The Third Construction and Strength of C: A Gradient Harmonic Grammar Approach

Gereon Müller

Abstract

This paper addresses the third construction in German, i.e., sentences that combine clause-internal movement from a control infinitive with extraposition of that infinitive. I argue that conflicting evidence regarding the degree of bi-/mono-clausality of the extraposed infinitive (as evidenced by Santorini & Kroch's (1991) observation that long-distance scrambling is possible whereas wide scope of negation is not) is best captured by assuming that it qualifies as a CP with a C head that has more strength than the C of a preverbal restructuring infinitive embedded under a control verb, but less strength than the C of a non-restructuring infinitive (or a finite clause). This presupposes an approach to syntax in which a number of different strength assignments to a given type of category (like C) can be postulated, and can have a direct effect on the (non-) application of syntactic operations. I will show that a version of minimalist syntax incorporating the Phase Impenetrability Condition (Chomsky 2001, 2013) that is embedded in a Gradient Harmonic Grammar approach (Smolensky & Goldrick 2016) can account for the variable strength of C in a principled way.

1. A Paradox

The third construction in German involves a combination of scrambling or unstressed pronoun fronting from an infinitive embedded by a restructuring control verb on the one hand, and extraposition of that infinitive on the other hand; see Besten & Rutten (1989), Geilfuß (1991), Santorini & Kroch (1991), Wöllstein-Leisten (2001), Wurmbrand (2001, 2007), Reis & Sternefeld (2004), and Lee-Schoenfeld (2007), among others. A relevant example illustrating the transparency of the extraposed infinitive (Γ_1) for fronting of an unstressed pronoun (ihn_2) is given in (1a); (1b) is a minimally different example of such movement with the restructuring infinitive Γ_1 in situ.

(1) a. dass sie   ihn$_2$ t$_1$ versucht [$_{\Gamma_1}$ PRO t$_2$ zu küssen]
         that she$_{nom}$ him$_{acc}$ tries   to kiss

b. dass sie   ihn$_2$ [$_{\Gamma_1}$ PRO t$_2$ zu küssen] versucht
   that she$_{nom}$ him$_{acc}$ to kiss   tries

Given that scrambling from a (finite or non-restructuring, non-finite) CP (unlike, say, wh-movement) is impossible in German (see Ross (1967)), the transparency of the extraposed infinitive for this movement operation is often taken to indicate that Γ_1 is not a CP in either (1a) or (1b). However, there is also conflicting evidence that supports a CP status of Γ_1 in the third construction. An indirect argument for this is that lower projections in the clausal spine (TPs, vPs, VPs) can otherwise never undergo extraposition in German (see Müller (2017)), with the *Ersatzinfinitiv* construction an exception that, under closer inspection, proves the rule (see Schmid (2005)). And a very clear and direct argument for a CP status of the extraposed infinitive Γ_1 is that scope of negation is strictly clause-bound in the third construction, in stark contrast to what is the case with non-extraposed restructuring infinitives. This observation goes back to Santorini & Kroch (1991). This is illustrated in (2a) (with only narrow scope of negation available in the third construction) vs. (2b) (where wide scope of negation is possible with standard restructuring infinitives).

(2)   a. dass ich seinen neuesten Roman$_2$ t$_1$ beschlossen habe [$_{\Gamma_1}$ PRO t$_2$
       that I   his   newest   novel$_{acc}$ decided   have
       nicht zu lesen]
       not   to read

                                                    (only narrow scope)
       b. dass ich seinen neuesten Roman$_2$ [$_{\Gamma_1}$ PRO t$_2$ nicht zu lesen]
       that I   his   newest   novel$_{acc}$ not   to read
       beschlossen habe
       decided   have

                                                    (wide scope possible)

Thus, a paradox arises: The availability of unstressed pronoun fronting and scrambling in the third construction in (1a) and (2a) suggests that Γ_1 is not a CP; and the unavailability of wide scope of negation in (2a) suggests that Γ_1 is a CP. It is the main goal of the present study to resolve this paradox in a principled way, by postulating that C is somewhat weaker in the third construction than in non-restructuring (and finite) contexts (so that scrambling and unstressed pronoun fronting from it are possible), but slightly stronger than in standard

restructuring contexts (so that it can undergo extraposition in the first place, and wide scope of negation becomes impossible).

2. Background: Strength in Grammar

It is an old idea in syntactic theory that a functional category X can be strong or weak (see, e.g., Rizzi (1986) and Koster (1986)). On this view, some syntactic operations may require a strong X, and others may require a weak X; yet others are compatible with any X. A more recent application of this general hypothesis involves complementizer-trace effects. Wh-Movement of a subject DP from a declarative clause embedded by a bridge verb is ungrammatical in English if it takes place across a C realized as *that* (see (3a)), but is possible if C is phonologically zero (see (3b)).

(3)   a. $[_{CP}$ Who$_1$ do you think $[_{CP}$ t$_1'$ $[_C \varnothing]$ t$_1$ saw John $]]$?
        b.*$[_{CP}$ Who$_1$ do you think $[_{CP}$ t$_1'$ $[_C$ that$]$ t$_1$ saw John $]]$?

To account for this, Chomsky (2013) suggests that "deletion of *that* [...] might leave only a *weakened* form of C" (my emphasis); this implies that the non-overt realization of C makes it possible to satisfy a constraint on movement that must be violated if the overt realization of C as *that* is chosen. Notwithstanding the issue of how such an idea is to be formally implemented, it can be noted that it raises a problem if a post-syntactic morphological realization of (at least) functional categories is adopted, as is the case in Distributed Morphology (see Halle & Marantz (1993)). On the one hand, a complementizer *that* cannot be assumed to be deleted in the syntax – *that* is in fact only inserted post-syntactically. On the other hand, if the difference between (3a) and (3b) only arises post-syntactically, how can it be the crucial factor for extraction?

   There are many other areas where strength of functional categories has been invoked. A well-known example involves subject pro-drop; see, e.g., (4a) in Spanish vs. (4b) in English.

(4)   a. $[_{TP}$ Hemos $[_{vP}$ *pro* trabajado todo el   día $]]$
                have-3.PL   worked   all   the day
        b.*$[_{TP}$ *pro*$_1$ Have $[_{vP}$ t$_1$ worked all day $]]$

A traditional assumption has been that the strength of T is decisive for allowing *pro* (see Rizzi (1986)): A strong T licenses *pro*, a weak T does not. More recently,

Chomsky (2015) makes use of essentially the same distinction, when he claims that in some languages, "T is too weak to serve as a label", and that "Italian T, with rich agreement, can label TP [...] for English, with weak agreement, it cannot".

A further widespread assumption instantiating the very same idea of strength concerns V-to-T movement; see, e.g. (5a) in English vs. (5b) in French.

(5)   a. John often kisses$_1$ Mary
       b. Jean embrasse$_1$ souvent t$_1$ Marie

In what is arguably still the standard approach (see Pollock (1989), Roberts (1993), Vikner (1997, 2001*a*,*b*), Holmberg & Platzack (1995), Rohrbacher (1999)), it is postulated that a strong T licenses V-to-T movement (as in French), whereas a weak T (as in English) does not.

In all these cases, it is typically assumed that strength correlates in one way or another with the extent of morphological realization (with zero realization as the limiting case). However, as observed by Bobaljik (2002), all these analyses face the problem of being incompatible with post-syntactic morphology that I have illustrated for complementizer-trace effects above. For instance, as regards V-to-T movement, properties of the morphological inventory cannot be held responsible for whether such movement can apply in the syntax or not if inflectional morphology is post-syntactic.

I conclude from all this, first, that there is some evidence that functional categories can have different degrees of syntactic strength; and second, that such strength cannot be determined on the basis of morphological realization if this latter information is not yet present in the syntax. Given this state of affairs, it looks as though two ways out suggest themselves naturally. One is to abandon the idea of post-syntactic morphological realization. The other one is to conclude that strength is an *abstract inherent property* of functional categories that (i) determines whether or not syntactic operations can apply, and that (ii) also determines post-syntactic morphological realization. I will pursue this latter approach in what follows. From this perspective, the task at hand is to show how syntactic building blocks (in the sense of operations, constraints, or rules) can be sensitive to different degrees of strength. *Gradient Harmonic Grammar* (see Smolensky & Goldrick (2016)) is a new grammatical theory designed to implement effects of this type. The particular minimalist version that I will adopt is laid out in the next section.

