**Discontinuous Constituents**

§1. Introduction: Definition and Examples.

Let \( A, B, C, D \) be four constituents of some word or phrase. \( A \) is said to be **discontinuous** if and only if (i) \( B \) is a constituent of \( A \), (ii) \( C \) is a constituent of \( A \), (iii) \( D \) is not a constituent of \( A \), and (iv) \( D \) is linearly ordered between \( B \) and \( C \). Take for example the verb phrase *woke your friend up*. Suppose this phrase contained three constituents *woke*, *your friend*, and *up*. If *woke up* is a constituent of this phrase as well, then it would be discontinuous, as (i) *woke* would be a constituent of *woke up*, (ii) *up* would be a constituent of *woke up*, (iii) *your friend* would not be a constituent of *woke up*, and (iv) *your friend* would be linearly ordered between *woke* and *up*. One natural way of visualizing discontinuous constituents is as trees with crossing branches:

![Tree Diagram](image)

Phrasal verbs aside, constructions that have been argued to contain discontinuous constituents involve parentheticals (2) and other nonrestrictive relative clauses; extraposed (3) and other shifted constructions; scrambling in Latin (4) and other free word order languages; degree modifiers (5)—and so on.\(^1\)

(2) \[ \text{We } [B \text{ talked } [D \text{ of course } [C \text{ about politics}]]] \]

(3) \[ [B \text{ Men } [D \text{ entered } [C \text{ who were wearing black suits}]]] \]

(4) \[ [B \text{ Huic } [D_1 \text{ ego } [D_2 \text{ me: } [C \text{ bello: } \text{ ducem profiteor.}]}}} \]

\[ \text{this } \text{I myself war leader proclaim} \]
\[ \text{‘For this war I proclaim myself leader.’} \]

(5) \[ \text{John was } [B \text{ so } [D \text{ tall } [C \text{ that he hit his head}]]] \]

Moreover, *prima facie* examples of discontinuous morphological constituents come from infixation (cf. the English expletive *Cali-friggin'-fornia*), circumfixation

---

\(^1\) To enhance perspicuity, the categories of these examples have been relabeled in accordance with the definition of discontinuous constituent given in the text.
(cf. the German participle ge-frag-t ‘asked’), and templatic morphology like that of Arabic verbs whose interleaved \( V_1 V_2 \) and \( C_1 C_2 C_3 \) morphological tiers are two constituents that render each other discontinuous:

(6) \( [C_1 k] [V_1 t] [C_2 a] [V_2 a] [C_3 b] \) ‘he wrote’

\[\text{§2. Early Transformational Grammars}\]

Discontinuous constituents were first acknowledged by American structuralists. See for example Pike (1943, §4.12), Harris (1945), Wells (1947, §§55-62), and Hockett (1958, §§17.6, 21.4). And discontinuous constituents were allowed in early generative grammars as well. For one thing, Chomsky (1955, 405) admitted permutations—elementary transformations whose only effect on a string was to reorder the terms of that string. Take for example the permutation in (7).

(7) \((\text{ROOT, PARTICLE, NP}) \Rightarrow (\text{ROOT, NP, PARTICLE})\)

This permutation will carry the string \( \text{woke up your friend} \) into the string \( \text{woke your friend up} \). And, since permutations do nothing but reorder constituents, the constituents of the input will be preserved in the output. The permutation in (7) will therefore map the continuous tree in (8) into the discontinuous one in (1).

\[\text{VP} \]

\[\text{VERB} \]

\[\text{ROOT} \quad \text{PARTICLE} \]

\[\text{woke} \quad \text{up} \quad \text{your friend} \]

Similarly, the early generative grammar developed in Yngve (1960, 448ff) allowed constituent structure rules of the following form.

(9) \( A \rightarrow B \sim C \)

Rules of the form in (9) were intended to map strings of the form \( XADY \) into strings of the form \( XBDCY \), thus “wrapping” a phrase \( [A \ B \ C] \) around a phrase \( D \)—but without turning \( D \) into a new constituent of \( A \). Take for example the following rule.

