Crossing dependencies, weak generative capacity, and complexity in language∗

0 Introduction

The question of the formal computational complexity of human language has been debated since the earliest days of generative grammar (Chomsky 1957), and in various forms remains at the center of ongoing inquiry in linguistics, neurobiology, and anthropology (cf. Hauser et al. 2002, Everett 2005, Fitch et al. 2005, Jackendoff & Pinker 2005, Sauerland & Trotzke 2011). One major area of investigation concerns the question of how best to formally characterize the complexity of the sets of grammatical sentences (weak generative capacity) and the associated sets of syntactic structures (strong generative capacity) of natural language. For the syntactic theorist, it is naturally the study of structure and derivation – strong generative capacity – that is of principal interest. Indeed, because weak generative capacity concerns only the surface forms of language, without reference to or consideration of the associated structure, it has thus been of considerably less relevance for theory building. This mathematical unambiguity and theory neutrality, however, have made weak generative capacity an essential tool for differentiating competing theories of grammar.

The first and most famous example which employs weak generative capacity to this aim appears in Syntactic Structures, where mathematical arguments are provided that the non-hierarchical regular grammars are mathematically insufficient for generating the surface strings of English. Of course Chomsky also argues against employing context-free grammars as models for English syntax, although in this case he does not provide a corresponding argument from weak generative capacity (no such argument exists to this day), relying instead on intuitive and conceptual justifications. In fact, an unequivocal demonstration that human language is non-context-free proved so elusive that it was not until the mid nineteen eighties, nearly thirty years after the publishing of Syntactic Structures, that the first such arguments appeared. This was not for lack of trying: Chomsky (1959), Bar-Hillel & Shamir (1960), Postal (1964), and Elster (1978) all offered various proposals, the inadequacies of which are pointed out by Pullum & Gazdar (1982). It was Huybregts (1984) and Shieber (1985) who finally delivered definitive non-context-freeness arguments, both based on a quirk of word order in successively embedded object-control clauses in Dutch and Swiss-German, respectively. Subsequent Huybregts and Shieber’s discoveries, studies of weak generative capacity shifted away from non-context-freeness to finding the appropriate upper bound on the complexity of natural language. Miller (1991) thus shows that examples involving arbitrary iteration of wh-island violations in Swedish are not generable with a tree-joining grammar (TAG), and Radzinski (1990) and Becker et al. (1991) show that Chinese numeral names and German scrambling, respectively, are beyond the generative capacity of even the set-local multicomponent tree-joining grammars (Local MC-TAG). Becker et al. (1992) goes on to show that set-local MC-TAGs are weakly equivalent to the more general linear context-free rewriting systems (LCFRS), so that Chinese numeral names and German scrambling exceed these grammars as well.1 Similar results have been argued for on the basis of Old Georgian case stacking (Michaelis & Kracht 1997) and Yoruba VP copying (Kobele 2006).

What these various studies on weak generative capacity have in common, with the exception perhaps of Kobele’s (2006) work, is that they are concerned chiefly with determining whether or not a given formal system is capable of generating a particular set of data. The actual linguistic factors at work behind the derivation of the relevant structures and strings are thus always of secondary concern. Given that weak generative capacity has historically served primarily as a mathematical diagnostic for differentiating theories, this makes sense. The ultimate aim of syntactic theory, however, must be to not only provide the

∗ I omit acknowledgements at this stage for the purposes of anonymity

1 That said, see Gazdar (1982) for a critique of arguments like the one offered for Chinese by Radzinski (1990).
mathematical characteristics or even the structure that underlies natural language, but also to furnish a set of principles at work in the derivation of such structure. This guiding principal calls for an investigation into the nature of complexity in the grammar that asks both what the minimal formal grammar required to derive this or that set of data is and what mechanisms natural language syntax actually invokes in the derivation of such data. To this end, work on weak generative capacity has identified exactly those constructions in natural language that are unequivocally complex in the formal sense, inviting investigation into the derivational origins of such complexity. It is to this issue that I turn my attention in this paper, taking up the following questions: what are those syntactic operations responsible for the crossing dependencies that underly the various instances of non-context-free (and non-LCFRS) complexity in language? Do they vary from one instance of non-context-freeness to another? Why aren’t they obviously present in every language, if they exist in some cases?

My admittedly modest consideration of these questions will take as its starting point a novel instance of crossed dependencies that lead to non-context-free complexity, multiple *wh*-questions in Romanian. After establishing the basic veracity of the data, I will investigate, from the perspective of a feature-driven minimalist syntactic theory, those syntactic mechanisms responsible for the derivation of the data. I will then consider a few other cases of crossed-dependencies and associated complexity and ask whether these are due to similar mechanisms. The claim that I will defend is that the essential syntactic mechanism overtly visible in Romanian multiple *wh*-questions, multiple-attraction of like elements by a single functional head, may in fact be the single syntactic device responsible for crossed-dependencies – and the associated complexity – in natural language.

The remainder of the paper is structured as follows. First, I demonstrate that Romanian is weakly non-context-free on the basis of its treatment of multiple *wh*-questions.\(^2\) This will constitute Section 1 of the paper: I present the relevant facts on Romanian in Section 1.1, the argument in Section 1.2, and a refutation of several counterarguments in Section 1.3. Second, in Section 2.1, I show that under a leading analysis of the Romanian facts, Richards’ (1997) theory of *tucking-in*, the essential operation in the grammar that yields non-context-freeness is multiple-attraction by a single functional head. In Section 2.2, I extend this analysis to other cases of non-context-free complexity, and in Section 2.3, I consider some cases to which the analysis does not obviously extend. In Section 3 I contextualize the result by considering alternatives and by discussing why, given the broad availability of the mechanism of multiple-attraction, non-context-free complexity is not more obviously common in the surface forms of the world’s languages. I offer a conclusion in Section 4.

1 The non-context-freeness of Romanian

1.1 Properties of Romanian *wh*-elements

In this section, I present the properties of Romanian required for the non-context-freeness proof. All discussion of potential complications is deferred to Section 1.3. The argument will be based upon the following three properties exhibited by the morphology of Romanian *wh*-words and their distribution in multiple *wh*-questions.

\[(1)\]
\[\text{a. Accusative and dative case is distinguished by overt morphology on *wh*-elements, and there exist object control verbs which subcategorize for accusative and dative DP complements.}\]
\[\text{b. Multiple *wh*-froniting of elements originating in different clauses is permitted.}\]
\[\text{c. The linear order of multiple fronted *wh*-elements may be identical to that of their selecting}\]

\(^2\) Modulo some poorly understood discourse and argument/adjunct asymmetries, the non-context-freeness result in Romanian can be extended along the lines suggested by Becker et al. (1992) to show Romanian is also not generable by a LCFRS. I will sketch the proof of this in Section 1.2. For the purposes of this paper, non-context-freeness will suffice, so I won’t spend much time investigating the ordering constraints on fronted *wh*-elements necessary to fully substantiate the non-LCFRS argument.
verbs.

1.1.1 Accusative and dative case morphology

Case on Romanian interrogative pronouns is distinguished with overt morphology. Nominative and accusative share the same basic form and may be differentiated by the obligatory presence of the preposition pe with accusative arguments. This preposition obligatorily pied-pipes with its complement in all cases of \(wh\)-extraction. Dative and genitive share the same form. The paradigm is given below for the interrogative pronouns that are relevant to the discussion.

<table>
<thead>
<tr>
<th>Nominative/Accusative</th>
<th>Dative/Genitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>who</td>
<td>cui</td>
</tr>
<tr>
<td>(pe) cine</td>
<td></td>
</tr>
<tr>
<td>which</td>
<td>căruia(m.sg)/carei(f.sg)/caror(pl.)</td>
</tr>
<tr>
<td>(pe) care</td>
<td></td>
</tr>
<tr>
<td>what</td>
<td>ce</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3 The pronoun pe appears with specific DPs that refer to humans, with pronouns, and with proper names (Farkas 1978 and Steriade 1980).

4 The pronoun ce, “what,” does not inflect for case and may not be used in situations requiring dative or genitive case. The pronoun căruia, “which,” inflects for number and gender in agreement with its complement in the dative/genitive.

5 In Romanian, like many other languages of the Balkans, obligatory control manifests in subjunctive complement clauses to traditional control verbs (Dohrovie-Sorin 1994, Dohrovie-Sorin 2000, Motapanyane 1995, Alboiu 2007, Alexiadou et al. 2010 for further discussion). Such complement clauses may be diagnosed by the presence of the subjunctive complementizer să.

6 In many of the examples below, a clitic associated with the complement of the verb remains attached to either the verb or the tense head. For the present discussion it suffices to note that the clitic is required in the presence of a specific, D-linked, or dative DP complement to the verb. The clitic agrees in number and gender with the associated verbal complement. The presence or absence of such clitics has no bearing on the non-context-freeness argument, so I leave this issue here (see, for example, Dohrovie-Sorin 1990 for more information).
“Which man did you convince to tell the story?”

c. Pe cine/*cui (l-)ai lăsat să viziteze Bucureștiul?
Who.ACC CL-have allowed to visit Bucharest
“Who did you allow to visit Bucharest?”

Among the verbs that subcategorize for dative complements are the object control verbs spune, ‘to tell’, zice, ‘to say, to tell’. Use of an accusative wh-element as the object of these verbs results in ungrammaticality.

(6) a. Cui/*pe cine i-ai spus că Ion a cumpărat brânză?
who.DAT/who.ACC CL-have.2SG told that Ion have.3SG bought cheese
“Who did you tell that John bought cheese?”

b. Cărei fete/*pe care i/l-ai zis că Ion a scris scrisoarea?
which.DAT girl.DAT/which.ACC girl.DAT CL.F/CL.M-have said that Ion have.3SG written letter
“Which girl did you tell that John wrote the letter?”

1.1.2 Multiple wh-fronting

Multiple wh-fronting is broadly permissible in Romanian. Matrix clause wh-arguments and adjuncts may appear together sentence initially, as in other multiple wh-fronting languages. There is no constraint on the number of such elements that may appear clause initially.

(7) a. Cine pe cine cum a prins?
who.NOM who.ACC how have.3SG caught?
“Who caught whom how?”

b. Cine unde ce a adus?
who.NOM where what has bought
“Who bought what where?”
(Bošković 2002: 369)

Multiple extraction is permitted in the case that the relevant wh-elements originate in an embedded clause (Rudin 1988: 452). Again, there is no limit on the number of wh-elements that may extract in such cases.

(8) a. Cine cui ce ziceai că i-a promis?
who.NOM who.DAT what say.2s.PAST that CL-have.3SG promised
“Who did you say promised what to whom?”
(Rudin 1988: 452)

b. Cine ce a spus Ion că a cumpărat?
who.NOM what have.3SG said Ion that have.3SG bought
“Who did John say bought what?”