3. Serial Gradient Harmonic Grammar

I would like to contend that Gradient Harmonic Grammar, which is introduced in Smolensky & Goldrick (2016) mainly on the basis of phonology, offers a new perspective on how to derive three different types of asymmetries as they can be observed with long-distance dependencies in the world's languages: first, asymmetries between movement types (e.g., movement types that are clause-bound vs. movement types that can apply long-distance); second, asymmetries between types of moved items (e.g., subjects vs. objects, or arguments vs. adjuncts); and third (and most importantly in the present context), asymmetries between types of local domain (e.g., VP typically permits extraction from it, CP often does not – and certain types of CPs will be shown to be different from certain other types of CPs, too). More specifically, the version of Gradient Harmonic Grammar that will be relevant in what follows combines properties of three subtheories: (i) Harmonic Grammar; (ii) Gradient Symbolic Representations; and (iii) Harmonic Serialism. I will address these in turn.

3.1. Harmonic Grammar

Harmonic Grammar (see Smolensky & Legendre (2006), Pater (2016)) is a version of optimality theory (see Prince & Smolensky (1993)) that abandons the strict domination property (according to which no number of violations of lower-ranked constraints can outweigh a single violation of a higher-ranked constraint) and replaces harmony evaluation by constraint ranking with harmony evaluation based on weight assignment to constraints. This makes it possible to derive some (but not all) kinds of cumulative effects in syntax. The central notion of harmony is defined in (6) (see Pater (2009)).

(6)   *Harmony*:

$$\mathcal{H} = \sum_{k=1}^{K} s_k \, w_k \qquad w_k = \text{weight of a constraint; } s_k = \text{violation}$$

score of a candidate

Thus, the weight of a constraint is multiplied with the violation score of a candidate for that constraint, and all the resulting numbers are added up, thereby determining the harmony score of a candidate. For present purposes,

we can assume that constraints assign negative scores throughout (e.g., −1 if the candidate violates a constraint once), and that constraint weights are always nonnegative (e.g. 2, or 3). Thus, if a candidate violates constraint A (with weight 2.0) once (−1) and constraint B (with weight 3.0) twice (−2), the harmony score of the candidate would be −8 if there were no further constraints in the grammar. Finally, an output qualifies as optimal if it is the candidate with maximal harmony in its candidate set; i.e., if it has the value closest to zero (or the lowest penalty).

3.2. Gradient Harmonic Grammar

Against this background, the main innovation of Gradient Harmonic Grammar is that Smolensky & Goldrick (2016) postulate that it is not just the constraints that are assigned weights. Rather, symbols in linguistic representations are also assigned weights; i.e., they are not categorical either. The weights in question are encoded by assigning some real number between 0 and 1. This way, the concept of varying strength of syntactic categories can be formally implemented in the grammar. For example, suppose that some category X can have three different kinds of weights in a given grammar: X:[0.4], X:[0.7], and X:[1.0]. Suppose further that X violates some constraint Γ that is associated with a weight of 2, and that it does so once (−1). Then, the first X will give rise to a −0.4 violation of Γ, yielding a (partial) harmony score of −0.8; the second X induces a −0.7 violation of Γ, which results in a (partial) harmony score of −1.4; and the third X triggers a −1.0 violation of Γ, which produces a (partial) harmony score of −2.0. Of course, there will be constraints counter-acting Γ, which may then imply that the violation of Γ incurred by X is tolerable in an optimal candidate if X has a weight of [0.4] but not tolerable in an optimal candidate if X has a weight of [1.0].

So far, most of the work on Gradient Harmonic Grammar has been in phonology; but cf. Smolensky (2017), Lee (2018), and Müller (2019) for applications in syntax.[1]

---

[1] As it turns out, there is a fairly obvious predecessor of Gradient Harmonic Grammar in syntax (not mentioned in Smolensky & Goldrick (2016)), viz., Squishy Grammar, which was developed by Ross (1973*a*,*b*, 1975). Ross argues that there is constituent class membership to a degree, and presupposes that instead of standard category symbols like [X], there are weighted category symbols like [αX] (where α ranges over the real numbers in [0,1]). Rules, filters, and other syntactic building blocks are given upper and lower threshold values of α between which

3.3. Harmonic Serialism

In addition to Harmonic Grammar and Gradient Representations, Harmonic Serialism is a third important ingredient of the present approach. Harmonic serialism is a strictly derivational version of optimality theory. (7) illustrates how it works (see McCarthy (2008) and Heck & Müller (2013), for phonology and syntax, respectively).

(7)   *Harmonic serialism*:
   a. Given some input I_i, the candidate set $CS_i = \{O_{i1}, O_{i2}, ... O_{in}\}$ is generated by applying at most one operation to I_i.
   b. The output O_{ij} with the best constraint profile is selected as optimal.
   c. O_{ij} forms the input I_{ij} for the next generation step producing a new candidate set $CS_j = \{O_{ij1}, O_{ij2}, ... O_{ijn}\}$.
   d. The output O_{ijk} with the best constraint profile is selected as optimal.
   e. Candidate set generation stops (i.e., the derivation converges) when the output of an optimization procedure is identical to the input (i.e., when the constraint profile cannot be improved anymore).

Harmonic Serialism was already identified as a possible alternative to standard parallel optimization in Prince & Smolensky (1993). However, it has been pursued in depth only over the last decade or so (see, e.g., McCarthy (2008, 2016), Torres-Tamarit (2016), Elfner (2016) for phonology; Caballero & Inkelas (2013), Müller (2018) for morphology; and Heck & Müller (2013), Georgi (2012), Assmann et al. (2015), Murphy (2017) for syntax). As shown in McCarthy & Pater (2016) and Murphy (2017), the combination of Harmonic Grammar and Harmonic Serialism is a natural one. As far as syntax is concerned, Harmonic Serialism can be viewed as a version of minimalist approaches employing sequential bottom-up structure-building (see Chomsky (1995, 2001, 2014)) that incorporates optimization procedures (like Merge over Move). The main

---

they operate. And indeed, closer inspection reveals that Ross's (1975) concept of "clausematiness" is extremely similar in all respects to the concept of "strength of C" that the present paper will focus on in its account of the properties of the third construction in German.

   Incidentally, it seems that among those who remember it, Squishy Grammar is widely perceived to have been proven to be on the wrong track (see, e.g., Newmeyer (1986)). However, closer scrutiny reveals that the literature contains hardly any substantive criticism; and what little there is (see in particular Gazdar & Klein (1978)) is far from convincing from the perspective of current grammatical theory.

empirical arguments here concern phenomena which provide evidence that (i) there is syntactic optimization, but (ii) this optimization can only take into account information that is accessible in an extremely local syntactic domain (from the current root down to the closest phase edge), and it can only distinguish between a finite (and small) number of operations that can in principle be carried out at any given step. In the present context, a Harmonic Serialism perspective ensures that the scores of constraint violations resulting from combining the weights of the constraints and the weights assigned to the linguistic expressions are consistently fairly small and managable, and are forgotten again once the derivation moves on to the next cycle.

Taken together, the three subtheories can be referred to as *Serial Gradient Harmonic Grammar*.

4. Proposal

4.1. Constraints and Weights

In the analysis of extraction from CP to be developed below, three constraints turn out to be important. First, there is the Phase Impenetrability Condition (PIC; Chomsky (2001, 2008, 2013)), which demands that all operations involving some item α_i in a phase and some other item outside the phase requires α_i to be in the edge (specifier or head) domain of the phase. In (8), the PIC is formulated as a constraint on heads.

(8)   *Phase Impenetrability Condition* (PIC):
      For all heads Y: *Y that c-commands α_i of a dependency δ but does not m-command α_{i-1} of δ.

The PIC in (8) is a strengthened version of Chomsky's original PIC since it acknowledges a potential barrier status of *all* XPs: Every phrase is a phase. In this respect, it resembles concepts proposed by Riemsdijk (1978), Koster (1978, 1987), Sportiche (1989), and Legendre et al. (2006), among others.

For movement steps leaving a phase, the PIC in (8) thus demands that extraction takes place via the specifier of the phase head. Crucially, I assume that the PIC is an inviolable constraint of the Gᴇɴ component of the grammar (see Prince & Smolensky (1993)).[2]

---

[2]This follows without further ado if one follows Chomsky in assuming that the PIC is derivable

In contrast, the remaining two constraints are violable, and are assigned weights. These are the Merge Condition and the Anti-Locality Condition. The Merge Condition (MC) can be formulated as in (9) (see Chomsky (1995, 2001); and Heck & Müller (2013) for the particular [•F•] notation for features triggering structure-building.)

