Computational approaches to syntactic variation

Tim Hunter and Robert Frank

June 2, 2023

Revised version to appear in M. Polinsky et al (ed.), *The Cambridge Handbook of Comparative Syntax*

1 Introduction

In this chapter we outline some computationally-significant dimensions of variation in the syntactic patterns found in natural languages. The relevant patterns are, more specifically, best thought of as particular configurations of elements that are related by *syntactic dependencies*. We begin by carefully introducing the key notion of a syntactic dependency via some abstract examples in section 2, and showing how it relates to constituency. Against this backdrop, section 3 introduces the idea that many "linguistically interesting" configurations of syntactic dependencies can be described in terms of *discontinuous constituency*. This notion provides a descriptive framework for the core of the chapter, section 4, where we present four dimensions along which analyses set in this framework can be classified. These dimensions of variation are computationally significant in the sense that they bear on the question of what kinds of computational machinery or formal system is needed to appropriately describe linguistic phenomena; we mention these consequences briefly and provide relevant references, but our focus here is squarely on the critical generalizations themselves rather than the implicated formal system. Finally, in section 5, we briefly describe an alternative to the discontinuous-constituency perspective that we adopt through much of the chapter, which provides a basis for linking the issues we have discussed to a distinct tradition of grammatical frameworks.

The precise landscape of many of the empirical patterns that we discuss, especially in section 4, remains indeterminate to some degree. Most saliently, the question often arises as to whether a constraint on syntactic patterning is best explained in terms of linguistic competence or as the result of processing limitations. Nonetheless, we proceed tentatively, with the goal of providing an accessible overview of the arguments that have been made for various kinds of syntactic complexity, conveying the logical structure of the debates in a way that we hope will promote empirical investigations bearing on the computationally critical issues. In reviewing the arguments we of course report the empirical claims on which those arguments have been based, but we present these primarily as illustrations of the significance of certain issues, not necessarily as settled facts.

2 Syntactic dependencies

2.1 Formal analysis of syntactic dependency patterns

We begin with some abstract examples to illustrate key ideas. Consider three simple fictional languages, each of which uses a vocabulary of just four words: 'flip', 'flop', 'tick' and 'tock'. In all three languages, each occurrence of 'flip' must appear with a corresponding occurrence of 'flop', and each occurrence of 'tick' must appear with a corresponding occurrence of 'tock'. We will call the linkage between such co-varying elements a *syntactic dependency*. More generally, we can define a dependency as obtaining between elements X and Y in a sentence when a modification to X (e.g. changing 'flip' to 'tick') requires a concomitant modification to Y (changing 'flop' to 'tock') (Chomsky, 1956).

Our fictional languages differ in the linear arrangement of the pairs of elements that enter into syntactic dependencies. In the first language, each occurrence of 'flip' must be immediately followed by an occurrence of 'flop', and likewise for ‘tick’ and ‘tock’. Some example “sentences” from this language are shown in (1). We’ll describe this as a *serial dependencies* pattern.

(1)
a. flip flop
b. tick tock
c. flip flop tick tock
d. tick tock flip flop flip flop
e. tick tock flip flop tick tock tick tock

In the second fictional language, all occurrences of ‘flip’ and ‘tick’ must appear first, followed by all corresponding occurrences of ‘flop’ and ‘tock’ in a “mirror image” order: the first word in the ‘flip’/‘tick’ portion of a sentence is matched with the *last* word in the ‘flop’/‘tock’ portion. Some examples are shown in (2). We’ll describe this as a *nesting dependencies* pattern.

(2)
a. flip flop
b. tick tock
c. flip tick tock flop
d. tick flip flip flop flop tock
e. tick flip tick tick tock tock flop tock

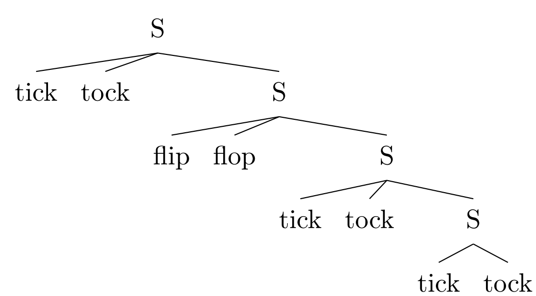
Finally, in the third language, all occurrences of ‘flip’ and ‘tick’ must again appear first, but here the corresponding occurrences of ‘flop’ and ‘tock’ are not mirrored: the first word in the ‘flip’/‘tick’ portion is matched with the *first* word in the ‘flop’/‘tock’ portion. Some examples are shown in (3). We’ll describe this as a *crossing dependencies* pattern.

(3)
a. flip flop
b. tick tock
c. flip tick flop tock
d. tick flip flip tock flop flop
e. tick flip tick tick tock flop tock tock

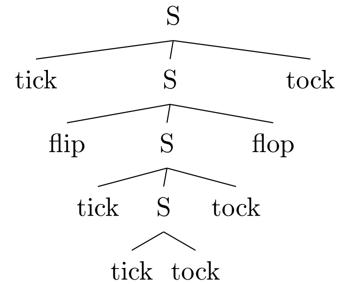
All three of these kinds of patterns are attested in natural language syntax, as we will see below, but there is an important difference between the serial and nesting patterns on the one hand, and the crossing pattern on the other. Broadly speaking, there is relatively wide agreement — sometimes hidden by differences in terminology and notation — about the nature of the grammatical mechanisms underlying the phenomena that exhibit serial and nesting patterns. The crossing dependencies pattern, however, requires invoking something extra beyond this common core of assumptions, and so it provides one useful way to focus attention on what is required of these additional mechanisms about which there is less widespread agreement.

One way to instantiate the common core of widely shared assumptions is to adopt the familiar framework of context-free phrase-structure grammars. Using this system, the serial and nesting dependency patterns can be straightforwardly accounted for as shown in (4) and (5), respectively.

(4) S → flip flop (S)
S → tick tock (S)



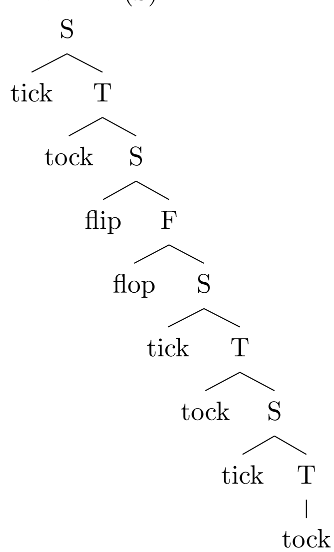
(5) S → flip (S) flop
S → tick (S) tock



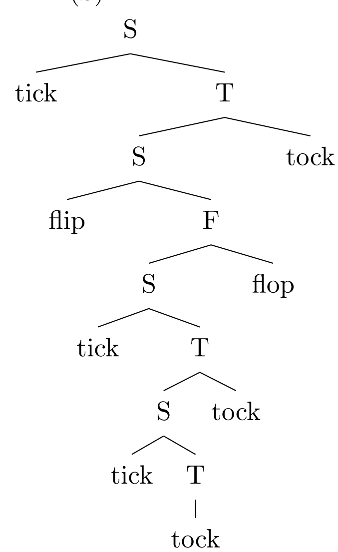
There is an obvious but important similarity between (4) and (5): each rule says either that an S can be comprised of one 'flip' and one 'flop' and another S, or that an S can be comprised of one 'tick' and one 'tock' and another S. This naturally creates the co-dependent 'flip'/'flop' and 'tick'/'tock' pairs. The difference is in how these elements are linearly arranged relative to their sister S subconstituent: serial dependencies arise if the S is peripheral (e.g. S → flip flop S), whereas nesting dependencies arise if the S is medial (e.g. S → flip S flop).

The phrase-structure rules in (4) and (5) are particularly simple in the sense that a *single rule* introduces each co-dependent pair of elements; the two members of each pair appear at exactly the same height in the tree. Analyses of natural language phenomena that exhibit these patterns will often involve additional constituent structure, and indeed must in order to adhere to constraints like binary branching. However, these additional details do not affect the key points. For example, (6) and (7) might seem more reminiscent of analyses of natural language than the very simple (4) and (5).

(6) S → flip F
F → flop (S)
S → tick T
T → tock (S)



(7) S → flip F
F → (S) flop
S → tick T
T → (S) tock



In (6) and (7), the F nonterminal symbol serves as a record of an unresolved 'flip'/'flop' dependency, and this record-keeping mechanism allows for applications of the separate rules that introduce 'flip' and 'flop' to be coordinated; likewise for the T nonterminal symbol, and 'tick' and 'tock'. But the coordinated effect of the two rules involving F is to put both 'flip' and 'flop' to one side of an S in (6), and to put them on either side of an S in (7), exactly as in (4) and (5). The simple picture conveyed by (4) and (5) might be thought of as a "zoomed out" view that focuses attention on the dependencies under discussion here, which is consistent with many analyses that incorporate finer-grained details.

This perspective lets us highlight the important difference between serial and nesting dependencies on the one hand, and crossing dependencies on the other. The standard and widely-held ideas embodied in phrase-structure grammars allow for the generation of the serial-dependency sentence (1e) and the nesting-dependency sentence (2e) by allowing for the combination of ‘tick’ and ‘tock’ with *some other smaller expression* in which all the appropriate dependencies are already enforced/resolved: in the renderings of (1e) and (2e) below, the crucial smaller expression is shown in bold. The property that sets apart a crossing-dependency sentence such as (3e) is the fact that the relevant smaller expression that a ‘tick’/‘tock’ pair needs to combine with does not correspond to a *contiguous* portion of the surface string.

