# Open Library of Humanities

# A principled exception to the Müller-Takano Generalization: A belated follow-up to Richards (2004)

**Stanislao Zompì,** University of Potsdam, DE, zompi@uni-potsdam.de

According to the Müller-Takano Generalization (Müller 1993; Takano 1994), a remnant that contains a trace resulting from a given type of movement cannot itself undergo that same type of movement. I argue that Richards' (2004) data from Bulgarian "Russian-doll" *wh*-questions constitute principled exceptions to this generalization under an approach that derives it from a minimality condition on movement paths. Alternative approaches to the generalization have a harder time deriving such exceptions.

## 1 Introduction

Movement is widely believed to obey the Müller-Takano Generalization (henceforth, MTG), named after Müller (1993) and Takano (1994).

(1) *The Müller-Takano Generalization (MTG)*
After phrase XP has moved from node $\alpha$ to node $\omega$, a remnant phrase YP that dominates $\alpha$ but not $\omega$ cannot move to any node c-commanding $\omega$ if movement of XP and movement of YP are of the same type (Müller 1993; Takano 1994).

(From here on out, I will refer to XP as the *subextractee*, to XP's movement from $\alpha$ to $\omega$ either as the *subextraction* step or as the *remnant-creating movement* step, to YP as the *remnant*, and to YP's movement to a position above $\omega$ as the *remnant movement* step.)

In (2) is an example of remnant movement allowed by the MTG.[1]

(2) [How proud $t_{XP}$]$_{YP}$ do you think [Mary will be $t_{YP}$ [of her sister]$_{XP}$ tomorrow]?

Here the remnant YP [*how proud $t_{XP}$*] ends up higher than the subextractee XP [*of her sister*].[2] However, since the subextractee and the remnant undergo different types of movement (extraposition and *wh*-movement, respectively), the result is MTG-compliant.

Contrast (2) with (3a), which involves an attempt to *wh*-move a remnant that itself contains a *wh*-trace. This violates the MTG, and the result is indeed jarringly unacceptable, in contrast with the comparatively acceptable remnant-movement–free counterpart in (3b) (cf. Saito 1989: 187, acknowledging a p.c. by Howard Lasnik).

(3) a. *I know you were wondering which countries to invite certain ministers of, but…*
\*[Which minister of $t_{wh_2}$]$_{wh_1}$ were you wondering [[which country]$_{wh_2}$ to invite $t_{wh_1}$]?
 b. *I know you were wondering which ministers of certain countries to invite, but…*
?[Which country]$_{wh_2}$ were you wondering [[which minister of $t_{wh_2}$]$_{wh_1}$ to invite $t_{wh_1}$]?

The MTG has remarkable empirical success across a variety of languages. However, it appears to run into a problem in Bulgarian, where, as Richards (2004) has observed, the kind of contrast illustrated in (3) is not replicated in (4).

---

[1] Here and throughout, I follow an extensive literature (reviewed in Grewendorf 2015 and Thiersch 2017) in taking remnant movement to generally be possible, and in assuming that the subextractee's trace gets to be interpreted (despite being unbound in its surface position) via semantic or syntactic reconstruction of a constituent that dominates it — either the whole remnant itself or an appropriate subconstituent of it.

[2] This must be so on the assumption that the constituent structure of YP is [*how* [*proud* [*of her sister*]]]. Readers who are willing to entertain alternatives like [[*how proud*] [*of her sister*]] will need more complex examples to convince themselves — e.g [*How different $t_{XP}$ a person*]$_{YP}$ *do you think* [*you were $t_{YP}$* [*from the rest of us*]$_{XP}$ *until yesterday*]?

(4)  Bulgarian                                                    (Richards 2004: 459)

a.  [Kolko      studenti $\mathbf{t}_{wh_2}$]$_{wh_1}$ se  opitvaš da razbereš  [[ot    koi    strani]$_{wh_2}$
    how.many students           REFL you.try to understand  from  which countries
    e    ubil     Ivan $\mathbf{t}_{wh_1}$]?
    AUX killed    Ivan
    '[How many students $\mathbf{t}_{wh_2}$]$_{wh_1}$ are you trying to understand [from which countries]$_{wh_2}$
    Ivan killed $\mathbf{t}_{wh_1}$?'

b.  [Ot    koi    strani]$_{wh_2}$ se  opitvaš da razbereš    [[kolko       studenti $\mathbf{t}_{wh_2}$]$_{wh_1}$
    from  which countries  REFL you.try to understand   how.many students
    e    ubil     Ivan $\mathbf{t}_{wh_1}$]?
    AUX killed    Ivan
    '[From which countries]$_{wh_2}$ are you trying to understand [how many students $\mathbf{t}_{wh_2}$]$_{wh_1}$
    Ivan killed $\mathbf{t}_{wh_1}$?'

Interestingly, Richards (2004) discussed (4) only as an argument for a particular analysis of another sentence, within a discussion of a different (albeit related) theoretical issue.[3] He therefore did not note that (4a) seemed to be (and, on his own analysis, really was) a counterexample to the MTG—nor, to my knowledge, has any of the literature noticed it since. The goal of this note, therefore, is to belatedly tackle the questions that the Bulgarian data pose for our understanding of Müller-Takano effects. Do those data constitute genuine counterexamples to the MTG? And if so, can we develop an account that will enforce the MTG in English (3) while also deriving Bulgarian (4) as a principled exception?

The paper explores these questions as follows. Section 2 first reviews an early account of the MTG based on the Minimal-Link Condition, and then shows that the success of that account can be preserved under a more powerful minimality constraint—Shortest. Section 3 then reviews Richards' (2004) analysis of the Bulgarian facts, which couples Shortest with a particular set of assumptions about the directionality of probing and movement. That analysis will turn out to contain all the necessary ingredients for a theory to explain both the Müller-Takano effects in English and the lack thereof in Bulgarian. Section 4 rounds out this theory's predictions with one important qualification, and Section 5 favorably contrasts these predictions with those of alternative approaches to the MTG. Finally, Section 6 wraps it all up.

---

[3] The other sentence at issue was the one reproduced below as (19a), and the theoretical question under discussion was whether the grammar should contain an explicit ban on "lowering" movements or whether such movements should just be (mostly) ruled out by some version of the Strict Cycle.

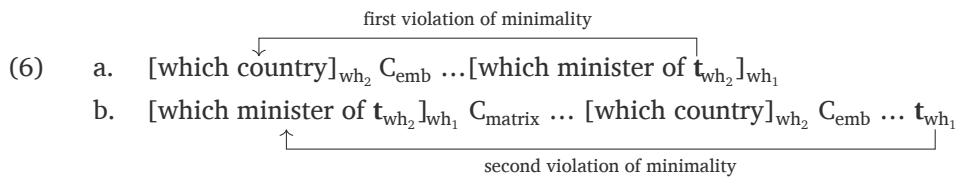## 2 Minimality and the Müller-Takano generalization

### 2.1 Kitahara's (1994; 1997) account

One of the first accounts of the MTG, due to Kitahara (1994; 1997: ch. 3), derives it from the movement minimality constraint in (5), on the assumption that $\beta$ is *closer* to K than $\alpha$ is if $\beta$ asymmetrically c-commands or properly dominates $\alpha$.

(5)     *Minimal-Link Condition*
"$\alpha$ can raise to target K only if there is no legitimate operation Move $\beta$ targeting K, where $\beta$ is closer to K [than $\alpha$ is]."          (Chomsky 1995: 272)

MTG-violating sentences like (3a) violate the Minimal-Link Condition twice. First, although initially the larger $wh_1$ properly dominates the smaller $wh_2$, it's the latter that raises to the embedded Spec,CP ((6a)); second, although the subextractee $wh_2$ now asymmetrically c-commands the $wh_1$-remnant as a result of the step in (6a), it's the $wh_1$-remnant that raises to the matrix Spec,CP in (6b).

<p align="center">first violation of minimality</p>

(6)     a.     [which country]$_{wh_2}$ C$_{emb}$ …[which minister of $\mathbf{t}_{wh_2}$]$_{wh_1}$
         b.     [which minister of $\mathbf{t}_{wh_2}$]$_{wh_1}$ C$_{matrix}$ … [which country]$_{wh_2}$ C$_{emb}$ … $\mathbf{t}_{wh_1}$

<p align="center">second violation of minimality</p>

By contrast, a mixed-movement sentence like (2) — [*How proud $\mathbf{t}_2$*]$_1$ *do you think* [*Mary will be $\mathbf{t}_1$* [*of her sister*]$_2$ *tomorrow*]? — never violates the Minimal-Link Condition because extraposable and *wh*-movable phrases simply don't compete for raising to the same target, and the slightly degraded remnant-movement–free (3b) violates the condition only once ((7)): things are done by the book in the embedded clause (where it's the larger $wh_1$ that raises to the embedded Spec,CP) but not in the matrix (where it's not that same constituent that raises further up, but rather its subconstituent $wh_2$).[4,5]

---

[4] Kitahara (1994; 1997) predates phase theory, but his account of the MTG is consistent with it — i.e. makes the same predictions even if other intermediate *wh*-attractors are posited in addition to embedded complementizers. (Parts of the rest of Kitahara's proposal are **not** trivially compatible with *v*P phases, but they fall outside of the scope of this paper.) See also fn. 15 below for how an updated theoretical context may slightly narrow down the range of our possible assumptions concerning phasal locality.

[5] A reviewer perceives a tension between the account's derivationalist commitment and the idea that constraint violations may lead to gradient decreases in acceptability. It is not obvious, however, why these two aspects should be considered mutually exclusive. In fact, certain generative research traditions have gone as far as to conceive of the entire grammar as "a system of *derivational constraints*, i.e., rules that specify either what may occur at some stage or other of derivations or how various stages in a derivation may or must differ from each other" (McCawley 1976: 14; emphasis his). If the Minimal-Link Condition is viewed as one such constraint, then it is reasonable to expect that, all else being equal, fewer violations of it may result in higher degrees of acceptability. This is what I take Kitahara (1994; 1997) to be implicitly assuming.

(7)  [Which country]$_{wh_2}$ C$_{matrix}$ ... [which minister of $\mathbf{t}_{wh_2}$]$_{wh_1}$ C$_{emb}$ ... $\mathbf{t}_{wh_1}$

<span style="display:block; text-align:center;">minimality-violating      minimality-compliant</span>

There are a couple of aspects of this analysis that are worth highlighting before moving on. First of all, it should be noticed that the account does not incorporate any constraints against movement out of previously moved constituents—such as, for example, Wexler & Culicover's (1980) Freezing Principle or its subsequent incarnations surveyed by Corver (2017). Consequently, the analysis attributes the deviance of the matrix movement step in (7) solely to a violation of the Minimal-Link Condition — not to the fact that $wh_2$ gets moved out of a $wh_1$ that has itself previously been moved. This explanation is at odds with the literature on so-called freezing effects that I've just referred to, but is in line with a growing body of research showing that movement out of previously moved constituents (what Sauerland 1999 calls 'surfing') is in fact not universally ruled out: see, among others, Saito (1985: 249) and Sauerland (1999: 180) on scrambling out of scrambled phrases in Japanese, McCloskey (2000) and Davis (2020: ch. 2–3) on intermediate stranding in successive-cyclic movement, Collins (2005a;b) and Belletti & Collins (2021) on so-called smuggling (a constituent moving and thereby "smuggling" one of its subconstituents out of the c-command domain of a potential intervener), and cf. especially the remarks by Abels (2007: 75), Neeleman & van de Koot (2010: 358), Corver (2017: §§ 5.2–5.3), and Keine (2020b: 162ff).