(9)   *Merge Condition* (MC):
       For all features [•F•] and XPs with a matching [F]: [•F•] triggers Merge of XP.

(9) presupposes that each head is associated with a set of structure-building features [•F•] which are discharged by individual Merge operations one at a time.[3] MC is formulated here as a constraint on two items: structure-building features on the one hand, and XPs with a matching feature on the other. This makes it possible to determine violations of the constraint (with its own weight) relative to the weights of these items (i.e., the attracting feature *and* the moved item).

The second violable constraint is the Anti-Locality Condition (see Bošković (1997), Abels (2003), Grohmann (2003a,b, 2011), Pesetsky (2016), and Erlewine (2016) for different implementations of this general idea), which is formulated in (10) in a maximally strict way that is made possible by assuming violability.

(10)   *Anti-Locality Condition* (AL):
        For all heads Y: *Y that c-commands α_i of a dependency δ and m-commands α_{i-1} of δ.

As regards links of movement dependencies, (10) is violated by all heads which c-command a (base or derived) position from which movement takes place and also m-command the landing site of this movement. The prototypical scenarios for this are (i) that movement has taken place from the specifier of some phrase ZP, across ZP's sister Y, to a specifier of Y, as in [$_{YP}$ α_{i-1} [$_{Y'}$ Y [$_{ZP}$ α_i [$_{Z'}$ ...]]]; or (ii) that movement has taken place from the complement of Y to Y's specifier, as in [$_{YP}$ α_{i-1} [$_{Y'}$ Y α_i]].[4] Given the PIC in (8) as a constraint

---

from cyclic spell-out of the phase head's complement after completion of the phase; under this assumption, material that is not in the edge domain is literally irrevocably gone after spell-out.

   [3]Alternatively, these features may be assumed to show up as members of a list (rather than a set); while ultimately important, this issue is negligible in the present context.

   [4]Strictly speaking, a third scenario might involve the configuration [$_{YP}$ Y [$_{ZP}$ α_{i-1} [$_{Z'}$ Z ... α_i

on all phrase heads, *all* movement violates AL (movement originates either in the complement position of some head Y, or in the specifier position of Y's complement). Thus, whereas MC is a trigger for movement, AL acts as a potential blocker: If AL cannot be violated in an optimal candidate, the PIC will subsequently ensure that movement cannot take place. Note that unlike a general economy constraint blocking movement (e.g., *Trace, as in Grimshaw (1997), Legendre et al. (2006)), AL has different effects depending on the nature of the head crossed in the course of movement. A head Y with a larger weight (i.e., more strength) will give rise to a more severe violation of AL than a head Y with a lower weight (i.e., less strength).

This approach depends on the availability of edge features that may trigger intermediate movement steps via MC. Following Abels (2012), I assume that intermediate movement steps are brought about by duplicates of criterial features, which can freely be assigned to any head Y. For instance, a feature like [•wh•] that is an inherent property of interrogative C in German can show up on all heads (C, T, V, v, etc.) intervening between the base position and the ultimate landing site SpecC$_{wh}$.

Summarizing so far, it emerges that weight (i.e., relative strength) plays a role for three different kinds of items that are subject to the constraints MC and AL. First, some Y heads give rise to stronger violations of AL than other Y heads if movement takes place across them. This derives asymmetries between types of local domain. For instance, VP typically permits extraction from it, and vP often does so; but CP in many cases does not. As will be shown below, this also accounts for the difference between restructuring and non-restructuring infinitival C in German, where the former but not the latter permits scrambling and unstressed pronoun fronting to the matrix domain. For concreteness, I will assume the following weights for Y heads involved in AL violations in German:

(11)   *Strength of Y*:

    a. V: [0.45]

    b. C$_{[+restr,-fin]}$: [0.6]

    c. C$_{[-restr,-fin]}$: [0.8]

    d. C$_{[-wh,+fin]}$: [0.9]

    e. C$_{[+wh,+fin]}$: [1.0]

...]]], where Y also c-commands α_i and m-commands α_{i-1}. However, it is not clear whether this scenario needs to be excluded by modifying AL (e.g., by adopting *minimal* c-command), given that α_i will never be accessible to Y because of the inviolable PIC (α_i will fail to be c-commanded by Y if it is not even part of the representation anymore at this point; see footnote 2 above).

Thus, V does not bear a lot of weight; consequently, an AL violation induced by movement to SpecV is usually tolerable in German.[5] Similar considerations apply for v and T (where the weights are not shown here). According to (11), C has more weight.[6] More generally, the underlying hypothesis is that the weight increases from bottom to top with functional heads in the clausal spine. Furthermore, all control infinitives in German are assumed to have CP status throughout. Abstracting away from the third construction for now, the infinital C head comes in two varieties, a non-restructuring version that has nearly the same weight as finite declarative C ([0.8]), and a restructuring version that has less weight ([0.6]). It is a property of restructuring control predicates that they can select either version of non-finite C (whereas other control predicates can only select the non-restructuring version).

Second, some movement-related features [•F•] give rise to stronger violations of MC (i.e., are stronger triggers of movement) than other movement-related features. This derives asymmetries between movement types. For instance, wh-movement can leave a finite CP in German whereas scrambling cannot do so. Concrete weights assigned to structure-building features that trigger movement in German include those in (12); [•wh•] is involved in wh-movement, and [•scr•] is involved in scrambling and unstressed pronoun fronting.[7] Again, the increase in strength corresponds to the relative position of the head(s) bearing the feature in the tree: The landing site of wh-movement is SpecC, the landing site of scrambling is Specv or SpecV.[8]

---

[5]See, however, Müller (2019), where I argue that the ban on splitting up particularly opaque kinds of idioms by certain kinds of movement can be traced back to an AL violation with movement to SpecV that is fatal in the presence of a moved item with extremely little strength (giving rise to a less severe MC violation if movement does not take place).

[6]Also, a finite interrogative C has more weight than a finite declarative C ([1.0] vs. [0.9]); this ultimately accounts for wh-islands; see Müller (2019).

[7]There are in fact several differences between scrambling of non-pronominal items, as in (2b), and unstressed pronoun fronting, as in (1b). Still, to simplify matters I pretend here that [•scr•] covers both movements; a more detailed analysis would postulate two separate features with sufficiently similar weights.

[8]Topicalization can leave wh-islands in German with objects (but not subjects), whereas wh-movement (or scrambling) cannot do so. In Müller (2019), this is modelled by assuming that the feature [•top•], which triggers topicalization, has more weight than the features triggering wh-movement and scrambling (viz., [0.65] vs. [0.5], [0.2]).

(12)   *Strength of* [•F•]:
    a. [•scr•]: [0.2]
    b. [•wh•]: [0.5]

Third, some XPs give rise to stronger violations of MC than other XPs if they
do not undergo movement. This accounts for asymmetries between moved
items (e.g., unmoved objects may induce stronger violations of MC than
unmoved subjects, and thus make MC violable less easily in optimal outputs).
For German, I assume that an object DP has a weight of [0.9], whereas a subject
DP only has a weight of [0.8]. However, I will be exclusively concerned with
object DPs in what follows.[9]

    With these assumptions in place, let me next illustrate the mechanics of the
resulting system on the basis of some data involving extraction from different
domains, and by different movement types.

4.2. Two Extraction Asymmetries in German

4.2.1. *Asymmetries between Types of Local Domain*

Scrambling can target SpecV in the German, either as a final landing site, or as
an intermediate escape hatch for further movement to Specv required by the
PIC; see (13a) and (13b), respectively.