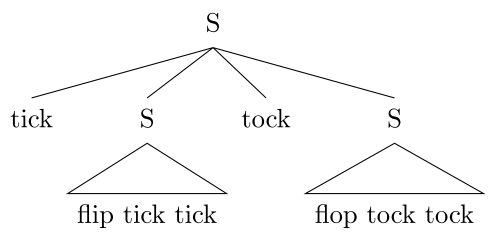
(1e) tick tock **flip flop tick tock tick tock**

(2e) tick **flip tick tick tock tock flop** tock

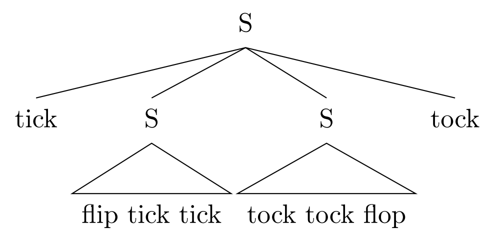
(3e) tick **flip tick tick** tock **flop tock tock**

It is instructive to notice that an attempt to generate crossing dependencies via a structure such as (8) will fail for the same reason that a structure like (9) fails to enforce nesting dependencies. The syntactically dependent elements reside in separate constituents, and the information needed to make them covary cannot be encoded in the categorization of the constituent itself (here encoded via the label S, but similarly true in any case where the content of the label is limited to a bounded amount of information). The ability of the nesting-dependencies grammars in (5) and (7) to “bundle up” the bolded portion of (2e) into a single constituent is crucial.

(8)



(9)



The intuition that is therefore suggested by (3e) is that our grammatical mechanisms need to go beyond those of simple (context-free) phrase-structure grammars in some way that effectively allows the discontiguous bolded portion ‘flip tick tick . . . flop tock tock’ to be somehow treated as a unit. This idea of *discontinuous constituency* provides a useful unifying perspective from which to consider the extent and limits of cross-linguistic syntactic variation, which we will adopt for the rest of this chapter.

2.2 Dependency patterns in natural language

All three of the patterns introduced above are attested in natural languages. A comparison between English, German and Dutch provides a neat illustrative “minimal triplet”, where the subject–verb dependencies appear in serial, nesting and crossing configurations respectively.

(10) John saw Peter let Marie swim (English, serial dependencies)

(11) . . . dass Hans Peter Marie schwimmen lassen sah
that Hans Peter Marie swim let saw
“. . . that Hans saw Peter let Marie swim” (German, nesting dependencies)

(12) . . . dat Jan Piet Marie zag helpen zwemmen
that Jan Piet Marie saw help swim
“. . . that Jan saw Piet help Marie swim” (Dutch, crossing dependencies)

For the reasons outlined above, this kind of Dutch sentence provides a particularly direct argument that the simple contiguous-constituency mechanisms of context-free phrase-structure grammars are an inadequate model of natural language syntax.[1] But it is certainly not the only argument that has been made for this conclusion.

The simple contiguous-constituency notion is also called into question by examples like (13), (14) and (15). In each pair, many of the considerations that lead us to treat the underlined portion of the 'a.' sentence as a grammatically-relevant unit could similarly be used to argue that the underlined portion of the 'b.' sentence is a grammatically-relevant unit, despite the fact that the latter is not contiguous.

(13) a. We bought a book about linguistics yesterday
     b. We bought a book yesterday about linguistics

(14) a. We know John thinks the girl bought the book
     b. We know which girl John thinks bought the book

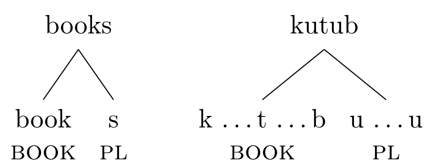
(15) a. It seems John will win
     b. John seems to win

Observations of essentially this sort underlie Chomsky's move from phrase-structure grammars to transformational grammars; Chomsky (1956, §4.1) observes that where simple phrase-structure grammars fall short is that they provide no way of "selecting as elements certain discontinuous strings".

3 Analyses in terms of discontinuous constituency

Holding fixed the idea that *all* the underlined portions of the sentences in (13), (14) and (15) correspond to grammatically-relevant units — along with the 'Piet Marie . . . helpen zwemmen' portion of (12) — we are led to the conclusion that these units must be able to combine in ways that are *non-concatenative*. This may sound somewhat unfamiliar when described in these terms, but it is the same basic idea that might be invoked in analyses of certain morphological patterns. The difference between the formation of the simple English plural 'books' and the Arabic templatic plural 'kutub' can be expressed as in (16).

(16)

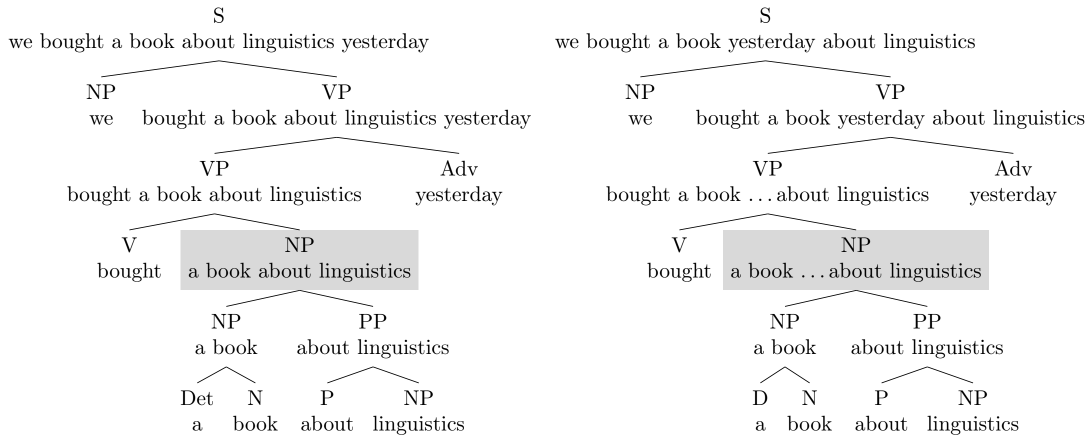


In both of these structures, the nominal root and the plural morpheme correspond to separate structural units that are combined in the derivation of the complex word. In the English case, the combination takes place purely concatenatively. In Arabic the morphemes include points of discontinuity, denoted by '. . . '. Upon combination, these are interpolated, and each one ends up corresponding to a discontinuous portion of the surface string.

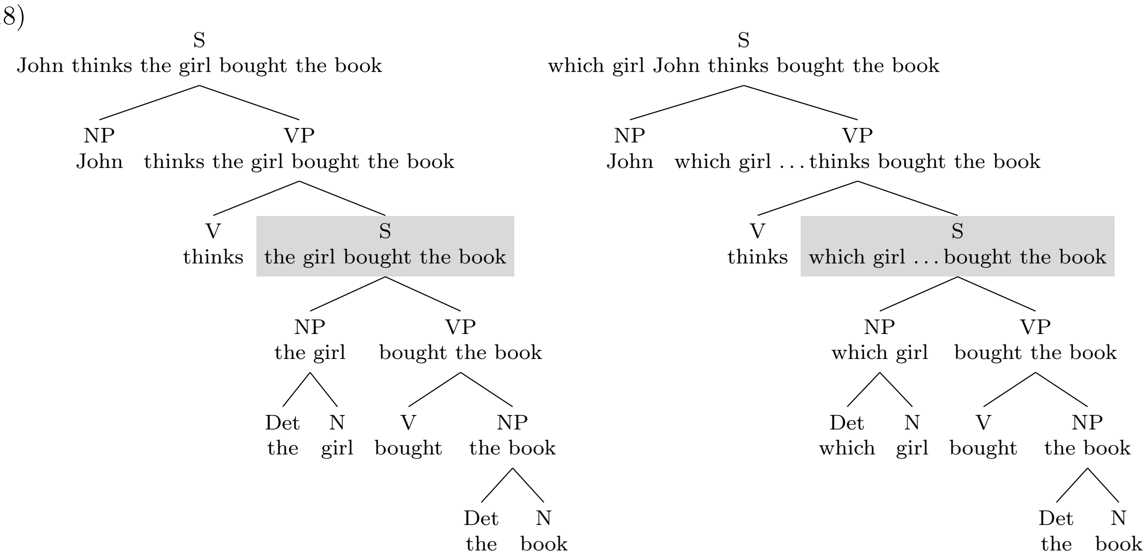
The application of this perspective to the sentences in (13)–(15) is illustrated by the pairs of trees in (17)–(19). In each pair, the first tree shows the 'a.' sentence derived in the familiar, purely concatenative manner, but we show explicitly the string associated with each constituent; this draws attention to the distinctive non-concatenative character of the second tree in each pair, which shows an analysis of the 'b.' sentence in which some constituents are associated with discontinuous strings. The highlighted nodes represent the underlined constituents in (13)–(15).

[1] Nesting dependencies themselves provide a direct argument for the inadequacy of a restricted kind of grammar where (contiguous) subexpressions must be peripheral (as in (4)) rather than medial (as in (5)). This is the point famously made by Chomsky (1956, §2.3) regarding *finite-state* grammars. As in the cases we discuss, this argument requires assumptions about the proper theoretical interpretation of the empirical landscape, given the difficulty of assessing the grammatical well-formedness of sentences with multiple nested embeddings.

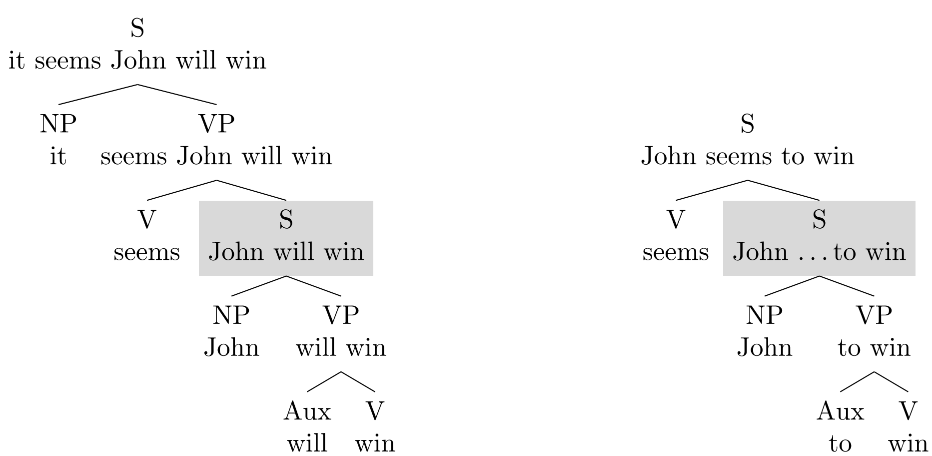
(17)



(18)



(19)



With this descriptive apparatus in place, we now turn to identifying four dimensions of variation that can be used to classify the syntactic patterns observed across languages, which we address in the next section.