Crucially for the non-context-freeness argument, Romanian also permits wh-elements from distinct clauses to extract to sentence initial position, at least in some cases (see Comorovski 1986: 175, Kraskow 1987, Rudin 1988: 452, for more discussion). Extraction is possible for multiple subjects (see (9)) and multiple

7 Two anonymous reviewers express doubt that this behavior is unique to Romanian. I agree that the acceptability of multiple wh-fronting of matrix clause arguments and adjuncts is a cross-linguistically widespread phenomenon, attested in at least Serbo-Croatian, Polish, Czech, Russian, Bulgarian, Hindi, and Malagasy (see Rudin 1988: 453ff. and Bošković 2002 for the Slavic languages and Bulgarian; see Mahajan 1990: 138f. for Hindi; see Sabel 2003, Potsdam 2006 for Malagasy). Malagasy differs from Romanian in that multiple wh-extraction is limited to two elements, which must always be an adjunct and an argument in that linear order (Sabel 2003; Potsdam 2006). Among the languages listed, only Bulgarian and Hindi allow multiple extractions from embedded clauses (Rudin 1988: 453ff., Mahajan 1990, Boris Harazinov p.c.). Among Hindi and Bulgarian, multiple wh-fronting of elements from different clauses seems to be ungrammatical in Bulgarian (Rudin 1988: 452 fn.7; Kraskow 1987; Boris Harizanov p.c.). Indeed, Bulgarian features a strict and pervasive clause-mate restriction on multiple A′-extraction, including multiple wh-movement, topic fronting, and negative fronting. Further investigation is needed to determine if Hindi patterns with Romanian in this regard, but the lack of an overt morphological distinction between accusative and dative case renders such a discussion irrelevant for present purposes, as I will show in the next section.
objects (see (10)) objects, with the latter generally rated as less acceptable (see Section 1.3). D-linking does not effect grammaticality (see (11)). There is some speaker variation in the acceptability of these examples, and those cases involving more embedded clauses and more wh-elements are consistently rated worse than corresponding examples with less structure. I return to issues of acceptability in Section 1.3, where I argue that the variation reported by speakers fits nicely into a well established experimental paradigm of wh-dependency processing.

(9) a. Cine pe cine a încercat sǎ prindǎ?
   who.NOM what.ACC have.3SG tried to catch
   “Who tried to catch whom?”

   b. Cine ce zicea cǎ isi închipuie cǎ ai descoperit?
   who.NOM what said.2SG that to himself imagines that have.2SG discovered
   “Who did you say imagines you discovered what?” (Comorovski 1986: 175)

(10) a. Pe cine ce ai convins sǎ cumpere?
    who.ACC what have.2SG convinced to buy
    “Who have you convinced to buy what?”

   b. ?Pe cine cui ai rugat sǎ-i spunǎ povestea?
    who.ACC who.DAT have.2SG asked-to-CL story
    “Who did you ask to tell who the story?”

(11) a. Pe care coleg cǎrui muzeu l-ai convins sǎ-i scrie scrisoarea?
    which.ACC colleague.ACC which.DAT museum.DAT have persuaded-to-CL write letter
    “Which colleague have you persuaded to write which museum the letter?”

   b. Pe care bǎiat cǎrei fete l-ai rugat sǎ-i scrie scrisarea?
    which.ACC boy.ACC which.DAT girl.DAT have.2SG asked-to-CL write letter
    “Which boy have you asked to write which girl the letter?”

1.1.3 Order of extraction & Superiority effects

Among sentences with multiply-extracted wh-elements, these elements may appear in the same linear order as their selecting verbs. This is on display in all of the examples thus far and is reported by speakers to be the most natural order.8 That said, the empirical picture is slightly more complicated. As Bošković (1997), Richards (1997) argue at length, while the wh-element corresponding to the hierarchically highest gap must come linearly first, a relatively freer order may obtain among the other elements. In the case of two extracted wh-elements, then, the order of the fronted wh-elements must match the order of the selecting verbs (see (12)). With three wh-elements, however, the lower two elements may take on either order with respect to each other (see (13)).

(12) a. Pe cine ce ai rugat sǎ scrie?
    who.ACC what have.2SG allow to write
    “Who have you allowed to write what?”

   b. *Ce pe cine ai rugat sǎ scrie?

(13) a. Cine pe cine cum a prins?
    who.NOM who.DAT how have.3SG caught
    “Who caught whom how?”

   b. ?Cine cum pe cine a prins?

With D-linked wh-phrases, the order requirement is even freer. Thus most speakers permit, at least in some cases, a totally free ordering among fronted wh-elements.

(De Hoop & Narasimhan 2005: 331, see also Pullum & Gazdar’s 1982 discussion of Dutch for an explanation as to why overt case matters).

8 Examples with adjuncts (see (7), (8)) complicate this picture slightly, but the non-context-freeness argument does not rely on adjuncts so I leave the issue here.
(14) a. Pe care băiat cărei fete l-ai rugat să-i scrie scrisarea?
Which.ACC boy.ACC which.DAT girl.DAT have.2SG asked to-CL write letter
“Which boy have you asked to write which girl the letter?”
b. Cărei fete pe care băiat l-ai rugat să-i scrie scrisarea?

The topic of exactly what principles factor into the available orders in a given context has been the subject
of much debate (cf. at least Rudin 1988; Billings & Rudin 1996; Bošković 1997, 2002; Richards 1997;
Grewendorf 2001), and appears to be influenced by the argument/adjunct status of the fronted phrases,
their salience in the discourse, and other factors. It is widely accepted, however, that the linear order
which reflects underlying hierarchical relations is always available, and that D-linked wh-phrases exhibit
considerably more freedom than their non-D-linked counterparts. I leave this issue here.9.

1.2 Account

The prerequisite information on Romanian syntax for the non-context-freeness argument is now in place.
To summarize briefly, Romanian distinguishes accusative and dative case on wh-elements with overt
morphology, and there exist ECM and object control verbs that subcategorize for only accusative or only
dative complement DPs. Moreover, Romanian allows multiple wh-elements to extract from distinct clauses,
and the linear order of such extracted elements may be the same as the linear order of their selecting verbs.
Such extraction patterns obtain with both non-D-linked and D-linked wh-elements.

It follows that by repeatedly embedding object control verbs which select for accusative or dative DP
complements, and by making these DP complements wh-elements, it is possible to construct grammatical
sentences where there are arbitrarily many extracted wh-elements and where the number of extracted
accusative and dative wh-elements exactly matches the number of verbs that subcategorize for accusative
and dative complements, respectively. I provide the general structure in (15), followed by two explicit
examples in Romanian.

(15) wh-element.ACCm wh-element.DATn you verb1.ACCm verb2.DATn DP.

(16) a. ?Pe cine cui ai rugat să-i spună povestea?
Who.ACC who.DAT have.2SG asked to-CL story
“Who did you ask to tell who the story?”
b. Pe care coleg cărui muzeu l-ai convins să-i scrie scrisoarea?
which.ACC colleague.ACC which.DAT museum.DAT have persuaded to-CL write letter
“Which colleague have you persuaded to write which museum the letter?”

Because ruga and spune subcategorize for only accusative and only dative complements, respectively, the
number accusative and dative marked wh-elements must exactly match the number of accusative and dative
subcategorizing verbs, respectively.

At this point, denote the language consisting of all of the strings of Romanian by ROMANIAN. Consider
the following set of sentences, which according to the discussion above is a subset of ROMANIAN:

(17) (Pe cine)m+1 cui n ai rugat (să rogăm)n (să spună)n povestea?
who.ACCm who.DATn have.2SG asked (to ask)m (to tell)n story
“Who have you asked to allow who . . . to tell who to tell who the story?”

Consider the following homomorphism, where w is an arbitrary string:

9 I should note that in some contexts, it may be acceptable to leave some wh-elements in situ. That said, full extraction
appears, modulo island effects, to always be possible, as the examples above make clear. As I will show in Section 1.2,
the non-context-freeness argument depends only on the full wh-extraction examples being among the set of grammatical
alternatives, not it being exhaustive of such a set. See Pesetsky 1987, Comorovski 1996, Bošković 2002 for more detailed
discussion.
Given the regular language \( L = a^*b^*c^*d^* \), and given that Romanian admits as grammatical the sentences represented in (17) with arbitrarily many of each particular verb–wh-element pair, the intersection of \( f(\text{romanian}) \) and \( L \) results in the following language:

(19) \[ L' = L \cap f(\text{romanian}) = a^{m+1}b^nc^md^n \]

This language \( L' \) can be shown by a pumping lemma argument not to be context free (Hopcroft & Ullman 1979: 128). Therefore the fact that context free languages are closed under homomorphism and intersection with regular languages entails that Romanian is not a context-free language.

As mentioned above, if we grant that arbitrary ordering obtains among multiply fronted wh-elements, it is possible to extend the non-context-freeness argument to show that Romanian is not generable by a linear context-free re-writing system. Becker et al. (1992) show that the following language has this property:

(20) \[ SCR = \{ \sigma(n^{[0]}, \ldots, n^{[m]}), v^{[0]} \ldots v^{[m]} : m \geq 0 \text{ and } \sigma \text{ a permutation} \}, \text{ where } n^{[i]} \text{ is an argument of } v^{[i]} \]

Again, assuming that at least D-linked, fronted wh-elements may take on any order, the subset of Romanian involving such sentences is exactly equivalent to \( SCR \), modulo a suitable homomorphism.\(^{10}\)

1.3 The acceptability of arbitrary wh-extraction

For the argument thus presented to obtain, there must be no bound on the number of embedded infinitival clauses permitted in Romanian, nor on the number of wh-elements that may be extracted from the object position of these clauses. In this section I argue that while more complex data are rated as less than acceptable, the degradation in acceptability is due wholly to factors in the performance system, so that such data is grammatical. Because my argument makes a claim about the set of grammatical strings of Romanian, not those that are easy to process, the argument goes through. Readers willing to accept this claim \textit{prima facie} can move on to Section 2.

I assume that the main claim outlined by Miller & Chomsky (1963) and refined in the processing literature since is correct: the acceptability of a given sentence is governed both by whether that sentence

\(^{10}\) The possibility in multiple wh-fronting examples of i) leaving some wh-elements in \textit{situ}, and ii) permuting the order of fronted wh-elements does not affect the original non-context-freeness argument. The availability of these two options means that the image of \text{romanian} under the homomorphism in (18) contains the following string set:

(21) \[ S = wxy \]

a. \( w \) contains \( i \) occurrences of \( a \) and \( j \) occurrences of \( b \) in any order (reflecting the ability of extracted wh-elements to take on other orders)

b. \( x \) contains \( i \) occurrences of \( c \) alternating with \( m \) occurrences of \( c \) (reflecting accusative wh-elements in \textit{situ})

c. \( y \) contains \( j \) occurrences of \( d \) alternating with \( n \) occurrences of \( d \) (reflecting dative wh-elements in \textit{situ})

Note, however, the strings that do not conform to the explicit condition that all wh-elements be extracted and all accusative elements precede all dative elements are not contained in the language \( L = a^*b^*c^*d^* \). In other words, \( S \cap L = a^{m+1}b^nc^md^n \). All of these strings are therefore ruled out from consideration via intersection with \( L \) and hence are unproblematic for the argument.
is grammatical and by the degree to which it taxes the processing system. Grammatical sentences may thus be judged as unacceptable, as with the canonical example from English center-embedding (Chomsky 1957, Partee et al. 1993: 477-480).

(22) ??The man\textsubscript{1} [the boy\textsubscript{2} [the girl liked e\textsubscript{2} was friends with e\textsubscript{1}]] gave a speech.

With this in mind, I turn to the task of isolating those factors that lead to the relative unacceptability of increasingly complex Romanian \textit{wh}-extraction examples. I argue that these factors are purely performance based, so that more complex examples are grammatical, if not wholly acceptable.

1.3.1 Implicating performance factors

The examples (9), (10), (11) are accepted as grammatical, if stilted, by even my most caviling consultants. I take as a point of departure the claim that extraction of distinct \textit{wh}-elements from two clauses is grammatical. That said, speakers universally have more trouble with sentences involving three clauses and three corresponding extracted \textit{wh}-elements.