(13)   a. dass sie [$_{VP}$ [$_{DP_2}$ das Buch] [$_{V'}$ [$_{DP_1}$ dem Karl] [$_{V'}$ t$_2$ [$_V$
          that she   the book$_{acc}$ the Karl$_{dat}$
          gegeben hat]]]]
          given   has
    b. dass [$_{vP}$ [$_{DP_2}$ das Buch] [$_{v'}$ [$_{DP_1}$ keiner]   [$_{v'}$ [$_{VP}$ t$'_2$ [$_{V'}$ t$_2$
          that   the book$_{acc}$ no-one$_{nom}$
          gelesen hat]] v]]]
          read   has

However, as noted above, scrambling is clause-bound in German (see Ross
(1967)): A finite CP cannot be crossed. From the present, PIC-based perspective,

---

[9] See Müller (2019) for discussion of asymmetries between types of moved items.

this can be taken to indicate that SpecC cannot be targetted as an intermediate landing site by this movement operation; see (14).[10]

(14) *dass sie [$_{DP_2}$ das Buch] gesagt hat [$_{CP}$ t$'_2$ [$_{C'}$ dass] [$_{TP}$ t$_2$ sie
 that she the book$_{acc}$ said has that she
 gelesen hat]]
 read has

This asymmetry between VP and CP with respect to scrambling follows from the current assumptions about weight assignments. On the one hand, given that what is moved is an object DP ([0.9]), and given that the feature responsible for the (intermediate or final) movement step is ([•scr•] ([0.2], a relatively weak trigger), there will be a −1.1 violation of MC in both environments if movement does not take place. Assuming MC itself to have a weight of 2.0, this produces a harmony score of −2.2. On the other hand, if movement takes place, an AL violation will be generated. Suppose that the intrinsic weight of AL is 3.0. Then, since V, by assumption, has a weight of [0.45] (see (11)), movement of any item to SpecV gives rise to a −0.45 violation of AL, and thus (abstracting away from other constraint violations that are irrelevant in the present context) to a harmony score of −1.35. Consequently, the output candidate O$_2$ employing a local scrambling step to SpecV emerges as optimal, and the output candidate O$_1$ which fails to carry out movement is suboptimal. This is illustrated by the tableau in (15) (where $\mathcal{H}$ stands for the overall harmony score of a candidate).

(15)   *Object scrambling via VP*:

I: [$_{VP}$ ... DP$_{obj:[0.9]}$ V$_{[0.45]}$,[•scr•]:[0.2]]	MC	AL	$\mathcal{H}$
	w = 2.0	w = 3.0	
O$_1$: [$_{VP}$ ... DP$_{obj:[0.9]}$ V$_{[0.45]}$,[•scr•]:[0.2]]	−1.1		−2.2
☞O$_2$: [$_{VP}$ DP$_{obj:[0.9]}$ [$_{V'}$ ... t$_{obj}$ V$_{[0.45]}$,[•scr•]:[0.2]]]		−0.45	−1.35

In contrast, if object scrambling wants to leave a finite declarative CP, intermediate movement to SpecC, across an intervening C with weight [0.9], produces a much more severe violation of AL: This time there is a −0.9 violation of AL, which ceteris paribus leads to a harmony score of −2.7. The candidate without movement (in the presence of [•scr•] and an object DP) has a harmony score

---

[10] In contrast, there would be nothing wrong as such with the subsequent movement step to matrix SpecV. Such a step is often excluded by some specific constraints against improper movement (see Müller (2014), Keine (2016) for recent overviews), but in the present approach based on variable weights, such constraints can be dispensed with; cf. 4.2.2 below.

of -2.2, exactly as before; but this MC violation now emerges as optimal, and intermediate scrambling to SpecC is therefore blocked. Ultimately, the PIC then ensures that long-distance scrambling cannot take place from the lower SpecT position in the embedded clause that we can assume to have been reached by prior intermediate scrambling-movement. This competition is shown in (16).

(16)　　*Object scrambling via finite declarative CP*:

	MC	AL	$\mathcal{H}$
I: $[_{CP}$ C$_{[0.9]}$,$[\bullet scr\bullet]:[0.2]$ $[_{TP}$ DP$_{obj:[0.9]}$ $[_{T'}$ ... T $]]]$	w = 2.0	w = 3.0	
☞O$_1$: $[_{CP}$ C$_{[0.9]}$,$[\bullet scr\bullet]:[0.2]$ $[_{TP}$ DP$_{obj:[0.9]}$ $[_{T'}$ ... T $]]]$	-1.1		-2.2
O$_2$: $[_{CP}$ DP$_{obj:[0.9]}$ $[_{C'}$ C$_{[0.9]}$,$[\bullet scr\bullet]:[0.2]$ $[_{TP}$ t$_2$ $[_{T'}$ ... T $]]]]$		-0.9	-2.7

Next, if different kinds of Cs ([±finite], [±restructuring], [±wh], etc.) can have different weights, it can be derived that one and the same movement type (e.g., scrambling) may leave CPs with a weak C head (restructuring infinitives) but not CP with a stronger C head (finite clauses or non-restructuring infinitives). A relevant pair of examples illustrating the lexically governed restructuring effect with control infinitives in German is given in (17).

(17)　　a. dass $[_{DP_{obj}}$ das Buch]　keiner　　　$[_{CP}$ t$'_2$ $[_{C'}$ C $[_{TP}$ PRO t$_2$ zu
　　　　　　that　　　the book$_{acc}$ no-one$_{nom}$　　　　　　　　　　　　to
　　　　　　lesen]]] versucht hat
　　　　　　read　　　tried　　has

　　　　b. *dass $[_{DP_{obj}}$ das Buch]　keiner　　　$[_{CP}$ t$'_2$ $[_{C'}$ C $[_{TP}$ PRO t$_2$ zu
　　　　　　that　　　the book$_{acc}$ no-one$_{nom}$　　　　　　　　　　　　to
　　　　　　lesen]]] abgelehnt hat
　　　　　　read　　　rejected　has

By assumption, restructuring C in (17a) has a weight of [0.6], whereas non-restructuring C in (17b) has a weight of [0.8]. Consequently, non-restructuring infinitival C blocks scrambling from it in basically the same way as finite declarative C in (16) (with a suboptimal harmony score of -2.4 if movement applies, violating AL); but with restructuring C, the AL violation incurred by movement is not so severe anymore (the overall harmony score is -1.8), and successfully blocks the candidate that fails to carry out movement (in violation of MC, with a harmony score of -2.2); see (18).[11]

---

[11] There is considerable variation among speakers of German as to which matrix control

(18) *Object scrambling via restructuring infinitive CP*:

	MC $w = 2.0$	AL $w = 3.0$	$\mathcal{H}$
I: [$_{CP}$ C$_{[0.6]}$,$_{[\bullet scr\bullet]:[0.2]}$ [$_{TP}$ DP$_{obj:[0.90]}$ [$_{T'}$ ... T]]]			
O$_1$: [$_{CP}$ C$_{[0.6]}$,$_{[\bullet scr\bullet]:[0.2]}$ [$_{TP}$ DP$_{obj:[0.9]}$ [$_{T'}$ ... T]]]	−1.1		−2.2
☞O$_2$: [$_{CP}$ DP$_{obj:[0.9]}$ [$_{C'}$ C$_{[0.6]}$,$_{[\bullet scr\bullet]:[0.2]}$ [$_{TP}$ t$_{obj}$ [$_{T'}$ ... T]]]]		−0.6	−1.8

The present approach makes it possible to uniformly assume a CP status of restructuring infinitives embedded under control verbs. This is arguably conceptually attractive in view of the implicational generalization that there is no control verb that permits restructuring which would not also permit a non-restructuring clausal complement. In approaches where the two complement types have a different categorial status (e.g., vP vs. CP; see Haider (1993, 2010), Wurmbrand (2001)), this state of affairs is purely accidental; in the present approach, it only requires the assumption that there is an unmarked strength of infinitival C items (viz., [0.8]) which can optionally be reduced (and which then is tolerated only by a subset of control predicates). However, there is also empirical evidence for CP in restructuring infinitives embedded by control verbs; see Baker (1988), Sternefeld (1990), Müller & Sternefeld (1995), Sabel (1996), Koopman & Szabolcsi (2000), and Müller (2017). For instance, one argument from the last-mentioned study relies on the generalization that unstressed pronoun fronting to the left edge of vP (which is obligatory in German) must be licensed by a higher C phase head. And whereas such movement is impossible in structures clearly lacking a CP (verb-auxiliary combinations as in (19a), raising environments as in (19b)), it *is* possible in restructuring contexts embedded by control verbs (as in (19c)).[12]

(19) a. *dass sie   mir$_1$ schon   letzte Woche [$_{vP}$ es$_2$ t$_1$ t$_2$ gegeben]
          that she$_{nom}$ me$_{dat}$ already last   week   it$_{acc}$ given
          hat
          has

---

predicates permit restructuring, and which ones do not. For some speakers, (17b) may be possible, but this does not affect the analysis: *ablehnen* just tolerates a weaker C here.