4 Dimensions of variation

4.1 Bounded vs. unbounded components

One question we can ask, about a particular linguistic pattern's description in terms of discontinuous constituency, is how many components discontinuous constituents need to be comprised of. In the trees in (17)–(19), no constituent is "split" into more than two components.

First notice that there are many cases where a sentence contains, intuitively, more than one instance of the discontinuities exemplified in (17)–(19), but no individual constituent is comprised of more than two components. For example, while (20) includes a displaced wh-phrase ('which girl') and a raised subject ('Mary'), no single constituent must keep these two displaced elements separate from other material. Certain deeply-embedded constituents, such as (21b), comprise two discontinuous pieces, with 'which girl' being kept separate, until it is incorporated into the contiguous (i.e. one-component) constituent in (21c); subsequently, 'Mary' is incorporated as a separate component in the larger discontinuous constituent (21e) that feeds raising.

(20) Mary seemed to know which girl John thinks bought the book

(21) a. which girl . . . bought the book
b. which girl . . . thinks bought the book
c. which girl John thinks bought the book
d. know which girl John thinks bought the book
e. Mary . . . know which girl John thinks bought the book

In other configurations, distinct discontinuities of the sorts exemplified in (17)–(19) "overlap" in a way that introduces three-component constituents. For example, an analysis of the sentence in (22) would involve a three-component constituent corresponding to the clause headed by 'think', shown in (23).

(22) Which girl does John seem to think bought the book?

(23) which girl . . . John . . . think bought the book

We can ask, then, whether the "size" (measured in terms of separate components) of discontinuous constituents is limited, either in a single language or across languages.[2] In a language like English, it does not seem to be possible for the number of distinct components in a discontinuous constituent to grow without bound. Such a property has been derived grammatically from principles like Relativized Minimality (Rizzi, 1990) or its more modern instantiations: there is a limited number of distinct features that can drive movement and some limit on the co-occurrence of such features, creating competition among potential movers for the available slots.

In contrast, other languages have been claimed to allow for exactly this kind of unbounded discontinuity. For example, Rambow (1994, pp.13–16) argues that scrambling in German can create discontinuities of the critical sort (see also Becker et al. 1992). A simple example of scrambling is given in (24), where the object of the embedded verb ('repair') can be scrambled out of its clause to a position to the left of the higher clause's subject ('no one'). The embedded clause therefore surfaces as a discontinuous constituent, as indicated by the underlining.

(24) a. . . . dass niemand [den Kühlschrank zu reparieren] versprochen hat
. . . that no one-NOM the refrigerator-ACC to repair promised has
. . . that no one has promised to repair the refrigerator

b. . . . dass den Kühlschrank niemand zu reparieren versprochen hat
. . . that the refrigerator-ACC no one-NOM to repair promised has
. . . that no one has promised to repair the refrigerator

Importantly, scrambling of one element does not seem to preclude scrambling of others. In (25) there is an additional argument ('the client') of the highest verb has also been scrambled to the left of the highest subject.

[2] In the dependency grammar literature (e.g. Kuhlmann, 2013), essentially the same notion of degree of discontinuity is known as "gap degree".

(25) ...dass dem Kunden den Kühlschrank bisher noch niemand zu reparieren versprochen hat
...that the client-DAT the refrigerator-ACC so far as yet no one-NOM to repair promised has
...that so far, no one yet has promised the client to repair the refrigerator

This example of "double scrambling" is suggestive of an analysis in terms of a three-piece discontinuous constituent, like (23), since two elements are displaced. But since the two scrambled elements are adjacent in the surface word order in (25), one might also imagine that in fact there are only ever two discontinuous pieces, one of which being a "cluster" of scrambled elements. This would not be compatible with examples like (26), however, where the two scrambled elements (both arguments of the embedded clause headed by 'indict') appear interspersed with arguments of the higher clause (Joshi et al., 2000).

(26) a. ...dass der Detektiv dem Klienten [den Verdächtigen des Verbrechens zu überführen]
...that the detective-NOM the client-DAT the suspect-ACC the crime-GEN to indict
versprochen hat
promised has
"...that the detective has promised the client to indict the suspect of the crime"

b. ...dass des Verbrechens der Detektiv den Verdächtigen dem Klienten zu überführen
...that the crime-GEN the detective-NOM the suspect-ACC the client-DAT to indict
versprochen hat
promised has
"...that the detective has promised the client to indict the suspect of the crime"

Here there is no alternative but to treat the complement of 'promise' as a three-piece discontinuous constituent.

Rambow (1994) and Becker et al. (1992) conclude from these patterns that German scrambling is "doubly unbounded", in the sense that (i) there is no limit to how many clause boundaries an argument can be scrambled across, and (ii) there is no limit to the number of arguments that can scramble across any particular clause boundary. If this descriptive generalization is correct, then it will be possible to construct larger grammatical examples along the lines of (25) and (26) with four-piece constituents, five-piece constituents, and so on without bound. Unsurprisingly, the relevant sentences become very difficult to comprehend, and there is no clear consensus on whether any limitations on the unboundedness of discontinuity in the sense of (ii) reflects a grammatical constraint (e.g., Fanselow 2001) or is the result of processing complexity (e.g., Bader and Meng 2023); see Joshi et al. (2000) for important discussion of this issue.[3]

An analogous argument to Rambow's might be made on the basis of languages that allow extraction from embedded questions. Maling and Zaenen (1982) give examples such as (27): here, the embedded 'had written' clause corresponds to a three-component constituent, because two of its arguments end up displaced to the edges of separate clauses. If extraction is indeed allowed from embedded questions in general, then we would expect this to generalize up to cases where an arbitrarily number of wh-phrases are extracted from some clause, all of which surface in separate clause-edge positions.

(27) Hvilke bøker spurte Jon hvem som hadde skrevet?
what books asked John who that had written
"What books did John ask who had written?" (Maling and Zaenen, 1982)

Bulgarian, and related "multiple wh-fronting" languages, allow for similar examples such as (28). The potential for extending these configurations, apparently without bound, is indicated by the longer examples in (29). When four separate wh-phrases are extracted, there is a five-component constituent, as indicated by the underlining.

(28) Koj kontinent te popita učitelja koj e otkril ?
which continent you asked teacher who has discovered
"Which continent did the teacher ask you who discovered?" (Rudin, 1988)

[3] A reviewer points out that the DPs in (26) all bear distinct case markings, and examples without this property are less acceptable. This too could be put down to processing effects or to a grammatical constraint that bans scrambling of a DP past another DP that bears the same case (Flack, 2007; Erlewine and Lim, 2023).

(29) a. Koj donor se čudi Ivan koja kartina znae Maria na koj muzej dari
which donor refl wonder Ivan which painting know Maria to which museum donated

b. Koj donor se čudi Ivan koja kartina znae Maria na koj muzej pita Silvija koga dari
which donor refl wonder Ivan which painting know Maria to which museum asked Silvia when donated

(R. Pancheva, p.c.)

In all of these examples, the extracted elements originate in the same clause. Unboundedness of the relevant sort would follow if it were possible to add arbitrarily many wh-adjuncts to a single clause (like 'when' in (29)), or if simultaneous extractions from arbitrarily many separate clauses were possible. The latter possibility is suggested by (30), where a natural analysis would invoke simultaneous extraction from the object position of 'sell' and the subject position of 'know', although this is arguably obscured by the string-vacuity of the subject displacement.

(30) Vidjah edna kniga, kojato se čudja koj znae koj prodava
saw-1S a book which wonder-1S who knows who sells
"I saw a book which I wonder who knows who sells (it)" (Rudin, 1988)

Rudin (1988, p.457) also notes that when extracting out of an embedded question, relativization as in (30) is generally better than movement of an interrogative, although (28) indicates that the latter is possible.

As in the case of German scrambling, some of the details of the relevant patterns remain controversial, and for similar reasons. Nonetheless, our aim here is to present clearly-framed questions in a way that provides a link between specific empirical details and their significance for the formal and computational properties of grammar. As emphasized by Rambow for German scrambling, a language that allows arbitrarily "large" constituents (as measured by distinct components) cannot be described by any of the large class of grammar formalisms known as Linear Context-Free Rewrite Systems (LCFRS; Weir, 1988).[4] This class includes Minimalist Grammars (Stabler, 1997), Multiple Context-Free Grammars (Seki et al., 1991) and Multi-Component Tree Adjoining Grammar (Weir, 1988).

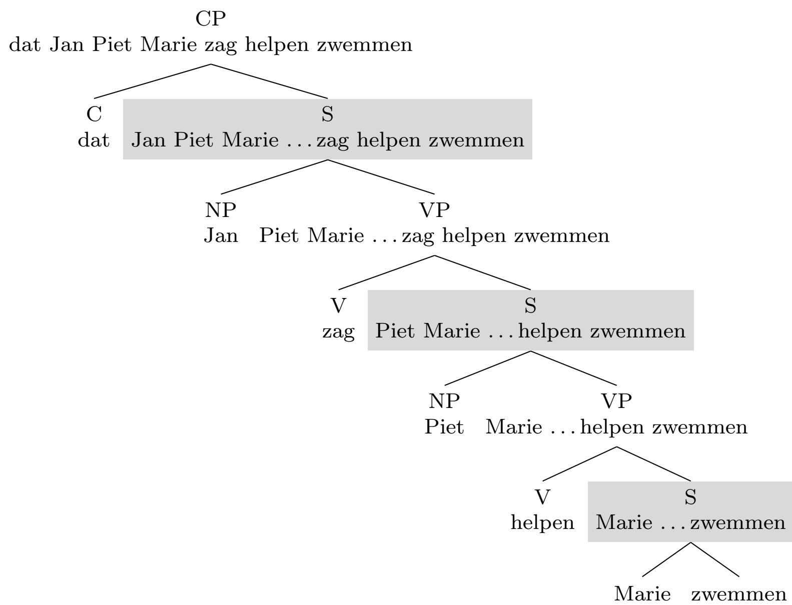
4.2 Unbounded parallel growth

There is a different sense of unboundedness that highlights an important distinguishing property of the Dutch crossing-dependencies example in (12). Rather than unboundedness in the discontinuous "width" of a particular constituent, as in the previous subsection, the relevant issue here concerns the way discontinuity can extend "upwards" through a tree of the sort introduced in (17)–(19).