[(23) a. Cine pe cine ce a vrut să roage să cumpere?
   who.NOM who.ACC what have.3SG wanted to ask to buy
   “Who wanted to ask who to buy what?”
   b. Pe cine cui ce ai rugat să-i spună să cumpere?
   who.ACC who.DAT what have.2SG asked to-CL tell to buy
   “Who did you ask to tell who to buy what?”
   c. *Pe cine cui pe cine ai rugat să-i spună să lase să viziteze Bucureștiul?
   who.ACC who.DAT who.ACC have.2SG asked to-CL tell to allow to visit Bucharest
   “Who did you ask to tell who to allow who to visit Bucharest?”]

The validity of the non-context-freeness argument therefore depends on the attribution of this deterioration in acceptability to extra-grammatical phenomena. It is well known in the processing literature that there is an inverse relationship between ratings of acceptability and structural complexity, even with examples that are canonically accepted as grammatical. Indeed there are at least four factors that have been independently linked to degraded acceptability that come to bear on the current data, thus lending credence to an extra-grammatical treatment of the degraded acceptability above. In the ensuing discussion, I refer to those data involving extraction of two \textit{wh}-elements from two clauses as two-extraction cases, and those involving extraction of three \textit{wh}-elements from three clauses as three-extraction cases.

To begin, Kluender & Kutas (1993) and Kluender (1998) have shown that \textit{wh}-questions lead to lower acceptability scores (and significantly longer reading times and differential ERP responses) as compared to analogous yes/no questions. All of the data included here are thus at an immediate disadvantage from an acceptability standpoint. Second, building on the work of Alexopoulou & Keller (2003), Hofmeister et al. (2007) show that at least in some cases, when bare \textit{wh}-elements intervene between a \textit{wh}-filler and its associated gap, acceptability decreases. Taking this result at face value, the sheer number of fronted \textit{wh}-elements may thus degrade the acceptability of the three-extraction examples compared with corresponding two-extraction cases.

Third, grammatical \textit{wh}-questions with a dependency involving an object-gap have been shown to be rated as significantly less acceptable than comparable questions with subject-gaps (Kluender & Kutas 1993; Kluender & Cowles 1997). This result has been replicated in studies based on behavioral and neuro-imaging experiments in language families as diverse as Indo-European, Mayan, and Austronesian (see Kwon et al. 2010, Clemens et al. 2015, Longenbaugh & Polinsky to appear for an overview). Indeed, across all of my examples, those with more \textit{wh}-objects tend to be reported as worse than similar examples with fewer (cf. (23a) vs. (23c); (9) vs. (10), (11); (10), (11) vs. (23b), (23c)). That this effect is at work in the degradation
of the data in (23), specifically in (23a) versus (23c), can be seen by comparing the simpler data below, which minimally modifies these examples to eliminate the second level of embedding. While both examples are accepted by my informants, (24b) is rated as significantly worse.

(24) a. Cine cui ce a încercat să-i trimită?
   who.NOM who.DAT what have.2SG tried to-CL send
   “Who tried to send what to whom?”

b. ?Pe cine cui ce ai convins să-i trimită?
   who.NOM who.DAT what have.2SG convinced to send
   “Who did you convince to send what to whom?”

The difference among these two examples is exactly analogous to the variation between (23a) and (23c) – (24a) involves double-extraction past a matrix subject and (24b) double-extraction past a matrix object. This effect is likely responsible for the difficulty, pointed out by an anonymous reviewer, associated with even some two-extraction examples involving the extraction of two objects, as in (10a).

Finally, canonically grammatical where-dependencies are rated as increasingly unacceptable as the number of embeddings they span increases (see Alexopoulou & Keller 2007 on English, Beltrama 2014 on Italian, and Xiang & Cui 2015 on Chinese, where the effect obtains even though there is no overt wh-displacement). Again, this comes to bear directly on the variable acceptability between the two and three extraction examples, which differ in the number of embedded clauses being extracted out of. To corroborate that this effect is in play in Romanian, and thus degrades the three-extraction examples, I again consider a simpler case: when the grammatical two-extraction example (10b) is embedded under another verb, acceptability is notably degraded. A similar result obtains in the case of the D-linked example in (11b).

(25) a. Pe cine cui ai rugat să-i spună povestea?
   who.ACC who.DAT have.2SG asked to-CL story
   “Who did you ask to tell who the story?”

b. ?Pe cine cui ai vrut să-l rogi să-i scrie scrisoarea?
   who.ACC who.DAT have.2SG wanted to ask to-CL write letter
   “Who have you wanted to ask to write the letter to who?”

Given these results, the degraded acceptability of the examples in (23) compared to the two-extraction examples is not surprising. Following the order of factors above, there are wh-questions involving multiple sentence initial wh-elements in which the relevant dependencies all involve object gaps that in some cases span three clauses.

While it is therefore clear that performance based factors intervene to degrade the three-extraction examples, this fact alone does not itself exclude the existence of some grammatical constraint against extracting three wh-elements from three separate clauses. Thus, to further substantiate the claim that there is no such restriction, let us suppose temporarily that the difference in acceptability between the acceptable (9), (10), and (11) and the three-extraction examples in (23) is because the latter are ungrammatical. If this is correct, it should be possible to isolate the source of such ungrammaticality by examining how the two and three extraction sentences differ syntactically.

The first major difference, the number of fronted wh-elements, can be easily dismissed as a source of ungrammaticality: the grammatical examples in (7b) and (8a) have three sentence initial wh-elements in which the relevant dependencies all involve object gaps that in some cases span three clauses.

While it is therefore clear that performance based factors intervene to degrade the three-extraction examples, this fact alone does not itself exclude the existence of some grammatical constraint against extracting three wh-elements from three separate clauses. Thus, to further substantiate the claim that there is no such restriction, let us suppose temporarily that the difference in acceptability between the acceptable (9), (10), and (11) and the three-extraction examples in (23) is because the latter are ungrammatical. If this is correct, it should be possible to isolate the source of such ungrammaticality by examining how the two and three extraction sentences differ syntactically.

The first major difference, the number of fronted wh-elements, can be easily dismissed as a source of ungrammaticality: the grammatical examples in (7b) and (8a) have three sentence initial wh-elements, so that there is no intrinsic restriction against such a configuration. That said, the wh-elements in all of the acceptable examples are case distinct, something that does not hold in (23). To rule this out as a source of ungrammaticality, note that it is perfectly acceptable for two accusative or two dative wh-elements to appear sentence initially, as (27a) and (27b) demonstrate. A similar result obtains when the wh-elements

11 I’d like to thank an anonymous reviewer for pointing this out.
are D-linked. I thus dismiss any effect of the number of fronted *wh*-elements.\(^\text{12}\)

\[(27)\] a. Pe cine ce ai rugat să scrie?  
who.NOM who.ACC what.ACC have asked to write  
“Who have you asked to write what?”

b. ?Cui carui muzeu i-ai spus să doneze tabloul?  
who.DAT which.DAT museum.DAT have told to donate painting  
“Who have you told to donate the painting to which museum?”

Beyond the sheer number of sentence initial *wh*-elements, the other major syntactic difference among the two-extraction examples and those in (23) is that the latter involve three *wh*-extractions from three distinct clauses. It may be possible, then, that there is simply some arbitrary bound on the number of such extractions that may obtain or the number of embeddings from which extraction may take place. Such a view would be at odds to explain why there are three-extraction examples, like (23a), repeated below, and (28b), that are rated as more or less acceptable by most speakers.

\[(28)\] a. ?Cine pe cine ce a vrut să roage să cumpere?  
who.NOM who.ACC what.3SG wanted to ask to buy  
“Who wanted to ask who to buy what?”

b. ?Cine pe cine cui a vrut să roage să-i spună povestea?  
who.NOM who.ACC who.DAT have.3SG wanted to ask to-cl tell story  
“Who wanted to ask who to tell who the story?”

If three-extraction is generally grammatical, however, the contrast between the relatively acceptable (23a), (28b) and the worse (23b), (23c) falls out immediately from the performance based factors implicated above: the more acceptable examples involve the extraction of one subject and two objects, a configuration predicted to be easier on the basis of well established cross-linguistic patterns, as opposed to three objects. Further bolstering this case is that fact that the three extraction examples in (23a) and (28b) are significantly better than truly ungrammatical data involving case-mismatch, agreement violations, and strong-island violations (cf. (29a), (29b), (30)).

\[(29)\] a. Cui/*Pe cine ai spus povestea?  
who.DAT/who.ACC tell.2S.PAST story  
“Who did you tell the story to?”

b. Ion a/*au ajuns  
Ion have.3SG/have.3PL arrived  
“John told Maria to visit Bucharest.”

\[(30)\] a. *Pe cine ai plecat înainte să-l examineze profesorul?  
who.ACC have.2SG left before to-cl examine doctor  

\(^\text{12}\) Relatedly, Bošković (2002) reports that there is a “low-level phonological” constraint against repeated identical *wh*-elements in Romanian. This restriction is strong enough to sometimes override multiple *wh*-fronting that would otherwise be strictly required (although as Bošković (2002: 365: fn.23) mentions in a footnote, consultants do not uniformly agree on the presence of such a constraint and there are exceptions).

\[(26)\] a. Ce ce precede?  
what precedes what  
“What precedes what”

b. ??Ce ce precede?  
(Bošković 2002)

The constraint does not come to bear on the acceptability of the data in (23), where none of the examples have repeated identical *wh*-elements. However, if the restriction is a strict grammatical constraint, the non-context-freeness argument does not obtain in its current form. Based on the responses from my consultants, I do not consider the restriction to be this strong. Even if it is, repeating D-linked elements is grammatical with or without the anti-homophony constraint (cf. ??, ??). The argument can thus be saved by substituting the D-linked expressions pe care baiat, “which.ACC boy.ACC” and carei fete, “which.DAT girl.DAT” for pe cine and cui, respectively, in examples like (17), and appropriately modifying the homomorphism.
“Who did you leave before the doctor examined?”

b. *Pe cine ai întiluit fata care l-a vazut anul trecut la cinema?
who.ACC have.2SG met girl REL CL-have.3SG seen year last the cinema
“Who did you meet the girl who saw e last year at the cinema?”

Finally, it is often assumed that performance factors can also improve ungrammatical examples, with
the obvious example being the case of resumptive pronouns in strong-islands. No such recourse can be
made here: even the relatively acceptable three-extraction examples instantiate patterns well known to
induce degraded acceptability. Proponents of the ungrammaticality account thus face the double burden of
explaining how the purportedly ungrammatical (23a), (28b) come to be rated as mostly acceptable and
how they do so in the presence of factors independently shown to induce severe performance-based acceptability
degradation.

To summarize, I have independently considered the effects of increasing numbers of clause-initial
wh-elements, of extraction from increasing levels of embedding, and of subject versus object wh-extraction
in the context of multiple wh-extraction. In each case, I showed that as the number of relevant operations
increases, acceptability decreases, but that such degradation is related to performance based factors
previously identified in the literature. Just as with the case of English center-embedding, it is always
possible to decree that this or that operation is limited to one or two iterations per sentence. Here, this
amounts to allowing wh-dependencies between two but not three (or more) clause initial wh-elements
and gaps in two but not three successively embedded clauses. Such a position not only faces a severe
encumbrance in explaining the relatively acceptable three-extraction data, but also imposes an artificial
and arbitrary bound that is wholly uncharacteristic of the language faculty as we know it. Given the force
of the arguments thus stated, and in the absence of a compelling theoretical argument to the contrary, I
therefore conclude that the degradation of increasingly complex instances of multiple wh-extraction is due
to limitations of the performance system and strong theoretical evidence that it is not due to the violation
of a grammatical constraint.