   Next, it should be noticed that, in order for the deviance of (7) to arise from a minimality violation, movement of $wh_1$ from the embedded Spec,CP to the matrix Spec,CP must crucially count as a legitimate operation for the purposes of the Minimal-Link Condition, so as to outcompete movement of the less close $wh_2$. This must be so, in particular, despite the fact that movement of $wh_1$ would not itself yield a grammatical string (due to what Rizzi 2006 calls *criterial freezing*, which I'll refrain from going into in any depth here).

(8)  *[Which minister of [which country]$_{wh_2}$]$_{wh_1}$ were you wondering $\mathbf{t}_{wh_1}$ to invite $\mathbf{t}_{wh_1}$?

The notion of "legitimate operation" featuring in (5) must therefore be understood broadly enough to include any movement of a *wh*-phrase to a specifier of a c-commanding *wh*-complementizer, regardless of whether such movement would ultimately lead to a grammatical outcome or not. For similar contrasts pointing to the same understanding of this notion (e.g. mild *vs* strong deviance in English nesting-paths *vs* crossing-paths *wh*-island violations, respectively), see again Kitahara (1994; 1997).

---

   Be that as it may, and whatever the merits of this approach for the treatment of English contrasts like (3), I should also emphasize that the Bulgarian data at the heart of this paper do not involve any such gradient acceptability contrasts, and that the account of those data will therefore not need to invoke any similar contrasts between single and multiple constraint violations.

While both of these aspects of the account could conceivably be tweaked with or dispensed with, we will see shortly that they both make it particularly easy to extend a version of Kitahara's intuition to Bulgarian.

## 2.2 Replacing the Minimal-Link Condition with Shortest

The success of Kitahara's (1994; 1997) account can also be preserved under slightly different conceptions of minimality. Particularly relevant to this note are alternatives based on the notion of a movement's *path*, which we may define, following Collins (1994), as the set of nodes properly dominating only one end of the movement chain.

(9)     "Let $\alpha$ and $\omega$ be two nodes in a tree, let $S_\alpha$ be the set of nodes properly dominating $\alpha$, and let $S_\omega$ be the set of nodes properly dominating $\omega$. The *path* between $\alpha$ and $\omega$ is defined as follows:   $\text{Path}(\alpha, \omega) = (S_\alpha \cup S_\omega) \setminus (S_\alpha \cap S_\omega)$"                    (Collins 1994: 56)

With this notion in place, we can now rethink the Minimal-Link Condition in (5) as a minimality condition on movement paths, along the lines explored by Kitahara (1993), Collins (1994), Nakamura (1998), and especially Richards (1997) and Müller (1998). Richards (1997; 2004), in particular, replaces the Minimal-Link Condition with a constraint he names *Shortest*, which I adapt in (10).[6]

(10)    *Shortest*

Let head H, at a given point in the derivation, be a *potential trigger* both for phrasal movement from $\alpha$ to $\omega$ and for phrasal movement from $\alpha'$ (possibly identical to $\alpha$) to $\omega'$

---

[6] Richards' own formulations are in (i–ii).

(i)    "An attractor K attracts a feature F, creating a copy $\alpha'$ of an element $\alpha$ containing F, and merging $\alpha'$ with K. The relations between $\alpha'$, K, and F must all obey Shortest."                    (Richards 1997: 111)

(ii)   "The relation between $\alpha$ and $\beta$ obeys Shortest iff there is a path $\pi$ between $\alpha$ and $\beta$ such that for any $\gamma$ distinct from both $\alpha$ and $\beta$, $\pi$ is a subset of the path $\pi'$ created by substituting $\gamma$ for either $\alpha$ or $\beta$."

(Richards 2004: 460)

(In (ii), the intended meaning is clearly that the only relevant $\gamma$'s to be considered are those whose substitution for $\alpha$ or $\beta$ would yield an otherwise well-formed dependency; cf. Richards 1997: 111, ex. (95).)

In (10), I'm departing from Richards and following Fitzpatrick (2002: 449fn12) in "conflat[ing] all locality of movement into one metric, rather than computing attractor-attractee (choice of mover) and departure site–landing site (choice of landing site) relations separately." This makes a difference if Shortest is coupled with Richards' (1997: 228) Principle of Minimal Compliance (see especially Richards 1997: § 2.6.2.1), but is of no consequence for our concerns.

(possibly identical to $\omega$). Then, if $|\mathrm{Path}(\alpha, \omega)| < |\mathrm{Path}(\alpha', \omega')|$, H cannot trigger phrasal movement from $\alpha'$ to $\omega'$ at that point in the derivation.[7]

(11)    At a given point in the derivation, head H is a *potential trigger* for moving phrase XP from $\alpha$ to $\omega$ if

    a.    $\omega$ is a specifier of HP

    and, at that point in the derivation,

    b.    H is projecting;

    c.    and H has an active attractor feature that XP matches;

    d.    and H c-commands $\alpha$;                     $\leftarrow$ *(to be revised in (26d))*

    e.    and $\alpha$ is not (reflexively) dominated by a previously deleted movement copy.

Shortest derives all the effects of the Minimal-Link Condition, including the account of the MTG we saw in § 2.1 (cf. in particular Müller 1998: 273ff). For example, just as the MTG-offending (3a)/(12) would violate the Minimal-Link Condition twice, it also violates Shortest twice: Shortest too would require the embedded C to prefer (13a) to (13b), and then, following the dispreferred choice (13b), it too would require the matrix C to prefer (14a) to (14b). (Here and throughout the rest of this note, I'll highlight the paths compared by Shortest in light blue.)

(12)    *[Which minister of $\mathbf{t}_{\mathrm{wh}_2}$]$_{\mathrm{wh}_1}$ were you wondering [which country]$_{\mathrm{wh}_2}$ to invite $\mathbf{t}_{\mathrm{wh}_1}$?

(13)    a.



---
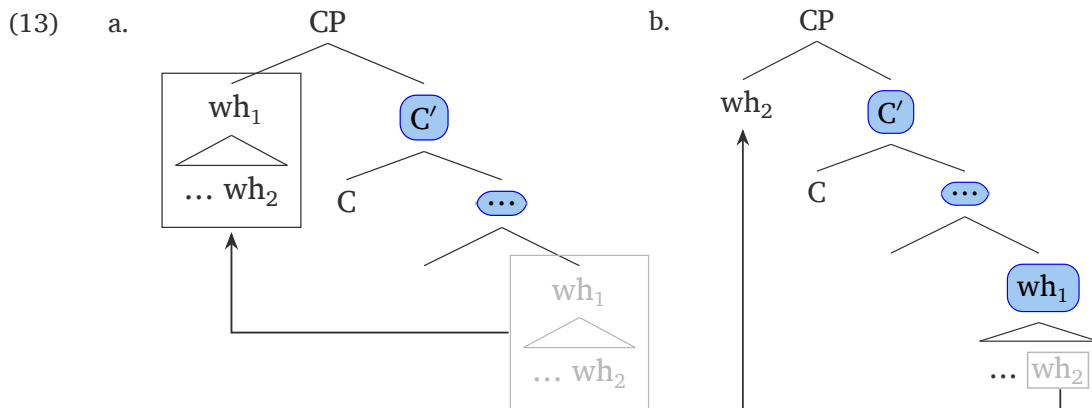
[7] An alternative equally in line with the facts, but which has fewer antecedents in the literature (and would raise potentially tricky questions about node identity across different derivational options), would be to replace "$|\mathrm{Path}(\alpha, \omega)| < |\mathrm{Path}(\alpha', \omega')|$" in (10) with "$\mathrm{Path}(\alpha, \omega) \subset \mathrm{Path}(\alpha', \omega')$" — an approach that Fitzpatrick (2002: 450) christens *SubPaths* and retraces to discussion in Bošković (1997: 251). As Fitzpatrick notes, both Shortest and SubPaths are transderivational constraints — a concern that I can only partially assuage by noting that the transderivational comparison in question is restricted to the single next derivational step, and is therefore an instance of local (rather than global) transderivational economy in Collins' (1997) sense.

(14)  a.



b.
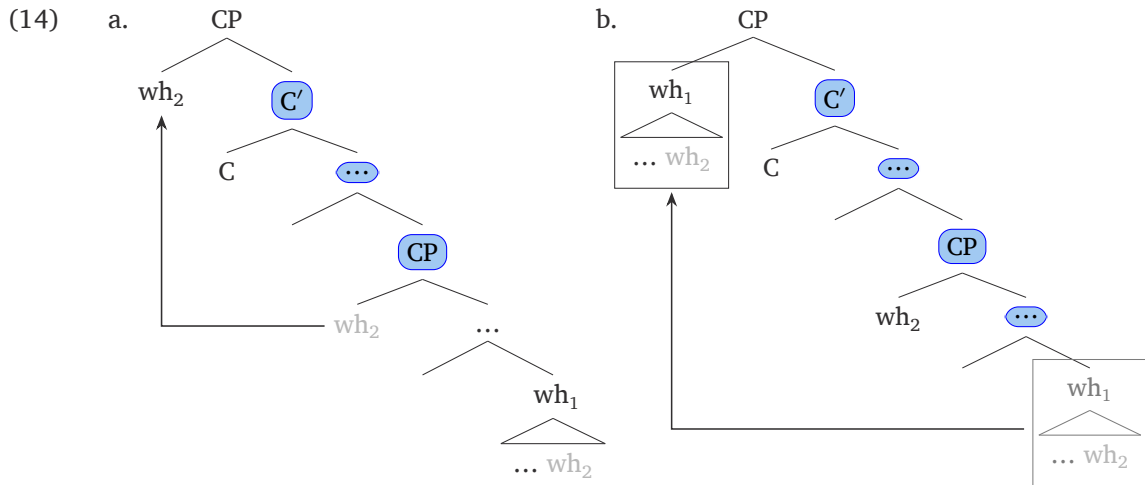
But Shortest also comes with a bonus, as pointed out by Richards (1997): if one abandons Chomsky's (1995) Extension Condition in (15), one can use Shortest to derive the order-preserving pattern in movement to multiple specifiers of a single probe, as illustrated by Bulgarian multiple *wh*-fronting in (16).[8]

(15)  *The Extension Condition* (to be abandoned!)