A discontinuous-constituency analysis of the Dutch sentence in (12) is shown in (31). As foreshadowed earlier in discussion of (3e), the key point is that the embedded clauses can be represented by discontinuous constituents, e.g. 'Marie . . . zwemmen' and 'Piet Marie . . . helpen zwemmen'. This allows for the repeated addition of subjects and verbs to be "kept in sync" in the way that we noted was not possible with (8). Also notice that the *structure* of this tree, ignoring the way the surface string is composed, is identical to what would be assumed for the equivalent English and German sentences in (10) and (11).

[4] See also Miller (1991) for related arguments based on Scandinavian examples such as (27). Maling and Zaenen (1982) make similar arguments against the feasibility of accounting for such patterns in a framework that adds enriched category symbols to the machinery of simple phrase-structure grammars, and thereby is able to maintain the contiguous-constituency property rather than adopting discontinuous constituents. See section 5 for discussion of this kind of alternative to the discontinuous-constituency approach.

(31)



This example includes three subjects and three verbs, abstractly of the form $N_1N_2N_3V_1V_2V_3$; and these are assembled via a sequence of three S constituents of the forms $N_3 \ldots V_3$, $N_2N_3 \ldots V_2V_3$ and $N_1N_2N_3 \ldots V_1V_2V_3$. But if the pattern generalizes as is typically assumed for center-embedding (Miller and Chomsky, 1963), the construction is not limited to depth three (or any other depth). As a result, the steps that add a noun to one component and add a verb to the other component of these two-component constituents can be applied without bound. This means that the derivation of discontinuous constituents highlighted in (31) might extend unboundedly, with each step in the derivation *growing both components* of this two-piece constituent.[5]

Note the distinction between this kind of unbounded *parallel growth* and the notion of (unbounded) *degree of discontinuity* discussed in the previous subsection. Even in the face of the unbounded parallel growth needed for the Dutch construction in (31), no more than two components are ever required in any of the discontinuous constituents.

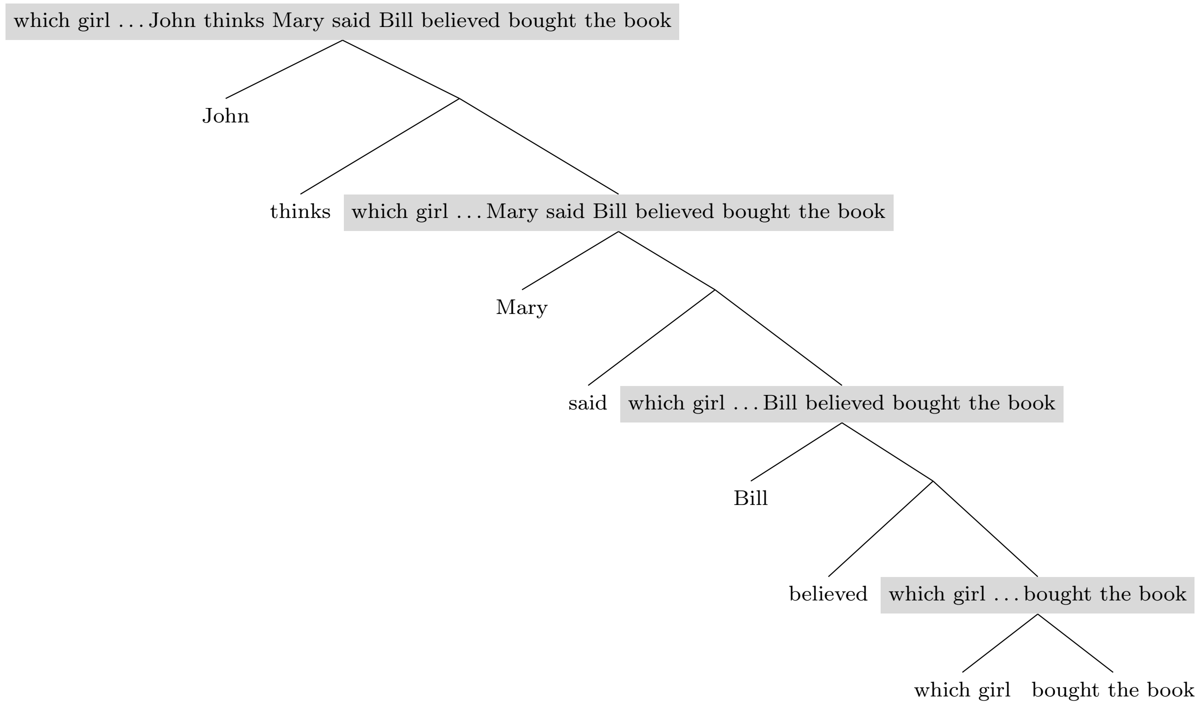
Note also the distinction between each of the two kinds of unboundedness considered so far, and the notion often called "unbounded displacement" that is exemplified in (32).

(32) a. I wonder which girl Bill believed bought the book
b. I wonder which girl Mary said Bill believed bought the book
c. I wonder which girl John thinks Mary said Bill believed bought the book

What this extensible pattern shows is that the displaced phrase 'which girl' can be kept separate across an unboundedly long sequence of two-component constituents; but it remains unchanged throughout, while only the other component grows, unlike (31) where *both* components grow.

[5]But see Manaster-Ramer (1987) for some relevant empirical details that we are glossing over here.

(33)



The unbounded discontinuity degree discussed in the previous subsection, and the unbounded parallel-growth exhibited by the Dutch construction, are both fundamentally incompatible, in a precise sense, with simple context-free phrase-structure grammars (CFGs). The more limited notion of "unbounded displacement" in (32), on the other hand, can be handled by a mere finite enrichment of the category system employed by a CFG, without altering the core computational mechanisms in any way.[6] The GPSG formalism (Gazdar et al., 1985) embodied essentially the hypothesis that this kind of finite enrichment would suffice for the description of all natural languages; it could therefore handle patterns like (32), but not the more computationally distinctive configurations exemplified by unbounded-width constructions such as German (26) and Bulgarian (29), nor parallel-growth constructions such as (31). See Maling and Zaenen (1982) and Bresnan et al. (1982) for discussion of these consequences for GPSG.

4.3 Well-nested vs. ill-nested composition

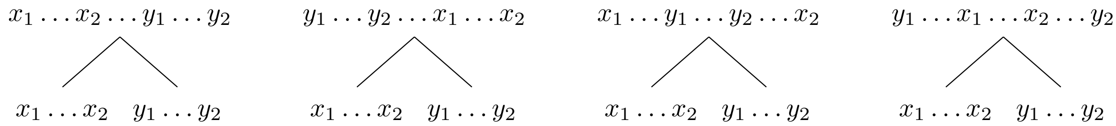
We now turn to a third point of variation on which discontinuous-constituency analyses can be classified: certain analyses only compose discontinuous constituents in ways that are known as *well-nested*. This is a condition on the way the components of two sister constituents can be intermingled.[7] Abstractly, if we consider composing a discontinuous constituent $x_1 \ldots x_2$ with a discontinuous constituent $y_1 \ldots y_2$, there are six possible ways that these four components might be arranged linearly in the parent constituent; as shown in (34), four of these six options satisfy the well-nestedness condition.[8] The well-nested arrangements are those where the two daughter constituents are arranged either (i) one entirely to the left and the other entirely to the right, or (ii) one entirely on the "inside" and the other entirely on the "outside". The ill-nested arrangements can be understood as those where the two x components end up separated by one but not both of the y components.

[6] As pointed out by an anonymous reviewer, analyses of coordination in terms of unboundedly wide flat structure are also compatible with CFGs if they are extended slightly to allow the Kleene star on the right-hand sides of rules. This extension is orthogonal to both the important notions of unboundedness discussed in the main text.

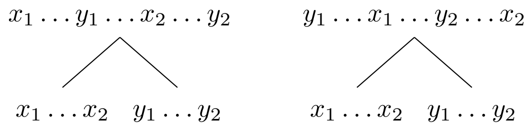
[7] The notion of well-nestedness originated in the dependency-parsing literature (Bodirsky et al., 2005; Kuhlmann and Nivre, 2006), and was adapted to the slightly different perspective on discontinuous constituency that we adopt here by Kanazawa (2009).

[8] There are only six possibilities since we assume, here and throughout the chapter, that the components of a discontinuous constituent are never "reordered", i.e., any derivation involving the discontinuous constituent $x_1 \ldots x_2$ will end up with x_1 to the left of x_2. This restriction to "non-permuting rules" is just for convenience and does not affect any of the formal issues that we discuss.

(34) a. Well-nested combinations of $x_1 \dots x_2$ and $y_1 \dots y_2$:



b. Ill-nested combinations of $x_1 \dots x_2$ and $y_1 \dots y_2$:



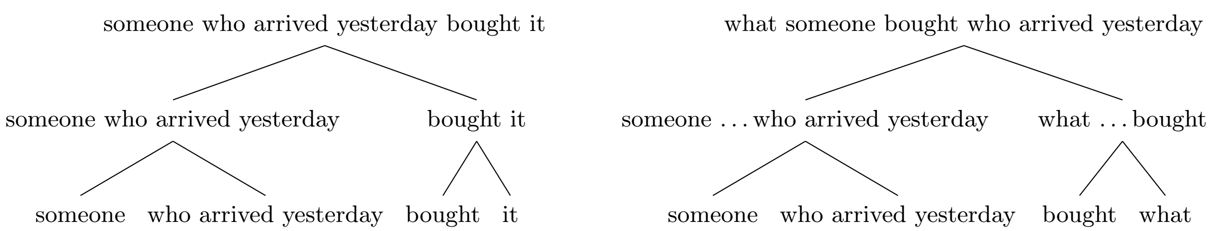
Note that the presence (or absence) of any discontinuities in the parent constituent are irrelevant to the classification in (34): when we write $y_1 \dots x_1 \dots x_2 \dots y_2$ at the root of node the last well-nested configuration, for example, this covers a range of possibilities including not only the four-piece constituent as shown but also the one-piece constituent $y_1x_1x_2y_2$, the two-piece constituents $y_1x_1 \dots x_2y_2$, $y_1 \dots x_1x_2y_2$ and $y_1x_1x_2 \dots y_2$, and the three-piece constituents $y_1 \dots x_1 \dots x_2y_2$, $y_1 \dots x_1x_2 \dots y_2$ and $y_1x_1 \dots x_2 \dots y_2$.