[12]In these examples, *mir$_1$* undergoes fronting to the matrix domain, thereby indicating transparency of the complement of the higher verb; *es$_1$* is fronted string-vacuously in the complement.

    b. *dass sie   mir   schon letzte Woche [$_{vP}$ es$_2$ t$_2$ zu lesen]
       that she$_{nom}$ me$_{dat}$ already last   week   it$_{acc}$ to read
       schien
       seemed

    c. dass sie   mir$_1$ schon letzte Woche [$_{CP}$ es$_2$ PRO t$_1$ t$_2$ zu
       that she$_{nom}$ me$_{dat}$ already last   week   it$_{acc}$ to
       geben] versucht hat
       give   tried   has

From a slightly more general perspective, under present assumptions there can be a lot of variation as far as the transparency of projections in the clausal spine for extraction is concerned (depending on the weights assigned to the heads in the extended projection of V). However, the variation is principled in the sense that it must obey an implicational universal: If an XP α can undergo Σ-movement across a Y head δ_1, and δ_1 has more weight than another Y head δ_2, then α can ceteris paribus also undergo Σ-movement across δ_2. Given the ancillary assumption that weight increases from bottom to top in the clausal spine, it is then predicted that if a given movement type affecting some particular item can take place across CP, it can also take place across TP; if it can leave TP, it can ceteris paribus leave vP; and similarly for vP and VP. I take this prediction to be correct.

4.2.2. *Asymmetries between Movement Types*

If a given head Y blocks a movement type triggered by a (intermediate or final) feature Σ_1 because the AL violation incurred by movement has a lower harmony score than the relatively weak MC violation incurred by not moving the item, this does not necessarily mean that Y will also block another movement type triggered by a different feature Σ_2: Not satisfying Σ_2's demand by leaving the item in place may give rise to a much more severe violation of MC if Σ_2 has greater strength than Σ_1, and this can then make the AL violation optimal. Such a situation obtains with wh-movement (triggered by [•wh•]) vs. scrambling (triggered by [•scr•]). Recall from (12) that the former feature is associated with a weight of [0.5] in German, and the latter with a weight of [0.2]. And indeed, for most speakers of German, wh-movement can leave a finite declarative CP where scrambling cannot (for reasons discussed in the previous subsection); see (20a) (with wh-movement) vs. (20b) (= (14)).

(20)　a.　(Ich weiß　nicht) [$_{CP}$ [$_{DP_2}$ welches Buch]　sie gesagt hat [$_{CP}$ t$'_2$
　　　　　I　know not　　　　　which　book$_{acc}$ she said　has
　　　　　[$_{C'}$ dass] [$_{TP}$ t$_2$ sie gelesen hat]]
　　　　　　that　　　　she read　　has
　　　b.*dass sie [$_{DP_2}$ das Buch]　gesagt hat [$_{CP}$ t$'_2$ [$_{C'}$ dass] [$_{TP}$ t$_2$ sie
　　　　　that she　　　the book$_{acc}$ said　has　　　　　　that　　　she
　　　　　gelesen hat]]
　　　　　read　　has

As shown in (21), wh-movement of an object DP via VP (as in O$_2$) is entirely unproblematic; as was the case with scrambling (see (15)), an AL violation is tolerable because the overall harmony score is closer to zero than that of a candidate that does not carry out movement in violation of MC (cf. O$_1$).

(21)　*Object wh-movement via VP*:

I: [$_{VP}$ ... DP$_{obj:[0.9]}$ $V_{[0.45]›[•wh•]:[0.5]}$]	MC	AL	$\mathcal{H}$
	w = 2.0	w = 3.0	
O$_1$: [$_{VP}$ ... DP$_{obj:[0.9]}$ $V_{[0.45]›[•wh•]:[0.5]}$]	−1.4		−2.8
☞O$_2$: [$_{VP}$ DP$_{obj:[0.9]}$ [$_{V'}$ ... t$_{obj}$ $V_{[0.45]›[•wh•]:[0.5]}$]]		−0.45	−1.35

However, things are different when it comes to extraction via CP. As shown in (22), the output candidate that moves the object DP to SpecC (i.e., O$_2$) now still has a better constraint profile than the candidate that does without such movement (i.e., O$_1$): The reason is that C's [•wh•] feature in (22) (with a weight of [0.5]) ceteris paribus gives rise to a much stronger violation of MC if movement does not take place than C's [•scr•] feature in (16) (with a weight of [0.2]) does.

(22)　*Object wh-movement via finite declarative CP*:

I: [$_{CP}$ C$_{[0.9]›[•wh•]:[0.5]}$ [$_{TP}$ DP$_{obj:[0.9]}$ [$_{T'}$ ... T]]]	MC	AL	$\mathcal{H}$
	w = 2.0	w = 3.0	
O$_1$: [$_{CP}$ C$_{[0.9]›[•wh•]:[0.5]}$ [$_{TP}$ DP$_{obj:[0.9]}$ [$_{T'}$ ... T]]]	−1.4		−2.8
☞O$_2$: [$_{CP}$ DP$_{obj:[0.9]}$ [$_{C'}$ C$_{[0.9]›[•wh•]:[0.5]}$ [$_{TP}$ t$_{obj}$ [$_{T'}$ ... T]]]]		−0.9	−2.7

Again, the approach predicts a lot of variation, but as before, such variation is principled: A second implicational universal can be derived which states that if an XP α can undergo Σ$_1$-movement across a Y head δ, and Σ$_1$ has less weight than another movement type Σ$_2$, then α can also undergo Σ$_2$-movement across δ, other things being equal. And, also as before, the relative weight of the

features that bring about movement via MC is not arbitrary but corresponds to the relative position of the heads bearing the features in the tree.[13]

Needless to say, the approach to extraction in German sketched so far needs to be extended in many directions, and with a broader empirical coverage, it must be subject to many further ramifications. However, I will leave it at that here. Instead, I will now turn to the main goal of the present paper, which is to solve the paradox with the third construction outlined in section 1 above.

5. The Third Construction

In many respects, the extraposed infinitival complement in the third construction in German behaves like the non-extraposed restructuring infinitive counterpart analyzed in subsection 4.2.1 above. First, as noted in section 1, the extraposed infinitival complement is transparent for scrambling and unstressed pronoun fronting if it would be transparent for these movement types in the pre-verbal base position – i.e., if the matrix predicate licenses restructuring.[14] Some relevant examples that document this are given in (23a), (23b) (= (1a)), and (23c).

(23)   a. dass das Buch$_2$ keiner t_1 versucht hat [$_{CP_1}$ PRO t_2 zu lesen]
           that the book$_{acc}$ no-one$_{nom}$ tried   has   to read
      b. dass sie   ihn$_2$ t_1 versucht [$_{CP_1}$ PRO t_2 zu küssen]
           that she$_{nom}$ him$_{acc}$ tries   to kiss
      c. dass es$_2$ Fritz   ihr t_1 empfohlen   hat [$_{CP_1}$ PRO im
           that it$_{acc}$ Fritz$_{nom}$ her$_{dat}$ recommended has   on the
           Zug t_2 zu lesen]
           train   to read

As with restructuring infinitives in situ, this might initially be taken to suggest that extraposed restructuring infinitives in the third construction do not have CP status. But as before, there are conceptual and empirical arguments for the presence of a CP shell here. For instance, the third construction provides a

---

[13]Concerning variation, it is also worth noting that by slightly increasing the weight of finite declarative C, wh-movement from CP will become impossible. As a matter of fact, such a scenario comes close to the situation in certain Northern varieties of German, which do not easily permit wh-movement from finite declarative clauses headed by a C with *that*.

[14]Of course, this holds true virtually by definition – movement from an extraposed restructuring infinitive is *the* constitutive property of the third construction.

C-licensed landing site (at the left edge of the embedded vP) for unstressed pronoun fronting, just like restructuring infinitives in situ do (cf. (19)); see (24) (where fronting of mir_3 into the matrix domain indicates transparency of the extraposed infinitive, and string-vacuous movement of es_2 indicates the presence of C as a licensor for unstressed pronoun fronting in the infinitive).

(24)   dass sie mir_1 schon letzte Woche versucht hat [$_{CP}$ es_2 PRO t_1
        that she$_{nom}$ me$_{dat}$ already last   week   tried   has   it$_{acc}$
        t_2 zu geben]
           to give

However, there are also differences between standard (i.e., pre-verbal) restructuring control infinitives and the third construction. In particular, there is Santorini & Kroch's (1991) observation that a negation showing up in the extraposed infinitive can never take wide scope; cf. (2), repeated here in (25) (with CP_1 replacing the original Γ_1 as the label of the infinitive, and some other information added).