(10) \( \text{VERB} \rightarrow \text{ROOT} \sim \text{PARTICLE} \)
The rule in (10) can wrap the constituents of VERB around an ensuing NP as indicated in (11).

\[
(11) \quad \begin{array}{c}
VP \\
\text{VERB} \quad \text{NP} \\
\rightarrow \\
\text{VERB} \quad \text{ROOT} \quad \text{NP} \quad \text{PARTICLE}
\end{array}
\]

In so doing, it will generate a discontinuous constituent.

§3. Classical Transformational Grammars

Although early generative grammars could accommodate discontinuous constituents, Postal (1964, 67-70) made an early case that analyses involving discontinuous constituents could—and should—be replaced by accounts that appealed only to continuous ones. In short, his case was that only analyses in terms of continuous constituents could (i) account for the intuitive relations between phrases, (ii) avoid stating selectional restrictions separately for phrases, and were (iii) generated by well-defined rules which (iv) provided a well-defined algorithm for the assignment of derived structure. Moreover, analyses in terms of discontinuous constituents called for (v) grammars of unknown generative power.

Consider once again the verb phrase woke your friend up. Notice that this phrase can be derived from (8) by deleting the particle up from within the verb and then restoring it, as shown in (12), after the direct object (and as an immediate constituent of the verb phrase).

\[
(12) \quad \begin{array}{c}
VP \\
\text{VERB} \quad \text{NP} \quad \text{PARTICLE} \\
\text{ROOT} \quad \text{woke} \quad \text{your friend} \quad \text{up}
\end{array}
\]

Crucially, this “cut-and-paste” analysis of the verb phrase woke your friend up appeals only to continuous constituents. Moreover, since it is derived from woke up your friend, it accounts for the intuitive relation between these two phrases and preserves, in so doing, the selectional restrictions of its source (thus eliminating
the need for stating the same selectional restrictions twice). Furthermore, it does this, as all transformations do, through a well defined algorithm for the assignment of derived structure.

Postal’s case against discontinuous constituents carried the day. Consequently, neither the permutations of Chomsky (1955) nor the wrapping rules of Yngve (1960) were seriously pursued in the ensuing literature. And one year after the publication of Postal’s work, Chomsky (1965, 144) went on to suggest the elimination of permutations in favor of deletions, substitutions, and adjunctions. In addition to being redundant, permutations were thought to lead to unnecessary complications in the theory of derived constituent structure (Chomsky, 1965, 226).

Now, if the notion of discontinuous constituents is idle, then constituent structure trees with crossing branches should be banned together with the constituents they were intended to represent. This ban on discontinuous constituent trees was not long in coming. As defined in Wall (1972, 149), constituent structure trees satisfy both an Exclusivity Condition and a Nontangling Condition. The Exclusivity Condition requires that every two nodes of a tree be ordered either in terms of dominance or in terms of precedence (but never in terms of both). The Nontangling Condition requires that, if some node precedes another, then all the nodes dominated by the former must precede all the nodes dominated by the latter.

Taken in conjunction, these conditions ban the tree in (1). For take the nodes VERB and NP found therein. Since they are not related in terms of dominance, they must be related in terms of precedence (Exclusivity). Thus, all the nodes they dominate must be related in terms of precedence as well (Nontangling). Yet, they are not; whereas one of the nodes dominated by VERB precedes the NP node, the other one follows it. (1) is therefore incompatible with the definition of constituent structure tree.

It is a testament to the success of Postal’s arguments that Exclusivity and Nontangling continued to be part of the definition of constituent structure trees for more than two decades (see Partee et al., 1993, §16.3.3). And the incompatibility between the notion of discontinuous constituency and the notion of constituent structure tree reinforced the general resistance to discontinuous constituents that has continued to this day.

Yet, contrary to what one might think, none of the above banished the term discontinuous constituent from the vocabulary of theoretical linguistics; all it did was justify assigning it a new sense. To be more specific, let A, B, C, D be four con-
stituents of a word or phrase. Let $B$, $D$, $C$ be three constituents of $A$. Let $B$, $D$, $C$ occur in that linear order. $A$ is discontinuous (in the new sense) if and only if $A$ is derived, transformationally, from a constituent $A'$ such that (i) $B$ is a constituent of $A'$, (ii) $C$ is a constituent of $A'$, (iii) $B$ precedes $C$ under $A'$ and (iv) $D$ is not a constituent of $A'$ (cf. Postal 1964, 67). The new sense of the term discontinuous constituent can be illustrated with the VP in (12). Just let $A = VP$, $B = VERB$, $D = NP$, $C = PARTICLE$, and assume, with Postal, that the VP in (12) is derived, transformationally, from the VP in (8) (this VP would therefore be the constituent $A'$ of the definition).