As a first illustration, consider the pair of sentences in (35). A natural analysis of this pair, and the relationship between them, might posit the trees in (36), with the second exhibiting both wh-movement and extraposition of the relative clause.

(35) a. (I know) someone who arrived yesterday bought it

b. (I know) what someone bought who arrived yesterday

(36)

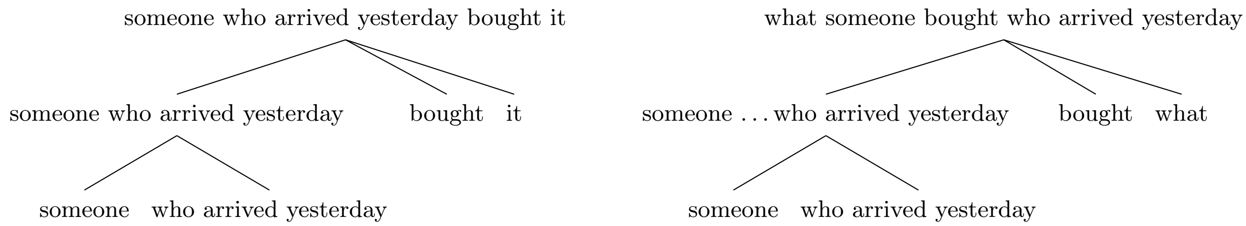


The combination of the two immediate daughters of the root node in the second tree is an instance of the ill-nested $y_1x_1y_2x_2$ pattern.

Analyses of natural language treebanks have discovered that a surprisingly small fraction (roughly in the range of 0.1% to 1%) of syntactic structures exhibit ill-nesting; even for languages with relatively free word order (e.g., Czech), and therefore significant numbers of discontinuous constituents in the relevant sense, there is an extremely strong tendency towards well-nestedness (Kuhlmann, 2013, p.379). The theoretical implications of these sorts of findings remain something of an open question. One possible explanation is that well-nestedness is a grammatically imposed constraint, or the result of the nature of the formal system used to derive syntactic structures. Alternatively, well-nestedness is not a grammatically significant property, but instead derives from speakers' attempts to ease cognitive load in production and/or processing. On this account, explored in Yadav et al. (2022), well-nested orders are more likely to be *chosen* by a speaker even if their grammar makes available a wide array of options, including ill-nested ones.

Another reason that the findings are difficult to interpret is that the question of whether a particular sentence exhibits ill-nesting depends significantly on relatively fine-grained details of how it is analyzed — more so than the distinctions that were the focus of the previous subsections. Returning to the two examples in (36), notice that the ill-nestedness relies crucially on the assumption that there is a constituent comprising (only) 'what' and 'bought', which is the sister of the discontinuous subject. If we instead adopted a ternary-branching structure for relating the verb to its subject and object, then there would be no ill-nesting.[9]

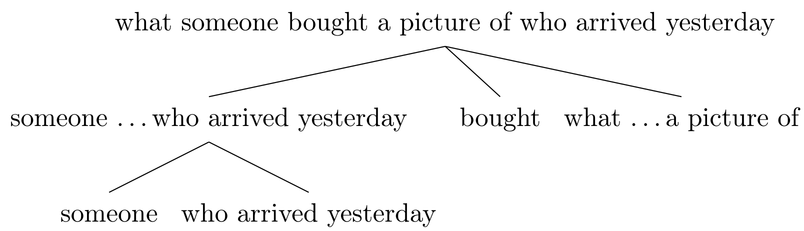
(37)



If we would like to find a more compelling illustration that English allows ill-nesting, with the same overall shape as (36) but without relying on the assumption that the verb and object form a constituent, one strategy is to construct a discontinuous *object* (rather than a discontinuous verb phrase).

(38) (I know) what someone bought a picture of who arrived yesterday

(39)

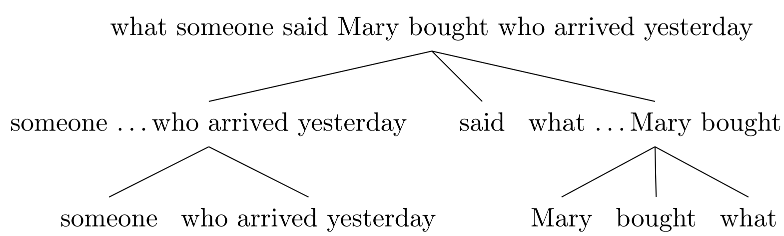


Here the two crucial sister constituents are the subject ('someone . . . who arrived yesterday') and the object ('what . . . a picture of').

A similar effect can be achieved by introducing extraction from an embedded clause.[10]

(40) (I know) what someone said Mary bought who arrived yesterday

(41)



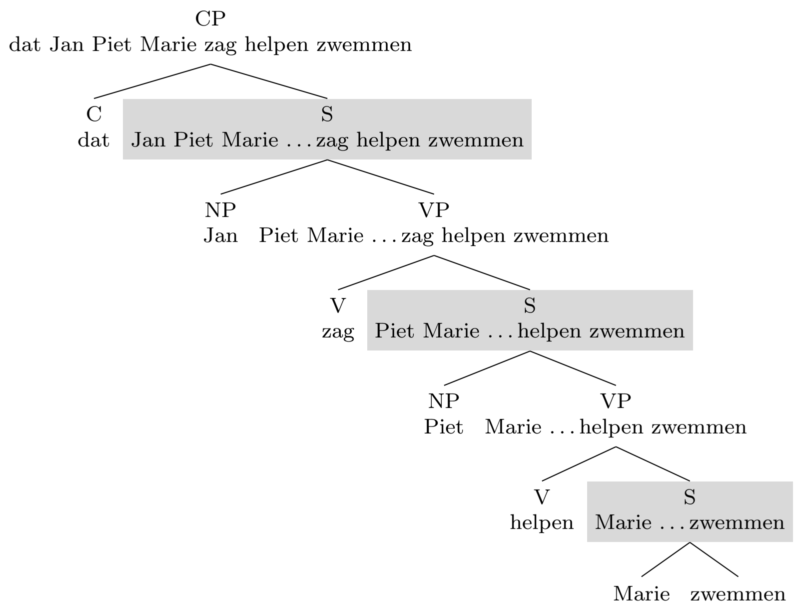
The main point of these examples is simply to illustrate the formal property known as well-nestedness and the kinds of analytical decisions about constituent structure that can interact with it. The issue of well-nestedness is a relatively recent addition to the landscape of formal classifications and as mentioned above its theoretical implications remain open.

Finally we should note an important distinction between the sense in which the ill-nested configurations in (34) exhibit "crossing" and the crossing-dependencies pattern discussed earlier from Dutch: the latter does *not* in fact exhibit ill-nestedness. To see this, consider the analysis we gave for the Dutch construction, repeated here.

[9] This is, in effect, the structure adopted both in TAG and in analyses based on dependency trees. In TAG, the fact that the verb and the object form a constituent to the exclusion of the subject is captured at the level of the derived tree, but not at the level of the derivation. In analyses based on dependency trees there is no such other level of representation, and the subject and the object are simply direct dependents of the verb. (In actual fact, neither a TAG derivation tree nor a dependency tree would show 'bought' as an explicit *daughter* of the relevant node here, but the crucial point is only that 'bought' is not treated as part of a discontinuous constituent with the object.)

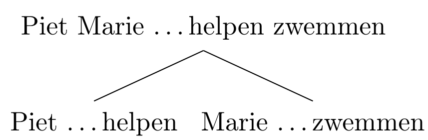
[10] Although we are making an assumption here that the matrix clause is "higher" in the derivation structure than the embedded clause; this is the reverse of the relationship assumed in analyses in TAG, which is one of the prominent formalisms that enforces well-nestedness.

(31)



Let us focus in particular on the bottom part of the tree, where 'Piet' and 'helpen' are added to the most embedded clause. Although these two added elements end up "intermingled" with 'Marie' and 'zwemmen' in a way that resembles the ill-nested configurations illustrated in (34), the crucial point is that 'Piet' and 'helpen' *are not components of a single discontinuous constituent*. A hypothetical composition step that *would* be ill-nested is the following.

(42)



But the two separate derivational steps in (31) that introduce 'Piet' and 'helpen' are well-nested: 'helpen' is placed entirely "inside" its sister S constituent, and 'Piet' is placed entirely to the left of its sister VP constituent. And this would not change if we adopted a ternary structure without a VP constituent: no two sister constituents in such a structure would combine in an ill-nested manner.

Well-nestedness makes an interesting cut in the space of grammatical formalisms that operate with some notion of discontinuous constituency: one class of formalisms including Tree Adjoining Grammar (Joshi, 1985) and Head Grammar (Pollard, 1984) can only express well-nested analyses, whereas the broader class of Linear Context-Free Rewrite Systems (Weir, 1988), which includes Minimalist Grammars (Stabler, 1997) and Multi-Component Tree Adjoining Grammar (Weir, 1988), allows ill-nestedness.[11]

Well-nestedness has also been proven to stand in an important relationship to a particular formalization of the idea of "completely free" word order. The formal language known as MIX (Bach, 1981) has a vocabulary of three words (say, 'flip', 'flop' and 'flap') and allows all and only those sentences where each word occurs the same number of times; this extreme freedom of word order was conjectured to be unlike any possible natural language in Joshi (1985). It was eventually proven that MIX *can* be generated by a grammar that allows two-piece discontinuous constituents, but *only* if constituents can be composed in ill-nested ways (Kanazawa and Salvati, 2012; Salvati, 2015). It seems very likely that the extreme freedom of word order exemplified by MIX is indeed unlike any possible human language, and if so then a ban on ill-nestedness may be part of the explanation for this.