"Move $\alpha$ extend[s phrase marker] K to [phrase marker] K*, which includes K as a proper part."                                                    (Chomsky 1995: 174)

(16)  Bulgarian (Rudin 1988: 481–482)

a.  Koj        kakvo pravi?
    who.NOM  what   does
    'Who is doing what?'

b. *Kakvo koj        pravi?
    what   who.NOM  does
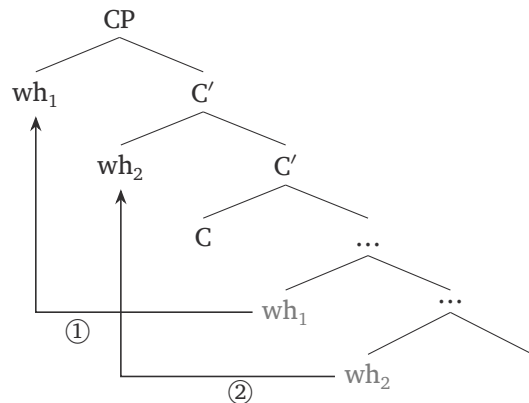


A key assumption here, tracing back to Rudin (1988), is that the Bulgarian complementizer differs from its English counterpart in having a *wh*-attractor feature that never ceases being active — an

---

[8] See Billings & Rudin (1996), Bošković (1997), Richards (1997: 272–277), and Krapova & Cinque (2008) for qualifications and discussion.

instance of *insatiability* in Deal's (2024) sense.[9] Building on this, Richards (1997) argues that the insatiable C in (16) first attracts the closer $wh_1$ to a specifier of CP (the only such specifier up to this point). Then C proceeds to attract the farther $wh_2$ and, absent a condition like (15), must decide where to move it to — whether to an inner Spec,CP below $wh_1$ ("tucking-in") or to an outer Spec,CP above $wh_1$ ("tucking-out"). As schematized in (17), Shortest dictates the former.
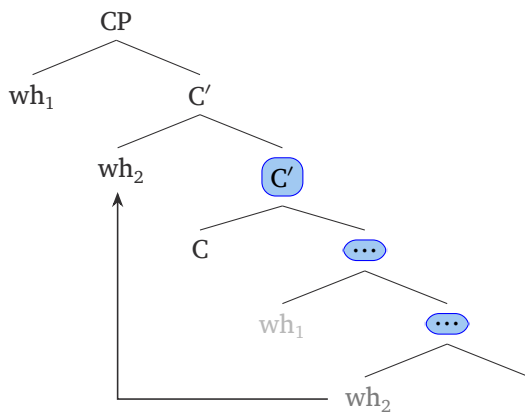
(17)  a. "Tucking-in"    b. "Tucking-out"
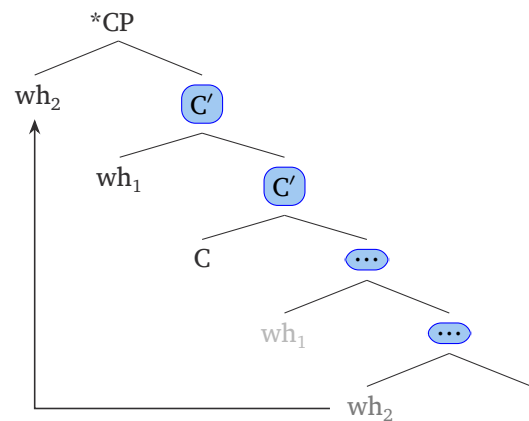


## 3 Bulgarian as a principled exception to the MTG

### 3.1 The data

Bulgarian — the same language that in § 2.2 provided key evidence for Shortest — also turns out to pose a challenge to the MTG as stated in (1). In particular, as previewed in Section 1, Richards (2004) notices that Bulgarian fails to replicate the English evidence for the MTG from (3). Particularly surprising is example (18a), superficially analogous to the unacceptable English example (3a), yet acceptable in apparent defiance of the MTG.

---

[9] As a reviewer points out, the literature on Bulgarian offers several alternatives to the notion of an insatiable attractor in C. Bošković (2007), for example, argues that in Bulgarian all *wh*-phrases must move to Spec,CP not due to any insatiable probing by C, but because the *wh*-phrases themselves each bear a feature in need of checking. Other analyses diverge even further from the approach outlined in the main text, in that they don't even take different *wh*-phrases to necessarily occupy distinct Spec,CP positions, but rather contend that all of the *wh*-phrases in a given CP's left periphery form a single constituent together (a "*wh*-cluster") as a result of one or more applications of sideward movement (Grewendorf 2001; Bailyn 2017; cf. also Citko & Gračanin-Yuksek 2013). Since the formulation of Shortest is in principle noncommittal as to whether the driving force of movement lies in the goal or in the probe, it seems eminently feasible to adapt the bulk of the current approach to integrate it with Bošković's proposal — provided that some version of the Strict Cycle were to replicate the effects attributed to (11b) in the current system. By contrast, it is far less clear how the approach could be reconciled with any of the *wh*-cluster accounts.

(18)   Bulgarian (Richards 2004: 459)

    a.   [Kolko       studenti  $t_{wh_2}$]$_{wh_1}$  se      opitvaš  da  razbereš    [[ot      koi
         how.many  students             REFL   you.try  to  understand  from   which
        strani]$_{wh_2}$  e      ubil      Ivan  $t_{wh_1}$]?
        countries  AUX  killed    Ivan
        '[How many students $t_{wh_2}$]$_{wh_1}$ are you trying to understand [from which countries]$_{wh_2}$
        Ivan killed $t_{wh_1}$?'

    b.   [Ot      koi      strani]$_{wh_2}$  se      opitvaš  da  razbereš     [[kolko         studenti
         from   which   countries  REFL   you.try  to  understand   how.many  students
        $t_{wh_2}$]$_{wh_1}$  e      ubil      Ivan  $t_{wh_1}$]?
              AUX  killed    Ivan
        '[From which countries]$_{wh_2}$ are you trying to understand [how many students
        $t_{wh_2}$]$_{wh_1}$ Ivan killed $t_{wh_1}$?'

Richards also notices that a similar effect can be observed even in Bulgarian monoclausal multiple-*wh*-fronting questions, the surprising example being (19a).

(19)   Bulgarian (Richards 2004: 456)

    a.   [Kolko       studenti  $t_{wh_2}$  [ot     Bulgaria]]$_{wh_1}$  [po kakvo]$_{wh_2}$  vidja     $t_{wh_1}$?
         how.many  students          from  Bulgaria         of  what       you.saw
        'How many students of what from Bulgaria did you see?'

    b.   [Po kakvo]$_{wh_2}$  [kolko       studenti  $t_{wh_2}$  [ot     Bulgaria]]$_{wh_1}$  vidja     $t_{wh_1}$?
         of  what       how.many  students          from  Bulgaria         you.saw

    c.   *[Kolko       studenti  [po kakvo]$_{wh_2}$  [ot     Bulgaria]]$_{wh_1}$  vidja     $t_{wh_1}$?
         how.many  students  of  what         from  Bulgaria         you.saw

Two quick notes are in order here concerning the basic characterization of the facts — especially (19) — before we move on towards an account.

    First, one may harbor doubts about the very existence of $t_{wh_2}$ in (19a). However, Richards (2004) argues for it based on a contrast with respect to the canonical constituent order in the absence of *wh*-phrases.

(20)   Bulgarian (Richards 2004: 455–456)

    a.   Vidja      [studenti  [po matematika]  [ot     Bulgaria]].
        you.saw  students  of  mathematics  from  Bulgaria
        'You saw students of mathematics from Bulgaria.'

    b.   *Vidja     [studenti  [ot     Bulgaria]  [po matematika]].
        you.saw  students  from  Bulgaria   of  mathematics

    c.   *Vidja     [[po matematika]  studenti  [ot     Bulgaria]].
        you.saw    of  mathematics  students  from  Bulgaria

Second, one might concede that $t_{wh_2}$ is there, but not that it results from the same type of movement as $t_{wh_1}$. For example, Bošković (2002) proposes, *contra* Rudin (1988) and Richards (1997), that only the topmost *wh*-phrase in Bulgarian undergoes real *wh*-movement to Spec,CP, with all other *wh*-phrases being focus-fronted to a lower projection. If true, this would make the two traces in (19a) the result of different types of movement, and would thereby void the issue for the MTG. The problem with this move, however, is that even for Bošković (2002) the second and third *wh*-phrases in a three-*wh* question do undergo the same type of movement, and yet they too replicate the contrast seen in (19).

(21)    Bulgarian (Snejana Iovtcheva, p.c.)

    a.   Koǐ$_{wh_1}$ [kolko studenti $\mathbf{t}_{wh_3}$ [ot Bulgaria]]$_{wh_2}$ [po kakvo]$_{wh_3}$ $\mathbf{t}_{wh_1}$
        who.NOM how.many students   from Bulgaria   of what
        vidja $\mathbf{t}_{wh_2}$?
        saw
        'Who saw how many students of what from Bulgaria?'

    b.   Koǐ$_{wh_1}$ [po kakvo]$_{wh_3}$ [kolko studenti $\mathbf{t}_{wh_3}$ [ot Bulgaria]]$_{wh_2}$ $\mathbf{t}_{wh_1}$
        who.NOM of what how.many students   from Bulgaria
        vidja $\mathbf{t}_{wh_2}$?
        saw

    c.   *Koǐ$_{wh_1}$ [kolko studenti [po kakvo]$_{wh_3}$ [ot Bulgaria]]$_{wh_2}$ $\mathbf{t}_{wh_1}$ vidja $\mathbf{t}_{wh_2}$?
        who.NOM how.many students of what   from Bulgaria   saw

In view of (20)–(21), I will therefore assume from here on that Richards' (2004) characterization of the facts in (18)–(19) is indeed correct.

## 3.2  Richards' (2004) analysis, or how to create pseudo-violations of the MTG

How can we derive a sentence like (19a)/(22) within the system we've set up so far?

(22)    [Kolko studenti $\mathbf{t}_{wh_2}$ [ot Bulgaria]]$_{wh_1}$ [po kakvo]$_{wh_2}$ vidja $\mathbf{t}_{wh_1}$?
      how.many students   from Bulgaria   of what   you.saw
      'How many students of what from Bulgaria did you see?'

Let's start by ruling out something that clearly won't work, namely a derivation in which C attracts $wh_2$ first and then attracts the $wh_1$-remnant in a tucking-out fashion (23). Both movements in this derivation would violate Shortest (cf. (13) for the first movement, and (17) for the second one)—which means that we would then expect the sentence to be just as bad as its MTG-violating English counterpart in (3a), contrary to fact.

(23)    Not what we want:
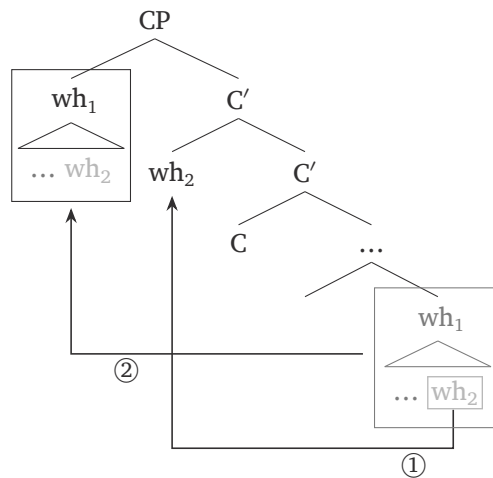


We may then hope to have better luck by first moving $wh_1$ instead. However, this also won't give us the desired result if we insist on the conditions in (11), repeated in (24).

(24)    At a given point in the derivation, head H is a *potential trigger* for moving phrase XP from $\alpha$ to $\omega$ if

   a.    $\omega$ is a specifier of HP

   and, at that point in the derivation,

   b.    H is projecting;
   c.    and H has an active attractor feature that XP matches;
   d.    and H c-commands $\alpha$;                    ← *(to be revised in (26d))*
   e.    and $\alpha$ is not (reflexively) dominated by a previously deleted movement copy.

Here's why. After moving $wh_1$ as in (25), we are bound to end up with two copies of $wh_2$, neither of which can be moved — the higher copy because it is not in the complementizer's c-command domain (see (24d)), and the lower one because it is dominated by an already-moved and -deleted constituent (see (24e)). We thus predict the derivation to stop, and $wh_2$ to never move in either a tucking-in or a tucking-out mode, again contrary to fact.

