusative or a dative can be the antecedent of the PG, and there does not appear to be variation in this regard for the two varieties of German that I will presently introduce.² Nevertheless, when it comes to which of the two scrambling DPs can license a PG in (1c, d), a puzzling asymmetry emerges when we carefully examine what has been reported in the literature (actual German examples given in Section 2). In recent work, Heck and Himmelreich (2017) provide data showing how only the accusative DP in (1c, d) can be the PG antecedent. Strikingly, German data reported in earlier literature (e.g. Müller 1995) reveal a different pattern. In this other dataset, it is the DP (either dative or accusative) that is linearly closer to the PG in (1c, d) that can serve as the PG's antecedent.

The systematic nature of these contrasting data patterns, when there are multiple scrambled DPs, has not been noticed before.³ Such a systematic contrast can be taken as evidence that there is grammatical variation across German speakers.

¹ Accordingly, if there were no such A'-movement, the PG would not be licensed. For examples of this, see the data in Webelhuth (1992: 175), Müller (1995: 173), Fanselow (2001: 411), and Himmelreich (2017: 109) where leaving the DP *in-situ* below the PG-containing adjunct (i.e. not scrambling the DP past the adjunct) is ungrammatical.

² As will be discussed shortly, this paper identifies for the first time two varieties of German for the data types under consideration. The examples in (1a, b) are from the variety of Heck and Himmelreich (2017). For grammatical examples from the other variety (i.e. that of Müller 1995) that parallel (1a, b), see Müller (1995: 173, 262) and Fanselow (1993: 34, 2001: 411). See also (4) and (10) below for further parallel examples that have one scrambling DP, and where judgments from the two varieties also converge.

³ Heck and Himmelreich (2017: 54) note differences in judgments in the literature, but do not identify the full extent of these differences, as well as their systematic divergence.

For descriptive purposes I will use the terms Accusative-antecedent Grammar (AG) to refer the grammar of the first pattern (i.e. that of Heck and Himmelreich) with only an accusative antecedent allowed for (1c, d), and Linear Order Grammar (LOG) for the second pattern (i.e. that of Müller), where only a linearly closer antecedent is allowed for (1c, d).⁴

This now raises the question of whether a unified analysis of this variation can be given. The fact that accusative case and not word order is critical for the AG data suggests that a structure-based analysis is relevant. This is indeed the approach of Heck and Himmelreich, who give an analysis of their dataset that crucially relies on the two DPs scrambling to the *vP* edge in a way that preserves their hierarchical relation with each other prior to movement. For Heck and Himmelreich (2017: 62), such hierarchy-preserving movement is implemented via a movement stack, which is a buffer/temporary storage area in the workspace where constituents that are attracted by a particular feature and copied from the tree are placed before they are re-merged, and which – when more than one constituent is attracted – provides an order according to which these constituents are re-merged in order to satisfy that feature (see details in Section 3, and cf. Stroik 2009 for more on buffers). However, little analysis has been proposed for the older LOG data from the literature, and a unified analysis seemingly faces a challenge with what appears to be a sensitivity to linear order, a sensitivity that appears to be at odds with Heck and Himmelreich's analysis that crucially relies on a movement stack and the hierarchy related to it.⁵ Indeed, the seeming relevance of linear order runs counter to the prevailing view that it is considerations of structural hierarchy that determine PG licensing.

In response to this challenge, I propose that Heck and Himmelreich's proposal can indeed form the basis of a unified analysis of the data with multiple scrambling dependencies, but with two key modifications (see (a) and (b) below).⁶ Thus,

4 I am not aware of these varieties correlating with any regional dialects, or of in any way being non-standard. That there could be such sub-varieties of standard German is consistent with the discussion of different acquisition trajectories in Section 5, where I also speculate on the possibility of additional variation.

5 Müller (1995: 264) is explicit in not attempting to analyze the older data in the literature. Fanselow (1993: 35) offers only a brief suggestion for such data, but assumes that scrambling involves base-generation of scrambled XPs (and not movement), an assumption I do not adopt, and which is challenged by the data from Heck and Himmelreich (2017) and the floating quantifier data in the Appendix.

6 Although the focus here is on scrambling (but see the Appendix for related non-PG data involving *wh*-movement), the discussion could in principle be extended to *wh*-movement. Heck and Himmelreich (2017: 52) provide additional PG data involving *wh*-movement that can be added to the scrambling paradigm for the AG in (3) below, and I follow them in assuming these examples' contrasts are captured in essentially the same way as those in (3), but I am not aware of relevant *wh*-data for the LOG. I leave a full investigation of *wh*-movement for future research. Also note that I

hierarchy and a movement stack (and not linear order) underlie resolving PG dependencies in both grammars. And although they do not suggest it at first, the LOG facts also provide further evidence (see e.g. McGinnis 1998; and Richards 2001) that multiple movement dependencies attracted by the same type of feature on a head (here: *v*) can be hierarchy preserving. Instead, the loci of grammatical variation concern (a) the availability of scrambling positions other than the *vP* edge, and (b) which phrase markers constitute domains for the linearization of syntactic structure. In short, the LOG contains a key position that can trigger scrambling within *vP* (by hypothesis in ApplP), but which cannot trigger scrambling within the AG. Further, the LOG contains a linearization domain within *vP* that effectively fixes the relative hierarchy of the scrambling DPs (as a result of fixing their relative word order), as per Fox and Pesetsky's (2005a) proposal for cyclic linearization. In contrast, the AG just has a larger linearization domain that properly contains *vP*. These differences between the grammars are outlined schematically in (2), where numbered positions indicate positions available for scrambling, and arcs indicate linearization domains.

(2) a. Accusative-antecedent Grammar (AG)



b. Linear Order Grammar (LOG)



A consequence of this proposal is to highlight two different ways that hierarchy preservation can emerge in grammar for constituents involved in multiple movement dependencies of the same type, even when (as we will see) this hierarchy can also sometimes be disrupted in these same dependencies. The first arises locally when multiple constituents are attracted by a head via the same movement stack, and involves maintaining a pre-movement hierarchical relation (cf. Richards 2001). This is what we see in both the AG and LOG when multiple DPs scramble to the *vP* edge. The second is the result of linearizing a chunk of structure (i.e. a linearization domain), which has the effect here of globally fixing the hierarchical relation of two constituents with respect to each other for all remaining iterations of linearization in the derivation

do not look at the one remaining PG data type in ditransitives considered in the literature, namely PGs licensed by pronouns. Pronouns appear to pattern differently from non-pronominals in the LOG (cf. Heck and Himmelreich 2017: 55; Müller 1995: 262–263), but the full empirical picture is not entirely clear, as the possibilities of weak and strong pronouns (i.e. non-stressed vs. stressed pronouns) behaving differently are not fully explored in these works. Nevertheless, an approach to such data that is consistent with the approach of this paper is in principle viable once the particular properties of pronouns are taken into consideration.

(cf. Fox and Pesetsky 2005a).⁷ To the extent that the discussion here motivates both local and global mechanisms, it can be taken as introducing an argument against the approaches in Müller (2001) or Müller (2007) that advocate a single (type of) mechanism for preservation effects more generally, and is in line with Fox and Pesetsky's (2005b: 256–258) suggestion that two such mechanisms might be necessary. Accordingly, given the differences proposed in (2), we will see different consequences stemming from how the two scrambling DPs have their relative structural height fixed. In the LOG, their relation to each other observed at the ApplP edge is maintained throughout higher phrase markers, whereas in the AG, their relation at the vP edge can subsequently be permuted higher in the structure. This rigid hierarchy we see with the scrambled DPs relative to each other in the LOG thus constitutes novel evidence (in addition to Heck and Himmelreich's data) against the view (e.g. Biskup 2017; Haider and Rosengren 2003) that scrambling is insensitive to hierarchy preservation and can freely permute the relative hierarchy of DPs within the German *Mittelfeld* (i.e., descriptively, the material that appears between a complementizer/fronted finite verb on the left and a clause-peripheral verbal complex on the right).

In what follows, I first describe the core German PG data in Section 2. I then briefly review in Section 3 Heck and Himmelreich's proposal for the AG and how it cannot capture the LOG facts, before discussing the modified proposal in Section 4. Section 5 concludes with further discussion of some theoretical implications of the analysis and some questions about language acquisition and further variation for future research, as well as how the analysis here could potentially shed light on these questions. An Appendix follows that presents further data (beyond PGs) that provide support for the current proposal.

2 Core data

In this section, I present the core data that illustrate the contrasting patterns of PG-licensing in German. The data here focus on ditransitive examples, in which the dative and accusative-marked non-pronominal internal arguments scramble past a vP-level adjunct containing a single PG.