[11] Kanazawa (2009) studies Well-Nested Multiple Context-Free Grammars (wn-MCFGs), which impose precisely the well-nestedness restriction on LCFRS, omitting the additional restriction to two-component discontinuous constituents in TAG and HG. Kanazawa et al. (2011) show that a variant of Minimalist Grammars where extraction from specifiers is disallowed generates a proper subclass of the languages of wn-MCFGs.

4.4 Copying operations

The last dimension of variation we discuss represents a slight departure from what we have looked at thus far, in the sense that it is not directly related to discontinuous constituency. But it is closely related: what discontinuous constituency draws our attention to is the possibility that the string yield of a certain (binary-branching, let us assume) constituent might be something other than simply the concatenation of a string that is the yield of one daughter constituent and another string that is the yield of the other daughter. What we have explored in the earlier sections of this chapter is the possibility that a daughter constituent might contribute multiple components, which surface as part of the parent constituent's yield in non-concatenative ways; a related possibility is that the parent constituent's yield might *use the contribution of some daughter constituent twice.*

One phenomenon that can be used to illustrate this possibility is Mandarin "A-not-A questions", which are formed by what looks intuitively like copying (Radzinski, 1990). A simple declarative sentence, such as (43a), can be negated by the addition of 'bu' before the predicate, as shown in (43b). The corresponding yes-no question includes two "copies" of the relevant predicate, separated by 'bu', as shown in (43c).

(43) a. ta zai jia
he/she at home
"He/she is at home"

b. ta bu zai jia
he/she not at home
"He/she is not at home"

c. ta **zai jia** bu **zai jia** ?
he/she at home not at home
"Is he/she at home?" (Radzinski, 1990, p.114)

According to Radzinski, there is no inherent limit on the size of the copied portion; (44) shows that phrases including multiple adjuncts can be copied in their entirety.

(44) a. ni **xihuan geng-da de pingguo** bu **xihuan geng-da de pingguo**
you like bigger GEN apple not like bigger GEN apple
"Would you like a bigger apple?"

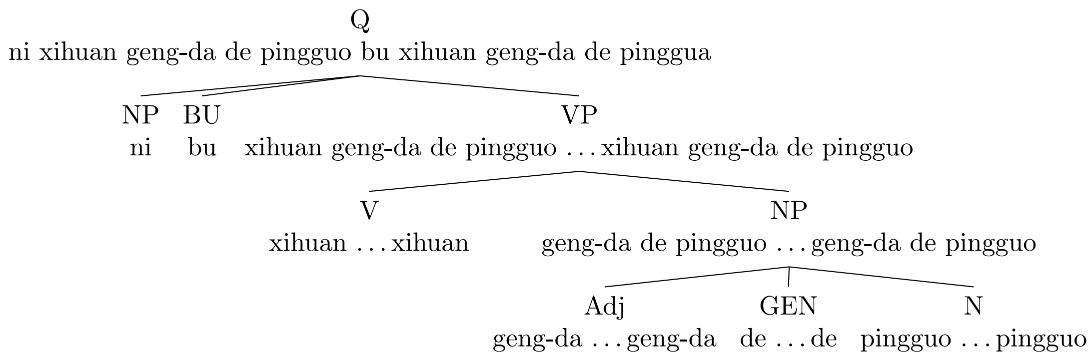
b. ni **xihuan geng-da geng-hao de pingguo** bu **xihuan geng-da geng-hao de pingguo**
you like bigger nicer GEN apple not like bigger nicer GEN apple
"Would you like a bigger nicer apple?" (Radzinski, 1990, p.117)

This kind of construction could be seen as analogous to the Dutch pattern discussed above, in the sense that if we consider the dependencies between the matched elements of each copy, they exhibit a crossing pattern.

(45) ni **xihuan geng-da de pingguo** bu **xihuan geng-da de pinggua**

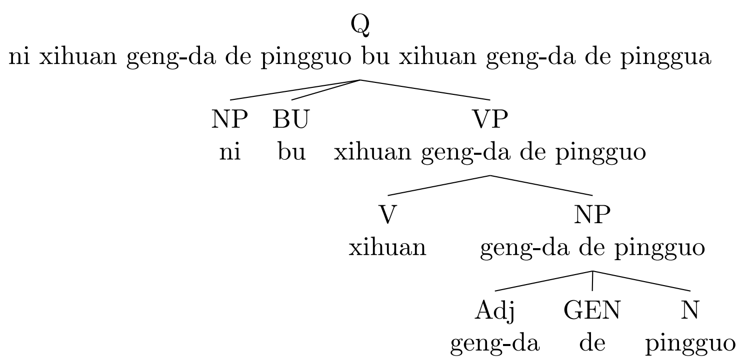
Pursuing this analogy, we might consider an analysis along the lines of (46), where the two copies are assembled in parallel as two components of a discontinuous constituent, just as the bundle of nouns and the bundle of verbs were assembled in parallel in (31).

(46)



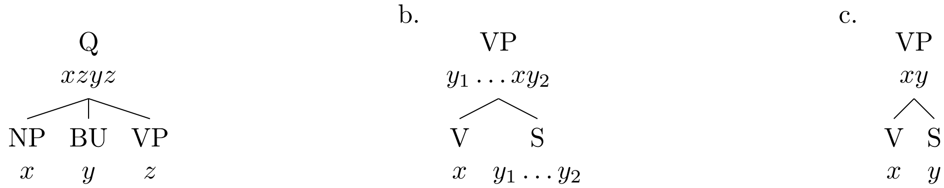
There are many details one might wonder about here, such as the "doubled" lexical items at the leaves of this tree.[12] Suppose though that there are reasonable ways to fill in those details. A more fundamental question is whether it makes sense to think of the VP constituent here as distinct from the VP constituent (meaning 'like a bigger apple') that would appear in a corresponding non-interrogative sentence. A more appealing way to think about (44) might be to suppose that the duplicated nature of the elements of the VP is purely a matter of the way the Q constituent is assembled out of an NP, a BU and a VP — and has nothing to do with the construction of the VP constituent itself. This line of thinking would lead to the tree in (47).[13]

(47)



The idea here is that the grammar of Mandarin includes a rule like (48a), which specifies that the string (z) contributed by the daughter VP node is used twice in constructing the string yield ($xzyz$) of the Q constituent. This is not dissimilar to the idea, implicit in the earlier parts of the chapter, that the grammar of Dutch, for example, includes a rule like (48b), which assembles a (discontinuous) VP constituent from a V and an S. What these two rules both have in common is that the "recipe" for constructing the yield of a parent constituent is *not* simply concatenating two daughter yields, as in the simple rule (48c) which might correspond to English.

(48) a. b. c.



Another phenomenon that is naturally described in terms of copying is that of "relativized predicates" in Yoruba (Kobele, 2006). Starting from a simple declarative such as (49a), copying the verb phrase to the front of the sentence produces an expression meaning roughly "the fact that X", as shown in (49b). The examples in (50)

[12]It is also interesting to note that this kind of derivation repeatedly combines such doubled elements in ill-nested ways, unlike the Dutch derivation in (31).

[13]See Stabler (2004, pp.708–709) for related discussion.

show the same operation applying to larger verb phrases, one augmented with a serial verb and the other with an adjective modifying the object.

(49) a. Jimọ ọ ra adię
Jimọ HTS buy chicken
"Jimọ bought a chicken"

b. Rira adię ti Jimọ ọ ra adię
buying chicken TI Jimọ HTS buy chicken
"the fact/way that Jimọ bought a chicken" (Kobele, 2006, p.214)

(50) a. Rira adię se ti Jimọ ọ ra adię se
buying chicken cook TI Jimọ HTS buy chicken cook
"the fact/way that Jimọ bought a chicken to cook" (Kobele, 2006, p.215)

b. Rira adię ti o gọ ti Olu ra adię ti o gọ
buying chicken TI 3S dumb TI Olu buy chicken TI 3S dumb
"the fact/way that Olu bought the stupid chicken" (Kobele, 2006, p.258)

The Yoruba construction provides a perhaps even more compelling case for a copying analysis like (47) rather than (46), in the sense that it is very natural to suppose that there is only one 'buy a chicken' phrase in (49b) given the interpretation of this construction. To the extent that the Mandarin examples in (44) might be argued to have a "contrastive coordination" analysis ("Would you like a bigger apple or not like a bigger apple?"), it might be less clearly the case that there is only one 'like a bigger apple' phrase there.[14] But no such alternative seems feasible for the interpretation of the Yoruba relativized predicates.

If we grant that copying operations sometimes occur in natural languages, then a subsequent question that is potentially significant from a computational perspective is whether a string that has been produced by copying can itself be copied. All else being equal, if indeed these sorts of constructions are handled via a copying operation rather than treated as a form of cross-serial dependencies, it might seem somewhat odd for copying to be restricted in a way that *prevents* it from applying to its own output[15] — this would require in some sense tracking a distinction between those constituents whose derivational history includes a copying operation, and those whose history does not. This is a part of the argument made by Kobele (2006), who suggests that, since the Yoruba construction provides good evidence for copying operations, we can conclude that the toolkit of natural language grammars includes this option, and we should therefore expect the recursive-copying option to not be ruled out by anything. This corresponds to arguing for the Parallel Multiple Context-Free Grammar (PMCFG) formalism rather than the simpler Multiple Context-Free Grammar (MCFG) alternative (Seki et al., 1991); the latter is equivalent to the LCFRS formalism mentioned above.

5 An alternative: Enriched categories instead of discontinuous constituency

In section 3 we identified the basic notion of *discontinuous constituency* as a way to meet the challenges posed by both the Dutch crossing-dependencies pattern in (12) and the "non-canonical" word-orders in (13), (14) and (15). We hope to have shown, through the rest of the chapter to this point, that this can be a useful perspective to adopt for classifying various kinds of "interesting" syntactic phenomena. But the discontinuous-constituency approach is not the only logically possible response to the relevant challenges; and although the various grammar formalisms that we have referred to in the preceding sections can be construed as specific implementations of the discontinuous-constituency idea, there are other well-known formalisms that take a different tack. To complete the picture, in this section we briefly describe the alternative route.