(25)   dass ich seinen neuesten $Roman_2$ t_1 beschlossen habe [$_{CP_1}$ PRO t_2
        that I   his   newest novel$_{acc}$ decided   have
        nicht zu lesen]
        not   to read

                                                        (only narrow scope)

Thus, we end up with the paradox that extraposed infinitives in restructuring contexts are transparent for scrambling but not transparent for scope of sentential negation. This paradox arguably poses a non-trivial problem for standard approaches.[15] From the present perspective, a simple solution suggests itself: The C head of the extraposed infinitive in the third construction has more

---

[15] One might think that *directionality* could be the relevant factor determining obligatorily narrow scope of negation in the third construction, especially since there is some evidence that pre- vs. postverbal position can play a role for scope assignment in German when focus particles are involved (see Bayer (1996)). However, for the case at hand, this seems unlikely. As shown in (i), a universal quantifier embedded in an extraposed PP can easily take wide scope (as a matter of fact, wide scope of the universal quantifier produces the only reading that is compatible with world knowledge).

(i)   dass der Polizist   eine Bombe t_1 gefunden hat [$_{PP_1}$ hinter jedem Haus]
       that the policeman$_{nom}$ a   bomb$_{acc}$ found   has   behind every house

strength than the C head of a restructuring infinitive in situ but less strength than the C head of a non-restructuring infinitive (or a finite C). More specifically, I would like to suggest that the C head of an extraposed infinitive in the third construction has a weight of $[0.7]$ (as opposed to $[0.8]$ for a non-restructuring C and $[0.6]$ for a regular restructuring C; cf. (11)).

A first consequence of this weight assignment to non-finite C in the third construction is that it patterns with restructuring C as far as scrambling or unstressed pronoun fronting to the matrix domain is concerned, rather than with non-restructuring (or finite) C. Thus, the outcome of the competition in (26) parallels that of (18) (where the optimal output candidate violates AL by applying the intermediate movement step to SpecC required by the PIC), and not that of (16) (where the optimal output candidate violates MC by not carrying out movement); see (26).

(26)   *Object scrambling via extraposed infinitive CP in the third construction*:

I: $[_{CP} C_{[0.7],[•scr•]:[0.2]} [_{TP} DP_{obj:[0.90]} [_{T'} \ldots T]]]$	MC	AL	$\mathcal{H}$
	w = 2.0	w = 3.0	
O_1: $[_{CP} C_{[0.7],[•scr•]:[0.2]} [_{TP} DP_{obj:[0.9]} [_{T'} \ldots T]]]$	-1.1		-2.2
☞O_2: $[_{CP} DP_{obj:[0.9]} [_{C'} C_{[0.7],[•scr•]:[0.2]} [_{TP} t_{obj} [_{T'} \ldots T]]]]$		-0.7	-2.1

The AL violation incurred by DP movement to SpecC in O_2 is more severe in (26) (-2.1) than it was in the case of restructuring infinitives in situ in (18) (-1.8), but the harmony score is still better than the harmony score of the competing output O_1 where movement fails to apply, and MC (with weight 2.0) gets a combined -1.1 violation incurred by the $[•scr•]$ feature ($[0.2]$) and the object DP ($[0.9]$), yielding a fatal -2.2 overall.

On the other hand, the larger weight of $[0.7]$ for this type of non-finite C can be held responsible for differences to standard restructuring infinitives. First of all, suppose that CP extraposition in German targets the next higher CP domain (a right-peripheral specifier or adjunct) if extraction from the extraposed CP needs to take place.[16] This implies that in order to permit a combination of CP extraposition and extraction from CP, an infinitive must

---

[16] See Müller (1998) for arguments to this effect. If there is no extraction from CP, extraposition can also target a lower position, and then participate in VP topicalization. This accounts for the contrast in (i-a) (without extraction from the extraposed infinitive) and (i-b) (without extraposition) vs. (i-c) (with extraction from the extraposed infinitive).

(i)   a. $[_{VP_3}$ t2 Versucht $[_{CP_2}$ dem Peter   das Buch2   zu geben]] hat sie   nicht t3
         tried   the Peter$_{dat}$ the book$_{acc}$ to give   has she$_{nom}$ not

have sufficient weight to outweigh the AL violation automatically incurred by all movement across a finite C; as we have seen, the latter has a harmony score of −2.7. Assuming a feature [•ex•] involved in extraposition to have a strength of [0.7], it is correctly predicted that an infinitival CP with a C head with strength [0.7] can undergo extraposition to the next higher CP domain, in optimal violation of AL: If movement does not take place, the resulting MC violation leads to a harmony score of −2.8. All of this is shown in (27).

(27)   *Infinitive extraposition in the third construction*:

I: $[_{CP}$ C$_{[0.9]\rangle[\bullet ex \bullet]:[0.7]}$ ... CP$_{[0.7]}$ V$_{restr}$]	MC	AL	$\mathcal{H}$
	w = 2.0	w = 3.0	
O$_1$: $[_{CP}$ C$_{[0.9]\rangle[\bullet ex \bullet]:[0.7]}$ ... CP$_{[0.7]}$ V$_{restr}$]	−1.4		−2.8
☞O$_2$: $[_{CP}$ C$_{[0.9]\rangle[\bullet ex \bullet]:[0.7]}$ ... t$_{cp}$ V$_{restr}$ CP$_{[0.7]}$]		−0.9	−2.7

Under these assumptions, it is clear that if the infinitival CP has a smaller weight of [0.6], it can never be affected by extraposition to the CP domain – in this latter case, the harmony score of −2.6 amassed by the MC-violating output is better than the harmony score of the AL-violating candidate that applies extraposition (which continues to be −2.7).

Finally, the lack of wide scope for negation in the third construction (and the concurrent availability of wide scope for negation in regular, preverbal restructuring infinitives) can also be tied to the different weights ([0.7] vs. [0.6]). I assume that scope of negation is in general the consequence of an Agree relation between an abstract operator position high in the clause and an overt negative item, which is typically in a much lower position in German (see Stechow (1993) and Zeijlstra (2004), among others). Agree is subject to an Agree Condition (AC; see Heck & Müller (2013)) that requires probe features ([*F*]) to participate in Agree with appropriate goal features ([F]). In the case at hand, there is a probe feature [*neg*] on the overt negation (*nicht* in (25)), and a goal feature [neg] in the left periphery of the matrix clause. Suppose furthermore that to bridge the distance in a local way that is compatible with the strict PIC employed here, Agree must take place cyclically (see Legate

---

b. $[_{VP_3}$ $[_{CP_2}$ Dem Peter   t$_1$ zu geben] versucht] hat sie   das Buch$_1$ nicht t$_3$
            the   Peter$_{dat}$ to give   tried   has she$_{nom}$ the book$_{acc}$ not

c.??$[_{VP_3}$ t$_2$ Versucht $[_{CP_2}$ dem Peter   t$_1$ zu geben]] hat sie   das Buch$_1$
        tried   the Peter$_{dat}$ to give   has she$_{nom}$ the book$_{acc}$

    nicht t$_3$
    not

(2005)). Such cyclic Agree will then also give rise to an AL violation for every head that it involves on the path to the ultimate target position in the matrix clause.[17] On this basis, it can be concluded that the harmony score of an output that does not carry out cyclic Agree for a [∗neg∗] feature across a CP and thereby violates AC must be better than -2.1 (so as be optimal vis-a-vis the harmony score of -2.1 resulting from AL if cyclic Agree across C applies in the third construction), but worse than -1.8 (so as to be suboptimal vis-a-vis the harmony score of -1.8 resulting from AL if cyclic Agree across C applies with regular restructuring infinitives). This result is achieved if, e.g., [∗neg∗] has a weight of [1.0], and AC has a weight of [2.0]. The competition underlying failed wide scope of negation in the third construction is illustrated in (28).