2 Deriving crossed-dependencies: multiple-attraction and tucking-in

Having established the basic veracity of the Romanian data, let us now turn our attention to outlining the
factors implicated in the derivation of the non-context-free crossed dependencies. I will begin with the case
of Romanian, where the details and operation of the analysis are clearest, then extended the result to other
cases. The pattern that will emerge is that in each case where it is possible to robustly implicate purely
syntactic factors in the derivation of the crossed-dependencies, the same essential mechanism appears to be
at work: multiple-attraction by a single functional head, and corresponding tucking-in of multiple specifiers.

2.1 Romanian multiple wh-fronting and tucking-in

There are at least two competing theories of multiple wh-dependencies, those of Richards (1997) and of
Grewendorf (2001), Sabel & Wolfgang (2001). Richards’ (1997) theory, which is based on multiple-attraction
and tucking-in, makes it possible to draw generalizations that cross-cut the various data sets we will be
considering, so that I will be working in this theory. I consider the implications of using Grewendorf’s

Richards’ (1997) theory is couched in a derivational, functional-head driven approach to syntax whereby
certain heads (V, v, T, C) are endowed with special properties, encapsulated in features, that trigger other
elements in the syntax to move into their local domain. Let us assume that this attraction is mediated
via an operation, agree, that relates a feature on a head with a corresponding feature on the head of a
phrase in its c-command domain. Thus movement of a phrase to the specifier of a head may only occur if
agree has obtained between the head and the phrase (or the head of the phrase). To be slightly more
explicit, I will assume, following Pesetsky & Torrego (2007), that features vary in two dimensions: valued vs. unvalued and interpretable vs. uninterpretable. AGREE is then driven by the need to provide a value to features that lack one, and is triggered by an unvalued feature which probes its c-command domain in search of a counterpart feature, a goal. The lower feature may itself be valued, in which case the value is copied to the higher head, or unvalued, in which case the two heads are linked in a chain such that a later valuation of either feature leads to valuation of both (Pesetsky & Torrego 2007). In the course of the syntactic derivation, all unvalued features must be valued, and all uninterpretable features must be deleted, with valuation a precondition on deletion. All derivational processes are thus driven by the need to value and delete features. In such a system, wh-movement is mediated via unvalued wh-features on C, which probe for and agree with valued counterparts on wh-elements in their c-command domain.\footnote{The [wh] feature on the highest C in the sentence is said to be interpretable, as it serves to encode that the utterance is a question. All intermediate C and v heads through which wh-movement passes can be thought of as having uninterpretable, unvalued features that can probe for the valued wh-elements but are not interpreted as introducing a separate question.}

Richards (1997) theory builds on these basic underpinnings by allowing functional heads to be endowed with two or more copies of a given feature. Under the assumption that each copy needs to undergo AGREE with a different corresponding feature, a single head can thus attract multiple phrases into its local domain. Richards (1997) proposes to derive Romanian multiple wh-fronting in exactly this way: in languages like Romanian and Bulgarian, C contains as many copies of a [wh] feature as there are wh-elements in the clause, and each wh-element is attracted overtly to Spec(CP) via agreement between C and the relevant phrase. Beyond this notion, of which versions have been in circulation since the multiple wh-fronting data was first brought to the attention of the field at large (cf. Rudin (1988)), the major innovation in Richards’s (1997) theory is the way in which it derives the mandatory ordering requirement on fronted elements. As we saw in Section 1.1, setting aside D-linked cases, the hierarchically highest wh-element must occupy the outermost specifier of the attracting head. In other words, if there are two wh-elements, the dependencies between the moved phrases and their associated gaps must cross, not nest. For Richards (1997), this follows from two basic, independently motivated principles of economy: attract closest and shortest move. Attract closest, which is just a version of Rizzi’s (1990) Relativized Minimality or Chomsky’s (1995) Minimal Link Condition, states that an unvalued feature on a given functional head must agree with the closest copy, where closeness is defined as in (33), of the associated feature in its c-command domain. Shortest move dictates that movement of a phrase to the specifier of a particular head must be as short as possible. In the standard case where a head attracts one phrase to its specifier, this constraint has no effect; however, if a phrase is undergoing movement to the specifier of a head that has already been the target of a previous movement step shortest move dictates that the second movement tuck-in under the first specifier. To see exactly what this means, consider the two derivations sketched below. Here H is a functional head bearing two copies of unvalued feature $f$ that attracts XP$_1$, XP$_2$ also bearing feature $f$.

\begin{align*}
\text{(31) Crossed-dependencies:} \\
a. & [\underbrace{\text{HP} \text{XP}_1, [f_{7}]}_{\text{HP} \text{XP}_2, [f_{8}]} [\underbrace{\text{H}, [f_{7}, s]}_{\text{YP}, \ldots} [\underbrace{\ell_{1}, \ldots} [\underbrace{z_{p}, \ldots} e_{2}]])])]
\end{align*}

\begin{align*}
\text{(32) Nested-dependencies:} \\
a. & [\underbrace{\text{HP} \text{XP}_2, [f_{8}]}_{\text{HP} \text{XP}_1, [f_{7}]} [\underbrace{\text{H}, [f_{7}, s]}_{\text{YP}, \ldots} [\underbrace{\ell_{1}, \ldots} [\underbrace{z_{p}, \ldots} e_{2}]])])]
\end{align*}

Attract closest mandates that AGREE obtain first between H and XP$_1$, which then moves to Spec(HP). Because H has a second copy of $f$, AGREE can obtain again, this time with XP$_2$. Assuming XP$_2$ then moves to Spec(HP), it can either tuck-in below XP$_1$ (see (31)) or merge above it (see (32)). Shortest move intervenes in this case in favor of the tucking-in option, which involves a shorter move, and is

13 The [wh] feature on the highest C in the sentence is said to be interpretable, as it serves to encode that the utterance is a question. All intermediate C and v heads through which wh-movement passes can be thought of as having uninterpretable, unvalued features that can probe for the valued wh-elements but are not interpreted as introducing a separate question.
thus preferred, *ceteris paribus*. This argument hinges on a notion of distance defined in terms of strict hierarchical dominance, as in (33) below. With this definition, the inner specifier of a head will always be the closest such specifier to all elements c-commanded by that head.

(33) **Closeness:**
Given nodes WP, XP, YP, where WP c-commands XP and YP, XP is closer to WP than YP iff there is some node ZP dominating YP that does not dominate XP.

Crucial for the ensuing discussion will be the behavior of multiple specifiers under additional attraction steps. The notion of distance employed in the above argument entails that the outer-most specifier of a given head is closest to all c-commanding functional heads. Thus, if a higher functional head bearing feature $f$ merges in the structure depicted in (31), ATTRACT CLOSEST will mandate that the element in the outer-most specifier, in this case XP$_1$, is agreed with and attracted before any other specifiers. As such, further instances of attraction will always preserve the crossed nature of the dependencies between XP$_1$ and its base position and XP$_2$ and its base position. This is true if the second attraction step involves single or multiple attraction: if the relevant attracting head has only one copy of the feature in question, the outermost specifier will be agreed with and attracted, leaving the other specifiers in place; if the attracting head has multiple copies of the relevant feature, the original inner specifier(s) will once again be tucked-in under the original outer-most specifier. I represent schematically the result of a second step of multiple- and single-attraction, respectively, on the structure in (31).

(34) Two multiple-attractions:

$$[H \ [HP \ XP_1 \ [H' \ [WP \ \ldots \ [HP \ \epsilon_1' \ [HP \ \epsilon_2' \ [H \ [YP \ \ldots \ [ZP \ \ldots \ [XP_2 \ [H \ldots \ [YP \ \ldots \ [ZP \ \ldots \ \epsilon_2 \ ]]])]]]]]]]$$

(35) Single-attraction following multiple-attraction:

$$[H \ [HP \ XP_1 \ [H' \ [WP \ \ldots \ [HP \ \epsilon_1' \ [HP \ \epsilon_2' \ [H \ [YP \ \ldots \ [ZP \ \ldots \ \epsilon_2 \ ]]])]]]]]$$

As alluded to above, Romanian multiple *wh*-movement is easily captured in this system. Each C head along the path of *wh*-movement is endowed with enough features to attract all of the *wh*-elements in its c-command domain, and ATTRACT CLOSEST and SHORTEST MOVE ensure the relevant dependency paths cross, not nest. A question with two matrix *wh*-elements is thus derived as in (31). If the *wh*-elements are more deeply embedded, attraction obtains along the lines of (34).

We have so far limited discussion to instances involving attraction of two phrases, deriving the empirical observable pattern that the hierarchically superior attracted element must occupy the outer specifier, and the inferior element the inner specifier. This same, strict ordering requirement does not appear empirically to obtain, at least in Romanian, if there are more than two attracted elements: when three or more *wh*-elements are involved, the ordering requirement is more flexible among the *wh*-elements occupying the inner specifiers. The theory does not yet predict this: the interaction of ATTRACT CLOSEST and SHORTEST MOVE should force crossed-dependencies in these cases as well, resulting in the same strict ordering requirement we observed above. To avoid this incorrect prediction, Richards (1997) invokes the so-called Principle of Minimal Compliance (Richards 1997, Richards 1998).

(36) **Principle of Minimal Compliance:**
For any dependency D that obeys constraint C, any elements that are relevant for determining whether D obeys C can be ignored for the rest of the derivation for purposes of determining whether any other dependency $D'$ obeys C.
The relevance of this principle for the present discussion is that it allows us to ignore ATTRACT CLOSEST for all agreement steps beyond the first. In other words, once the attracting head with unvalued feature f has satisfied ATTRACT CLOSEST by agreeing with the closest relevant candidate bearing feature f, all further instances of agreement with elements bearing feature f may violate ATTRACT CLOSEST (Richards 1997: 295ff.). The relatively freer order among the inner specifiers thus follows from the ability to attract lower phrases bearing feature f in any order.

(37) \[
[H_\text{P} \quad \text{XP}_1 \quad [H_\text{P} \quad \text{XP}_2 \quad [H_\text{P} \quad \text{XP}_3 \quad [H \quad [\text{WP} \quad \ldots \quad \text{e}_1 \quad [\text{VP} \quad \text{e}_2 \quad \ldots \quad [\text{ZP} \quad \ldots \quad \text{e}_3 \quad \ldots \quad ]]]]]]]
\]

(38) \[
[H_\text{P} \quad \text{XP}_1 \quad [H_\text{P} \quad \text{XP}_3 \quad [H_\text{P} \quad \text{XP}_2 \quad [H \quad [\text{WP} \quad \ldots \quad \text{e}_1 \quad [\text{VP} \quad \text{e}_2 \quad \ldots \quad [\text{ZP} \quad \ldots \quad \text{e}_3 \quad \ldots \quad ]]]]]]]
\]

Richards (1998) argues in depth that the PMC can be used to explain a variety of cross-linguistic phenomena, so that the principle is motivated independent of these data. Because the alternative orderings are of only peripheral importance to the overall discussion, I leave this here, referring interested readers to the original works for more details.

So far we have seen that by allowing complementizers to bear multiple unvalued copies of [wh] features, we can correctly predict the presence of crossing paths by making use only of independently motivated principles of economy. I have kept the discussion as general as possible, however, to emphasize that this theory and its predictions are not limited to wh-movement: indeed any instance of multiple-attraction is predicted to behave as does Romanian wh-fronting. I would now like to investigate several other instances of non-context-free crossed-dependencies in the literature to see if they, too, can be plausibly derived via a multiple-attraction step.