Heck and Himmelreich (2017) present the data in (3), which show that only the accusative DP can be the antecedent when a dative DP also moves past the adjunct. Regardless of whether the scrambled dative DP precedes the scrambled

⁷ These two ways can be referred to more generally as mechanisms for “shape conservation” (Williams 2002). However, as my focus here is more specifically on the mechanisms' roles with respect to syntactic hierarchy, I will talk about “hierarchy preservation” instead.

accusative DP, the dative is not a licit antecedent (3a–b). However, under either order, the accusative DP is a licit antecedent (3c–d).

(3) PG-licensing in the AG: core scrambling data (Heck and Himmelreich 2017: 53–54)

- a. * wenn jemand der Anette₂ das Buch₃
 if someone the Anette.DAT the book.ACC
 [ohne PG₂ zu vertrauen] ₂ ₃ ausleiht
 without to trust lends
 ‘if someone lends Anette the book without trusting her’
- b. * wenn jemand das Buch₃ der Anette₂
 if someone the book.ACC the Anette.DAT
 [ohne PG₂ zu vertrauen] ₂ ₃ ausleiht
 without to trust lends
 ‘if someone lends Anette the book without trusting her’
- c. dass Hans der Maria₂ das Buch₃
 that Hans the Maria.DAT the book.ACC
 [ohne PG₃ durchzulesen] ₂ ₃ zurückgibt
 without through.to.read back.gives
 ‘that Hans returns the book to Maria without reading it through’
- d. dass Hans das Buch₃ der Maria₂
 that Hans the book.ACC the Maria.DAT
 [ohne PG₃ durchzulesen] ₂ ₃ zurückgibt
 without through.to.read back.gives
 ‘that Hans returns the book to Maria without reading it through’

The conclusion is that scrambled accusative DPs block datives from acting as antecedent to the PG here, and this is supported by the observation that dative DPs are licit antecedents when the accusative stays low as we have already seen in (1b), or when there is no accusative DP, illustrated in (4) from Heck and Himmelreich.⁸

- (4) Dative-scrambling past the subject in the AG (Heck and Himmelreich 2017: 54)
- wenn der Anette₂ jemand [anstatt PG₂ zu gratulieren]
 if the Anette.DAT someone.NOM instead to congratulate
 kondoliert hat
 condoled has
 ‘if someone condoled with Anette (on something) instead of congratulating her (on it)’

⁸ Note that Heck and Himmelreich (2017: 54) provide additional data showing a dative DP scrambled below the subject acting as antecedent to a PG.

Supporting data partially illustrating the pattern above with multiple scrambled DPs can also be found in Lee and Santorini (1994:267).

Müller (1995) presents a contrasting paradigm. The examples in (5) show that when both dative and accusative DPs scramble past the adjunct, either one can be a licit antecedent in principle, but word order matters. Of the dative and accusative DPs, it is only the one that is linearly closer to the adjunct that can be the antecedent.

(5) PG-licensing in the LOG: core scrambling data (Müller 1995: 261–263)

- a. * dass Hans Maria₂ das Buch₃
 that Hans Maria.DAT the book.ACC
 [ohne PG₂ zu danken] __2 __3 zurückgegeben hat
 without to thank back.give has
 ‘that Hans returned the book to Maria without thanking her’
- b. dass Hans das Buch₃ Maria₂
 that Hans the book.ACC Maria.DAT
 [ohne PG₂ zu danken] __2 __3 zurückgegeben hat
 without to thank back.give has
 ‘that Hans returned the book to Maria without thanking her’
- c. dass Hans Maria₂ das Buch₃
 that Hans Maria.DAT the book.ACC
 [ohne PG₃ gelesen zu haben] __2 __3 zurückgegeben hat
 without read to have back.give has
 ‘that Hans returned the book to Maria without having read it’
- d. * dass Hans das Buch₃ Maria₂
 that Hans the book.ACC Maria.DAT
 [ohne PG₃ gelesen zu haben] __2 __3 zurückgegeben hat
 without read to have back.give has
 ‘that Hans returned the book to Maria without having read it’

Supporting data showing that it is the linearly closer dative or accusative DP that is the licit antecedent when both those DPs scramble past the adjunct containing the PG are given by Müller and Sternefeld (1994:375) and Fanselow (1993: 34), who gives examples parallel to (5a) and (5c–d) (though they involve more across-example lexical differences than in (5)), and who reports (p.c.) that (5b) is also grammatical.

To conclude, I take the systematic differences across the paradigms in (3) and (5) as evidence for variation across the grammars of German speakers (and for some discussion of the possibility of further variation in German see Section 5).⁹ I now turn to an analysis of the data.

⁹ A question arises regarding the German data and a generalization put forward by Nissenbaum (2000). Based largely on English data, Nissenbaum concludes that for every DP that moves to the vP edge, there must be a corresponding PG in any adjunct that is merged with vP below the vP-edge positions those DPs are merged at, and Nissenbaum derives this generalization, in part, via principles of semantic composition that crucially interact with the assumption that there is predicate-deriving movement of a semantically vacuous null operator from every PG site. Davis (2020) also provides data from English in support of this generalization, but note that he and Nissenbaum provide conflicting data patterns (cf., for example, Davis 2020: 221, 224; Nissenbaum 2000: 117; see also Williams 1990: 227 for further supporting data that again pattern somewhat differently). Now, the core German data here appear to be at odds with such a generalization: in (3) and (5), two scrambled DPs merge with vP, yet there is only one PG (instead of two) in the adjunct. Two PGs are possible in an adjunct in German, each with a separate internal argument antecedent (Fanselow 1993: 34; Heck and Himmelreich 2017: 54), but two PGs are not required in (3) and (5) (Note that the full empirical range of such two-PG examples remains to be explored, but I follow Heck and Himmelreich in assuming the examples cited immediately above are captured just like similar examples discussed here, modulo the discussion in Heck and Himmelreich 2017: 73 for multiple PG antecedents). The question, then, is to what extent German is compatible with Nissenbaum's generalization. Gould (2020) is the first work to look in detail at this type of apparent discrepancy (cf. Kathol 2001: 329), but discusses only the AG *wh*-movement PG data (mentioned in note 6) from Heck and Himmelreich, and leaves the discrepancy as an open puzzle.

Discussing this issue in detail goes far beyond what can be covered here (but cf. Gould 2020: 119, n. 10 for a proposal that could be modified to account for the AG data by allowing for, in principle, free attachment of the PG-containing adjunct to two different phrases – a possibility overlooked by Gould, although such free attachment is independently problematic given the issue of control in note 12). However, given the LOG data (including the floating quantifier data in the Appendix, which pattern like the PG data, and so, I assume should be analyzed in a similar way), it is not clear to me that there is a viable way of maintaining Nissenbaum's generalization here. The conclusion, then, would be that the generalization does not apply to German. This suggests that semantic composition involving PGs is slightly different in English and German. Indeed, considerations of semantic composition lead me to hypothesize that instead of predicate-deriving null operator movement (as in English), in German there is something more like canonical DP movement (which immediately saturates any derived semantic argument position it creates; see also note 12) from the site of the PG. This could be implemented by adopting Himmelreich's (2017) analysis of German PGs, according to which a pronoun-like constituent moves from the PG site to the edge of the adjunct (see note 13; cf. Postal 1998). Note that adopting Himmelreich's analysis regarding the moving element within the adjunct would leave the rest of the analysis proposed in this paper intact, as the analysis in Sections 3 and 4 concerns the syntax external to the adjunct. In sum, the difference between the English and German PG facts, and the scope of Nissenbaum's generalization, may ultimately hinge on whether a null operator or a semantically contentful DP moves from the PG site.

3 Heck and Himmelreich's (2017) proposal

In this section I first review how Heck and Himmelreich's (2017) proposal captures the AG pattern, and then show the difficulty this proposal faces with the LOG data.

Heck and Himmelreich assume that *v* in German can have an Edge Feature (cf. Chomsky 2007) that can trigger movement (such as scrambling) of one or more constituents.¹⁰ When, for example, the Edge Feature attracts both dative and accusative DPs from within its c-command domain, the two XPs are placed in an ordered buffer in the form of a (movement) stack. Recall that such a buffer functions as a storage area for attracted constituents and places them in an order for re-merge. Dative indirect objects c-command accusative direct objects within *v*'s sister for the data under consideration (see also Fanselow 2000; Müller 1995 on this point), and for concreteness, I follow Georgala (2011) in assuming an Applicative Phrase (ApplP) hosts the IO (6a). Given this c-command relation, because datives are encountered first by the Edge Feature in its search down the tree for its two goals within its c-command domain, datives are placed in the stack first (at the bottom), and accusatives placed in the stack last (at the top). This is illustrated schematically in (6b): from their base positions, the attracted DPs have been placed in a movement stack.¹¹ The DPs will subsequently be merged with *vP* iteratively, from the top of the stack to the bottom of the stack, in a first-in last-out manner. This also holds in cases where the derivation involves attaching an adjunct containing a PG to the *vP*; when the accusative DP from the top of the stack merges with this *vP*, the dative DP remains in the stack (6c).¹² Upon re-merging and occupying its position c-commanding the

10 For Heck and Himmelreich (2017), *v* can have an Edge Feature because it is a phase head, with the consequence that movement to the edge of *vP* is required when a constituent moves out of the sister of phase head *v* (cf. Chomsky 2000, 2001). This means that the edge of *vP* functions as a designated escape hatch. This leads to a point of divergence with Fox and Pesetsky's (2005a) proposal for linearization (see Section 4), which I adopt in this paper. For more discussion on this point, see note 17.