(28) *Wide scope of negation in the third construction*:

I: [$_{CP}$ C$_{[0.7]}$ ... [∗neg∗]:[1.0] ...]	AC	AL	$\mathcal{H}$
	w = 2.0	w = 3.0	
☞O$_1$: [$_{CP}$ C$_{[0.7]}$ ... [∗neg∗]:[1.0] ...]	−1.0		−2.0
O$_2$: [$_{CP}$ C$_{[0.7]}$[∗neg∗] ... [∗neg∗]:[1.0] ...]		−0.7	−2.1

Thus, the PIC will block any non-local transmittance of [∗neg∗], and an Agree relation with the target position in the matrix clause cannot be established in the third construction. Of course, with a lower C weight of [0.6] (as in regular restructuring infinitives), the candidate that carries out (intermediate) cyclic Agree with the C head (as required by AC) becomes optimal: Now the violation of AL is less severe (yielding a harmony score of -1.8).[18]

---

[17] Strictly speaking, given the definition of AL in (10), this presupposes that if there is a (cyclic) Agree dependency between [(*)neg(*)] on some head Y and [∗neg∗] on an item c-commanded by Y, the former feature must be *m-commanded* by Y to generate an AL violation. Depending on the exact nature of feature insertion in cyclic Agree contexts and the precise definition of m-command, this may either follow directly, or it may require a generalization of the concept of m-command (e.g., along the lines of Chomsky's (1995) notion of minimal residue).

[18] Ultimately, a bit more will have to be said. E.g., it is generally held that narrow scope of negation is in fact impossible in standard restructuring infinitives. This does not yet follow from the analysis; an obvious possibility here might be to assume that a certain strength of C is required to license an interpretable [neg] feature. In this context, it is worth pointing out that the present approach to scope of negation in terms of cyclic Agree is by far not the only one that can be entertained. One could, e.g., assume that AC-driven Agree does not obey the PIC (cf., e.g., Bošković (2007)), and then let the strength differences of the two infinitival C heads (restructuring vs. third construction) interact with a violable intervention constraint.

6. Strength and Morphological Realization

In section 2 above, I concluded that strength is an abstract property of heads that can have two different consequences: First, it determines whether or not syntactic operations can apply, and second, it also determines post-syntactic morphological realization. In the present study of strength of C in German I have focussed on the former issue; to end this paper, let me make a few remarks on the latter one.

In Lee (2018), it is argued that finite declarative C in English comes in two versions distinguished only by their strength. Strong C blocks wh-movement of subjects (but not of objects, which are themselves stronger than subjects); weak C does not. Transferring this analysis to the present approach in terms of MC and AL, this follows if weak C has a weight of [0.5] in English, strong C has a weight of [1.0], [●wh●] has a weight of [0.8], and subject and object DPs have weights of [0.4] and [0.8], respectively. Crucially, Lee (2018) shows that these different weight assignments to declarative finite C in English can also be assumed to govern post-syntactic morphological realization. A strong C:[1.0] gives rise to a severe (and fatal) violation of a constraint demanding vocabulary insertion if it is not post-syntactically realized by *that*; in contrast, with a weak C:[0.5], the violation of this constraint is not so severe anymore, and the violation of a DEP constraint prohibiting vocabulary insertion that is incurred by the presence of *that* becomes fatal. Thus, the complementizer-trace effect in (3b) (vs. (3a)) is derived without giving up the assumption that the morphological shape of C is determined only post-syntactically.

In the same way, the fact that finite declarative C can be morphologically realized by *dass* in German whereas the non-finite Cs of control infinitives are not realized by morphological exponents does not emerge as fully accidental under present assumptions: The former kind of C is stronger than the latter ones ([0.9] vs. [0.6], [0.7], [0.8]). Thus, whereas one might abstractly conceive of a variety of German where, e.g., Cs of non-restructuring infinitives are also overtly realized in some way whereas Cs of the third construction and restructuring infinitives are not, the prediction clearly is that it would ceteris paribus be impossible to have a variety of German where the Cs that are more transparent to movement are overtly realized, and Cs that are less transparent remain without morphological exponence. I take this to be a non-trivial and welcome result.

References

Abels, Klaus (2003). *Successive Cyclicity, Anti-Locality, and Adposition Stranding.* PhD thesis, University of Connecticut. Storrs, Connecticut

Abels, Klaus (2012). *Phases. An Essay on Cyclicity in Syntax*. Vol. 543. of *Linguistische Arbeiten*. De Gruyter: Berlin.

Assmann, Anke, Doreen Georgi, Fabian Heck, Gereon Müller & Philipp Weisser (2015). Ergatives Move Too Early. On an Instance of Opacity in Syntax. *Syntax* 18. 343–387.

Baker, Mark (1988). *Incorporation. A Theory of Grammatical Function Changing.* University of Chicago Press: Chicago.

Bayer, Josef (1996). *Directionality and Logical Form. On the Scope of Focussing Particles and Wh-in-situ*. Kluwer: Dordrecht.

Besten, Hans den & Jean Rutten (1989). On Verb Raising, Extraposition and Free Word Order in Dutch. In D. Jaspers, W. Klooster, Y. Putseys & P. Seuren (eds). *Sentential Complementation and the Lexicon. Studies in Honor of Wim de Geest*. Foris: Dordrecht. 41–56.

Bobaljik, Jonathan (2002). Realizing Germanic Inflection: Why Morphology Does Not Drive Syntax. *Journal of Comparative Germanic Linguistics* 6. 129–167.

Bošković, Željko (1997). *The Syntax of Nonfinite Complementation. An Economy Approach*. MIT Press: Cambridge, Mass..

Bošković, Željko (2007). On the Locality and Motivation of Move and Agree: An Even More Minimal Theory. *Linguistic Inquiry* 38. 589–644.

Caballero, Gabriela & Sharon Inkelas (2013). Word Construction: Tracing an Optimal Path Through the Lexicon. *Morphology* 23. 103–143.

Chomsky, Noam (1995). *The Minimalist Program*. MIT Press: Cambridge, Mass..

Chomsky, Noam (2001). Derivation by Phase. In M. Kenstowicz (ed.). *Ken Hale. A Life in Language*. MIT Press: Cambridge, Mass.. 1–52.

Chomsky, Noam (2008). On Phases. In R. Freidin, C. Otero & M. L. Zubizarreta (eds). *Foundational Issues in Linguistic Theory*. MIT Press: Cambridge, Mass.. 133–166.

Chomsky, Noam (2013). Problems of Projection. *Lingua* 130. 33–49.

Chomsky, Noam (2014). Lecture 4. Class Lectures, MIT, May 19, 2014 (31:00-48:00). Available from: http://whamit.mit.edu/2014/06/03/recent-linguistics-talks-by-chomsky/.

Chomsky, Noam (2015). Problems of Projection: Extensions. In E. Di Domenico, C. Hamann & S. Matteini (eds). *Structures, Strategies and Beyond: Studies in Honour of Adriana Belletti*. Benjamins: Amsterdam. 1–16.

Elfner, Emily (2016). Stress-Epenthesis Interactions in Harmonic Serialism. In *Harmonic Grammar and Harmonic Serialism*. Equinox: Sheffield.

Erlewine, Michael (2016). Anti-Locality and Optimality in Kaqchikel Agent Focus. *Natural Language and Linguistic Theory* 34. 429–479.

Gazdar, Gerald & Ewan Klein (1978). Review of Ed Keenan, ed., Formal Semantics of Natural Language. *Language* 54. 661–667.

Geilfuß, Jochen (1991). Zur Syntax des Verbalkomplexes. Ms., Universität Tübingen.

Georgi, Doreen (2012). A Local Derivation of Global Case Splits. In A. Alexiadou, T. Kiss & G. Müller (eds). *Local Modelling of Non-Local Dependencies in Syntax*. Linguistische Arbeiten De Gruyter: Berlin. 306–336.

Grimshaw, Jane (1997). Projection, Heads, and Optimality. *Linguistic Inquiry* 28. 373–422.

Grohmann, Kleanthes K. (2003*a*). *Prolific Domains: On the Anti-Locality of Movement Dependencies*. John Benjamins: Amsterdam & Philadelphia.

Grohmann, Kleanthes K. (2003*b*). Successive Cyclicity under (Anti-)Local Considerations. *Syntax* 6. 260–312.

Grohmann, Kleanthes K. (2011). Anti-Locality: Too-Close Relations in Grammar. In C. Boeckx (ed.). *The Oxford Handbook of Linguistic Minimalism*. Oxford University Press: . 260–290.

Haider, Hubert (1993). *Deutsche Syntax – generativ*. Narr: Tübingen.

Haider, Hubert (2010). *The Syntax of German*. Cambridge University Press: Cambridge.

Halle, Morris & Alec Marantz (1993). Distributed Morphology and the Pieces of Inflection. In K. Hale & S. J. Keyser (eds). *The View from Building 20*. MIT Press: Cambridge, Mass.. 111–176.

Heck, Fabian & Gereon Müller (2013). Extremely Local Optimization. In H. Broekhuis & R. Vogel (eds). *Linguistic Derivations and Filtering*. Equinox: Sheffield. 135–166.