We thus arrive at two distinct senses for the term discontinuous constituent. If necessary, we can distinguish between them by saying that a constituent is truly discontinuous if it is discontinuous in the original sense and seemingly discontinuous if it is discontinuous in the sense described in the preceding paragraph. Note that this choice of terms is justified by Postal’s analysis, which regards truly discontinuous constituents as illusory.

Seen in retrospect, Postal’s arguments against truly discontinuous constituents warranted closer scrutiny. Consider first discontinuous constituents generated by the permutation transformations of Chomsky (1955). Being transformations, permutations can assert relations between their inputs and outputs (and thus account for the intuitive relations between phrases); being equivalent to cut-and-paste transformations, permutations can ensure that any selectional restrictions preserved by said transformations will be preserved by permutations as well (so selectional restrictions do not have to be stated separately for the inputs and the outputs of a permutation); being among the elementary transformations of Chomsky (1955), permutations are well-understood formal devices with well-defined algorithms for the assignment of derived structure. And, since permutations have the same effect on strings as cut-and-paste transformations, the weak generative power of permutations will be the same as the weak generative power of the equivalent cut-and-paste transformations. In short, the truly discontinuous constituents generated by permutation transformations do not lead to any of the shortcomings Postal feared.

And similar points can be made about the truly discontinuous constituents generated by the wrapping rules of Yngve (1960), as these rules are string-equivalent to permutations (and hence to cut-and-paste transformations). The only thing to keep in mind is that wrapping rules are couched in a general framework that requires them to expand the nonterminal symbols of a tree (or a derivation) from left to right. Thus, as shown in (11), VERB must wrap around NP before the latter is expanded—in this case, as a determiner followed by a nominal. This en-
sures that the wrapping rules are formal devices with well-defined algorithms for derived structure.

Beyond this, notice that Postal’s cut-and-paste analysis requires that any constituent that is ordered between two constituents of a seemingly discontinuous constituent A must itself be a constituent of A. Take for example the NP in (12). It is ordered between two constituents of VP and is, accordingly, itself a constituent of VP. Naturally, no such requirement holds vis-à-vis truly discontinuous constituents. See for example the NP in (1), which is ordered between two constituents of a VP without being a constituent of that VP. It follows that the external constituent structure of the material intruding in discontinuous constituents may provide grounds for deciding whether these constituents are truly discontinuous or only seemingly so.

Evidence bearing on such a constituent structure was presented in McCawley (1982). Consider for example the VP anaphora in (13).

(13) John talked, of course, about politics…
    a. …and Mary, you’ll be surprised to hear, did too.
    b. …but Mary, you know, would never do that.
    c. …which is something that Mary would perhaps never do.

In all these examples, the antecedent is only *talk about politics*; it is not *talked, of course, about politics* (nor is it just *talk*, for that matter). This means that the parenthetical *of course* is ordered between the constituents of a VP *talk about politics* without being itself a constituent of that VP—or that *talk about politics* forms a truly discontinuous constituent.

Further evidence for truly discontinuous constituents comes from constraints on extraction. For, as is well known, island constraints prevent material from being extracted from complex noun phrases:

(14) a. A man who was wearing a black suit entered.
    b. *What kind of clothing was a man who was wearing entered?

But such extraction is inadmissible even after the relative clause of the complex noun phrase has been extraposed:

(15) a. A man entered who was wearing a black suit.
    b. *What kind of clothing was a man entered who was wearing?
Clearly, we can provide a unified analysis of these facts if relative clauses formed a constituent with their head nominals after extraposition, as only then would we have a complex noun phrase in both (14) and (15). But if a relative clause forms a constituent with its head nominal after extraposition, then the verb *entered* is ordered between the constituents of the complex noun phrase (without being a constituent of the complex noun phrase—which would then be truly discontinuous.\(^2\)

To generate the truly discontinuous constituents involved in parentheticals, extraposed relative clauses, and other constructions, McCawley (1982) in effect appeals to the permutation transformations of Chomsky (1955). To do so McCawley had to adopt a theory of constituent structure that did not abide by the conditions of Exclusivity and Nontangling mentioned above. One way to develop such a theory so would be simply to define constituent structure as *the output of syntactic rules* rather than use constituent structure as *a constraint on syntactic rules*. Permutations could then be included in the set of admissible syntactic rules.