(25)    What we want to start with
        but not to end with:
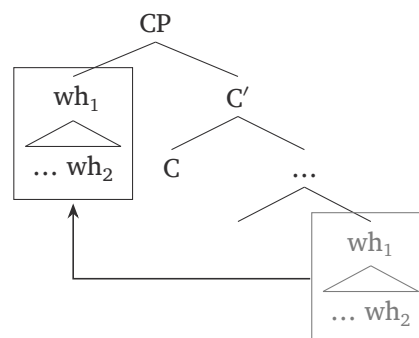
What Richards (2004) proposes, therefore, is effectively to replace "c-command" with "m-command" in (24d), i.e. to let probes search into their specifiers as well as into their complement. This idea, which has since gained widespread currency in the agreement literature under the name of "Cyclic Agree" (Rezac 2003; Béjar & Rezac 2009; Keine & Dash 2022; Clem 2023), has been specifically defended by Rezac (2003) as a consequence of Chomsky's (1995) Bare Phrase Structure: if what projects is not just the categorial label of the head but rather its whole feature bundle, then any of the head's still-active features with a potential to trigger agreement or movement operations should continue to trigger such operations at the intermediate-projection level, thereby cyclically expanding the search domain to specifiers. The proposal under discussion can therefore be viewed just as an application of this idea to the domain of movement-triggering within a system that independently allows for tucking-in.

(26)     At a given point in the derivation, head H is a *potential trigger* for moving phrase XP from $\alpha$ to $\omega$ if

a.     $\omega$ is a specifier of HP

and, at that point in the derivation,

b.     H is projecting;

c.     and H has an active attractor feature that XP matches;

d.     and HP **dominates** $\alpha$;                           ← *(previously "H c-commands $\alpha$")*

e.     and $\alpha$ is not (reflexively) dominated by a previously deleted movement copy.

Within this revised system, Richards (2004) argues that in monoclausal (19) the complementizer first attracts $wh_1$ (in compliance with Shortest) and then attracts $wh_2$ from within the already-created Spec,CP to a new Spec,CP. Crucially, for this latter movement, Shortest does not adjudicate between tucking-out and tucking-in.[10],[11] In the tucking-out option schematized in (27a), the path contains C′, $wh_1$, and other $wh_1$-internal nodes, because they all dominate $wh_2$'s pre-movement

---

[10] Both (27a) and (27b) involve movement out of a previously moved phrase — something I've already assumed to be licit, *contra* the Freezing Principle and similar constraints (cf. § 2.1 and references cited there).

[11] There's a question I've still left open — namely, what in this system prevents perpetual string-vacuous movement of $wh_1$ from its earliest Spec,CP landing site to the immediately inner/outer Spec,CP position, and from there to the next inner/outer Spec,CP, and so on *ad infinitum* — and why such movement doesn't block movement of $wh_2$. (Notice that the question is not specific to Russian-doll multiple-*wh* questions; it also arises in simpler cases like (16).) A quick and dirty fix would be to add another conjunct to (26a–e), to the effect that $\alpha$ must not itself be a specifier of HP. However, as a reviewer suggests, the condition is probably more principled and general than that, along the lines of Bobaljik's (1995) and Abels' (2003; 2012) theory of antilocality as a consequence of movement being *last-resort*: the string-vacuous movement option we need to rule out is an instance of "movement from one position to another both of which are checking relationships with the same elements" (Bobaljik 1995: 269), and "[i]f movement is indeed of a last resort character — i.e. solely for the purposes of feature checking — then it should in fact be prohibited" (*ibid.*).

copy but not $wh_2$'s post-movement copy. In the tucking-in option schematized in (27b), the path contains $wh_1$ and other $wh_1$-internal nodes (once again because they dominate only $wh_2$'s pre-movement copy), but it also contains C′, which now dominates only $wh_2$'s post-movement copy. Neither path contains fewer nodes than the other.[12,13]

(27)   a.



   b.
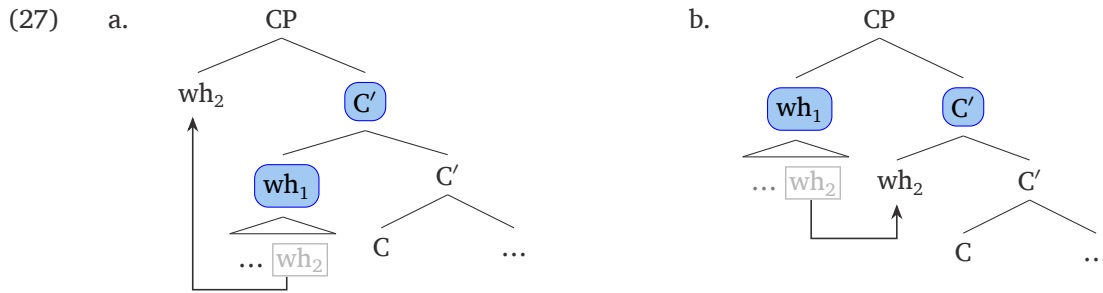
Tucking-in and tucking-out are therefore predicted to be equally viable options. The prediction is borne out, as we have seen, by the comparable acceptability of (28) and (29), both of which we already saw in (19).[14]

---

[12] An alternative, mentioned in passing by Richards (2004: 462fn10), would be to keep (24d) unchanged and abandon (24e)/(26e) instead, so as to permit extraction of $wh_2$ out of the already-deleted copy of $wh_1$. On such an account, however, we would predict that the tucking-in structure $[[\ldots t_{wh_2}]_{wh_1} \ wh_2 \ C \ldots t_{wh_1}]$ should violate Shortest one less time than the tucking-out structure $[wh_2 \ [\ldots t_{wh_2}]_{wh_1} \ C \ldots t_{wh_1}]$ — a conclusion at odds with the fact that the two structures are equally acceptable. Frampton (2004) observes that the dilemma as to whether to extract $wh_2$ out of $wh_1$'s higher or lower copy just dissolves within a system that treats movement as multidominance. The main points of this paper could indeed be reworked within such a system; I sketch out a possible way to do so in the appendix.

[13] Richards (2004) adopts a different definition of *path* — the one in (i) — which derives the optionality in (27) in a slightly different way. According to (i), the movement in (27b) has no path at all, and so is excluded from any Shortest competition and cannot lose out to the alternative in (27a).

(i)   Path$(\alpha, \beta)$ is the nonempty set (if any) of nodes $x$ such that $\alpha$ c-commands $x$ and $x$ dominates $\beta$.

A potential issue with (i), however, is that it leaves Path$(\alpha, \beta)$ undefined whenever $\alpha$ does not c-command $\beta$; (i) therefore prevents Shortest from choosing between two movements if neither of them lands in a position c-commanding the trace (e.g. any two instances of head movement). Cf. Fitzpatrick (2002: 453–454).

[14] As for three-*wh* questions like (21), Shortest correctly derives both the orders in (21a–b), as well as the unacceptability of (21c). The insatiable C first attracts $wh_1$, then obligatorily tucks in $wh_2$, and then obligatorily extracts $wh_3$ out of $wh_2$ — via either tucking-in or tucking-out, as in (27).

(i)   $wh_1 \ \langle wh_3 \rangle \ [_{wh_2} t_3 \ldots] \ \langle wh_3 \rangle \ C \ldots t_1 \ldots t_2$
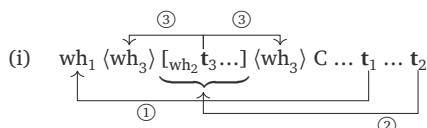
The problem is that those in (21a–b) are the **only** two orders that Shortest permits: for example, tucking-out of $wh_2$ should be blocked by the tucking-in option in ②, and tucking-out of $wh_3$ to a specifier higher than $wh_1$ should be blocked by both the options in ③. However, judgements tend to be more liberal (Snejana Iovtcheva, p.c.):

(28)  [Po kakvo]$_{wh_2}$ [kolko     studenti $\mathbf{t}_{wh_2}$ [ot     Bulgaria]]$_{wh_1}$ vidja    $\mathbf{t}_{wh_1}$?
      of  what       how.many students        from Bulgaria        you.saw
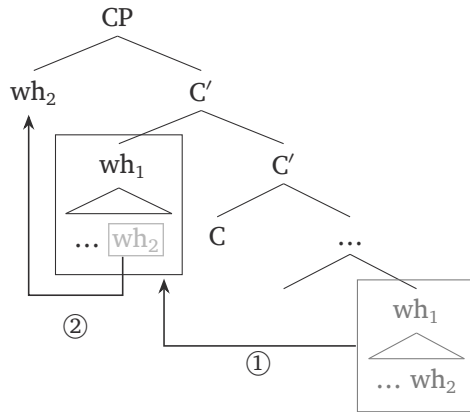


(29)  [Kolko      studenti $\mathbf{t}_{wh_2}$ [ot     Bulgaria]]$_{wh_1}$ [po kakvo]$_{wh_2}$ vidja    $\mathbf{t}_{wh_1}$?
      how.many students        from Bulgaria        of  what       you.saw
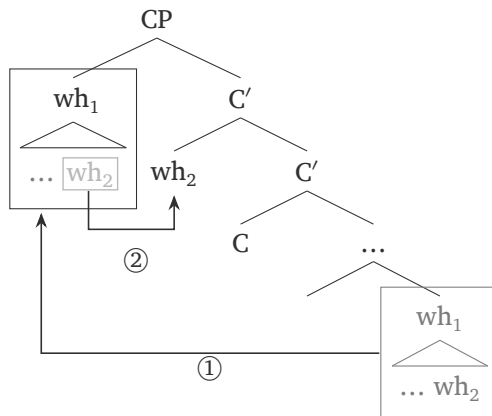


(ii)  a.  [Kolko      studenti $\mathbf{t}_{wh_3}$ [ot     Bulgaria]]$_{wh_2}$ koǐ$_{wh_1}$     [po kakvo]$_{wh_3}$ $\mathbf{t}_{wh_1}$ vidja $\mathbf{t}_{wh_2}$?
          how.many students        from Bulgaria        who.NOM    of  what        saw
          'Who saw how many students of what from Bulgaria?'

      b.  [Po kakvo]$_{wh_3}$ koǐ$_{wh_1}$     [kolko      studenti $\mathbf{t}_{wh_3}$ [ot     Bulgaria]]$_{wh_2}$ $\mathbf{t}_{wh_1}$ vidja $\mathbf{t}_{wh_2}$?
          of  what        who.NOM how.many students        from Bulgaria        saw

      c.  *[Kolko      studenti $\mathbf{t}_{wh_3}$ [ot     Bulgaria]]$_{wh_2}$ [po kakvo]$_{wh_3}$ koǐ$_{wh_1}$     $\mathbf{t}_{wh_1}$ vidja $\mathbf{t}_{wh_2}$?
          how.many students        from Bulgaria        of  what        who.NOM    saw

      d.  *[Po kakvo]$_{wh_3}$ [kolko      studenti $\mathbf{t}_{wh_3}$ [ot     Bulgaria]]$_{wh_2}$ koǐ$_{wh_1}$     $\mathbf{t}_{wh_1}$ vidja $\mathbf{t}_{wh_2}$?
          of  what        how.many students        from Bulgaria        who.NOM    saw

These facts, unexpected on my account and on anyone else's, appear to be part of a broader range of cases in which the presence of a third *wh* seemingly relaxes certain Shortest-related restrictions; cf. Bošković (1997: 238ff) and Richards (1997: 272ff) on Bulgarian, and Kayne (1983: 235) and Pesetsky (2000: 44) on English (e.g. *What did who give to {whom/*Mary}?*). Notice, however, that even if Shortest turned out to just not apply to sentences with three or more *wh*-phrases, the demonstration in (21) that the second and third *wh* can violate the MTG would then still militate in support of the idea that the MTG is a by-product of Shortest.