Holmberg, Anders & Christer Platzack (1995). *The Role of Inflection in the Syntax of the Scandinavian Languages*. Oxford University Press: Oxford.

Keine, Stefan (2016). *Probes and Their Horizons.* PhD thesis, University of Massachusetts, Amherst.

Koopman, Hilda & Anna Szabolcsi (2000). *Verbal Complexes*. MIT Press: Cambridge, Mass..

Koster, Jan (1978). *Locality Principles in Syntax*. Foris: Dordrecht.

Koster, Jan (1986). The Relation Between Pro-Drop, Scrambling, and Verb Movements. Ms., Rijksuniversiteit Groningen.

Koster, Jan (1987). *Domains and Dynasties*. Foris: Dordrecht.

Lee, Hyunjung (2018). Generalized Complementizer-Trace Effects in Gradient Harmonic Grammar: Deriving Extraction Asymmetries. Ms., Universität Leipzig. (Talk, 40th DGfS Meeting, Stuttgart, March 2018).

Lee-Schoenfeld, Vera (2007). *Beyond Coherence*. Benjamins: Amsterdam.

Legate, Julie Anne (2005). Phases and Cyclic Agreement. *MITWPL* 49. 147–156. . Perspectives on Phases

Legendre, Géraldine, Colin Wilson, Paul Smolensky, Kristin Homer & William Raymond (2006). Optimality in Syntax II: Wh-Questions. In P. Smolensky & G. Legendre (eds). *The Harmonic Mind*. Vol. II. MIT Press: Cambridge, Mass.. chapter 14. 183–230.

McCarthy, John (2008). The Serial Interaction of Stress and Syncope. *Natural Language and Linguistic Theory* 26. 499–546.

McCarthy, John (2016). The Theory and Practice of Harmonic Serialism. In J. McCarthy & J. Pater (eds). *Harmonic Grammar and Harmonic Serialism*. Equinox: Sheffield. 47–87.

McCarthy, John & Joe Pater (eds) (2016). *Harmonic Grammar and Harmonic Serialism*. Equinox: Sheffield.

Müller, Gereon (1998). *Incomplete Category Fronting*. Kluwer: Dordrecht.

Müller, Gereon (2014). *Syntactic Buffers*. Vol. 91. Linguistische Arbeits Berichte: Universität Leipzig.

Müller, Gereon (2017). Rethinking Restructuring. Ms., Universität Leipzig. To appear in András Bárány, Theresa Biberauer, Jamie Douglas, and Sten Vikner (eds.), *Clausal Architecture and its Consequences: Synchronic and Diachronic Perspectives*. Language Science Press.

Müller, Gereon (2018). *Inflectional Morphology in Harmonic Serialism*. Book ms., Universität Leipzig. To appear in the series *Advances in Optimality Theory*, London: Equinox. (http://home.uni-leipzig.de/muellerg/harm/imhs.pdf).

Müller, Gereon (2019). Anti-Locality and Gradient Harmonic Grammar. Ms., Universität Leipzig.

Müller, Gereon & Wolfgang Sternefeld (1995). Extraction, Lexical Variation, and the Theory of Barriers. In U. Egli, P. E. Pause, C. Schwarze, A. v. Stechow & G. Wienold (eds). *Lexical Knowledge in the Organization of Language*. Benjamins: Amsterdam. 35–80.

Murphy, Andrew (2017). *Cumulativity in Syntactic Derivations*. PhD thesis, Universität Leipzig.

Newmeyer, Frederick (1986). *Linguistic Theory in America*. The University of Chicago Press: Chicago.

Pater, Joe (2009). Weighted Constraints in Generative Linguistics. *Cognitive Science* 33. 999–1035.

Pater, Joe (2016). Universal Grammar with Weighted Constraints. In J. McCarthy & J. Pater (eds). *Harmonic Grammar and Harmonic Serialism*. Equinox: Sheffield. 1–46.

Pesetsky, David (2016). Exfoliation: Towards a Derivational Theory of Clause Size. Ms., MIT, Cambridge, Mass..

Pollock, Jean-Yves (1989). Verb Movement, Universal Grammar, and the Structure of IP. *Linguistic Inquiry* 20. 365–424.

Prince, Alan & Paul Smolensky (1993). Optimality Theory. Constraint Interaction in Generative Grammar. Book ms., Rutgers University.

Reis, Marga & Wolfgang Sternefeld (2004). Review of Wurmbrand, Susanne: Infinitives. *Linguistics* 42. 469–508.

Riemsdijk, Henk van (1978). *A Case Study in Syntactic Markedness: The Binding Nature of Prepositional Phrases*. Foris: Dordrecht.

Rizzi, Luigi (1986). Null Objects in Italian and the Theory of 'pro'. *Linguistic Inquiry* 17. 501–557.

Roberts, Ian (1993). *Verbs and Diachronic Syntax*. Kluwer: Dordrecht.

Rohrbacher, Bernhard (1999). *Morphology-Driven Syntax: A Theory of V to I Raising and Pro-Drop*. Benjamins: Amsterdam.

Ross, John (1967). *Constraints on Variables in Syntax*. PhD thesis, MIT. Cambridge, Mass.

Ross, John (1973*a*). A Fake NP Squish. In C.-J. Bailey & R. Shuy (eds). *New Ways of Analyzing Variation in English*. Georgetown University Press: Washington, DC. 96–140.

Ross, John (1973*b*). Nouniness. In O. Fujimura (ed.). *Three Dimensions of Linguistic Research*. TEC Company Ltd: Tokyo. 137–257. . Reprinted in Bas Aarts et al. (eds.): *Fuzzy Grammar: A Reader*, 351-422. Oxford: Oxford University Press

Ross, John (1975). Clausematiness. In E. Keenan (ed.). *Formal Semantics of Natural Language*. Cambridge University Press: New York and Cambridge. 422–475.

Sabel, Joachim (1996). *Restrukturierung und Lokalität*. Akademie-Verlag: Berlin.

Santorini, Beatrice & Anthony Kroch (1991). Remnant Extraposition in German. Ms., University of Pennsylvania.

Schmid, Tanja (2005). *Infinitival Syntax. Infinitivus Pro Participio as a Repair Strategy*. Benjamins: Amsterdam.

Smolensky, Paul (2017). Gradient Representations. Tutorial, Universität Leipzig; November 10-12, 2017.

Smolensky, Paul & Geraldine Legendre (2006). *The Harmonic Mind*. MIT Press: Cambridge, Mass..

Smolensky, Paul & Matthew Goldrick (2016). Gradient Symbolic Representations in Grammar: The Case of French Liaison. ROA 1286.

Sportiche, Dominique (1989). Le Mouvement Syntaxique: Contraintes et Paramètres. *Langages* 95. 35–80.

Stechow, Arnim von (1993). Die Aufgaben der Syntax. In J. Jacobs, A. v. Stechow, W. Sternefeld & T. Vennemann (eds). *Syntax, vol. I*. De Gruyter: Berlin. 1–88.

Sternefeld, Wolfgang (1990). Scrambling and Minimality. In G. Grewendorf & W. Sternefeld (eds). *Scrambling and Barriers*. Benjamins: Amsterdam. 239–257.

Torres-Tamarit, Francesc (2016). Compensatory and Opaque Vowel Lengthening in

Harmonic Serialism. In *Harmonic Grammar and Harmonic Serialism*. Equinox: Sheffield.

Vikner, Sten (1997). V-to-I Movement and Inflection for Person in All Tenses. In L. Haegeman (ed.). *The New Comparative Syntax*. Longman: London. 189–213.

Vikner, Sten (2001*a*). V-to-I Movement and Do-Insertion in Optimality Theory. In G. Legendre, J. Grimshaw & S. Vikner (eds). *Optimality-Theoretic Syntax*. MIT Press: Cambridge, Mass.. 424–464.

Vikner, Sten (2001*b*). Verb Movement Variation in Germanic and Optimality Theory. Habilitation thesis, Universität Tübingen.

Wöllstein-Leisten, Angelika (2001). *Die Syntax der dritten Konstruktion*. Stauffenburg Verlag: Tübingen.

Wurmbrand, Susanne (2001). *Infinitives. Restructuring and Clause Structure*. Mouton de Gruyter: Berlin.

Wurmbrand, Susi (2007). How Complex are Complex Predicates?. *Syntax* 10. 243–288.

Zeijlstra, Hedde (2004). *Sentential Negation and Negative Concord.* PhD thesis, University of Amsterdam.