Alternatively, one could define constituent structure in a way that is consistent with truly discontinuous constituents. This is the approach followed in McCawley (1982), who weakened the Exclusivity Condition so that two nodes—say *VERB* and *NP* in (1) above—could be related neither by dominance nor by precedence. On the other hand, McCawley required (i) that all terminal nodes be linearly ordered, and (ii) that nonterminal nodes stand in an ordering relation if and only if all their descendents stand in the same relation.\(^3\)

Finally, it should be pointed out that McCawley constrained discontinuous constituency by allowing only those discontinuous constituents that arise from movement transformations that preserve grammatical relations.

**§4. Generalized Phrase Structure Grammars**

As McCawley was developing his case for discontinuous constituency, a renewed interest arose in (nontransformational) phrase structure grammars (Gazdar et al.

---

\(^3\) It has been claimed that the condition requiring terminals to be fully ordered may be “a consequence of the application of the laws of nature to the human mouth: you can, in point of fact, just make one sound at a time” (Higginbotham 1983, 151). If so, this condition would not have to be stated in linguistic theory proper. For further reformulations of constituent structure trees that allow for discontinuous constituents, see Ojeda (1988), Blevins (1990), and Bunt (1996).
Discontinuous constituents presented a challenge for these grammars. On the one hand, both Postal’s cut-and-paste account and McCawley’s permutation accounts were irreducibly transformational. As such, they were barred from nontransformational grammars. On the other hand, phrase structure grammars cannot generate truly discontinuous constituents. For, let A be any constituent of a word or phrase. If A is discontinuous, then it has to combine with another constituent D of that word or phrase—and has to do so in such a way that part of A precedes D while part of A follows it. But this form of nonconcatenative composition cannot be carried out by a series of phrase structure rules of the form in (16), as these rules may only combine constituents $B_1, \ldots, B_n$ by placing them either fully before or fully after each other (in the context of $C_1$ and $C_2$).

$$A \rightarrow B_1 \ldots B_n / C_1 \_\_ C_2$$

Thus, if phrase structure rules are to yield discontinuous constituents, they have to undergo substantial generalization. One such generalization was proposed by Yngve (1960) and has been presented above. Another generalization that rewrites symbols around nonconstituents is given in Bunt (1996). A generalization in a different direction is proposed in Ojeda (1988). The point of departure of the latter proposal is the fact (see Gazdar et al. 1985) that a phrase structure rule like (17a) can be split into a statement about immediate dominance (17b) and statement about linear precedence (17b),

$$\begin{align*}
(17) & \quad a. \quad S \rightarrow NP \ VP \\
& \quad b. \quad S \text{ may immediately dominate } NP \text{ and } VP. \\
& \quad c. \quad \text{If } NP \text{ and } VP \text{ are sisters, then the } NP \text{ must precede the } VP.
\end{align*}$$

But suppose now that we had, in addition to statements (17c) of total linear precedence, statements (18) of partial linear precedence.

$$\begin{align*}
(18) & \quad \text{If } NP \text{ and } VP \text{ are sisters, then the head of the } VP \text{ must precede the } NP.
\end{align*}$$

Now, if constituent structures can be discontinuous, then the statement of immediate dominance in (17a) and the statement of partial linear precedence in (18) will license the truly discontinuous constituent VP in (19).
Notice that these statements provide a monostratal account of VSO word order that does not give up on VPs. Such an account may be applied to the description of VSO languages like Breton, as they have VPs but no underlying SVO orders (Anderson and Chung 1977). And it may apply also to the phenomenon of subject-auxiliary inversion (Ojeda 1987). Moreover, suppose we had statements of immediate partial precedence like (21).

(21) If X and Y are sisters, then (the head of) X must immediately precede (the head of) Y.

If we did, then we could also describe the cross-serial dependencies of Dutch in (22) and the cognate argument structure of English in (23) simply as what they appear to be—permutational variants of each other.