Notice that (29) might **look** like an MTG-violation: $[\ldots t_{wh_2}\ldots]_{wh_1}$ c-commands $wh_2$, which in turn c-commands $t_{wh_1}$, and both traces result from movements of the same type. However, the MTG, at least as we formulated it in (1), is not a constraint on representations but on derivations, and the derivation schematized in (29) does not violate it: we did not **first** move $wh_2$ to **then** move the remnant $wh_1$ above it; rather, we first moved $wh_1$ to some head's specifier, and then subextracted and tucked in $wh_2$ to a lower specifier of the same head.

From here on, I will refer to this kind of licit movement from within an outer specifier of a head to an inner specifier of the same head as *tucking-down*, and will refer to the ensuing purportedly remnant-movement–like configurations as *pseudo-violations of the MTG*.

### 3.3 Pseudo-violations of the MTG can feed real violations

It is now trivial to extend Richards' (2004) account to biclausal questions like (18), repeated here in (30)–(31). Specifically, it can simply be assumed that the same two options introduced in (28)–(29) are also available to the embedded complementizer in these more complex examples, and that the derivation in the matrix will proceed accordingly in compliance with Shortest. If the embedded CP opts for tucking-out, Shortest will force the matrix complementizer to attract the now closer $wh_2$, as in (30), whereas, if the embedded CP opts for tucking-down, Shortest will force the matrix complementizer to attract the $wh_1$-remnant, as in (31).[15,16]

(30)  $[\text{Ot} \quad \text{koi} \quad \text{strani}]_{wh_2}$ se opitvaš da razbereš  $[[\text{kolko} \quad \text{studenti} \ \mathbf{t}_{wh_2}]_{wh_1}$
from which countries REFL you.try to understand  how.many students
e ubil Ivan $\mathbf{t}_{wh_1}]$
AUX killed Ivan

---

[15] The analysis is compatible with the positing of any number of additional successive-cyclic stop-offs, so long as all such stop-offs are also triggered by insatiable *wh*-attractors. (Satiable ones would incorrectly prevent $wh_1$ and $wh_2$ from moving together out of an embedded CP.) Conversely, for English, we will have to assume that whatever additional stop-offs are there must be triggered by satiable *wh*-attractors, so that in (3b) (or in *Which ministers of which countries did you invite?*) the phrase $[_{wh_1}\ldots[_{wh_2}\ldots]]$ can land unsplit into the closest Spec,CP (cf. (13a)). The potential awkwardness of these extra assumptions is reminiscent of other cases in which the literature has observed a tension between phase theory and constraints on complex movement interactions — cf. fn. 4 and especially Müller (2014: 33–38, 69–76) and Keine (2020b: 261–272). We may avoid these tensions by assuming, with Keine (2017; 2020a; 2020b), that Spec,CP is universally the only obligatory successive-cyclic stop-off point. See Müller (2014; 2015) and fn. 21 for a very different solution.

[16] A question arises as to what stops the insatiable matrix C in (30)–(31) from attracting **both** $wh_1$ and $wh_2$. I assume that the relevant restriction doesn't have to do with *wh*-movement locality. Rather, the matrix predicate in (30)–(31) selects for an interrogative CP complement, which in turn necessitates that at least one *wh*-phrase take scope and be *criterially frozen* in that CP complement's left periphery (Rizzi 2006). Cf. the brief discussion of English examples like (8) in Section 2.1.
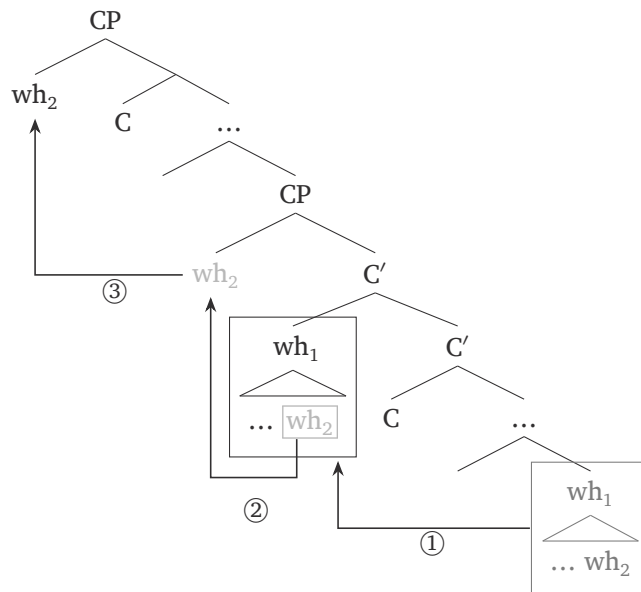
CP
wh₂
C
...
CP
wh₂
C′
wh₁
... wh₂
③
②
C
...
wh₁
... wh₂
①

(31)　[Kolko　　　studenti　**t**_{wh₂}]_{wh₁}　se　　opitvaš　da razbereš　　[[ot　　koi　　strani]_{wh₂}
　　　　 how.many　students　　　　　　　REFL　you.try　to understand　　from　which　countries
　　　　 e　　ubil　Ivan　**t**_{wh₁}]
　　　　 AUX　killed　Ivan

CP
wh₁
... wh₂
C
...
CP
wh₁
... wh₂
wh₂
C′
③
②
C′
C
...
wh₁
... wh₂
①
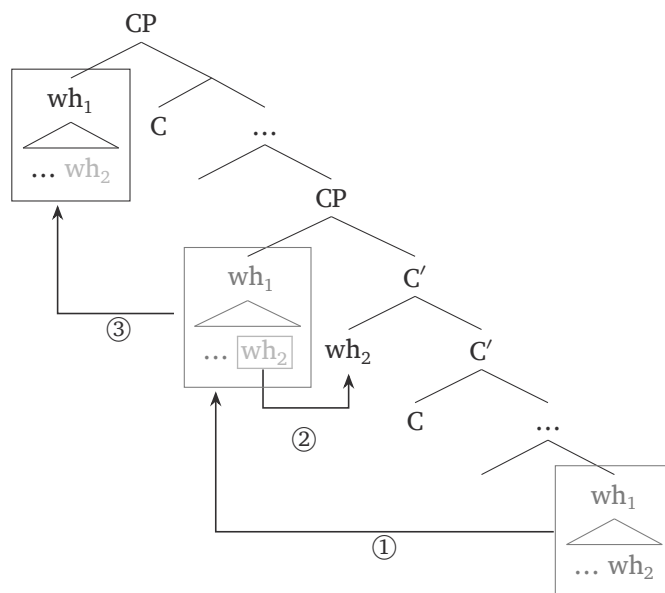
It is important to notice that (31) instantiates a genuine violation of the MTG as stated in (1): the $wh_1$-remnant *wh*-moves above $wh_2$, and does so **after** $wh_2$ has *wh*-moved out of it. However, if the MTG is replaced with a theory based purely on Shortest, (31) is correctly predicted to be acceptable, as it involves no Shortest violations.[17]

It also bears noting that in (31) the genuine MTG-violation can arise only because $wh_1$ and $wh_2$ give rise to a pseudo-violation of the MTG somewhere along the way below. This result holds more generally: pseudo-violations of the MTG (with the remnant and the subextractee occupying multiple specifiers of the **same** probe) are predicted to be a precondition to genuine MTG-violations (with the remnant and the subextractee occupying specifiers of **different** probes) — that is, deriving a sentence like (31) requires the ability to derive a sentence like (29).

## 4 Summary of the predictions, with one more qualification

If the account presented so far is correct, then Universal Grammar does not directly incorporate the MTG but only a minimality principle like Shortest, as well as general restrictions on the potential movements that this minimality principle adjudicates between. Müller-Takano effects are predicted to emerge as a by-product of this if the probes involved in a given movement interaction are each able to attract at most one relevant goal. However, if at least one of the probes in play is insatiable, and thus able to trigger tucking-down (movement from within an outer specifier to an inner specifier), then Müller-Takano restrictions are predicted to be circumventable without Shortest being multiply violated in the process.

Before closing in § 5 with some of the implications of these conclusions, I just need to round them out at this point with a straightforward but crucial qualification: Of course, an insatiable attractor will be able to trigger tucking-down of phrase XP from inside an outer specifier YP only if (non-tucking-down) movement of XP out of YP is in principle possible in the language; by contrast, if movement of XP out of YP is independently ruled out regardless of probing- and movement-directionality, the insatiable attractor will be unable to tuck XP down, and will therefore be of no use in circumventing the MTG in that case. I will briefly illustrate the importance of this qualification with two case studies focused on Romanian *wh*-movement and German scrambling.

Romanian is relevant here as the other best-studied language — alongside Bulgarian — showing the hallmarks of an insatiably *wh*-attracting complementizer. See, for example, the evidence for obligatory tucking-in in (32).

---

[17] Or, at least, no more Shortest violations than any multiple-*wh* question does — cf. the discussion in fn. 11.

(32)     Romanian (Rudin 1988: 474)

    a.  Cine  ce    a    spus?
         who  what  has  said
         'Who said what?'

    b.  *Ce   cine  a    spus?
         what  who  has  said

In light of (32) and Rudin's (1988) further evidence to the same effect, one might then expect Romanian *wh*-movement to pattern with Bulgarian across the board, including with respect to the ability to pseudo-violate or genuinely violate the MTG. This expectation, however, clashes with the unacceptability of sentences like those in (34).

(33)     Romanian (Andreea Cristina Nicolae, p.c.)

    a.  nişte  miniştri  de interne  din  Europa  centrală
         some  ministers  of  internals  from  Europe  central
         'some ministers of internal affairs from Central Europe.'

    b.  *nişte  miniştri  din  Europa  centrală  de interne.
         some  ministers  from  Europe  central  of  internals

(34)     Romanian (Andreea Cristina Nicolae, p.c.)

    a.  *[Care  miniştri  $t_{wh_2}$  din  Europa  centrală]$_{wh_1}$  [de care]$_{wh_2}$  ai
         which  ministers    from  Europe  central    of  which  have
         întâlnit  $t_{wh_1}$?
         met
         Intended: 'Which ministers of what from Central Europe did you meet?'

    b.  *[Care  miniştri  $t_{wh_2}$  din  Europa  centrală]$_{wh_1}$  te  întrebai  [de care]$_{wh_2}$  am
         which  ministers    from  Europe  central    2SG  asked  of  which  have
         putea  întâlni  $t_{wh_1}$?
         could  meet
         Intended: '[Which ministers $t_{wh_2}$ from Central Europe]$_{wh_1}$ were you wondering [of what]$_{wh_2}$ we might meet $t_{wh_1}$?'

As it turns out, however, the sentences in (34) are ungrammatical not because of any problems with tucking-down *per se*, but because the tucking-down step that each of them involves happens to violate an independently observable restriction on Romanian *wh*-movement: as pointed out by Steriade (1981), Romanian just prohibits *wh*-moving anything out of a DP. In view of this constraint, we thus expect that *wh*-movement of a *wh*$_2$-PP out of a *wh*$_1$-DP should be impossible not only when going from within an outer specifier to an inner specifier (tucking-down), but also going the other way around (tucking-out), and even going from the specifier of a lower CP up to

the specifier of a higher CP. These expectations are all borne out: the ungrammatical sentences in (34) turn out to be part of a larger paradigm including (35).