2.2 Extraction from embedded questions in Swedish

The non-context-free data in Swedish (Miller 1991) bears the closest resemblance to the data from Romanian, so we’ll begin our investigation here. Swedish – and the Scandinavian languages more generally – permits relativization and questioning out of embedded questions.

(39) a. Vilken lingvist1 frågade läraren dig vilket språk2 han1 studerade e2?
   which linguist asked teacher you which language he studied
   “Which linguist asked you which language he studied?”

b. Vem1 frågade läraren dig vad2 han1 studerade e2?
   who asked teacher you what he studied
   “Who asked teacher you what he studied?”

As seen in (39), the dependencies engendered by such extractions may cross, contingent on the appearance of resumptive pronouns in accordance with the following constraint.

(40) Given a gap G1 and its filler F1, G1 must be realized as a resumptive pronoun if there is a gap G2 following G1 such that the filler F2 of G2 follows F1.
    (Miller 1991)

The constraint effectively states that a crossed-dependency reading in a sequence of filler-gap relations is dependent on resumptive pronouns occupying all but the final gap. In (39), this has the effect of forcing a resumptive pronoun in the gap associated with the sentence initial wh-element. Miller (1991) extends the pattern in (39) arbitrarily, producing the following paradigm.
(41) a. Here is [the NP, that I wonder [wh2 John wonders [wh3 Bill wonders [wh4 pro1 thought [pro2 thought [pro3 had recommended e4 to the students]]]]]]
b. NP1 wh2 ... whn whn+1 pro1 pro2 ... pro2 en+1

Critically, (40) requires that all the gaps subsequent the final appearance of “wonders” be filled with resumptive pronouns: the empty gap following “recommended” serves as the $G_2$ for all the other gaps. The resulting pattern, schematized in (41b), along with the mandatory number and gender agreement between the filler and the resumptive pronoun in the associated gap, suffice to push Swedish beyond the weak generative capacity of context-free grammars.\(^\text{14}\)

On the surface, it is not at all clear that the Swedish involves multiple-attraction. Whereas in Romanian the relevant attracted elements all accumulate at the clause edge, the relevant Swedish wh-elements show up as the single specifier of different C heads throughout the sentence. These patterns are given schematically below.

(42) Romanian
a. wh1 wh2 wh3 ...[CP ...e1 ...[CP ...e2 ...[CP ...e3 ]]]

(43) Swedish
a. wh1 [CP ...[CP wh2 ...[CP wh3 ...e1 ...e2 ...e3]]]

Recall from the previous section that the definition of distance we employed had as a consequence the fact that outer-specifiers are closer to higher functional heads than inner specifiers. Consequently, higher attractors are predicted to attract the outer-specifier before any inner specifiers. Armed with this theoretical consequence, it becomes plausible to analyze extraction from embedded questions as involving multiple-attraction. Specifically, we can view such extractions as involving an initial step of multiple-attraction, which creates a series of layered specifiers as in Romanian. Unlike in the Romanian cases discussed so far, however, in Swedish ensuing attraction steps undo the layered specifier structure, so to speak, leaving behind one wh-element in each embedded clause. This can be seen to follow as a consequence of the fact that the embedded C’s in the Swedish examples are themselves endowed with an interpretable [wh] feature, given that they denote a question, and are thus licensed/required to host an overt wh-element in their specifier. A derivation along these lines is represented abstractly in (35), where multiple-attraction feeds an ensuing step of single-attraction that removes the embedded outer-, but not inner-, specifier. I provide a schematic for the derivation of ?? below.

(44) Who1 did the teacher ask what2 he1 studied e2?
a. [CP who1 [C [TP teacher ask [CP e1 [CP what2 [C [TP e’ studied e2]]]]]]]

This analysis crucially depends on allowing C to host multiple specifiers. Structures as in (35) are at therefore at least in principle possible in languages like Bulgarian and Romanian, which overtly allow multiple specifiers of C. In English, where it is not obvious that this is true, corresponding examples are predicted to be bad. This is indeed borne out.

(45) Bulgarian

\(^{14}\) The non-context-freeness proof hinges on the fact that the number of masculine, feminine, and plural resumptive pronouns must exactly match the number of corresponding masculine, feminine, and plural wh-elements. I refer the interested reader to Miller (1991) for the proof.
a. Koj se opitvat da razberat kogo e1 e ubil e2?
who self try to find out whom AUX killed

“Who did he try to find out whom killed?” (Richards 1997: 295)

(46) a. *The boy who you didn’t realize which game played...
b. *Which man did you ask what ate?

Unlike Romanian and Bulgarian, Swedish, and the Scandinavian languages more generally, do not allow multiple overt specifiers of C. Faced with this confound, there are at least two paths forward. Richards (1997), who also assimilated the Scandinavian data into his framework, pursues the first in proposing that the relevant multiple-attraction step is not triggered by C, but rather by the same functional head that is responsible for the object-shift phenomena in these languages. Under this analysis, clause-mate wh-arguments move first in a multiple-attraction fashion complete with the tucking-in of the second argument. I demonstrate the relevant dependencies for (48b) below, denoting the abstract functional head responsible for object-shift as H.

(47) [CP wh1 [C [ asked teacher you [CP wh2 [C [HP e′1 [HP e′2 [H [e1 killed e2 ]]]]]]]]]

This account has several theoretical and empirical drawbacks. On the empirical side, note that only pronouns can undergo object-shift in Swedish. Richards (1997) suggests that perhaps D-linked wh-phrases, owing to their discourse salience, are also licensed to participate. But as (39) makes clear, non-D-linked wh-phrases are clearly licensed in the construction. Second, given that object-shift is clause-bounded, this sort of extraction should not be licensed for non-clause-mate wh-elements, again counter to fact ((48b) is from Norwegian, which behaves identically to Swedish with respect to object shift).

(48) a. Det var Lisa1 läraren frågade vem2 vi trodde hon1 ville (att) Johan skulle hälsa on “It was Lisa1 that the teacher asked who2 we thought she1 wanted John to visit.”
b. Det er politiamannen1 som jeg lurer på hvilke piker2 dommeren villa hvilke droger3 han1 trodde de2 hadde solgt e3 til barne.

“This is the policeman1 that I wonder which girls2 the judge will want to know which drugs3 he1 thought they2 had sold to the children3.” (Maling & Zaenen 1982: 237)

On the theoretical side, the derivation in (47) requires a rejection of successive cyclicity, among other dubious claims related to the timing of wh-movement.15

---

15 Richards (p.c.) suggests that Swedish and Bulgarian differ in that crossed dependencies are allowed as a marked option in Swedish, but that they are preferred and in some cases required in the Bulgarian. The object shift analysis above can potentially attribute this fact to the general observation that non-subject arguments appear to be more accessible to a variety of operations in object-shift languages (see at least Reinhart 1979, Comorovski 1986, Rudin 1988, Ura 1995, Richards 1997). That said, Engdahl (1982, 1986) reports based on experimental data that in Scandinavian, whether a particular sentence gets a nested or crossed interpretation is strongly tied to the context of utterance. While some Bulgarian speakers reject nested readings in some cases involving non-D-linked wh-elements (Richards 1997: 41), in certain contexts it is possible to get this effect in Swedish too, at least among some of my informants.

(49) a. *?Vad2 frågade läraren dig vem3 e2 studerade e1?
what asked teacher the you who studied

“What1 did the teacher ask you who2 studied?”
b. Vem1 frågade läraren dig vad2 e1/han1 studerade e2?

Moreover, in both sets of languages, dependencies involving D-linked wh-elements tend to permit both nested and crossed interpretations (see Richards 1997: 40 for Bulgarian, Engdahl 1982: 169 for Swedish).
In light of these shortcomings, the second approach to handling the Swedish data has at least the benefits of empirical coverage and theoretical simplicity. Specifically, the relevant data can be derived by allowing multiple specifiers of C in Swedish so long as the phrases occupying these specifiers are not both the heads of their respective movement chains. This proposal has the effect of allowing configurations as in (35) – and hence derivations along the lines of (35) – but forbidding overt multiple wh-movement. This proposal can be seen as a generalization of the assumptions that underly successive cyclic wh-movement: in English, for example, C heads with an uninterpretable [wh] feature can host a wh-elements only if that element is not the head of its movement chain. Perhaps then the ability to covertly host multiple specifiers is simply a parameter just as the ability to host them overtly.

Given that the claim I am defending is that Richards’s (1997) theory provides a uniform mechanism for deriving the constructions implicated in the non-context-freeness of Romanian and Swedish, the resolution of this debate is not strictly relevant. That is, whether or not the initial multiple-attraction step is triggered by C or some other, lower, head, the fact remains that the same fundamental mechanism is at work in Romanian and Swedish: a crossed-dependency inducing multiple-attraction step.

2.3 German scrambling

The next case of crossed-dependencies we will consider is German scrambling. As Becker et al. (1991), Becker et al. (1992) show, the dependencies involved in scrambling are beyond the generative capacity of the context-free grammars, the set-local multi-component tree adjoining grammars (local MC-TAG), and the linear context-free re-writing systems (LCFRS). Consider first the empirical paradigm. German is known to permit a relatively free word order clause-internally, even once V2 phenomena are controlled for. Thus, in an embedded clause the subject, object, and various modifiers may take on a variety of orders with respect to one another.

(50) a. dass die beiden immer noch einander lieben.
   that the both,nom still each-other,acc love
   “The two still love each other.”

   b. dass einander die beiden immer noch lieben
   (Grewendorf & Sabel 1999: 7)

This phenomenon, known as scrambling, is clause-bounded in German, so that a subject or object may not scramble out of a finite clause.

(51) a. dass [Hans dem Studenten gesagt hat [dass Maria dieses Buch besitzt]]
   that Hans,nom [the student,dat] told has that Maria,nom [this book,acc] owns
   “Hans told the student that Mary owns this book.”

   b. *dass [[dieses Buch],1 Hans [dem Studenten] gesagt hat [dass Maria e1 besitzt]]
   (Grewendorf & Sabel 1999: 11)

With a subset of infinitive taking verbs – the so-called restructuring verbs of Rizzi (1982), Wurmbrand (1998) – however, scrambling appears to be able to cross clause boundaries, a phenomenon termed long scrambling. The basic empirical picture is as below. Embedded objects and indirect objects may appear in various orders, including before and after the higher clause subject.

16 I would like to thank an anonymous reviewer for bringing these data to my attention.
17 Long scrambling is usually thought to involve some kind of clause reduction, where I use this term as neutrally as possible. Thus Wurmbrand (1998 et seq.) develops an account where the relevant infinitives are bare-VPs, so long scrambling is never truly cross-clausal. Grewendorf & Sabel (1994), Grewendorf & Sabel (1999), and Fanselow (2001) alternatively derive long scrambling analyses based on verb-incorporation, which either obviates the need for cross-clausal movement, as Fanselow (2001) argues, or simply renders such movement licit, as Grewendorf & Sabel (1994) argue.
(52) a. dass ich₁ [dem Kunden]₂ [PRO₁ [den Kühlschrank]₃ zu rearieren] versprochen habe that I the client.DAT the refrigerator.ACC to repair promised have “. . . that I have promised the client to repair the refrigerator.”
b. dass ich₁ [den Kühlschrank]₃ [dem Kunden]₂ [PRO₁ e₃ zu rearieren] versprochen habe “. . . that I have promised the client to repair the refrigerator.”
c. dass [den Kühlschrank]₃ ich₁ [dem Kunden]₂ [PRO₁ e₃ zu rearieren] versprochen habe “. . . that I have promised the client to repair the refrigerator.”