Relative to the starting point of simple context-free phrase-structure grammar, what we have described as the discontinuous-constituency approach modifies the relationship between the groupings of words into constituents

[14] Although see Radzinski (1990, pp.120–121) for arguments against this.

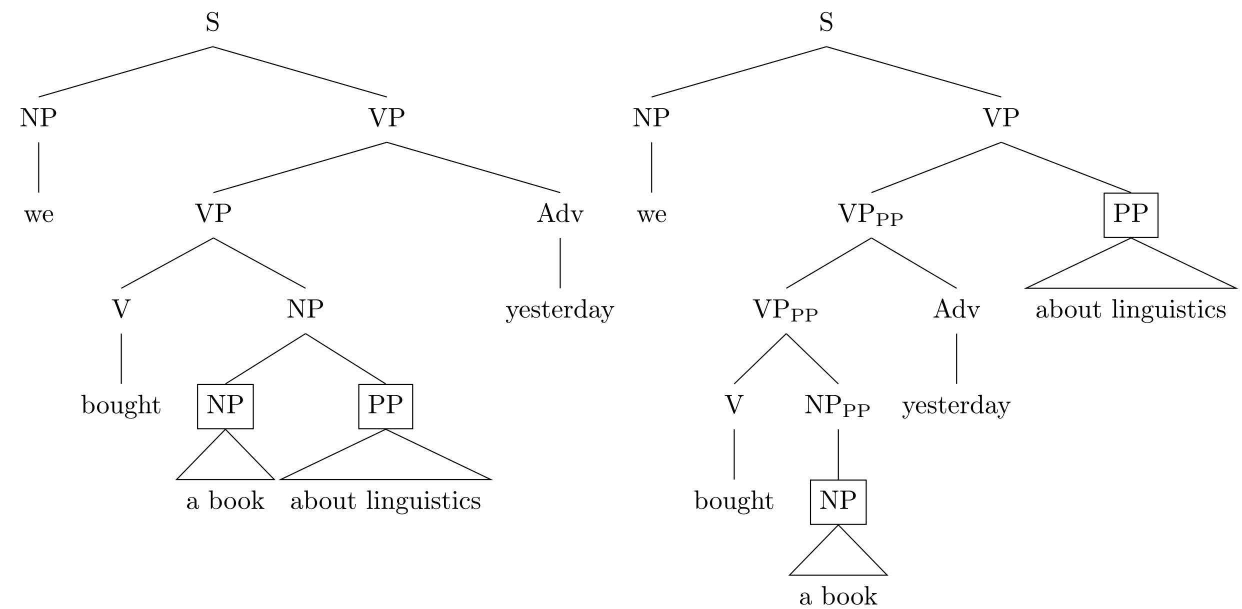
[15] Though see Johnson (1972) for a similar constraint that has been assumed to hold in phonological derivations. As Johnson observes, the imposition of this constraint has the effect of limiting the power of apparently context-sensitive rewriting to finite state transductions.

and surface word order; what remains unchanged is the relationship between constituency and syntactic dependencies, which was first introduced in discussion of (4) and (5). The common constituency of the two trees in (18), for example, reflects the commonality in the dependencies entered into by ‘the girl’ and ‘which girl’ in the two corresponding sentences (leaving aside the different ways in which these phrases contribute to the surface word order).

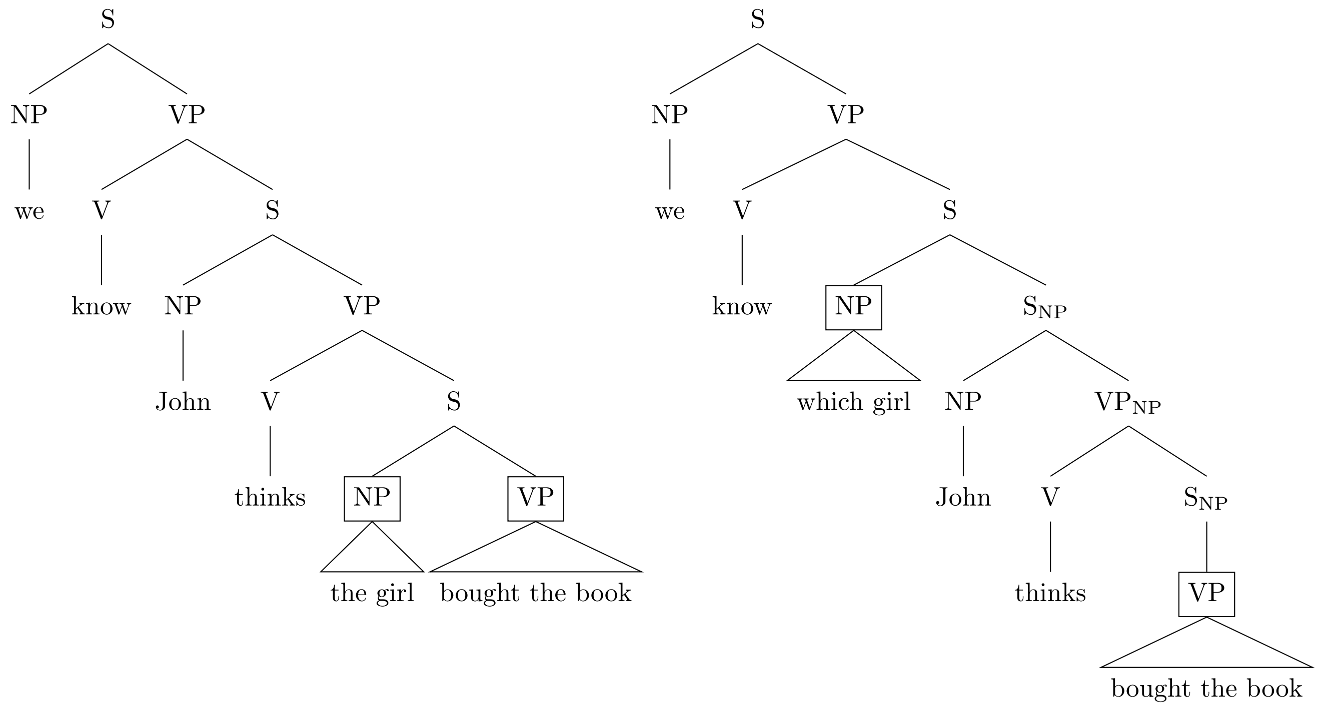
The alternative is what we will call the *enriched-category* approach. This approach leaves unchanged the relationship between constituency and surface word order, so that sister constituents are always combined through simple concatenation. Instead the relationship between constituency and syntactic dependencies is modified: the “non-canonical” word-orders in (13), (14) and (15) must involve constituency structures that diverge from those of their canonical counterparts where the underlined portions are contiguous.

One way to flesh out this approach would be to assign structures like the following; compare with (17), (18) and (19).

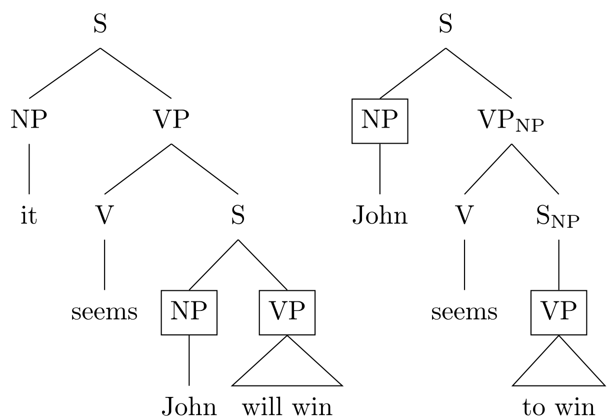
(51)



(52)



(53)



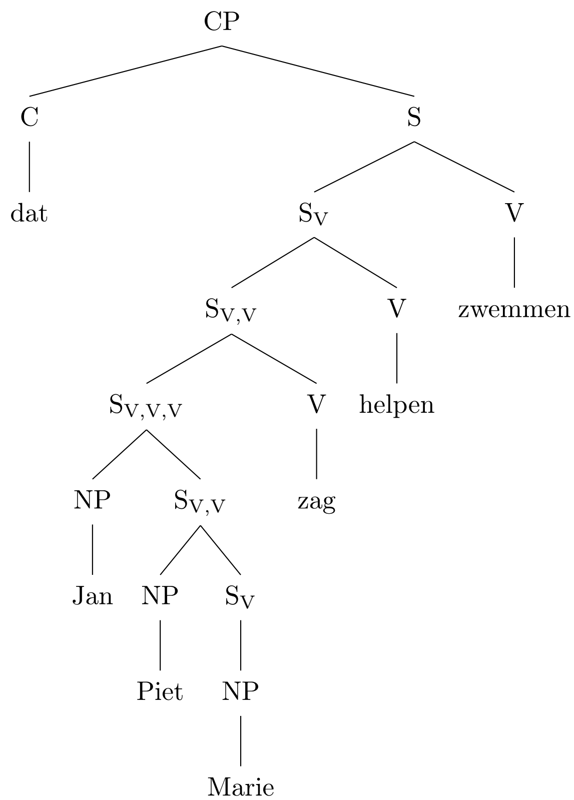
Unlike the discontinuous-constituency trees throughout this chapter, these maintain the familiar relationship between constituency and word order. But in each pair of trees, there is a certain relationship between two constituents that is expressed differently in the first tree than in the second tree; the two relevant constituents are boxed in each tree. The enriched categories — here using subscripts, but the specific notation does not matter — allow for information to be "passed along" through the tree, and this information is what enables the important relationship between 'which girl' and 'bought the book', for example, to be "reconstructed"; this in turn brings the relationship between these two constituents into line with that between 'the girl' and 'bought the book' in the non-interrogative counterpart.

In the discontinuous-constituent analyses in (18), on the other hand, the commonality between the two subject–verb relationships is expressed transparently in the trees (which therefore do not require enriched categories), and the difference in their surface realizations requires some additional machinery (namely non-concatenative operations).