(35)     Romanian (Andreea Cristina Nicolae, p.c.)

     a. *[De care]$_i$  ai    întălnit [un ministru $t_i$ din   Europa centrală]?
        of   which have met    a    minister    from Europe central
        Intended: 'What did you meet a minister of from Central Europe?'

     b. *[De care]$_{wh_2}$ [care  miniştri   $t_{wh_2}$ din   Europa centrală]$_{wh_1}$ ai
        of    which   which ministers      from Europe central       have
        întălnit $t_{wh_1}$?
        met
        Intended: 'Which ministers of what from Central Europe did you meet?'

     c. *[De care]$_{wh_2}$ te   întrebai [care  miniştri   $t_{wh_2}$ din   Europa centrală]$_{wh_1}$ am
        of    which  2SG asked    which ministers      from Europe central       have
        putea întălni $t_{wh_1}$?
        could  meet
        Intended: '[Of what]$_{wh_2}$ were you wondering [which ministers $t_{wh_2}$ from Central Europe]$_{wh_1}$ we might meet $t_{wh_1}$?'

Romanian thus teaches us that insatiability is a necessary but not a sufficient condition for successful use of tucking-down as a way to elude the MTG: the tucking-down step must not be independently forbidden by the grammar.

   This type of reasoning applies even to cases where involvement of an insatiable attractor is less securely established to begin with: if the tucking-down step is independently ruled out, then the question of whether the relevant attractor is insatiable or not just becomes moot for the purposes of our predictions, as that attractor would not be able to circumvent the MTG either way. One such case, brought to my attention by a reviewer, has to do with German scrambling. Scrambling in German is famously constrained by the MTG (cf. (36))—it is, in fact, one of the constructions that led Müller (1993) to formulate his version of the generalization.[18]

---

[18] Müller (1998: 227–228) also notes an apparent exception to the generalization: "cases of remnant infinitive scrambling become acceptable if the antecedent of the unbound trace is not a full NP that has undergone scrambling, but rather a weak pronoun […] or a pronominal clitic […] in a Wackernagel(-like), i.e., pre-subject, position." I follow him in taking "this as an indication that pronoun movement to a pre-SpecI position in German is not an instance of scrambling; rather, some other movement type seems to be involved" (*ibid.*).

(i)   German (Müller 1998: 227)
     dass [$t_1$ zu lesen]$_3$ {es$_1$/'s$_1$} keiner   $t_3$ versucht hat
     that    to  read   it       nobody    tried    has
     'that nobody has tried to read it'

(36) German (Müller 1998: 24)

    a. *… dass [$t_2$ zu lesen]$_1$ [das Buch]$_2$ keiner       $t_1$ versucht hat

        that     to read   the  book  nobody.NOM   tried     has

        Intended: '… that nobody tried to read the book.'

    b. … dass [das Buch zu lesen]$_1$ keiner       $t_1$ versucht hat

        that  the  book  to  read   nobody.NOM   tried    has

    c. … dass [das Buch]$_2$ keiner      $t_2$ zu lesen versucht hat

        that  the  book   nobody.NOM    to  read  tried     has

The construction has, however, sometimes been assumed to involve at least one insatiable attractor (e.g. Müller 2014: 25), although the evidence for obligatory tucking-in remains somewhat murky.[19] However, even if the evidence for insatiability were crystal-clear, the current theory would not necessarily be in trouble, because there turns out to be, here too, an independent constraint that would block the tucking-down step anyway: as Müller (1998: 251ff) and Sauerland (1999: 180) observe, German just prohibits scrambling anything out of a scrambled constituent (see e.g. (37a)).[20]

---

[19] Consider (i), for example, where an intraposed restructured infinitive is used to make sure $t_1$ asymmetrically c-commands $t_2$. Although speakers vary in how easily they accept the examples (multiple scrambling and intraposed infinitives both being marked phenomena), none of them seem to perceive a contrast.

(i) German (Kai von Fintel, Martin Hackl, and Verena Hehl, p.c.)

    a. ?… dass [dem     Johann]$_1$ [die     Torte]$_2$ keiner  $t_1$ [$t_2$ zu essen] erlaubte.

        that the.DAT J.      the.ACC cake   nobody       to eat    allowed

    b. ?… dass [die     Torte]$_2$ [dem    Johann]$_1$ keiner  $t_1$ [$t_2$ zu essen] erlaubte.

        that the.ACC cake    the.DAT J.      nobody       to eat    allowed

        '… that no-one allowed Johann to eat the cake'

[20] Stating the constraint does not, of course, amount to explaining why it holds. Müller (1998) takes it to follow from the Condition on Extraction Domain, but we have already seen that that condition is arguably overly restrictive (cf. § 2.1 and fn. 10). In particular, even when it comes specifically to scrambling, Saito (1985: 249) provides evidence that Japanese, unlike German, does permit scrambling out of scrambled constituents—an option that Universal Grammar should therefore not contain a blanket ban on. The latter fact also entails, incidentally, that if Japanese scrambling were to be securely traced back to an insatiable attractor, then we could not replicate the same strategy we used for German in order to explain why the construction still obeys the MTG (which Takano 1994 shows it does). It should be kept in mind, however, that the empirical picture in Japanese is muddled by the joint presence of A- and Ā- (or at least clause-internal and cross-clausal) scrambling, with mixed evidence as to whether the two deal in the same features or not. On the one hand, Richards (1997: 77ff) shows that multiple A-scrambling displays obligatory tucking-in effects, while mixtures of A- and Ā-scrambling allow greater ordering flexibility—which suggests that the two types of scrambling are triggered by different features. On the other hand, if that were so, we would expect A- and Ā-scrambling to be able to interact without Müller-Takano effects (just like, e.g., *wh*-movement and extraposition), contrary to fact. The picture strikes me as too murky to pose a serious threat to the account at this time. More work on Japanese scrambling from the current perspective is needed.

(37) German (modeled after Sauerland 1999: 180)

    a. *?… dass [das Buch]$_2$ vergeblich     [t$_2$ zu lesen]$_1$ keiner      t$_1$ versuchte
           that the book    unsuccessfully    to read   nobody.NOM    tried
        Intended: '…that nobody unsuccessfully tried to read the book.'

    b.  … dass [das Buch]$_2$ vergeblich     keiner       [t$_2$ zu lesen] versuchte
           that the book    unsuccessfully nobody.NOM     to read   tried

    c.  … dass vergeblich     [das Buch zu lesen]$_1$ keiner      t$_1$ versuchte
           that unsuccessfully the   book to read    nobody.NOM    tried

As a result, a scrambling-triggering probe — which, following Grewendorf & Sabel (1999), we may refer to as Σ — may attract to its own Spec,ΣP the VP *das Buch zu lesen* 'to read the book', but then may not further scramble *das Buch* 'the book' out of that VP, regardless of (in)satiability and directionality. That Σ probe will therefore be unable to give rise to any MTG-violations, just as desired.

In summary, the theory presented so far might sometimes appear to overgenerate MTG-violations, but if the cases of Romanian and German are any indication, the problem disappears once independent strictures on movement (especially on the subextraction step) are properly factored in.

## 5 A challenge to alternative approaches

A Shortest-based theory along the lines sketched here thus holds promise as an account of both MTG-effects (as with *wh*-movement in English or, with a twist, in Romanian) and apparent or genuine MTG-violations (as with *wh*-movement in Bulgarian). In this final section, I would like to argue that, by contrast, alternative approaches to the MTG have a hard time replicating these results.

Most such approaches, exemplified by Williams (2003; 2011), Grewendorf (2003; 2015), and Abels (2007),[21] treat the MTG as the limiting case of a larger set of constraints on remnant

---

[21] One approach I am not discussing in the main text is Müller's (2014; 2015), which presupposes somewhat different empirical *explananda* from the ones I have assumed here — especially regarding movement out of previously moved phrases (cf. § 2.1 and fn. 10), which Müller believes should always be ruled out. The key idea behind Müller's approach is that remnant movement obeys a special licensing condition in the course of the derivation: immediately upon reaching its own criterial position, the remnant must c-command the criterial position of the subextractee. Müller (2014: 80ff) then develops a set of principles to the effect that, whenever the remnant and the subextractee criterially move into multiple specifiers of a single head, the remnant must always move first, and therefore reach its criterial position too early to meet its licensing condition. As Müller (2014: 92) points out, however, this account leaves a loophole open: "remnant movement should be possible, in violation of the [MTG] as it is formulated above, if the remnant […] has the same movement-related feature as [the subextractee], but checks this with some higher head in the clause." While this loophole might come in handy to account for the acceptability of (18a)/(31) in Bulgarian (cf. also Müller 2014: 93 for prosodic evidence from German scrambling), it appears elsewhere to be both too strong and too weak: on the one hand, it still isn't enough to allow for the grammaticality of (19a)/(29) in Bulgarian; on the other hand, it's unclear why the same loophole shouldn't be available to English examples like (3a).

movement. Foundational to these approaches is the observation that, as far as minimality is concerned, remnant movement (barring tucking-down) should be restricted **only** by the MTG — i.e. it should be fine as long as the two interacting movements are of different types, regardless of which movement type applies first (the remnant-creating movement) and which one applies second (the movement of the whole remnant). However, as proponents of these approaches have pointed out, this is not the case: it actually does matter which movement type feeds which. More specifically, the two types appear to be constrained along the lines of (38) (cf. especially Grewendorf 2003: 67).

(38)    After phrase XP has moved from node $\alpha$ to node $\omega$, a remnant phrase YP that dominates $\alpha$ but not $\omega$ can move to a node c-commanding $\omega$ only if movement of YP is of a higher type than movement of XP according to the following hierarchy:

mvt to subject position $\ll$ clause-bounded scrambling $\ll$ *wh*-mvt $\ll$ topicalization

The purview of the MTG thus turns out to be a proper subset of the purview of the putative generalization in (38): not only must remnant movement not be exactly as high in type as the remnant-creating movement; it must, in fact, be strictly **higher**. By only ruling out the MTG subcases, the minimality approach thus puts us in danger of missing a broader pattern — i.e. of artificially divorcing the ban on certain same-movement-type interactions (e.g. on *wh*-moving a *wh*-trace-containing remnant) from the ban on certain **mixed**-movement-types interactions (e.g. on A-moving a *wh*-trace-containing remnant).[22]

What I wish to suggest, however, is that, for all the initial appeal of the broader pattern, Bulgarian provides us with an empirical argument for the superiority of a Shortest-based account of the MTG over any alternative in terms of (38). Remember that, by adopting a Shortest-based approach, we just had to open the door to tucking-down (following Richards 2004) in order to derive the grammaticality of both (39a) (a single CP involving tucking-down) and (39b)

---

[22] A potential argument for divorcing the two bans in just this way, however, comes from Abels (2007: 71ff). Building on Sakai (1994), Grewendorf (2003), and Williams (2003), Abels argues that the constraint in question should regulate not only remnant movement but all feeding interactions between movements — including multiple movements of one and the same constituent (proper/improper movement), as well as so-called 'surfing' interactions (Sauerland 1999) whereby the first movement step targets a larger phrase YP and the second step moves a subconstituent of YP from inside YP's landing site. If that is correct, then the constraint in (38) might have to be weakened along the lines of (i) in order to allow, for example, for 'surfing' *wh*-movement out of a *wh*-moved constituent, as exemplified in (3b) (cf. also Grewendorf 2015: 26). The weaker version in (i) would then leave Müller-Takano effects out of the constraint's purview.

(i)    After phrase XP has moved from node $\alpha$ to node $\omega$, a remnant phrase YP that dominates $\alpha$ but not $\omega$ can move to a node c-commanding $\omega$ only if movement of YP is **not of a lower** type than movement of XP according to the following hierarchy:

movement to subject position $\ll$ clause-bounded scrambling $\ll$ *wh*-movement $\ll$ topicalization

For further issues with (38), see Keine (2020b: ch. 6) and Kobayashi (2020).