(Becker et al. 1991: 22)
The elements subscripted above may take on any of the possible ordering permutations for three elements, although each order carries its own special discourse and information-structural ramifications. Iterating this phenomenon arbitrarily, we derive that the following language is a grammatical subset of German, modulo an appropriate homomorphism.

SCR = \{σ(n₀, . . . , nₘ), . . . , ν₀ . . . νₘ : m ≥ 0 and σ a permutation\}, where nᵢ is an argument of νᵢ

Becker et al. (1991) show that generating a linguistically realistic representation of SCR requires a non-set-local multi-component tree adjoining grammar to generate a linguistically realistic representation of the language.\(^1\)

In the explication of the theory surrounding multiple-attraction offered in Section 2.1, the hallmark property characterizing this phenomenon was the obligatory crossing of paths. This is immediately obvious in Romanian, and only slightly less obscure in Swedish. With the German data above, there appears to be no such requirement at all. With appropriate emphasis, or in appropriate discourse situations, any order can obtain among the scrambled elements. On the surface, this calls into question the feasibility of analyzing these data as involving multiple attraction at all. That said, there are a number of intervening factors that might be at work in obscuring the ordering requirements characteristic of multiple-attraction.

Following Grewendorf & Sabel (1999), among others, I assume that scrambling involves feature driven movement, where the triggering head(s) is somewhere in the so-called *mittelfeld* between v and C. It is well known that scrambling is sensitive to such varied properties as animacy, definiteness, specificity, discourse salience. Given that the theory we are working in adopts features as the locus of interpretability, we might then expect that different surface orderings under scrambling correspond to the activity of different syntactic heads (and their associated features). It may well be the case, then, that certain orders are derived via the independent action of two completely different heads; perhaps some orderings involve a multiple-attraction step that is obscured by a subsequent movement driven by a feature not relevant for the original attraction. Of course it is far beyond the scope of this paper to develop an involved theory of scrambling. Instead, I would like to argue that when the factors above are controlled for, the strict ordering requirement characteristic of multiple-attraction re-emerges. This lends credence to the notion that the crossed-dependencies upon which the non-context-free complexity hinges are indeed derived here, as in the other cases, via multiple attraction.

Considering the behavior of idiom chunks under scrambling makes it possible to control for almost every factor listed above all at once. Idiom chunks are non-referential, so that effects of animacy, specificity, definiteness and discourse salience should be nullified. Informations structure should be likewise irrelevant.

\(^1\) For these authors, linguistically realistic means that each verb projects an elementary tree containing slots for all and only its arguments.

\(^2\) Strictly speaking, both of the studies referenced above make claims about something other than weak generative capacity. The definition of SCR depends on being able to identify which argument corresponds to which verb, something that is not weakly possible. At best, the robust morphological case distinction in German across nominative, accusative, and dative case makes it possible to filter arguments and verbs three ways, much as we relied on in the non-context-freeness argument for Romanian. Indeed it is possible to trivially construct a parrot argument in German on the basis of these data, so that German scrambling is at least weakly non-context-free.
given that idiom chunks are, on their own, semantically vacuous. By selecting for an idiom that makes use of two non-subject idiom chunks, we can test for the presence of multiple-attraction by considering the permitted order of the idiom chunks under scrambling. The results are exactly as predicted under a multiple-attraction account. One or both of the idioms chunks may scramble as long as the relative order among them preserves the order they take on \textit{in situ}.

(54) a. dass [der Hans\textsubscript{1}] nicht versucht hat [[die Katze\textsubscript{2}] [aus dem Sack\textsubscript{3}] zu lassen].
   that the Hans not tried has.acc cat out the.dat bag to let
   Lit. “That he has not tried to let the cat out of the bag.”
   “That he has not tried to tell the secret.”
   b. dass [der Hans\textsubscript{1} [die Katze\textsubscript{2} [aus dem Sack\textsubscript{3}]] nicht versucht hat [e\textsubscript{2} e\textsubscript{3} zu lassen]
   c. *dass [der Hans\textsubscript{1} [aus dem Sack\textsubscript{3} [die Katze\textsubscript{2}]] nicht versucht hat [e\textsubscript{2} e\textsubscript{3} zu lassen]
   d. dass [der Hans\textsubscript{1} [die Katze\textsubscript{2}] nicht versucht hat [e\textsubscript{2} [aus dem Sack\textsubscript{3} zu lassen]
   e. *dass [der Hans\textsubscript{1} [aus dem Sack\textsubscript{3}] nicht versucht hat [[die Katze\textsubscript{2}] e\textsubscript{3} zu lassen]

What these data suggest is that the relationship between the idiom chunk examples and the corresponding general case of scrambling is not unlike the relationship between multiple \textit{wh}-movement of non-D-linked and D-linked elements, respectively. The general machinery involved in all four operations is multiple-attraction, although in a given scenario other, poorly understood and possibly non-syntactic factors can intervene to distort or eliminate the appearance of crossed dependencies on the surface (see Richards (1997): 106ff. for a multiple-attraction account of D-linked multiple \textit{wh}-fronting in Bulgarian).

In addition to Romanian, we have now seen that two other ostensibly unrelated instances of non-context-free complexity in the literature can be identified as being derived from multiple-attraction in the manner of Richards (1997).\textsuperscript{20} Given this result, we might ask whether multiple-attraction is the only way to derive the crossing paths that characterize non-context-freeness in language. When the dependencies involved are derived via feature-driven syntactic movement, it is difficult to see how a theory that admits a principle like \textit{attract closest} (or the related notion of \textit{relativized minimal}, or the \textit{minimal link condition}, etc.) could derive the data otherwise without admitting special mechanisms (see Section 3.1 for more discussion). As we will see in the next section, however, there do exist instances of crossed-dependencies that are not obviously derived via multiple-attraction. In at least one instance, although quite possibly both, it is clear that the mechanisms involved in dependency formation are extra-syntactic, lending credence to notion that multiple-attraction is the single syntactic locus of crossed-dependency formation.

\textsuperscript{20} An additional case that might yield to a multiple-attraction analysis is Spanish clitic climbing, which Bleam (2000) shows to be beyond the generative capacity of all but the set local MC-TAGs. Richards (1997) has explicitly argued in favor of a multiple-attraction derivation for other instances of clitic movement, and indeed the construction shares the properties we have associated with this operation. Thus the order of fronted clitics must respect their underlying hierarchical relationship.

(55) a. Quiero permitir-te hacer-lo
   want.1.sg.to.allow-2.sg.to.do-3.sg.masc
   “I want to allow you to do it.”
   b. Te quiero permitir hacer-lo.
   c. Quiero permitir-te-lo hacer.
   d. Te-lo quiero permitir hacer.
   e. *Lo quiero permitir-te hacer.
   f. *Lo-te quiero permitir hacer.
   g. *Quiero permitir-lo-te hacer.

That said, the proper treatment of clitic climbing – the features which motivate it, the type of movement involved, etc. – remains sufficiently unclear as to render definitive conclusions illusive (see at least Kayne 1989, Sportiche 1992, Roberts 1994, Roberts 1997). I therefore leave this data set for further investigation.
2.4 Bambara and Swiss German

In an odd coincidence, it is the two original instances of non-context-freeness documented in the literature that fail to yield to a multiple-attraction analysis. Consider first the case of Bambara. The non-context-freeness argument is based on two morphological processes: noun reduplication and an agglutinative operation for forming agentive compounds. The first process is a construction of the form “noun o noun,” which translates as “whichever/whatever noun.” The same noun must be repeated on either side of the o morpheme.

(56) a. wulu o wulu
   dog  dog
   “whichever dog”

b. *malo o wulu
   rice  dog
(Culy 1985)

Next, Bambara features an agentive construction whereby a noun (N) and a transitive verb (TV) agglutinate to form a compound best translated as “one who TVs Ns”:

(57) a. wulu + nyini + la = wulunyinina
   dog  search for
   “One who searches for dogs”

b. malo + fil` e + la = malofil` ela
   rice  watch
   “one who watches rice” (Culy 1985)

This construction is recursive, so that the relevant noun involved in the construction may itself be a compound formed via the same agglutinative process:

(58) a. wulunyinina + nyini + la = wulunyininanyinina
   dog searcher  search for
   “one who searches for dog searchers”

b. malofil` ela + fil` e + la = malofil` elafil` ela
   rice watcher  watch
   “One who watches rice watchers” (Culy 1985)

Given that it is also possible to use the output of the agentive operation as the input to the “noun o noun” construction, compounds of the sort given below are licensed. It is this structure that serves as the basis of the non-context-freeness argument. I omit the proof.

(59) wulu(fil` ela)m(nyinina)n o wulu(fil` ela)m(nyinina)n

Culy (1985) makes the case that both the “noun o noun” construction and the agentive-forming operation are morphological, not syntactic, in nature, so that non-context-freeness arises in the vocabulary system of the language. The dearth of literature on Bambara makes it difficult to dispute or corroborate Culy’s (1985)
With this in mind, I will tentatively accept that Bambara attests to the non-context-freeness of the vocabulary forming system of natural language, not the syntax, so that the mechanisms at play in the previous section are absent here.

Consider now the case of Dutch and Swiss-German. As early as the mid-1970s, Huybregts (1976) had already identified the inherent non-context-freeness of Dutch embedded infinitives. His original argument suffered from a slight inadequacy (see Gazdar 1982), which was revised by Huybregts (1984). Shieber (1985) revises Huybrecht’s original argument as well, employing the overt case morphology of Swiss-German, absent in Dutch.

The basic empirical picture is as follows. Embedded clauses in the West-Germanic languages (except English) exhibit OV word order, so that the successive embedding of multiple clauses results in a sequence of sentence final verbs, usually referred to as a verb-cluster in the literature. In Swiss-German and Dutch, the linear order of verbs in such clusters does not straightforwardly reflect the underlying argument structure. Both languages are usually considered to be head final, (although see Zwart (1996)), so that the order of verbs expected by simply reading the nodes off the phrase structure has the hierarchically lowest verb coming first, with subsequent verbs ordered based on increasing hierarchical position. In the structure schematized below, this is the order $V_n - \ldots - V_2 - V_1$.

(60) 

However, in both Swiss-German and Dutch, the verb clusters formed by successively embedding clauses, as above, surface with the inverse of the expected order, $V_1 - V_2 - \ldots - V_n$. If we successively embed verbs that take both DP and clausal arguments, surface strings of the following general sort result.

(62) $\text{DP}_1 \text{DP}_2 \ldots \text{DP}_n \ V_1 \ V_2 \ldots \ V_n$

(63) Dutch

a. De mannen hebben Hans$_1$ [de paarden$_2$] leren$_1$ voeren$_2$.
the men have Hans the horses teach.$\text{INF}$ feed.$\text{INF}$
“The men have taught Hans to feed the horses.”
(Bach et al. 1986)

(64) Swiss-German

a. ...mer em Hans$_1$ es huss$_2$ hälfe$1$ aastiiche$_2$
...we Hans.$\text{DAT}$ the house.$\text{ACC}$ help paint
“...we helped Hans paint the house.”
(Shieber 1985: 334)

22 To the best of my knowledge the extant sources of primary Bambara data on which all subsequent discussion relies comprise only Bird’s (1966) unpublished PhD dissertation, two text books geared at aid workers in Mali (Bird & Kante 1976, Bird et al. 1977) and Koopman’s (1992) analysis of (the lack of) case chains.