11 Note that in principle, a scrambling trigger such as an Edge Feature can skip over a higher potential goal and attract a lower goal. Accordingly, a scrambling trigger can in principle always cause an accusative DP to scramble over a dative DP that it does not attract. The assumption, though, is that in its search for goals, a probe cannot backtrack up the tree to attract a goal that has already been skipped. Thus if both dative and accusative DPs are to be attracted as goals, then the structurally higher dative cannot be skipped over and must be placed in the movement stack first. See Heck and Himmelreich (2017: 60–61, 63) for further discussion.

12 Heck and Himmelreich's (2017) assumption is that this adjunct (along with floating quantifier *alles* 'all' discussed in the Appendix) must attach to *vP* before the external argument merges in and before any re-merging happens. Note that the analysis in this paper can apply in a highly comparable way to Heck and Himmelreich's data, as well as the LOG data here, if instead the adjunct (and floating quantifier *alles*) must simply attach to *vP* before any re-merging happens, but after merging in the

adjunct, the accusative DP immediately establishes an Agree relation (Chomsky 2000, 2001) with the PG, thereby licensing the PG.¹³ Only at this point in the derivation does

external argument (For relevant discussion on the order of attachment to an XP between a selected argument in its base-generated specifier position and additional specifiers derived by scrambling, see Heck and Himmelreich 2017: 59, 77). I leave as an open question the exact position of the adjunct (either above or below the external argument), but the key point that the adjunct is attached to vP is supported by the fact that these PG-containing adjuncts are all non-finite adverbial clauses with a null subject that is controlled by the matrix subject (Assmann 2010: 118; Kathol 2001). Building on the scope-based adjunction theories of earlier work (such as Frey and Pittner (1999) and Pittner (1999)), Fischer and Høyem (2022) claim that these types of adjunct clauses, independently of whether they contain PGs, attach no higher than vP. Fischer and Høyem also argue that they involve obligatory control, and consequently that obligatory control of the external argument into the adjunct clause (such as in the PG data) requires a local relationship between the two (i.e. attaching the adjunct to vP). Specifically they propose attaching the adjunct above the external argument, but the exact position of vP-attachment is less clear to me. Incorporating control into the discussion here goes beyond the scope of this paper, and I believe it is ultimately orthogonal to present concerns given the distinct A/A'-licensing conditions relevant for control and PGs respectively (cf. Chomsky 1982). Still, we can observe the following. With a high vP-attachment of the adjunct above the external argument, there is perhaps a more straightforward semantic composition of the adjunct (having a saturated, sentential meaning) with a saturated vP (cf. note 9 and Nissenbaum 1998: 510). High attachment, though, would mean that the base position of the external argument does not c-command the adjunct, and a reasonable expectation would then be that the external argument could license a PG (cf. note 1), even though this does not appear to be attested in German (Heck and Himmelreich 2017: 52, n. 6). In contrast, attachment of the adjunct below the external argument's first-merge position has the advantage of providing a configuration that is clearly inconsistent with the external argument licensing a PG. Low attachment, though, would seemingly complicate semantic composition of the adjunct with the unsaturated vP. One direction to pursue for low attachment could involve assuming that the null subject in the adjunct indicates that it is a predicate; obligatory control would then be a result of the external argument taking both the vP and the adjunct as predicates along the lines of, for example, Chierchia's (1984) approach to control (cf. also Kathol's 2001: 332 approach).

13 More precisely, the antecedent of the PG agrees locally with an element that moves from the PG to the edge of its adjunct. Heck and Himmelreich (2017: 69–74) follow Chomsky (1986), as well as the brief discussion in Contreras (1984), in assuming that this element is a null operator, which ostensibly is semantically vacuous (cf. Chomsky 1982: 31, 1986: 56; Nissenbaum 2000: 28). They propose that Chomsky's (1986: 56) operation of chain composition for PGs can be implemented via Agree providing values for semantic indices. Note that the idea that PGs are licensed via Agree, although not standard, can be viewed as recasting the general idea from various older proposals that PG licensing is via binding (e.g. Chomsky 1982; Engdahl 1983; Postal 1998) along the lines of more recent works that analyze binding phenomena by means of Agree (e.g. Fischer 2006; Rooryck and Vanden Wyngaerd 2011). The exact status of this moving element and the features involved in Agree do not figure in the main discussion here, and I largely abstract away from them, but see note 9 for a suggested refinement to the analysis, where following Himmelreich (2017), the moving element is not taken to be a null operator. Himmelreich also develops an Agree-based approach to PGs and proposes that the moving element is a constituent that is derived by splitting features off the antecedent (and is

the dative DP merge with vP, emptying the stack (6d).¹⁴ But this step of merge is too late for the dative to license the PG, as this has already been done by the accusative DP. This accounts for the pattern we saw above in (3a) and (3c). Then, after the Edge Feature is no longer active, the external argument merges with vP (6e).

- (6) Licensing a PG with a DO_{ACC} antecedent
- Partially build vP
 $[_{vP} \nu \dots [_{AppIP} IO \text{ Appl } [\dots DO \dots]]]$
 - Place DPs in movement stack
 $[_{vP} \nu \dots [_{AppIP} \text{IO} \text{ Appl } [\dots \text{DO} \dots]]]$ Stack: $\begin{bmatrix} DO \\ IO \end{bmatrix}$
 - Re-merge DO; DO agrees with PG
 $[_{vP} DO_2 [_{Adjunct} \dots PG_2 \dots] \nu \dots [_{AppIP} \text{IO} \text{ Appl } [\dots \text{DO} \dots]]]$ Stack: $\begin{bmatrix} \text{DO} \\ IO \end{bmatrix}$
 - Re-merge IO
 $[_{vP} IO DO_2 [_{Adjunct} \dots PG_2 \dots] \nu \dots [_{AppIP} \text{IO} \text{ Appl } [\dots \text{DO} \dots]]]$ Stack: $\begin{bmatrix} \text{DO} \\ IO \end{bmatrix}$
 - Merge in external argument
 $[_{vP} S IO DO_2 [_{Adjunct} \dots PG_2 \dots] \nu \dots [_{AppIP} \text{IO} \text{ Appl } [\dots \text{DO} \dots]]]$
 - Scramble S and DO to TP level
 $[_{TP} S DO_2 \dots [_{vP} S IO \text{DO}_2 [_{Adjunct} \dots PG_2 \dots] \nu \dots [_{AppIP} \text{IO} \text{ Appl } [\dots \text{DO} \dots]]]]]$

Further, if the accusative DP subsequently scrambles past the dative DP as in (6f), which also involves movement of the subject, the Agree relation between the accusative and the PG, and the attendant licensing by the accusative of the PG is maintained. This is what we see given the word order in (3b) and (3d), where even though the ultimate DO_{ACC} < IO_{DAT} hierarchy (a < indicating c-command when discussing hierarchy) does not parallel the hierarchical IO_{DAT} < DO_{ACC} relation below ν in (6a), that hierarchy is maintained among the multiple specifiers at the vP level in (6d), resulting in only the accusative DP licensing the PG.¹⁵ And if there is no

crucially involved in agreement for case features with the antecedent, and not semantic indices). See Himmelreich also for further motivation for an Agree-based approach to PGs.

14 Thus Heck and Himmelreich do not adopt tucking-in as a part of movement dependencies, in contrast to Richards (2001), although the result after moving multiple targets to the same edge is hierarchy preserving under either account.

15 Heck and Himmelreich (2017: 72) suggest that T can optionally bear a [scr] feature in German, but do not discuss scrambling/optional subject raising to the TP level in detail. Note that as T is not a phase head, it cannot bear an Edge Feature, and thus needs some other feature to trigger movement. Here I first point out that we could assume that T can have multiple [scr] features, each one an attractor of a single XP (unlike the Edge Feature, which can attract multiple XPs). Thus each [scr] feature could find, in principle, any potential DP to scramble and place in its own movement stack (cf. note 11), resulting in two independent XP movements to the TP level that need not be hierarchy preserving in relation to those XP's positions at the vP level. Such lack of hierarchy preservation would be noticeable in (3d), where the IO_{DAT} < S_{NOM} hierarchy at the vP level suggested in note 9 would be reversed at the TP level. Second, I also assume T can have an optional EPP feature that can trigger optional subject raising in

scrambling accusative, as in (1b) or (4), the stack will contain only the dative DP, and upon merging with vP, the dative is free to agree with, and license, the PG.