A formalism that adopts the enriched-categories route only extends the generative capacity of simple context-free grammars if the enriched categories that it introduces are unbounded in number. As mentioned in discussion of (32), the GPSG formalism (Gazdar et al., 1985) allows only a finite enrichment of the category system, and therefore cannot generate the Dutch crossing-dependencies pattern, for example. Generating these non-context-free patterns without adopting discontinuous constituency necessarily requires a system that allows for infinitely many categories.

The tree in (54) illustrates the essentials of how this could work for the Dutch pattern: an unbounded number of enriched categories comes from allowing subscripts to *accumulate*, and the multiple separate subscripts on $S_{V,V,V}$ allow for the "reconstruction" of the multiple subject–verb dependencies, all of which require passing information along the tree.[16]

(54)



Formalisms that allow for *unbounded* enrichment of categories and therefore (unlike GPSG) can handle these strictly non-context-free phenomena include Combinatory Categorial Grammar (Steedman, 2000), Linear Indexed Grammar (Gazdar, 1988) and Lexical Functional Grammar (Bresnan et al., 1982).

6 Conclusion

This chapter has provided an overview of the kinds of patterns that "push the boundaries" of formal syntactic complexity in various ways. We adopted *discontinuous constituency* as a tool for characterizing computationally significant dimensions of grammatical variation in a unified but relatively theory-neutral manner. In particular, we have identified four such dimensions:

Number of Components How many components does a language allow to occur in a single discontinuous constituent? Is there a finite bound on the number of such components?

Bounded vs. Unbounded Parallel Growth Can multiple components of a discontinuous consistuent grow without bound, or is unbounded growth limited to a distinguished component?

Well-nested vs. Ill-nested Composition Must discontinuous constituents compose in a well-nested manner?

[16] This tree is intended only as a rough illustration of the key idea; many details of the relevant structures are controversial, although the tree here is based on the analysis of Kroch and Santorini (1991).

Copying Operations Are cases of apparent syntactic reduplication the result of a basic copying operation or of parallel derivations?

Each of these dimensions might be resolved in a language-universal manner (e.g., it may be that all grammars impose well-nestedness) or they might constitute a locus of cross-linguistic variation. In either case, these dimensions bear on the suitability of various grammar formalisms as models of natural language syntax; we have pointed out these consequences along the way.

One of our main goals here has been to frame the relevant computational issues in an accessible way that brings them closer to everyday syntactic analysis. We hope that our formulations facilitate productive interaction between the empirical and formal sides of work on natural language syntax, as will be necessary to resolve the status of these dimensions. For example, the *number of components* dimension is roughly parallel to the widely studied question of whether there is a bound on the number of "simultaneous extractions" that a language permits. Nonetheless, there remain numerous critical open empirical questions. It would be useful to document extraction possibilities in terms like our number of components dimension. Similarly, determining the status of *well- vs. ill-nesting* will require careful atteption to fine-grained aspects of the syntactic analysis of the relevant constructions, and whether a given relationship is syntactically enforced or the result of an interpretive relation (Abeillé, 1994). Finally, there is the question of whether natural language syntax invokes *copying* operations of the sort often invoked in Mandarin A-not-A constructions, or whether these are better handled as cross-serial dependencies that result from *unbounded parallel growth*. It will be important to seek out empirical signatures of these two kinds of derivations, and to understand their relationships to proposals in syntactic theory regarding the linearization of multidominance structures (e.g. Citko, 2005) and structures involving copies (e.g. Nunes, 2004).

References

Abeillé, A. (1994). Syntax or Semantics? Handling Nonlocal Dependencies with MCTAGs or Synchronous TAGs. *Computational Intelligence*, 10(4):471–485.

Bach, E. (1981). Discontinuous constituents in generalized categorial grammars. In *Proceedings of the 11th Annual Meeting of the North East Linguistic Society*, pages 1—12.

Bader, M. and Meng, M. (2023). Processing noncanonical sentences: effects of context on online processing and (mis)interpretation. *Glossa Psycholinguistics*, 2(1):1–45.

Becker, T., Rambow, O., and Niv, M. (1992). The Derivational Generative Power of Formal Systems or Scrambling is Beyond LCFRS. Technical Report IRCS-92-38, University of Pennsylvania.

Bodirsky, M., Kuhlmann, M., and Möhl, M. (2005). Well-nested drawings as models of syntactic structure. In *10th Conference on Formal Grammar and 9th Meeting on Mathematics of Language.*

Bresnan, J., Kaplan, R. M., Peters, S., and Zaenen, A. (1982). Cross-serial Dependencies in Dutch. *Linguistic Inquiry*, 13(4):613–635.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, 2(3):113–124.

Citko, B. (2005). On the nature of merge: External merge, internal merge, and parallel merge. *Linguistic Inquiry*, 36(4):475–496.

Erlewine, M. Y. and Lim, J. (2023). Deriving scrambling with word order freezing. *Glossa*, 8(1).

Fanselow, G. (2001). Features, θ-roles and free constituent order. *Linguistic Inquiry*, 32(3):405–437.

Flack, K. (2007). Ambiguity avoidance as contrast preservation: Case and word order freezing in japanese. In *UMass occasional papers in linguistics 32: Papers in optimality theory III*, pages 57–89. GLSA, University of Massachusetts.

Gazdar, G. (1988). Applicability of indexed grammars to natural languages. In Reyle, U. and Rohrer, C., editors, *Natural Language Parsing and Linguistic Theories*, pages 69–94. Springer.

Gazdar, G., Klein, E. H., Pullum, G. K., and Sag, I. A. (1985). *Generalized Phrase Structure Grammar*. Harvard University Press, Cambridge, MA.

Johnson, C. D. (1972). *Formal Aspects of Phonological Description*. Mouton, The Hague.

Joshi, A. (1985). How much context-sensitivity is necessary for characterizing structural descriptions? In Dowty, D., Karttunen, L., and Zwicky, A., editors, *Natural Language Processing: Theoretical, Computational and Psychological Perspectives*, pages 206–250. Cambridge University Press, New York.

Joshi, A. K., Becker, T., and Rambow, O. (2000). Complexity of scrambling: a new twist to the competence-performance distinction. In Abeillé, A. and Rambow, O., editors, *Tree Adjoining Grammars*, pages 167–181. CSLI Publications, Stanford, CA.

Kanazawa, M. (2009). The Pumping Lemma for Well-Nested Multiple Context-Free Languages. In Diekert, V. and Nowotka, D., editors, *Developments in Language Theory*, volume 5583 of *Lecture Notes in Computer Science*. Springer, Berlin Heidelberg.

Kanazawa, M., Michaelis, J., Salvati, S., and Yoshinaka, R. (2011). Well-Nestedness Properly Subsumes Strict Derivational Minimalism. In Pogodalla, S. and Prost, J.-P., editors, *Logical Aspects of Computational Linguistics*, pages 112–128. Springer, Berlin Heidelberg.

Kanazawa, M. and Salvati, S. (2012). MIX Is Not a Tree-Adjoining Language. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 666–674.

Kobele, G. M. (2006). *Generating Copies: An investigation into Structural Identity in Language and Grammar*. PhD thesis, UCLA.

Kroch, A. and Santorini, B. (1991). The Derived Constituent Structure of the West Germanic Verb-Raising Construction. In Freidin, R., editor, *Principles and Parameters in Comparative Grammar*, pages 269–338. MIT Press.

Kuhlmann, M. (2013). Mildly non-projective dependency grammar. *Computational Linguistics*, 39(2):355–387.

Kuhlmann, M. and Nivre, J. (2006). Mildly non-projective dependency structures. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 507–514, Sydney, Australia. Association for Computational Linguistics.

Maling, J. and Zaenen, A. (1982). A Phrase Structure Account of Scandinavian Extraction Phenomena. In Jacobson, P. and Pullum, G. K., editors, *The Nature of Syntactic Representations*, pages 229–282. Reidel.

Manaster-Ramer, A. (1987). Dutch as a Formal Language. *Linguistics and Philosophy*, 10(2):221–246.

Miller, G. and Chomsky, N. (1963). Finitary models of language users. In Luce, R. D., Bush, R. R., and Galanter, E., editors, *Handbook of Mathematical Psychology, Volume 2*, pages 419–491. John Wiley and Sons, New York.

Miller, P. H. (1991). Scandinavian extraction phenomena revisited: Weak and strong generative capacity. *Linguistics and Philosophy*, 14(1):101–113.

Nunes, J. (2004). *Linearization of Chains and Sideward Movement*. MIT Press, Cambridge, MA.

Pollard, C. (1984). *Generalized phrase structure grammars, head grammars and natural language*. PhD thesis, Stanford University.

Radzinski, D. (1990). Unbounded Syntactic Copying in Mandarin Chinese. *Linguistics and Philosophy*, 13:113–127.

Rambow, O. (1994). *Formal and computational aspects of natural language syntax*. PhD thesis, University of Pennsylvania.

Rizzi, L. (1990). *Relativized Minimality*. MIT Press, Cambridge, MA.

Rudin, C. (1988). On multiple questions and multiple wh-fronting. *Natural Language and Linguistic Theory*, 6:445–501.

Salvati, S. (2015). MIX is a 2-MCFL and the word problem in $\mathbf{Z}^2$ is captured by the IO and OI hierarchies. *Journal of Computer and System Sciences*, 81(7):1252–1277.

Seki, H., Matsumara, T., Fujii, M., and Kasami, T. (1991). On multiple context-free grammars. *Theoretical Computer Science*, 88:191–229.

Stabler, E. P. (1997). Derivational minimalism. In Retoré, C., editor, *Logical Aspects of Computational Linguistics*, volume 1328 of *LNCS*, pages 68–95, Berlin Heidelberg. Springer.

Stabler, E. P. (2004). Varieties of crossing dependencies: structure dependence and mild context sensitivity. *Cognitive Science*, 28:699–720.

Steedman, M. (2000). *The Syntactic Process.* MIT Press, Cambridge, MA.

Weir, D. (1988). *Characterizing mildly context-sensitive grammar formalisms.* PhD thesis, University of Pennsylvania.

Yadav, H., Husain, S., and Futrell, R. (2022). Assessing Corpus Evidence for Formal and Psycholinguistic Constraints on Nonprojectivity. *Computational Linguistics*, 48(2):375–401.