23 Standard German exhibits the expected word order $V_n - \ldots - V_2 - V_1$. Thus the Dutch example in (63) surfaces as (61) in German.

(61) Die Männer haben Hans die Pferde füttern lehren.
the men have hans the horses feed teach
“The men have taught Hans to feed the horses.”
The dependencies between verbs and their associated DP arguments above are therefore crossed, a fact which Huybregts (1984) and Shieber (1985) capitalize on to deliver non-context-freeness arguments. In a manner identical to the one we employed for Romanian, Shieber (1985) exploits the overt morphological distinction between accusative and dative case on DPs, as well as the existence of verbs that subcategorizes for only accusative or only dative complements, to codify the dependencies in a way that permits a weak non-context-freeness argument. In Dutch, which lacks such overt case morphology, Huybregts (1984) relies instead on the fact that some verbs subcategorize on the basis of animacy, thus allowing a similar codification and corresponding argument.

For ease of reference, let us denote as the “copy order” the arrangement of verbs associated with crossed-dependencies, and as the “mirror order” the arrangement expected from the underlying phrase structure. The existence of the copy order has long defied easy explanation, and there exists a detailed and complex literature with several opposing views. Let us begin with some general arguments that the data do not involve multiple-attraction, then I’ll sketch some of the extant analyses of the phenomenon and discuss their ramifications for our investigation.

Perhaps the most telling evidence that we are dealing with something other than feature-driven multiple-attraction is the permissibility of a wide variety of different orderings among the relevant constituents. Shieber (1985) reports that (64a) may alternatively surface with any of the below orders, where I repeat the numbering from the previous example.

(65) a. . . . mer [em Hans]₁ [es huss]₂ hälfd₁ aastriče₂
   b. . . . mer [em Hans]₁ hälfd₁ [es huss]₂ aastriče₂
   c. . . . mer [em Hans]₁ [es huus]₂ aastriče₂ lälfd₁

With triply embedded cases, it has been reported that in addition to the copy (1-2-3) and mirror (3-2-1) orders, the verbs in the cluster may surface in the order 1-3-2 and 2-3-1 (Bobaljik 2004: 132, Wurmbrand 2006). This of course bears an immediate resemblance to the scrambling data we saw above, with two crucial differences that render a multiple-attraction analysis untenable. First, whereas each order in the scrambling cases carries different information-structural content, the various verb cluster orders do not appear to correspond to even subtle differences in meaning (this is a widely reported fact; see Wurmbrand 2006 and work cited therein). Second, while a wide variety of orders are permissible, the verb clusters do not exhibit the total freedom of ordering characteristic of scrambling. Thus in the three-verb case, the orders 2-1-3 and 3-1-2 are never observed. Recall that in order to argue for a multiple-attraction analysis of scrambling, we showed that when the various information-structural and discourse-related factors known to influence scrambling were controlled for, the freedom of ordering dissolved and the familiar obligatory crossed-dependency patterns re-emerged. Such an approach does not seem possible here. Given the semantic equivalence of the various orders, there do not appear to be any intervening factors at work.

Another empirical fact that casts doubt on the feasibility of the multiple-attraction analysis is the penetrability of the verb clusters to adjuncts. The layer specifier structure characteristic of multiple-attraction is, in general, impenetrable to adverbials and other adjuncts. This is easiest to demonstrate in cases like Romanian, where the layered specifier structure is unequivocally present on the surface and where discourse and related factors do not intervene. Rudin (1988: 467-471) provides numerous examples like the one below.

(66) *Cine, după păreră ta, ce a făcut?
    who.NOM after opinion your what have.3.SG do.PART
    “Who, in your opinion, did what?”

West-Germanic verb clusters display a more complicated pattern with respect to such intervention. Clusters in the mirror order, and indeed even subsequences of clusters whose elements appear in the mirror
order, are impenetrable to adverbs and other elements. The copy-order relevant here, however, permits intervention both by adjuncts and even DP arguments (Kiss & van Riemsdijk 2004: 23). We have already seen an example of this in the Swiss-German, repeated below.

(67) ... mer [em Hans]₁ hälled₁ [es huss]₂ aastriiche₂ we Hans.DAT help the house.ACC paint “We let Hans help paint the house.”

It follows that whatever the mechanisms involved in the derivation of the copy-order (if such an order is derived at all, and not simply underlying), they do not appear to result in the sorts of clusters of elements that we observe as resulting from multiple-attraction.

For these and other reasons, none of the various analyses of the West-Germanic verb clustering phenomena, of which the copy-order is an instantiation, have invoked anything like multiple-attraction. In fact, of the three major types of approaches to these phenomena, two eschew familiar syntactic devices altogether. The first type of approach is popular in non-derivational frameworks like HPSG, where they can be stated most naturally. These approaches reject the underlying constituency of a given verb in the cluster and its argument, instead holding that the verbs in the cluster coalesce by some mechanism into a single complex predicate. The various DP arguments are then selected not by the verb which they serve as the semantic argument to, but rather by the complex verbal constituent comprising the verb cluster (Steedman 1985, Hinrichs & Nakazawa 1994, Kathol 2000). Under these approaches, there are no discontinuous constituents and hence no crossed-dependencies in the syntax; each object enters into a dependency with the entire verb cluster, or at least with some non-singleton subset.

The second type of approach fits more naturally in the framework we are working with here, and is captured best in the work of Haegeman & Van Riemsdijk (1986) and Wurmbrand (2004, 2006, 2012). Setting aside technical details for the moment, these approaches treat Swiss-German, Dutch, and German as underlying uniform, with a phrase-structure along the lines of (60). While such approaches are thus forced to treat the various non-mirror orders as derived, they all reject familiar mechanisms of head- or phrasal-movement in this derivation. Rather, these theories make use of some kind of reanalysis operation that yields a re-bracketed and re-ordered version of the verb cluster. In many such analysis, especially that of Wurmbrand (2004 et seq.), this re-ordering is taken to apply post-syntactically, in the component of the derivation that maps the hierarchical syntactic representation to the surface phonological string. Crucially, in such theories the crossed-dependencies that materialize on the surface are not underlying crossed in the syntax; it is only via the intervention of a post-syntactic reordering operation that this behavior manifests.

A third class of approaches likewise embraces a uniform underlying source for Swiss-German, Dutch, and German, and treats all but the mirror order as derived (Evers 1975; Zwart 1996; Brody 1997, 2000; Koopman & Szabolcsi 2000). Unlike the analyses mentioned above, however, these approaches rely only on standard syntactic mechanisms of head- and phrasal-movement. Most modern takes on this idea, first introduced by Evers (1975), are couched in the Anti-symmetry (Kayne 1994) approach to syntax and make extensive use of so-called remnant movement (Zwart 1996, Koopman & Szabolcsi 2000). None of these accounts that explicitly embrace syntactic movement rely on anything akin to multiple-attraction.

To summarize, the crossed-dependencies in Bambara and Swiss-German are not derived via multiple-attraction. In fact, in both cases it is unclear that the crossed-dependencies are derived in the syntax at all. For Bambara, this is Culy’s (1985) explicit claim. For Dutch/Swiss-German, the case is less clear. While there do exist purely syntactic analyses for the derivation of the relevant data, the massive VP evacuation and roll-up remnant movement employed by these accounts raise a number of difficult questions. While remnant movement is an oft applied tool, it is unclear why the scale and complexity of such movement

24 Thus in a cluster ordered 1-3-2, an adverb or object may intervene between 1 and 3 but not between 3 and 2 (Kiss & van Riemsdijk 2004: 132, Wurmbrand 2006)
necessary to derive the verb cluster data is not more linguistically widespread. The verb-cluster phenomena are of course cross-linguistically unique, being limited to West-Germanic and Hungarian, but whether this should warrant the complexity of analysis employed remains debatable. Analyses like that of Wurmbrand (2004) of course also make use of novel and otherwise unattested mechanisms, although in this case the unusual behavior takes place outside of syntax proper. It is of course something of an aesthetic choice where such language-specific rules should be located in the grammar. That said, if the conjecture that crossed-dependencies have a uniform mechanism in the syntax, multiple-attraction, is on the right track, the failure of West-Germanic verb clusters to yield to such an analysis might suggest that theories along the lines of Wurmbrand (2004) are to be preferred.

3 Multiple-attraction: alternatives and perspective

Let us now turn to contextualizing the result obtained so far by considering alternatives and related issues. For the remainder of the paper, I will accept the working hypothesis that multiple-attraction is the unique mechanism furnished by the syntax for deriving crossed-dependencies.

3.1 Alternatives

If we consider closely the framework in which we are working, the fact that multiple-attraction underlies the various instances of crossed-dependencies is not surprising. Under the assumptions that movement is triggered by the need to value and delete features and that such valuation is constrained by the basic idea of locality encapsulated in attract closest, we are left with relatively few options for deriving crossing paths. Consider the general case from (31), repeated below.

\[(31) \quad [H_P \cdot X_P_1 \cdot H_P \cdot X_P_2 \cdot [H \cdot [Y_P \cdot \ldots \cdot e_1 \cdot \ldots \cdot Z_P \cdot \ldots \cdot e_2 \cdot \ldots]]] \]

Assuming that X_P_1 and X_P_2 are attracted by the same functional head, X_P_1 will always be attracted first, so that we are forced to invoke the tucking-in of X_P_2 to derive the dependency. That said, there are alternative mechanisms at our disposal provided we are willing to relax the requirement that both phrases be attracted by the same head. Let us consider some of these alternatives.

First, we might posit that each of X_P_1 and X_P_2 are attracted by a different functional heads in the vicinity of H above, so that both attractors c-command X_P_1 and X_P_2. In order to avoid the same issue with minimality that arises with a single attractor, we’ll have to assume that each attraction step is motivated by a different feature. This raises a number of difficult issues: how do we ensure that the head with the feature corresponding to the hierarchically highest XP is itself hierarchically superior to the other attracting heads? If the fronted elements are separated by various functional heads, why can adverbs and other adjuncts not attach between the phrases? Absent persuasive answers to these questions, this approach is problematic. A much more robust (and ultimately empirically successful) approach to the multiple-attractors idea is posit that the heads of the attracted phrases are themselves the attractors of lower phrases. In the case above, then, the head of X_P_1, X_1, would bear an unvalued feature that probes for the corresponding feature on X_P_2. Subsequent movement of X_P_2 to X_P_1, the entire cluster can then be attracted by H.

\[(68) \quad [H_P \cdot [X_P_1 \cdot X_P_2] \cdot [H \cdot [Y_P \cdot \ldots \cdot e_1 \cdot \ldots \cdot Z_P \cdot \ldots \cdot e_2 \cdot \ldots]]] \]

25 These questions correspond empirically to the behavior that we actually observe with, for example, the general case of scrambling or of D-linked multiple wh-fronting, where a relatively free order obtains between displaced elements and where adverbs and other adjuncts can intervene (see ?? for examples from scrambling, Bošković 2002 for intervention data on wh-fronting). This has in fact lead Richards (1997) and others to posit that both the general case of scrambling and multiple wh-fronting of D-linked elements involves the action of at least two heads.
This exact mechanism is at the heart of the theory of multiple *wh*-movement developed by Grewendorf (2001). In this theory, each *wh*-element in a multiple-attraction scenario contains a D head that bears an unvalued [*wh*] feature. An instance of AGREE, and an associated movement step, can then obtain between this feature and a lower *wh*-element. The internal structure of a given *wh*-element is thus as below. Note that the DP itself is presumed to carry an uninterpretable, valued [*wh*] feature that may be the target of an AGREE relation with a higher probe (Grewendorf 2001: 96).