In sum, to capture the AG, Heck and Himmelreich's proposal of using a stack for multiple movement dependencies ensures hierarchy-preserving movement at the vP level triggered by the Edge Feature, and has the result of the accusative DP blocking the dative from licensing an adjunct containing a single PG. I will follow their proposal, then, for the analysis of PG licensing in the AG data here (but see notes 9, 12, and 13 for qualifications).

However, it is clear that as it stands, Heck and Himmelreich's proposal makes the wrong predictions for the LOG. Given the structure in (6c), where the accusative DP agrees with the PG, Heck and Himmelreich predict that in (5) the accusative antecedent should always be good, and the dative antecedent should never be good. In contrast to Heck and Himmelreich's proposal, in (5) the hierarchical relation between the internal arguments prior to moving to the vP edge – and thus the role of any movement stack (or hierarchy-preserving movement) – is seemingly irrelevant for PG-licensing. In this way the LOG grammar poses a challenge to any unified account of PG-licensing in German. To see how these factors can be relevant, though, in the next section I will consider an important assumption embedded in Heck and Himmelreich's proposal, which involves the number of positions a DP can scramble to.

4 Modifying the proposal

In this section I show that positing grammatical variability in the positions that scrambling targets and variability in which phrase markers are linearization

German (cf. Grewendorf 1989; Müller 1999; Heck and Himmelreich 2017: 69), again with its own movement stack that is separate from those of [scr] features (and is thus not necessarily hierarchy preserving in relation to another XP scrambling to the TP level). In principle, then, subject movement targeting TP could be due to [scr] or the EPP. As for the impossibility of *v* also having [scr] in German, see Heck and Himmelreich (2017: 77), where they discuss how the presence of both an Edge Feature and [scr] on *v* would allow (incorrectly here) for any hierarchy of DP movements to the vP level. They also argue that such a [scr] attractor is not compatible with *wh*-associates of floating quantifier *alles* 'all' (see the Appendix for more on *alles*). If that is correct, then the scrambling below *v* of such associates that I propose in the Appendix can be taken to involve a comparable, though different, kind of scrambling feature. Finally note that Heck and Himmelreich (2017: 79) assume that the different ordering of XPs between (6e) and (6f) must involve movement from the vP to a position higher than the vP domain (such as TP), and this is naturally understood as stemming from considerations of anti-locality.

domains allows for a unified analysis of both grammars under Heck and Himmelreich's (2017) basic framework.

For Heck and Himmelreich, scrambling in the German *Mittelfeld* crucially targets only the vP edge within the verbal domain, with further scrambling possible at the TP level. Consider now a minimally modified proposal for scrambling within the *Mittelfeld* that will capture the LOG.

Instead of scrambling targeting just TP and vP, as per the AG, I propose that in the LOG, scrambling could additionally target some other phrasal marker dominated by vP. For the sake of concreteness, let us assume this vP-internal position is at the edge of ApplP. Similar to how T can optionally attract a DP via a scrambling feature, Appl would be able to optionally attract a DP to its edge via a comparable feature. A schematic structure for ApplP prior to this movement is given in (7a), where we see the canonical $IO_{DAT} < DO_{ACC}$ hierarchy. (7b) then shows movement of the accusative DO to ApplP, resulting in the completed ApplP not being hierarchy preserving. Importantly, in the LOG, the mechanics of multiple attraction of DPs by the Edge Feature in v work just as in the AG. After merging in a v that will attract multiple DPs, those DPs again will be placed in a stack. However given (7b), it is now the accusative DP that will be placed in the stack first (at the bottom), because it is closer to the c-commanding v (7c). The stack will again be emptied by re-merging the DPs with vP in a first-in last-out manner. But given (7c), it is now the dative DP (on the top of the stack) that will first be re-merged with vP, allowing it to agree with, and license, the PG (7d). This agreement relation also blocks the accusative from subsequently agreeing with the PG when it merges with vP in (7e).

- (7) Licensing a PG with an IO_{DAT} antecedent
- Partially build ApplP
 $[_{ApplP} IO Appl [\dots DO \dots]]$
 - Scramble DO within ApplP
 $[_{ApplP} DO IO Appl [\dots \emptyset \emptyset \dots]]$
 - Partially build vP while placing DPs in movement stack
 $[_{vP} [_{Adjunct} \dots PG \dots] v \dots [_{ApplP} \emptyset \emptyset \emptyset Appl [\dots \emptyset \emptyset \dots]]]$ Stack: $\begin{bmatrix} IO \\ DO \end{bmatrix}$
 - Re-merge IO; IO agrees with PG
 $[_{vP} IO_2 [_{Adjunct} \dots PG_2 \dots] v \dots [_{ApplP} \emptyset \emptyset \emptyset Appl [\dots \emptyset \emptyset \dots]]]$ Stack: $\begin{bmatrix} IO \\ DO \end{bmatrix}$
 - Re-merge DO; Merge in external argument
 $[_{vP} S DO IO_2 [_{Adjunct} \dots PG_2 \dots] v \dots [_{ApplP} \emptyset \emptyset \emptyset Appl [\dots \emptyset \emptyset \dots]]]$ Stack: $\begin{bmatrix} IO \\ \emptyset \emptyset \end{bmatrix}$

Introducing a new position for scrambling provides the first step in capturing the two points of contrast between the LOG and the AG, namely (5b) and (5d): when the dative DP is linearly closer to the PG than the accusative, as is the case when the structure containing (7e) is linearized and no further IO/DO scrambling occurs (see below for more on linearization), the dative must be the licenser. Note that this analysis does

involve a movement stack and thus hierarchy-preserving movement to the vP edge triggered by the Edge Feature. This hierarchy is established at ApplP and can differ from that found in the AG at the vP edge because of the possibility of scrambling to ApplP in the LOG. Accordingly, because dative PG antecedents are impossible in (3), I assume that scrambling cannot target ApplP in the AG.

It is important to point out here that proposing that ApplP is a landing site for scrambling is in line with the null hypothesis that any phrase marker could in principle provide such a landing site. The null hypothesis is especially pertinent given the well-documented cross-linguistic variation in scrambling across languages. Now, this variation can be viewed as parametric variation regarding which heads can trigger scrambling. For example, the fact that scrambling cannot cross finite CP boundaries in German, but can in Japanese (Müller 1995), can simply reflect whether finite C is parametrically chosen as a scrambling head in the two languages. Thus introducing Appl as another scrambling head helps round out the typological picture, allowing for a more parsimonious view of scrambling, as per the null hypothesis – C, T, v, and Appl all having been identified as scrambling heads now – and the variation between the AG/LOG with respect to ApplP scrambling would be a parametric difference (cf. Section 5 for further discussion of parameter setting).

A question that now arises is why the structure in (7e) cannot feed the structure in (8). (8) is derived by scrambling the PG-antecedent dative IO in (7e) to TP and past the accusative DO. (8) should not be possible in the LOG, as it would give us the ungrammatical order in (5a) of a dative antecedent DP followed by the accusative (In the remaining discussion, I abstract away from the possibility of string vacuous scrambling of IO/DO to TP, though nothing substantively changes if such scenarios are considered).

(8) Illicit dative-scrambling to TP in the LOG (cf. (7e))

$$* [_{TP} S IO_2 \dots [_{vP} \& DO \text{ } \textcircled{2} [_{Adjunct} \dots PG_2 \dots] v ([\dots [_{ApplP} \textcircled{2} \textcircled{2} Appl [\dots \textcircled{2} \textcircled{2} \dots]]])]]]$$

The second step of the analysis is to assume that the LOG is distinguished from the AG by having a linearization domain within vP (in the sense of Fox and Pesetsky's 2005a "spell-out domain"; see Fox and Pesetsky 2005a for the possibility that different grammars can have different linearization domains).¹⁶ For concreteness, I assume that the sister of v constitutes this linearization domain, as indicated by the arc in (8). For (7c), this gives us the linearization domain in (9a). According to Fox and Pesetsky, building a linearization domain triggers generating a set of ordering statements that define the linear order of the elements within the linearization domain (for details on

¹⁶ Fox and Pesetsky (2005a) do not develop their proposal for what determines different linearization domains across languages. I will assume that it is subject to parametric variation (cf. Section 5).

this relation, see Fox and Pesetsky 2005a: 40–42).¹⁷ Being specifiers of the same phrase, with the DO c-commanding the IO, then translates into the schematic ordering statement in (9b) for (9a), which establishes the order of the DO as preceding the IO (a < indicating precedence when discussing word order).¹⁸

- (9) a. Linearization domain in (7c) = sister of v : ... [_{AppIP} DO IO Appl [... $\bar{D}\theta$...]
 b. DO < IO