(a biclausal structure where tucking-down in the embedded CP feeds Shortest-compliant attraction of the remnant into the matrix CP).

(39) Bulgarian (Richards 2004: 456, 459)

a. [Kolko studenti $t_{wh_2}$ [ot Bulgaria]]$_{wh_1}$ [po kakvo]$_{wh_2}$ vidja $t_{wh_1}$?
   how.many students     from Bulgaria     of what     you.saw
   'How many students of what from Bulgaria did you see?'     (= (19a))

b. [Kolko studenti $t_{wh_2}$]$_{wh_1}$ se opitvaš da razbereš [[ot koi
   how.many students     REFL you.try to understand     from which
   strani]$_{wh_2}$ e ubil Ivan $t_{wh_1}$]?
   countries AUX killed Ivan
   '[How many students $t_{wh_2}$]$_{wh_1}$ are you trying to understand [from which
   countries]$_{wh_2}$ Ivan killed $t_{wh_1}$?'     (= (18a))

By contrast, if we were to adopt a constraint like (38), we would have a problem deriving these data—and the introduction of tucking-down could only help us out so much.

Specifically, tucking-down can only reconcile the constraint in (38) with the acceptability of (39a). On a tucking-down analysis, (39a) involves movement of the subconstituent XP/$wh_2$ **following** movement of the larger phrase YP/$wh_1$ (cf. (29)/(40)), and therefore falls outside the remit of the constraint.

(40)
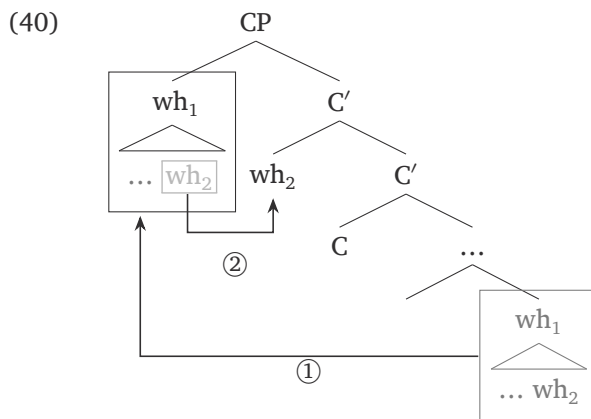


However, even tucking-down cannot make (38) consistent with the acceptability of (39b). That is because, after the tucking-down step in the embedded CP (the arrow marked as ② in (31)/(41)), the larger phrase YP/$wh_1$ does qualify as a constituent that something has *wh*-moved out of—and so, *per* (38), would be predicted to be unable to undergo any further *wh*-movement itself, contrary to fact.

(41)
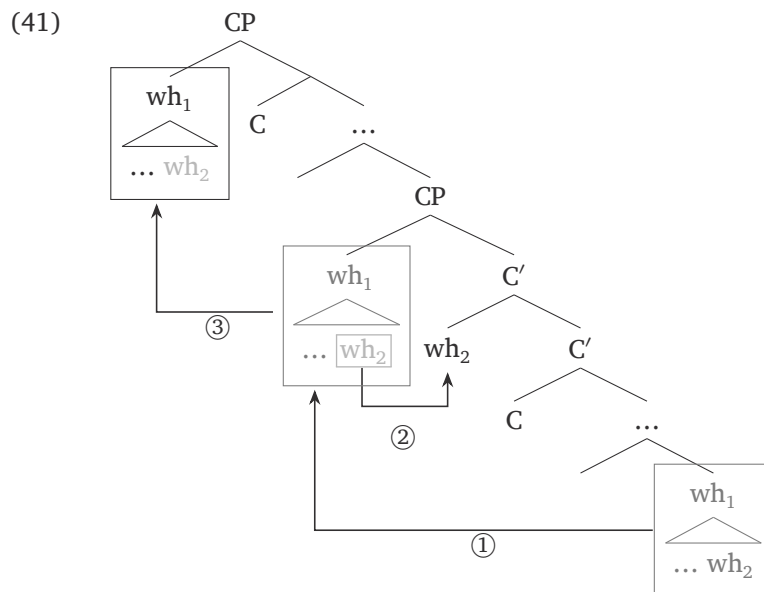


The only way to reconcile this approach with (39b), then, would be to not only countenance tucking-down, but to effectively hard-code into (38) an exception specifically devised for it — e.g. by replacing "a remnant phrase YP that dominates $\alpha$ **but not** $\omega$" with "a remnant phrase YP that dominates $\alpha$ **and is c-commanded by** $\omega$" in the wording of the constraint. This would amount, however, to stipulating a difference between tucking-down remnants and all other remnants sheerly by brute force — a difference that the Shortest-based account can instead derive on principled grounds.

## 6 Conclusions

In this paper, I have argued that Richards' (2004) data from Bulgarian nested-*wh* questions instantiate genuine exceptions to the MTG. I have also shown that such exceptions are actually predicted by an approach deriving the generalization from a minimality constraint on movement paths, as long as one adopts Richards' (2004) assumption that insatiable attractors can trigger "tucking-down" movement from inside their outer specifier to their inner specifier. Finally, I have argued that non-minimality-based approaches to the MTG, even coupled with the countenancing of tucking-down, have a hard time accounting for the Bulgarian facts.

## Appendix: A multidominance alternative (Frampton 2004)

Frampton (2004) observes that some of the conceptual awkwardness of tucking-down dissolves within a system that treats movement as remerge/multidominance — a system he advocates for and supplements with an elegant linearization algorithm. On this account, we may still maintain that probes only attract constituents they c-command (cf. (42)), under a multidominance-adjusted definition of *c-command* such as (43b).

(42)　At a given point in the derivation, head H is a *potential trigger* for remerging phrase XP as a daughter of $\omega$ if

a.　$\omega$ is a projection of H that does not immediately dominate H

and, at that point in the derivation,

b.　H is projecting;

c.　and H has an active attractor feature that XP matches;

d.　and **H c-commands XP**.

(43)　C-command & dominance in multidominance (notationally adapted from Abels 2012: 102–3)

a.　Given two nodes $\alpha$ and $\beta$ in a single-rooted directed graph,

(i)　$\alpha$ *totally dominates* $\beta$ iff $\alpha$ is contained in every directed path from the root to $\beta$, and

(ii)　$\alpha$ *partially dominates* $\beta$ iff $\alpha$ is contained in some directed path from the root to $\beta$,

where a *directed path* from the root to $\beta$ is an ordered set $\langle n_1, \ldots, n_k \rangle$ such that $n_1$ is the root, $n_k$ is $\beta$, and for $1 \leq i < k$, $n_i$ is a mother of $n_{i+1}$.

b.　$\alpha$ *c-commands* $\beta$ iff $\alpha$ does not dominate $\beta$, $\beta$ does not dominate $\alpha$, and some mother of $\alpha$ partially dominates $\beta$.

At the same time, we may also recast the Shortest-based account as follows.[23]

First, we adapt Collins' (1994) notion of a movement path (9) as in (44).

(44)　The *remerger-path* of $\alpha$, to be notated as $P_R(\alpha)$, is the set of all the nodes that partially dominate $\alpha$ but don't totally dominate $\alpha$.

Next, we dictate that remerger options whose remerged category ends up having a smaller remerger-path must block alternative remerger options whose remerged category ends up having a larger remerger-path.

---

[23] This is not, however, what Frampton (2004) himself does. Instead, he pursues an altogether different route, replacing Shortest with a condition that favors minimal disruption of previously established precedence relations (cf. Müller 2014: 81fn11) — an alternative that strikes me as theoretically costlier.

(45) Let head H be, at some point in the derivation, a potential trigger both for remerging $\alpha$ as a daughter of $\omega$ and for remerging $\alpha'$ (possibly identical to $\alpha$) as a daughter of $\omega'$ (possibly identical to $\omega$).

If remerging $\alpha$ as a daughter of $\omega$ would result in $|P_R(\alpha)| = n$, while remerging $\alpha'$ as a daughter of $\omega'$ would result in $|P_R(\alpha')| > n$, then $\alpha'$ cannot be remerged as a daughter of $\omega'$ at that point in the derivation.

I will leave it as an exercise for the reader to verify that the minimality principle in (45) replicates the results of Shortest from (13)–(14) and (17). As for the crucial Bulgarian case in (28)–(29), we can see that (45), too, requires C to remerge $wh_1$ first ((46)), and that it, too, does not single one option out for the subsequent step, as the remerger-paths in (47a) and (47b) have the same cardinality.

(46)



(47)
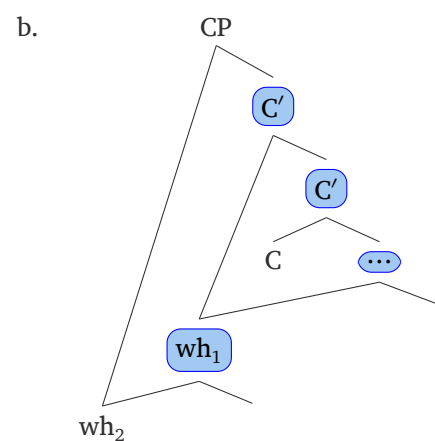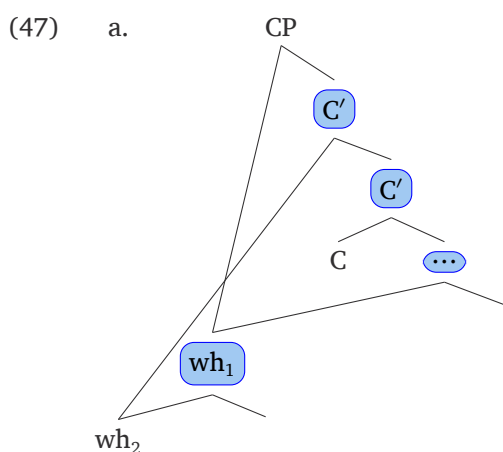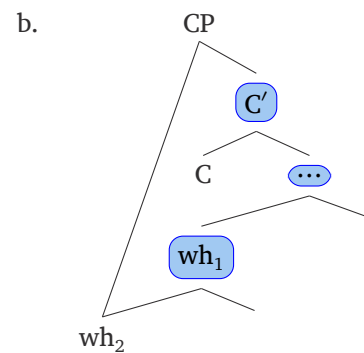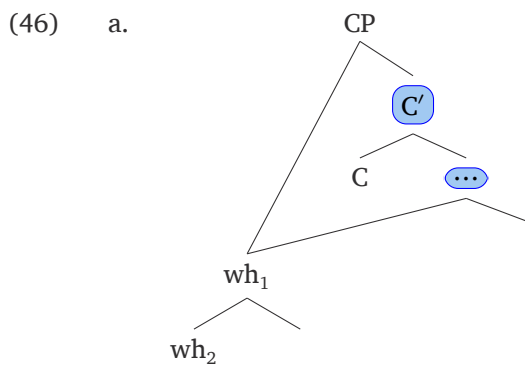


It thus appears that I could have reimplemented Richards' (2004) account of Bulgarian in terms of Frampton's (2004) system. I have refrained from doing so, however, both in the interest of expository convenience and because the agreement literature cited in §3.2 has offered independent arguments that probes can search into their non-complements (even

non-complements base-generated as such)—which detracts from the appeal of strictly-downward probing as an independent theoretical *desideratum*.