(69) \[
\text{DP}_{[u\text{wh}[\text{val}]]} \quad \text{D}_{[u\text{wh}[\text{val}]]} \quad \text{NP}
\]

While Grewendorf’s (2001) theory is a viable alternative for the basic Romanian facts, the theory faces a few immediate theoretical complications that should be pointed out. In order to allow the D head of a *wh*-DP to probe and enter in an AGREE relation with other features in the structure, we have to relax the c-command requirement normally thought to hold of AGREE: once a DP like the one in (69) has been merged in the structure, D does not c-command anything beyond its sister. We might alternatively propose that the unvalued, uninterpretable probing feature is percolated to DP, which then probes for an agreeing feature. But this requires us to allow both heads and maximal projections to probe for agreeing features. Moreover, its not clear how or if DP should be able to bear both a valued and unvalued copy of the same feature. Relatedly, the movement step required in this theory is technically an instance of so-called *sideward*-movement, given that the landing site of the displacement does not c-command the trace position (Nunes 1995, 2001). Such movement poses non-trivial problems for theories of chain-linearization and spell-out, among other things, and should be relied upon with caution. Finally, the *wh*-DP to *wh*-DP movement steps must involve a non-canonical landing site. Thus whereas movement to a head usually involves specifier formation, in this case the landing site must involve adjunction to some other position to derive the correct word order (under standard assumptions).

(70) \[
\text{DP}_1 \quad \text{precedes} \quad \text{DP}_2:
\]

\[
\begin{align*}
\text{a. } & *\text{[DP}_1 \text{[DP}_2 \text{D NP ] [DP}_2 \text{D NP ]]} \\
\text{b. } & \text{[DP}_1 \text{[DP}_2 \text{D [XP NP [DP}_2 \text{D NP ]][]]]}
\end{align*}
\]

Presuming these issues can be worked out, we might ask if this clustering approach can be extended to other cases of crossed-dependencies. Consider first the case of Swedish, which we argued was underlyingly very similar to Romanian. Let’s examine the derivation of an example like (39), repeated below.

(39) “Who did the teacher ask what he studied?”

Given that the two *wh*-elements must ultimately end up at disparate syntactic positions, any cluster formation that occurs must be undone at some point. Because of this, it’s unclear what cluster formation accomplishes. Recall that the sorts of clusters we’re dealing with involve successively embedded DPs, so that the only available mechanism for moving the elements apart is to extract the embedded *wh*-element(s) to higher projections. A sort of remnant movement might then ensue, displacing the remains of the cluster to a higher position. This series of operations is depicted schematically below: XP₁ and XP₂ form a cluster which then is attracted by H; XP₂ is then attracted out of the cluster by H′ and the remnant of the cluster is attracted by H″. Some series of operations like this must underly the Swedish data if we are to invoke cluster formation in the derivation.
In order to avoid minimality violations, the evacuation of the XP_2 from the cluster will have to be motivated by a feature that is not shared by XP_1. But if XP_1 and XP_2 are already made up of different features, we can accomplish the same derivational output without ever forming a cluster: each of XP_1 and XP_2 can just be attracted by the heads bearing the disparate features (H’ and H” above, respectively).

It follows that in order to derive data of the sort we saw in Swedish without availing ourselves of multiple-attraction, we will need to invoke separate attractors for each of the wh-elements involved. Cluster formation is irrelevant. Setting aside the various theoretical issues with this separate-attractors type theory, by denying that the same derivational mechanisms are at work in deriving multiple wh-movement and the wh-island violating crossed-dependencies in Swedish, the clustering theory misses a deep implicational universal. As Richards (1997) argues, in each of those multiple wh-fronting languages that exhibit the ordering effects we associated with a multiple-attraction step (his so-called CP-absorption languages), there are also observed the sort of wh-island violating crossed dependencies we saw in Swedish. While the Swedish data attest to the uni-directional nature of the implication, the multiple-attraction account can simply invoke the parameterization of overt multiple specifiers to unify Swedish with the other cases. Under the clustering theory, the connection between these data sets is a mystery: fundamentally different mechanisms are necessarily at work in the derivation of the relevant crossed-dependencies.

The clustering theory also faces a severe challenge in deriving the scrambling data. I limit discussion here to the idiom chunk examples, where the cross-dependency inducing step is easiest to observe. Technically, we can derive the data via cluster formation if we allow D heads of non-wh-elements to probe as well. This raises a number of questions that are beyond the scope of this paper to address. For example, what is the motivating feature for this movement? Is the movement of the A or A’-type? Why does cluster formation of this sort only happen in scrambling examples?

This discussion makes clear the severe limitations imposed by a principle like attract closest on the formation of crossed-dependencies. In the abstract case, some special mechanism, be it tucking-in, cluster formation, separate attractors (of the non-cluster forming variety), or some other mechanism entirely, must be appealed to in the course of the derivation. As we have seen, the mechanism of multiple-attraction/tucking-in avoids the theoretical shortcomings of the other two accounts and permits us to unify the three instances of crossed-dependencies with unequivocal syntactic origins. Of course, the fundamental assumptions that undergird our analysis are not universally held. By rejecting any one of these assumptions, that syntax is fundamentally derivational, that structure building and displacement are accomplished by merge, that displacement is motivated by feature valuation, etc., we might happen upon a different account of these phenomena.

3.2 Multiple-attraction and non-context-freeness in perspective

As mentioned in the introduction, the number of instances of proved non-context-free complexity in natural language remains exceptionally small. Given that the essential syntactic mechanism responsible for such complexity appears to be fairly pervasive, we might ask why this is so. The first reason is that crossed-dependencies do not imply non-context-freeness unless they can be arbitrarily iterated. As such, absent the capacity for unbounded dependency formation that characterizes the A’-system, some mechanism of clause-reduction, or a related device, crossed-dependencies are not weakly non-context-free. Richards (1997) identifies a number of other instances of clause-bounded crossed-dependencies of exactly this sort:

---

26 It is interesting to note that both the multiple-attraction theory and the clustering theory make recourse to DP movement that targets a non-root position in the tree. If we conceive of the structure building operation merge in the simplest way possible, namely as taking two elements and forming a set from them, both accounts involve undervivable structure (see, for example, Epstein et al. 2012 for discussion).
Varieties of Japanese scrambling; Serbo-Croatian clitic movement; West-Flemish negative fronting. This harks back to the relatively uninformative nature of weak generative capacity for the purposes of syntactic theorizing. Although these cases do not instantiate non-context-free string sets, the essential mechanisms underlying their derivation are the same as the ones involved in the non-context-free cases discussed above.

The multiple wh-fronting examples from Romanian suggest another factor that may intercede to block ready observation of the inherent non-context-freeness of natural language. A pervasive view of wh-dependencies holds that what Romanian and Bulgarian do in overt syntax actually occurs covertly in multiple-questions cross-linguistically (see at least Karttunen 1977, Huang 1982, Hagstrom 1998; Pesetsky 2000; Cable 2007). While care must be taken to differentiate the different types of multiple-wh-fronting languages (see Richards 1997 and Bošković 2002, in particular), it is possible that English exhibits identical crossed-dependencies to Romanian at a level of the derivation invisible on the surface. Thus at some stage of the derivation, the English sentence in (72a) may be represented as in (72b).

(72) a. Who bought what?
   b. [CP who1 [CP what2 [C [TP e1 bought e2]]]]

There are various proposals for how this might be effected, most notably post spell-out movement in the Y-derivational model (Chomsky 1995) and mandatory pronunciation of lower copies of all but the highest wh-element in single-output syntax (Lidz & Idsardi 1998). If such approaches are on the right track, presumably every language, or at least a sizable subset, makes use of the complexity inducing operations overtly manifest in Romanian and other cases.

On a related note, a cohort of researchers have recently advanced the idea that weak generative capacity ought to be formally delimited as a parameter of natural language syntax, much like headedness (see Deutscher (2010), Everett (2005), Givón (2009), Heine & Kuteva (2007), Wray & Grace (2007) for various versions of this idea). Under this view, languages may be meaningfully classified according to the formal complexity of their surface strings. This claim is difficult to maintain in light of our discussion. First, while I have amassed evidence that some instances of non-context-freeness have a uniform derivational source, there remain cases like Swiss-German where the same mechanisms are obviously not involved. It’s unclear what insights may be gleaned by parameterizing something that is so clearly the result of different operations in the grammar. Next, such a classification threatens to cross-cut languages that really ought to be grouped together under any reasonable parameterization of complexity. For example, as Miller (1991) points out, the resumptive pronouns which obligatorily fill the gaps associated with the cross-dependencies in Swedish do not appear in Norwegian, so that technically the weak non-context-freeness argument does not obtain in this case. However, the underlying structural configurations in the two languages are identical, as are the actual syntactic devices involved in the derivations. The more appropriate parameterization might be whether a language does or does not allow multiple-attraction/tucking-in, although as discussed above this is not always immediately obvious from surface strings.

4 Conclusion

Let us take stock. At the outset of this study, we set out to investigate the source of the crossed-dependencies that lead to the various instances of non-context-free complexity in language. Taking as a starting point Romanian, where the derivational mechanisms at work are readily observable on the surface, we explored a profusion of instances of crossed-dependencies that have been implicated in the complexity literature. From

27 Again though weak non-context-freeness has provided us with cases of unequivocal complexity, so that at least the data, if not the analysis, presented here is beyond theoretical reproach.

28 Problems remain with this approach, especially in light of the present discussion. For example, if English allows covert multiple-specifiers and tucking-in, why do we not observe the wh-island violating crossed dependencies that appear in such languages? As we saw in Swedish, the crucial ingredient in deriving these data is the ability to covertly hose multiple-specifiers, which English could presumably do under these accounts.
those cases investigated, a pervasive pattern has emerged: in all those cases where the dependencies can be reliably construed as involving syntactic movement, it is possible to isolate a multiple-attraction step à la Richards (1997). While the grammar clearly furnishes other mechanisms for deriving crossed dependencies, as we saw in Bambara and Swiss-German, it appears that multiple-attraction may constitute the unique syntactic device for deriving such dependencies.

This result fits nicely into the Minimalist framework in which we have been working. Specifically, the crossed-dependencies investigated herein have been shown to arise as the minimally deviant structures to basic, universal, and potentially extra-linguistic (see Chomsky 2005 for discussion) principles of economy – attract closest, shortest move – meant to optimize the satisfaction of interface conditions. Given the general heuristic that increasing complexity on the Chomsky hierarchy corresponds to an increasing demand on computational resources, both for production and processing (although see Bach et al. 1986), this result confirms the sentiment encoded in the Strong Minimalist Thesis (Chomsky 2005), namely that optimality and economy in the language faculty should be limited to those areas which directly concern the satisfaction of interface conditions.

References


Cable, Seth. 2007. *The grammar of Q: Q-particles and the nature of Wh-fronting, as revealed by the Wh-questions of Tlingit*; Massachusetts Institute of Technology dissertation.


Clemens, Lauren Eby, Jessica Coon, P Mateo Pedro, AM Morgan, Maria Polinsky, Gabrielle Tandet & Matt Wagers. 2015. Ergativity and the complexity of extraction: A view from Mayan. *Natural Language & Linguistic Theory*.


Evers, Arnold. 1975. *The transformational cycle in Dutch and German*; University of Utrecht dissertation.


Kraskow, Tina. 1987. Implications of Multiple Wh-Movement for Wh-island Violations. In *LSA annual meeting, San Francisco*,.