17 The idea I am exploring is that there are both phases and linearization domains, with distinct properties. Recall from note 10 that as vP is a phase, its edge functions as an escape hatch, enforcing cyclic movement out of vP . As Fox and Pesetsky discuss, linearization domains, in contrast, must not as a general property enforce cyclic movement, as in some cases movement must not pass through the edge of a linearization domain. However, Fox and Pesetsky aim to do away with designated escape hatches (including the use of such a notion in the understanding of syntactic islands with overt movement), attributing escape hatch effects to the more general design of linearization domains. If they are correct, then one can ask about the motivation of having both linearization domains and phases with escape hatches. Nevertheless, it is not clear that all cases of obligatory cyclic movement can be reduced to the logic of linearization domains. As pointed out by Cheng and Demirdache (2010), there are syntactic island violations due to covert movement, and these cannot be due to linearization concerns, as covert movement has no effect on linearization (cf. Fox and Pesetsky 2005b: 245, n. 14; see Cheng and Demirdache for examples). Yet these island effects with covert movement can be taken as evidence that there must be obligatory covert cyclic movement, if we follow the standard reasoning (cf. Chomsky 1973) given for obligatory overt cyclic movement that is based on island effects with overt movement. Accordingly, there must be something like the escape hatch property of phases that is distinct from linearization domains. It thus seems reasonable to propose the existence of phases in addition to linearization domains, that the boundaries of the two types of domains are not necessarily isomorphic (recall their conflicting properties mentioned above), and that (as the null hypothesis) phases can enforce both overt and covert cyclic movement. With this in mind, and given the empirical ground that can be covered, I will maintain some dissociation between phases and linearization domains. I will continue to assume that v 's status as a phase head enforces cyclic movement (as per Heck and Himmelreich 2017), but also that no linearization domain in German corresponds to a projection of v .

18 I do not focus here on linearization of the lexical verb. Müller (2005: 167–168) discusses the potential difficulty of linearizing verbs in an SOV language such as German, in which V2 structures can also lead to SVO word order. This difficulty is related to having a linearization domain lower than the position for finite verbs in V2 structures (as in the LOG, but not the AG). Such a lower linearization domain has the potential to always rule out one of the word orders above, when both orders are in fact possible in the language. A possible solution is proposed in Müller (2007: 75), though Müller does not ultimately adopt this proposal. Based on Müller's (2007) proposal, then, in V2 structures the finite verb in the LOG would move to a head above Appl, but below v (on its way to C) – i.e. to a position at the edge of the linearization domain within vP in the LOG – thereby preceding a non-first position object, and resulting in VO word order (In OVS-type clauses, first-position constituents such as O would also be attracted by this head, yielding an OV ordering; see Müller 2007: 75). In verb-final clauses, though, the finite verb would remain within the verbal domain in a clause-final position, resulting in OV word order. As this approach is viable with the linearization data under consideration, I will adopt it here, but I leave additional details of the approach for further study.

Crucially, once a set of ordering statements has been generated for a linearization domain, it cannot be contradicted by any further sets of ordering statements resulting from subsequent construction of higher linearization domains. This still allows the IO and DO to scramble to a higher linearization domain, as in (7d) and (7e). As such movement is hierarchy-preserving, creating specifiers of vP , it effectively recreates what we saw with the specifiers of $ApplP$ in (9). The result is that in this higher linearization domain, there will be a new ordering statement that is identical to (9b) (that is, if there is no further IO/DO scrambling in the domain containing (7e)), thereby preserving the relative order of the DPs that was established in the lower domain in (9a). However, linearizing the structure containing (8) so as to give us (5a) would involve contradicting the statement in (9b). As the c-command relation between the DO and IO specifiers has now been reversed (with IO now c-commanding DO) given (8), for the higher linearization domain we generate ordering statements such that the IO now precedes the DO, in contradiction of (9b). It is this ordering contradiction that I propose is the source of the ungrammaticality of (5a), given the derivation begun in (7).

The lower linearization domain in the LOG thus has the effect here of freezing the hierarchical relation of the IO with respect to the DO. This is the second type of hierarchy preservation mentioned in Section 1. Regardless of what subsequent movement dependencies are established later in the derivation by creating new specifiers, the relative order within lower linearization domains must be preserved upon the completion of all subsequent linearization domains.

Importantly, this proposal does not always block scrambling to TP in the LOG. The analysis here predicts that scrambling to TP is possible so long as it preserves the order established in the lower linearization domain. This prediction is borne out in (10a) (which parallels (4) of the AG), from Fanselow (1993: 34), who also provides data representative of the LOG (see also note 19 for further supporting data). In (10a), the dative DP first undergoes Edge Feature-driven scrambling to vP , where it establishes an Agree relation with the PG, before scrambling again because of T's scrambling feature, as shown in (10b).

(10) Licit dative-scrambling to TP in the LOG

- a. wenn der Anette₂ jemand [anstatt PG₂ zu gratulieren]
 if the Anette.DAT someone.NOM instead to congratulate
 kondoliert haben sollte
 condoled have should
 ‘if someone should have condoled with Anette (on something) instead of
 congratulating her (on it)’
 b. [_{TP} DP_{DAT2} ... [_{vP} S ~~DP_{DAT2}~~ [_{Adjunct} ... PG₂ ...] v [... ~~DP_{DAT}~~ ...]]]

Unlike what we saw in (8), this second instance of scrambling in (10) is now possible because there is no constituent (such as the other internal argument in (8)) from within the lower linearization domain that will trigger an ordering contradiction along with the dative DP. Further, examples such as (10) show that structural considerations are critical in PG-licensing, and not linear order by itself, as the linearly closer subject DP does not disrupt the dative DP's ability to participate in Agree and license the PG.

Let us now consider the other possibility available under the LOG, namely one where the accusative DP is the antecedent, and not the dative. I propose that this results in cases where the accusative DO does not scramble to ApplP, as in the AG. If the step of movement in (7b) does not occur, then movement to the vP edge will proceed as per the steps in (6a–d) above. As we saw there, in this type of derivation the accusative DO can license the PG, giving us (5c), but the dative IO cannot, resulting in (5a). As no scrambling has occurred within ApplP in (6), DP-hierarchy from this domain yields a $IO_{DAT} < DO_{ACC}$ word order, and this is not contradicted when linearizing a structure containing (6e) and no further IO/DO scrambling. And similar to (8), the additional scrambling in (6f) is not possible in the LOG, as shown in (11).

(11) Illicit accusative-scrambling to TP in the LOG (cf. (6f))

$$* [TP \ S \ DO_2 \ \dots \ [vP \ \& \ IO \ \cancel{DO}_2 \ [Adjunct \ \dots \ PG_2 \ \dots] \ v \ ([\ \dots \ [ApplP \ \cancel{IO} \ Appl \ [\dots \ \cancel{DO} \ \dots]]])]]$$

Again, as there is no scrambling within the lower linearization domain in (11), the $IO_{DAT} < DO_{ACC}$ word order from the lower domain must be preserved with subsequent scrambling in (11). But as the higher scrambling in (11) disrupts the $IO_{DAT} < DO_{ACC}$ hierarchy from the vP, linearizing the structure in (11) so as to give us (5d) would thus involve a contradictory $DO_{ACC} < IO_{DAT}$ word order. This then correctly rules out the accusative antecedent in (5d), given the derivation begun in (6).¹⁹

Having discussed the PG data, I point out here that further support for the analysis can be found with data involving floating quantifier *alles* 'all', which are discussed in the Appendix. Independent of PGs then, we see familiar AG and LOG

¹⁹ Note that a second step of scrambling for the IO instead of the DO in (11) is also predicted to be acceptable in the LOG (cf. (10)), as this conforms to the linearization statement of $IO < DO$ from the lower linearization domain; cf. (11'), where the IO scrambles above S. This is confirmed by the grammaticality of corresponding (5c') (Gisbert Fanselow, p.c.).

(11') $[TP \ IO \ \dots \ [vP \ S \ \cancel{IO} \ DO_3 \ [Adjunct \ \dots \ PG_3 \ \dots] \ v \ \dots \ [ApplP \ \cancel{IO} \ Appl \ [\dots \ \cancel{DO} \ \dots]]]]$
 (5c') dass (der) Maria₂ Hans das Buch₃
 that (the) Maria.DAT Hans the book.ACC
 [ohne PG₃ gelesen zu haben] $_2 _3$ zurückgegeben hat
 without read to have back.give has

patterns (in the AG, an accusative DO associate of *alles* instead of a dative IO; and in the LOG, a linear order effect involving the ditransitive's internal arguments), and this can be analyzed in a parallel way as per the preceding sections. I refer the reader to the Appendix, and in particular the contrast in (14).