Readers who value that *desideratum* more highly than I do are invited to assess the proposal under the Framptonian implementation outlined in this appendix.

## Acknowledgements

## Competing interests

The author has no competing interests to declare.

## References

Abels, Klaus. 2003. *Successive cyclicity, anti-locality, and adposition stranding*. Storrs, CT: University of Connecticut dissertation.

Abels, Klaus. 2007. Towards a restrictive theory of (remnant) movement. *Linguistic Variation Yearbook* 7. 53–120. DOI: https://doi.org/10.1075/livy.7.04abe.

Abels, Klaus. 2012. *Phases: An essay on cyclicity in syntax*. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110284225.

Bailyn, John Frederick. 2017. Bulgarian Superiority and Minimalist movement theory. In Oseki, Yohei & Esipova, Masha & Harves, Stephanie (eds.), *Proceedings of Formal Approaches to Slavic Linguistics 24*, 27–49. Ann Arbor, MI: Michigan Slavic Publications.

Béjar, Susana & Rezac, Milan. 2009. Cyclic Agree. *Linguistic Inquiry* 40(1). 35–73. DOI: https://doi.org/10.1162/ling.2009.40.1.35.

Belletti, Adriana & Collins, Chris (eds.). 2021. *Smuggling in Syntax*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780197509869.001.0001.

Billings, Loren & Rudin, Catherine. 1996. Optimality and Superiority: A new approach to overt multiple-*wh* ordering. In Toman, Jindřich (ed.), *Formal Approaches to Slavic Linguistics: The College Park meeting, 1994*, 35–60. Ann Arbor: Michigan Slavic Publications.

Bobaljik, Jonathan David. 1995. *Morphosyntax: The syntax of verbal inflection*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Bošković, Željko. 1997. On certain violations of the Superiority Condition, AgrO, and economy of derivation. *Journal of Linguistics* 33(2). 227–254. DOI: https://doi.org/10.1017/S0022226797006476.

Bošković, Željko. 2002. On multiple *wh*-fronting. *Linguistic Inquiry* 33(3). 351–383. DOI: https://doi.org/10.1162/002438902760168536.

Bošković, Željko. 2007. On the locality and motivation of Move and Agree: An even more minimal theory. *Linguistic Inquiry* 38(4). 589–644. DOI: https://doi.org/10.1162/ling.2007.38.4.589.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press. DOI: https://doi.org/10.7551/mitpress/9780262527347.001.0001.

Citko, Barbara & Gračanin-Yuksek, Martina. 2013. Towards a new typology of coordinated *wh*-questions. *Journal of Linguistics* 49(1). 1–32. DOI: https://doi.org/10.1017/S0022226712000175.

Clem, Emily. 2023. Cyclic expansion in Agree: Maximal projections as probes. *Linguistic Inquiry* 54(1). 39–78. DOI: https://doi.org/10.1162/ling_a_00432.

Collins, Chris. 1994. Economy of Derivation and the Generalized Proper Binding Condition. *Linguistic Inquiry* 25(1). 45–61.

Collins, Chris. 1997. *Local economy*. Cambridge, MA: MIT Press.

Collins, Chris. 2005a. A smuggling approach to raising in English. *Linguistic Inquiry* 36(2). 289–298. DOI: https://doi.org/10.1162/0024389053710701.

Collins, Chris. 2005b. A smuggling approach to the passive in English. *Syntax* 8(2). 81–120. DOI: https://doi.org/10.1111/j.1467-9612.2005.00076.x.

Corver, Norbert. 2017. Freezing effects. In Everaert, Martin & van Riemsdijk, Henk C. (eds.), *The Wiley Blackwell companion to syntax*, 2nd edn. Oxford: Wiley-Blackwell. DOI: https://doi.org/10.1002/9781118358733.wbsyncom055.

Davis, Colin. 2020. *The linear limitations of syntactic derivations*. Cambridge, MA: MIT dissertation.

Deal, Amy Rose. 2024. Interaction, satisfaction, and the PCC. *Linguistic Inquiry* 55(1). 39–94. DOI: https://doi.org/10.1162/ling_a_00455.

Fitzpatrick, Justin M. 2002. On minimalist approaches to the locality of movement. *Linguistic Inquiry* 33(3). 443–463. DOI: https://doi.org/10.1162/002438902760168563.

Frampton, John. 2004. Copies, traces, occurrences, and all that: Evidence from Bulgarian multiple *wh*-phenomena. Ms. http://mathserver.neu.edu/ling/pdf/CopiesTraces.pdf.

Grewendorf, Günther. 2001. Multiple *wh*-fronting. *Linguistic Inquiry* 32(1). 87–122. DOI: https://doi.org/10.1162/002438901554595.

Grewendorf, Günther. 2003. Improper remnant movement. *Gengo Kenkyo: The Journal of the Linguistic Society of Japan* 123. 47–94.

Grewendorf, Günther. 2015. Problems of remnant movement. In Grewendorf, Günther (ed.), *Remnant movement*, 3–31. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9781614516330-002

Grewendorf, Günther & Sabel, Joachim. 1999. Scrambling in German and Japanese: Adjunction versus multiple specifiers. *Natural Language and Linguistic Theory* 17(1). 1–65. DOI: https://doi.org/10.1023/A:1006068326583.

Kayne, Richard S. 1983. Connectedness. *Linguistic Inquiry* 14(2). 223–249.

Keine, Stefan. 2017. Agreement and vP phases. In LaCara, Nicholas & Moulton, Keir & Tessier, Anne-Michelle (eds.), *A schrift to fest Kyle Johnson*, 177–185. Amherst, MA: Linguistics Open Access Publications.

Keine, Stefan. 2020a. Locality domains in syntax: Evidence from sentence processing. *Syntax* 23(2). 105–151. DOI: https://doi.org/10.1111/synt.12195.

Keine, Stefan. 2020b. *Probes and their horizons*. Cambridge, MA: MIT Press. DOI: https://doi.org/10.7551/mitpress/12003.001.0001.

Keine, Stefan & Dash, Bhamati. 2022. Movement and cyclic Agree. *Natural Language and Linguistic Theory* 41. 679–732. DOI: https://doi.org/10.1007/s11049-022-09538-1.

Kitahara, Hisatsugu. 1993. Deducing 'Superiority' effects from the Shortest Chain Requirement. *Harvard Working Papers in Linguistics* 3. 109–119.

Kitahara, Hisatsugu. 1994. Restricting ambiguous rule-application: A unified analysis of movement. *MIT Working Papers in Linguistics* 24. 179–209.

Kitahara, Hisatsugu. 1997. *Elementary operations and optimal derivations*. Cambridge, MA: MIT Press.

Kobayashi, Filipe Hisao. 2020. Proper interleaving of A- & A'-movement: A Brazilian Portuguese case study. Ms. MIT. https://ling.auf.net/lingbuzz/005609.

Krapova, Iliyana & Cinque, Guglielmo. 2008. On the order of wh-phrases in Bulgarian multiple wh-fronting. In Zybatow, Gerhild & Szucsich, Luka & Junghanns, Uwe & Meyer, Roland (eds.), *Formal Description of Slavic Languages: The fifth conference, Leipzig 2003*, 318–336. Frankfurt am Main: Peter Lang.

McCawley, James D. 1976. Introduction. In McCawley, James D. (ed.), *Notes from the linguistic underground*, 1–19. New York: Academic Press. DOI: https://doi.org/10.1163/9789004368859_002

McCloskey, James. 2000. Quantifier float and *wh*-movement in an Irish English. *Linguistic Inquiry* 31(1). 57–84. DOI: https://doi.org/10.1162/002438900554299.

Müller, Gereon. 1993. *On deriving movement type asymmetries*. Tübingen: Universität Tübingen dissertation.

Müller, Gereon. 1998. *Incomplete category fronting: A derivational approach to remnant movement in German*. Dordrecht: Springer. DOI: https://doi.org/10.1007/978-94-017-1864-6.

Müller, Gereon. 2014. *Syntactic buffers*. Leipzig: Linguistische Arbeitsberichte. http://www.uni-leipzig.de/~muellerg/mu765.pdf.

Müller, Gereon. 2015. Remnant movement in a local derivational grammar. In Grewendorf, Günther (ed.), *Remnant movement*, 53–92. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9781614516330-004.

Nakamura, Masanori. 1998. Reference set, Minimal Link Condition, and parametrization. In Barbosa, Pilar & Fox, Danny & Hagstrom, Paul & McGinnis, Martha & Pesetsky, David (eds.), *Is the best good enough?*, 291–313. Cambridge, MA: MIT Press.

Neeleman, Ad & van de Koot, Hans. 2010. A local encoding of syntactic dependencies and its consequences for the theory of movement. *Syntax* 13(4). 331–372. DOI: https://doi.org/10.1111/j.1467-9612.2010.00143.x.

Pesetsky, David. 2000. *Phrasal movement and its kin*. Cambridge, MA: MIT Press. DOI: https://doi.org/10.7551/mitpress/5365.001.0001.

Rezac, Milan. 2003. The fine structure of Cyclic Agree. *Syntax* 6(2). 156–182. DOI: https://doi.org/10.1111/1467-9612.00059.

Richards, Norvin. 1997. *What moves where when in which language?* Cambridge, MA: MIT dissertation.

Richards, Norvin. 2004. Against bans on lowering. *Linguistic Inquiry* 35(3). 453–463. DOI: https://doi.org/10.1162/0024389041402643.

Rizzi, Luigi. 2006. On the form of chains: Criterial positions and ECP effects. In Cheng, Lisa Lai-Shen & Corver, Norbert (eds.), Wh-*movement: Moving on*, 97–133. Cambridge, MA: MIT Press. DOI: https://doi.org/10.7551/mitpress/7197.003.0010.

Rudin, Catherine. 1988. On multiple questions and multiple Wh fronting. *Natural Language and Linguistic Theory* 6(4). 445–501. DOI: https://doi.org/10.1007/BF00134489.

Saito, Mamoru. 1985. *Some asymmetries in Japanese and their theoretical implications*. Cambridge, MA: MIT dissertation.

Saito, Mamoru. 1989. Scrambling as semantically vacuous A′-movement. In Baltin, Mark & Kroch, Anthony (eds.), *Alternative conceptions of phrase structure*, 182–200. Chicago: University of Chicago Press.

Sakai, Hiromu. 1994. Derivational economy in long distance scrambling. *MIT Working Papers in Linguistics* 24. 295–314.

Sauerland, Uli. 1999. Erasability and interpretation. *Syntax* 2(3). 161–188. DOI: https://doi.org/10.1111/1467-9612.00019.

Steriade, Donca. 1981. Extraction from NP in Romance and possessor raising. Ms. MIT.

Takano, Yuji. 1994. Unbound traces and indeterminacy of derivation. In Nakamura, Masaru (ed.), *Current topics in English and Japanese*, 229–253. Tokyo: Hituzi Syobo.

Thiersch, Craig. 2017. Remnant movement. In Everaert, Martin & van Riemsdijk, Henk C. (eds.), *The Wiley Blackwell companion to syntax*, 2nd edn. Oxford: Wiley-Blackwell. DOI: https://doi.org/10.1002/9781118358733.wbsyncom123.

Wexler, Kenneth & Culicover, Peter W. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

Williams, Edwin. 2003. *Representation theory*. Cambridge, MA: MIT Press. DOI: https://doi.org/10.7551/mitpress/5893.001.0001.

Williams, Edwin. 2011. *Regimes of derivation in syntax and morphology*. London: Routledge. DOI: https://doi.org/10.4324/9780203830796.