5 Conclusion, implications, and future research

The analysis in Section 4 thus captures the data in (5) and, along with the analysis in Section 3, allows us to maintain the core of Heck and Himmelreich's (2017) proposal in accounting for both the AG and the LOG. For PG antecedents in both grammars, then, there is movement to the vP edge that is crucially dictated by an Edge Feature movement stack that preserves hierarchical relations below v. However, the relations of constituents that feed this stack can vary. In the data here, the LOG allows two different relations to precede movement to the vP edge, depending on whether Appl triggers scrambling. Further, the grammars vary in whether these relations can subsequently be disrupted after moving to the vP edge via scrambling to TP, with only the AG allowing such permutations. This is attributed to only the LOG having a linearization domain contained within vP, which acts to fix the final ordering, and as a consequence here, the hierarchy of the internal arguments. As the AG lacks such a low linearization domain, either ordering (and thus hierarchy) of the internal arguments is possible in the *Mittelfeld* (after scrambling to vP/TP) when linearization occurs with a higher domain (say, CP).

Thus, although any hierarchy of DPs involved in multiple scrambling dependencies within the *Mittelfeld* is in principle possible in both grammars, the analysis here advances novel evidence for hierarchy preservation as a constraining factor for scrambling. This view stands in contrast to accounts where scrambling in German is taken to be movement that more or less freely reconfigures constituents' relations with each other (e.g. Biskup 2017; Haider and Rosengren 2003), and instead helps support Heck and Himmelreich's proposal that restricts scrambling (and PG-licensing), in part, through feature-triggered movement and hierarchy-preserving movement stacks.

The analysis also highlights how hierarchy preservation phenomena involving multiple movement dependencies of the same type can stem from independent mechanisms in the grammar – namely the effects of movement stacks and linearization domains – although sometimes these mechanisms' effects overlap. Thus on the one hand, when the Edge Feature attracts two DPs from a lower linearization domain, as in (7d) and (7e) in the LOG for example, the hierarchy preservation resulting from the movement stack matches the hierarchy preservation requirements that can be tied to the low linearization domain. Yet on the other hand,

these mechanisms still have independently discernible effects. The first concerns how the hierarchy-preserving movement to vP of a movement stack feeds subsequent agreement outcomes, even when this phrase marker contains no linearization domain, as in (6) for the AG. And further, in the cases where scrambling is blocked in the LOG (cf. (8) and (11)), this is because of a hierarchy and consequent word order determined in a lower linearization domain, and is not related to multiple constituents being buffered in a movement stack. In sum, the data here have motivated an approach where such preservation effects are ultimately due to both local and global mechanisms: locally at particular derivational points via stacks, which no longer play a role in the derivation once they have been emptied; and globally via earlier linearization domains, which continuously play a role upon completion (as their ordering statements cannot be contradicted in the output of all subsequent domains). This is problematic for the approaches in Müller (2001 or 2007), which adopt only a single global or local mechanism respectively.

Next, an important question to consider is how child learners come to acquire the different German varieties. Properly addressing this goes far beyond what can be covered here, but I would like to point to an available direction for future research that is consistent with the proposal here that these varieties differ in terms of scrambling positions and linearization domains. The question of acquisition is especially pertinent given what is likely to be the total (or near total) absence of evidence for the learner that directly bears on distinguishing these varieties. Consider that the evidence here for the two varieties (and for the analytical differences proposed for them) relies on PGs, as well as floating quantifier *alles* ‘all’ in the Appendix, co-occurring with ditransitive verbs and very specific multiple movement dependencies. Such highly particular data points are likely to be vanishingly rare in the learner’s linguistic input. Indeed, a preliminary search through eight corpora in CHILDES (MacWhinney 2000) of child-directed speech in German (the Caroline, Leo, Manuela, Miller, Rigol, Szagun, Wagner, and Weissenborn databases) yielded no examples of PGs in the data (my thanks to Zhuqing Wang for informing me of these results). This is in line with what Pearl and Sprouse (2013: 54) report as an absence of PGs in their preliminary search through corpora of child-directed speech. Similarly, in a more detailed search for floating quantifier *alles* throughout the Caroline corpus, I identified 98 instances where *alles* was not adjacent to a potentially relevant *wh*-word associate. Forty five tokens involved copular constructions with *sein* ‘to be’, where *alles* is not clearly a floating quantifier. In the remaining 53 tokens, *alles* is clearly a floating quantifier, but none of these examples involve ditransitive verbs, let alone the necessary multiple movement dependencies.

The conclusion, then, is that there appears to be a poverty of the stimulus acquisition puzzle (cf. Chomsky 1980): regardless of the theory adopted for PGs or *alles*, how do learners end up acquiring such similar, yet distinct, varieties of German

given the paucity of evidence? If learners are not exposed to the kinds of data discussed in this paper, then to sharpen the issue in terms of the theory here, we can observe that the (ditransitive) input that the learners do end up receiving is in fact ambiguous as to whether there is scrambling within ApplP or whether there is a low linearization domain.

However, this ambiguity regarding scrambling positions and linearization domains might be key to understanding the variation we see. Gould (2017) presents a model for the acquisition of syntax that can crucially learn in systematic ways from ambiguous evidence. Gould shows that when given an input corpus of entirely ambiguous data for certain syntactic parameters, the model can arrive at different parameter settings for different learners (provided these parameters do not interact to a sufficient degree with other parameters), thereby providing an account of systematic variation that can be observed across speakers. An advantage of the proposal here is that the way it distinguishes the German varieties can be understood in terms of motivated sets of parameters that can be fed into this kind of learning model.

The full range of cross-linguistic variation (and thus the full details of these parameters for the model) remains to be explored. Still, as discussed in Section 4, (a) we can assume that one set of parameters will concern which phrase markers can allow for scrambling (including ApplP), which is motivated by the cross-linguistic variation we see regarding scrambling; and (b) given apparent cross-linguistic variation in what constitutes a linearization domain (cf. Fox and Pesetsky 2005a), we can also hypothesize that another set of parameters concerns which phrase markers delimit linearization domains. Different parameter settings for scrambling to ApplP and having a low linearization domain would then yield the AG and LOG along the lines of the analysis here.

Now, given these ingredients – (i) independently observed variation across languages as regards scrambling and linearization domains (represented via parameters), (ii) the ambiguous evidence facing German learners, and (iii) the independently motivated learning model in Gould (2017) – it could very well be the case that the variation in German is actually expected. And if such variation is indeed expected and can be modelled accordingly, then this would provide a different direction of support for the proposal here.

The details and results of actually applying the theory here to the learning model by running modeling simulations are a topic for future research (including more fine-grained details of the input corpus that the model learns from), but initial considerations indicate that this approach has the potential for success. That is, the ambiguous German input could result in the model learning the relevant parameter settings for the AG/LOG, thereby supporting the analysis here and addressing the acquisition puzzle regarding these varieties.

A final point here concerns the possibility of further variation in German. While I am not aware of additional variation regarding PGs and *alles* in German, this could be a fruitful area to investigate further, especially in light of the modeling considerations above. Indeed, given the ambiguity of the learner's input and the learning scenarios sketched above, one reasonable expectation is that some learners would acquire a grammar where there is scrambling within ApplP, but no low linearization domain (a kind of cross between the AG and LOG). This would be reflected by judgments for (3)/(5), for example, where either the displaced accusative or dative DP could be the PG antecedent, regardless of linear order. It remains to be seen, though, whether such a sub-population of German speakers exists.

Acknowledgments: In developing and writing this paper, I have benefited from the assistance of a number of people. I would like to express my thanks to Sam Alxatib, Michael Yoshitaka Erlewine, and Zhuqing Wang, as well as the editors and reviewers who have helped with this project. Finally, I thank Gisbert Fanselow, in whose memory this paper is dedicated.

Appendix: AG/LOG variation with floating quantifier *alles*

Some evidence in support of the proposal here comes from a novel empirical distinction between the AG and LOG that is independent of PGs. Heck and Himmelreich (2017) present a paradigm that illustrates a restriction on what the DP associate of the morphologically invariant floating quantifier *alles* 'all' can be. For Heck and Himmelreich, this restriction tracks the accusative/dative distinction we saw with PGs. We thus have an independent manifestation of the AG, and I follow Heck and Himmelreich in how they capture the distribution of *alles* in the AG (but again see note 12 for a qualification), which is along the lines of the AG analysis for PGs in Section 3. Given this AG analysis and the formal parallels between the *alles* and PG paradigms, the analysis in Section 4 leads us to predict a different restriction (based on linear order) for the associate of *alles* in the LOG. This prediction appears to be borne out, illustrating a further point of variation between the AG and LOG, as well as how the new ingredients proposed for the LOG can be fruitfully applied to capture data beyond PGs. In this Appendix, I briefly introduce the *alles* data and the analysis, which closely follow the contours of the previous sections.

In the data here, *alles* must associate with a *wh*-phrase.²⁰ The examples here are copied from Heck and Himmelreich (2017: 50–51). First, in the AG, an accusative or dative *wh*-word can associate with *alles* across an indefinite nominative subject (12a–b), however an indefinite accusative/dative DP functions as an intervener and blocks a nominative *wh*-subject from associating, resulting in ungrammaticality (12c–d) (note that nominative *wh*-subjects can independently associate with *alles* when there is no indefinite intervener; see Heck and Himmelreich for examples). The judgments for (12) are the same in the LOG.^{21, 22}

- (12) Association with *alles* in the AG/LOG: one DP moves to the vP edge
- | | | | | | | |
|----|--------------------------------------|-----|-------|---------------|--------------------|---------------|
| a. | Wem ₂ | hat | ein | Professor | alles ₂ | geholfen? |
| | who.DAT | has | a | professor.NOM | all | helped |
| | ‘Who all did a professor help?’ | | | | | |
| b. | Wen ₂ | hat | ein | Professor | alles ₂ | beleidigt? |
| | who.ACC | has | a | professor.NOM | all | insulted |
| | ‘Who all did a professor insult?’ | | | | | |
| c. | *Wer ₂ | hat | einem | Professor | alles ₂ | gratuliert? |
| | who.NOM | has | a | professor.DAT | all | congratulated |
| | ‘Who all congratulated a professor?’ | | | | | |
| d. | *Wer ₂ | hat | einen | Professor | alles ₂ | vergöttert? |
| | who.NOM | has | a | professor.ACC | all | idolized |
| | ‘Who all idolized a professor?’ | | | | | |

A difference between the AG and LOG is not expected in (12) given Heck and Himmelreich’s basic assumptions about association with *alles* (see Heck and Himmelreich for more detailed discussion). Similar to adjuncts containing PGs, *alles* is assumed to be a vP adjunct (similar to floating quantifier *all* in English being a vP adjunct; cf. Heck and Himmelreich 2017: 66 n. 21) that is in an Agree relation with the closest c-commanding (*wh*)-indefinite (note that *alles* can co-occur adjacent to a PG-containing adjunct, which is also consistent with both being attached to vP).²³

²⁰ For a detailed discussion of what *alles* can associate with, see Doliana (2021). Doliana (2021, 2022) presents an alternative analysis of *alles*, but it is unclear how it can account for the range of contrasts reported here, and so I do not discuss it further.

²¹ The LOG data reported in this Appendix come from Gisbert Fanselow (p.c.).

²² Note that Beck (1996) observes that intervention effects with *alles* disappear if the indefinite below the *wh*-phrase is generic or specific. Indeed, Gisbert Fanselow (p.c.) observes that (12c–d) and (14a–b) are all acceptable (in the LOG) with a generic conditional infinitival construction, as in ... *würde* ... *gratulieren* ‘... would ... congratulate’ for (12c). I set aside the disappearance of intervention effects with certain indefinites as a topic for future research; to my knowledge these patterns have not received a close analysis in the literature, and I abstract away from them here.

²³ See Rooryck and Vanden Wyngaerd (2011) for an earlier proposal that adverbial floating quantifiers are licensed via an Agree relation with their associates. Heck and Himmelreich’s (2017)

Non-*wh*-phrases can agree with *alles*, but if one does, the derivation will crash. Assuming that *wh*-movement proceeds through the *vP* edge via the Edge Feature on a par with scrambled DPs, then *alles* will have target agreement with a licit associate if the first DP to merge with *vP* is a *wh*-word. This happens in (12a–b): movement of the *wh*-internal argument brings it to the edge of *vP*, where it can successfully agree with *alles* (13a) before the indefinite external argument merges in (13b); subsequently *wh*-movement will target CP. In contrast in (12c–d), when a non-*wh*-indefinite first scrambles to the *vP* edge before the external argument merges in, the non-*wh*-indefinite first establishes a non-target Agree relation with *alles* (13a), effectively blocking target agreement with the higher *wh*-nominative (13b). Note that neither steps of scrambling internal to *v*'s complement, nor linearization statements within *v*'s complement are relevant for the structures in (13), and thus we account for the lack of an AG/LOG difference in (12).

- (13) Core *vP* structure for (12) in the AG/LOG
- a. Move one internal argument to *vP* edge; Internal argument agrees with *alles*

$$[_{vP} \text{DP}_{\text{ACC2/DAT2}} [\text{alles}_2] \text{ } v \dots [\dots \text{DP}_{\text{ACC/DAT}} \dots]]$$
 - b. Merge in external argument

$$\dots [_{vP} \text{DP}_{\text{NOM}} \text{DP}_{\text{ACC2/DAT2}} [\text{alles}_2] \text{ } v \dots [\dots \text{DP}_{\text{ACC/DAT}} \dots]]$$

As with the PG data, though, an AG/LOG difference emerges when we consider multiple movement dependencies with ditransitives that target *vP*, as illustrated in (14). Consider first the AG judgments from Heck and Himmelreich. A scrambled dative non-*wh*-indefinite IO does not block a *wh*-accusative DO from associating with *alles*, whereas a scrambled accusative non-*wh*-indefinite DO blocks association with a *wh*-dative IO. The analysis parallels the account of PGs in the AG: the accusative DO (regardless of its *wh*-status) always merges with *vP* before the dative IO because of the hierarchy-preserving effect of the movement stack associated with the Edge Feature of *v*. Thus just as the accusative DP in the AG must agree with the PG before the dative DP can, the accusative must also agree with *alles* before the dative can (successfully when the accusative is a *wh*-word, but unsuccessfully otherwise), as shown schematically in (15). To my knowledge, the data reported in the literature only reflect the AG judgments in (14). But given the discussion so far, we can now

proposal for *alles* is along similar lines, and an Agree proposal is well-suited for capturing the intervention effects we see with *alles*. Further note that along the lines of the discussion in note 13 for PGs, an Agree-based approach to *alles* could potentially be developed as a way of recasting what is proposed by Fitzpatrick (2006; cf. Doetjes 1997) as a binding relation between a pronoun within the floating quantifier and its associate. This would involve assuming there is a null nominal within the adverbial *alles* constituent, a possibility that I leave for future research.

make a prediction for a linear order effect involving *alles* in the LOG. As with PGs, when both internal arguments move to the vP edge via the Edge Feature in the LOG, the linearly closer one is the structurally closer one to a lower vP-adjoined target for agreement. Thus we predict both scrambled accusative and dative non-*wh*-indefinites (as linearly closer DPs) to act as interveners blocking association of the *wh*-accusative/dative with *alles*. In (14b) with an intervening scrambled accusative, the $IO_{DAT-wh} < DO_{ACC}$ word order reflects the same ungrammatical derivation as in the AG (15). But in (14a) with an intervening scrambled dative, the $DO_{ACC-wh} < IO_{DAT}$ word order indicates scrambling of the *wh*-accusative DP to ApplP (16a); consequently hierarchy-preserving movement to the vP edge will result in the *wh*-accusative DP merging too late with vP to successfully agree with *alles* (16c), as the non-*wh*-indefinite will have already merged with vP and agreed with *alles* (16b). Indeed, as noted in (14), these predictions appear to be borne out in the LOG.

- (14) Association with *alles* in the AG/LOG: two DPs move to the vP edge
- Wen₂ hat sie einem Professor alles₂ vorgstellt?
 who.ACC has she a professor.DAT all introduced
 ‘Who all did she introduce to a professor?’
 AG: ok; LOG: *
 - Wem₂ hat sie einen Professor alles₂ vorgstellt?
 who.DAT has she a professor.ACC all introduced
 ‘Who all did she introduce a professor to?’
 AG: *; LOG: *
- (15) Core vP structure for (14a–b) in the AG; (14b) in the LOG
- Place DPs in movement stack; Re-merge DO; DO agrees with *alles*
 $[_{vP} DO_2 [alles_2] v \dots [_{ApplP} \bar{I}\bar{\Theta} Appl [\dots \bar{\Theta}\bar{\Theta} \dots]]]$ Stack: $\begin{bmatrix} \bar{\Theta}\bar{\Theta} \\ IO \end{bmatrix}$
 - Re-Merge IO
 $\dots [_{vP} IO DO_2 [alles_2] v \dots [_{ApplP} \bar{I}\bar{\Theta} Appl [\dots \bar{\Theta}\bar{\Theta} \dots]]]$ Stack: $\begin{bmatrix} \bar{\Theta}\bar{\Theta} \\ \bar{I}\bar{\Theta} \end{bmatrix}$
- (16) Core vP structure for (14a) in the LOG
- Scramble DO within ApplP
 $[_{ApplP} DO IO Appl [\dots \bar{\Theta}\bar{\Theta} \dots]]$
 - Partially build vP while placing DPs in movement stack; Re-merge IO; IO agrees with *alles*
 $[_{vP} IO_2 [alles_2] v \dots [_{ApplP} \bar{\Theta}\bar{\Theta} \bar{I}\bar{\Theta} Appl [\dots \bar{\Theta}\bar{\Theta} \dots]]]$ Stack: $\begin{bmatrix} \bar{I}\bar{\Theta} \\ DO \end{bmatrix}$
 - Re-Merge DO
 $\dots [_{vP} DO IO_2 [_{Adjunct} \dots PG_2 \dots] v \dots [_{ApplP} \bar{\Theta}\bar{\Theta} \bar{I}\bar{\Theta} Appl [\dots \bar{\Theta}\bar{\Theta} \dots]]]$
 Stack: $\begin{bmatrix} \bar{I}\bar{\Theta} \\ \bar{\Theta}\bar{\Theta} \end{bmatrix}$

In sum, that judgments from the different phenomena of PGs and floating quantifier *alles* appear to cluster along AG/LOG lines as predicted here provides further support for positing different sets of scrambling positions and different linearization domains in the two grammars.

References

- Assmann, Anke. 2010. *Parasitic gaps in derivational grammar*. Universität Leipzig Master's thesis.
- Beck, Sigrid. 1996. Quantified structures as barriers for LF movement. *Natural Language and Linguistic Theory* 4. 1–56.
- Biskup, Petr. 2017. Labeling and other syntactic operations. In Leah Bauke & Andreas Blümel (eds.), *Labels and roots*, 91–116. Berlin: Walter de Gruyter.
- Cheng, Lisa Lai-Shen & Hamida Demirdache. 2010. Trapped at the edge: On long-distance pair-list readings. *Lingua* 120. 463–484.
- Chierchia, Gennaro. 1984. *Topics in the syntax and semantics of infinitives and gerunds*. University of Massachusetts Amherst Doctoral Dissertation.
- Chomsky, Noam. 1973. Conditions on transformations. In Stephen R. Anderson & Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 232–286. New York: Holt, Rinehart and Winston.
- Chomsky, Noam. 1980. *Rules and representations*. Oxford: Basil Blackwell.
- Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 1–52. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.), *Interfaces + recursion = language?: Chomsky's minimalism and the view from syntax-semantics*, 1–30. Berlin: Mouton de Gruyter.
- Contreras, Heles. 1984. A note on parasitic gaps. *Linguistic Inquiry* 15. 698–701.
- Davis, Colin P. 2020. *The linear limitations of syntactic derivation*. Massachusetts Institute of Technology Doctoral dissertation.
- Doetjes, Jenny. 1997. *Quantifiers and selection: On the distribution of quantifying expressions in French, Dutch and English*. Universiteit Leiden Doctoral dissertation.
- Doliana, Aaron. 2021. *All about alles: The syntax of wh-quantifier float in German*. The University of Maryland Doctoral dissertation.
- Doliana, Aaron. 2022. *Wh quantifier float in German*. *Syntax* 25. 335–378.
- Engdahl, Elisabet. 1983. Parasitic gaps. *Linguistics and Philosophy* 6. 5–34.
- Fanselow, Gisbert. 1993. Die Rückkehr der Basisgenerierer. *Groninger Arbeiten zur germanistischen Linguistik* 36. 1–74.
- Fanselow, Gisbert. 2000. Optimal exceptions. In Barbara Stiebels & Dieter Wunderlich (eds.), *The lexicon in focus*, 173–209. Berlin: Akademie-Verlag.
- Fanselow, Gisbert. 2001. Features, θ -roles, and free constituent order. *Linguistic Inquiry* 32. 405–437.

- Felix, Sascha W. 1985. Parasitic gaps in German. In Werner Abraham (ed.), *Erklärende Syntax des Deutschen*, 173–201. Tübingen: Narr.
- Fischer, Silke. 2006. Matrix unloaded: Binding in a local derivational approach. *Linguistics* 44. 913–935.
- Fischer, Silke & Inghild Flaate Høyem. 2022. Adjunct control in German, Norwegian, and English. *The Journal of Comparative Germanic Linguistics* 25. 1–41.
- Fitzpatrick, Justin. 2006. *The syntactic and semantic roots of floating quantification*. Massachusetts Institute of Technology Doctoral dissertation.
- Fox, Danny & David Pesetsky. 2005a. Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31. 1–45.
- Fox, Danny & David Pesetsky. 2005b. Cyclic linearization and its interaction with other aspects of grammar: A reply. *Theoretical Linguistics* 31. 235–262.
- Frey, Werner & Karin Pittner. 1999. Adverbialpositionen im deutsch-englischen Vergleich. In Monika Doherty (ed.), *Sprachspezifische Aspekte der Informationsverteilung*, 14–41. Berlin: Akademie-Verlag.
- Georgala, Effi. 2011. Why German is *not* an exception to the universal <IO, DO> base order of double object constructions. In Mary Byram Washburn (ed.), *Proceedings of the 28th west coast conference on formal linguistics*, 96–105. Somerville, MA: Cascadia Proceedings Project.
- Gould, Isaac. 2017. *Choosing a grammar: Learning paths and ambiguous evidence in the acquisition of syntax*. Amsterdam: John Benjamins.
- Gould, Isaac. 2020. Multiple movement dependencies and parasitic gaps. *Canadian Journal of Linguistics* 65. 110–121.
- Grewendorf, Günther. 1989. *Ergativity in German*. Dordrecht: Foris Publications.
- Haider, Hubert & Inger Rosengren. 2003. Scrambling: Nontriggered chain formation in OV languages. *Journal of Germanic Linguistics* 15. 203–267.
- Heck, Fabian & Anke Himmelreich. 2017. Opaque intervention. *Linguistic Inquiry* 48. 47–97.
- Himmelreich, Anke. 2017. *Case matching effects in free relatives and parasitic gaps: A study on the properties of agree*. Universität Leipzig Doctoral dissertation.
- Kathol, Andreas. 2001. On the nonexistence of true parasitic gaps in Standard German. In Peter W. Culicover & Paul M. Postal (eds.), *Parasitic gaps*, 315–338. Cambridge, MA: MIT Press.
- Lee, Young-Suk & Beatrice Santorini. 1994. Toward resolving Webelhuth's paradox: Evidence from German and Korean. In Norbert Corver & Henk van Riemsdijk (eds.), *Studies on scrambling*, 257–300. Berlin: Mouton de Gruyter.
- MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk*, 3rd edn. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Mahajan, Anoop. 1990. *The A/A-bar distinction and movement theory*. Massachusetts Institute of Technology Doctoral dissertation.
- McGinnis, Martha. 1998. *Locality in A-movement*. Massachusetts Institute of Technology Doctoral dissertation.
- Müller, Gereon. 1995. *A-bar syntax: A study in movement types*. Berlin: Mouton de Gruyter.
- Müller, Gereon. 1999. Optimality, markedness, and word order in German. *Linguistics* 37. 777–818.
- Müller, Gereon. 2001. Order preservation, parallel movement, and the emergence of the unmarked. In Géraldine Legendre, Jane Grimshaw & Sten Vikner (eds.), *Optimality-theoretic syntax*, 279–313. Cambridge, MA: MIT Press.
- Müller, Gereon. 2005. Remarks on cyclic linearization and order preservation. *Theoretical Linguistics* 31. 159–171.

- Müller, Gereon. 2007. Toward a relativized concept of cyclic linearization. In Uli Sauerland & Hans-Martin Gärtner (eds.), *Interfaces + recursion = language?: Chomsky's minimalism and the view from syntax-semantics*, 61–114. Berlin: Mouton de Gruyter.
- Müller, Gereon & Wolfgang Sternefeld. 1994. Scrambling as A-bar movement. In Norbert Corver & Henk van Riemsdijk (eds.), *Studies on scrambling*, 331–386. Berlin: Mouton de Gruyter.
- Nissenbaum, Jonathan. 1998. Derived predicates and the interpretation of parasitic gaps. In Kimary Shahin, Susan Blake & Eun-Sook Kim (eds.), *The proceedings of the seventeenth west coast conference on formal linguistics*, 507–521. Stanford: CSLI Publications.
- Nissenbaum, Jonathan. 2000. *Investigations of covert phrase movement*. Massachusetts Institute of Technology Doctoral dissertation.
- Pearl, Lisa & Jon Sprouse. 2013. Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition* 20. 23–68.
- Pittner, Karin. 1999. *Adverbiale im Deutschen*. Tübingen: Stauffenburg.
- Postal, Paul. 1998. *Three investigations of extraction*. Cambridge, MA: MIT Press.
- Richards, Norvin. 2001. *Movement in language*. Oxford: Oxford University Press.
- Rooryck, Johan & Guido Vanden Wyngaerd. 2011. *Dissolving binding theory*. Oxford: Oxford University Press.
- Stroik, Thomas. 2009. *Locality in minimalist syntax*. Cambridge, MA: MIT Press.
- Webelhuth, Gert. 1992. *Principles and parameters of syntactic saturation*. Oxford: Oxford University Press.
- Williams, Edwin. 1990. The ATB theory of parasitic gaps. *The Linguistic Review* 6. 265–279.
- Williams, Edwin. 2002. *Representation theory*. Cambridge, MA: MIT Press.