---

4 A statement to the effect that V must fully precede its sister NP will of course play a role in this as well.
LP rules:  
NP < VP  
\( V_2 \ll H(\text{VP}) \)  

(= NP precedes any sister VP)  
(= \( V_2 \) immediately precedes the head of any sister VP)

\[ (23) \]

\[ \text{John} \quad \text{saw} \quad \text{Pete} \quad \text{help} \quad \text{Marie} \quad \text{let} \quad \text{Cecilia} \quad \text{swim} \]

ID rules:  
S \rightarrow NP, VP  
(= S may consist of NP and VP)

VP \rightarrow V_2, NP, VP  
(= VP may consist of \( V_2 \), NP, VP)

VP \rightarrow V_1  
(= VP may consist of \( V_1 \))

LP rules:  
NP < VP  
(= NP precedes any sister VP)

V < X  
(= V precedes all its sisters)

It should be borne in mind that cross-serial dependencies of this type are beyond the weak generative power of context-free grammars (Huybregts 1984, Shieber 1985). Since rules of immediate partial precedence can generate them, these rules affect the expressive power of phrase structure grammars in the most fundamental of senses.

§5 Other Nontransformational Grammars

Drawing from the study of the formal languages of logic, Montague (1970) suggested that the syntactic rules of the natural languages of linguistics be triples of the form in (24), where \( F^n \) is an operation that builds new expressions out of old ones (taken \( n \) at a time), \( C_1 \) through \( C_n \) are (names of) categories of expressions, and \( C_{n+1} \) is likewise (a name of) a category of expressions.

\[ (24) \quad (F^n, (C_1, C_2, \ldots, C_n), C_{n+1}) \]
Taken as a syntactic rule, the triple in (24) asserts that, if $a_1$ through $a_n$ are expressions of categories $C_1$ through $C_n$ (respectively), then $F^n(a_1, a_2, \ldots, a_n)$ is an expression of category $C_{n+1}$. To illustrate, consider the triple in (25).

(25)  \((\text{CONCATENATE, } (\text{NP, VP}), S)\)

Taken as a syntactic rule, the triple in (25) states that, if $a_1$ is a noun phrase and $a_2$ is a verb phrase, then the binary concatenation of $a_1$ and $a_2$, taken in that order, is a sentence.

Now, if $n$-place concatenations were the only operations $F^n$ that could appear as the first terms of triples (24), then the syntactic rules defined in this way would be but notational variants of context-free phrase structure rules. As such, they would be incapable of generating truly discontinuous constituents. But there is nothing in Montague’s proposals that requires concatenations to be the only operations available to syntactic rules. Indeed, Bach (1979) proposed a nonconcatenative binary operation WRAP such that $\text{WRAP}(a_1, a_2)$ is the expression that results from placing expression $a_1$ after the first word of expression $a_2$. Take, for example, the rule in (26).

(26)  \((\text{WRAP}, (\text{NP, VP}), S)\)

If a VP is a constituent that consists of a verb followed by an object, then (26) will form a sentence by placing its subject between its verb and its object, thus producing a sentence with a VSO word order. Moreover, if we define the constituents of an expression generated by a rule $R$ as the expressions $R$ operated on (here an NP and a VP), then any VSO sentence produced by (26) would contain a VP constituent as well, thus accounting for the VPs of canonical sentences of VSO languages like Breton (see Dowty 1982, §5.2).

Now, if a constituent structure tree is to represent the process whereby an expression is generated, then (27) is a constituent structure tree which represents the process whereby the S \textit{Johnny will eat the crepes} is generated, by (25), from a NP \textit{Johnny} and a VP \textit{will eat the crepes}.

\[
\begin{array}{c}
\text{Johnny} \\
\text{will eat the crepes} \\
\end{array}
\]

\[
\begin{array}{c}
\text{NP} \\
\text{VP} \\
\end{array}
\]

(27)  \[
\begin{array}{c}
\text{Johnny} \\
\text{will eat the crepes} \\
\end{array}
\]

And, similarly, (28) is a constituent structure tree that represents the process whereby the Breton sentence \textit{e-tebro Yannig krampouezh} ‘Johnny will eat the
crepes’ is generated, by (26), from a noun phrase *Yannig ‘Johnny’* and a verb phrase *e-tebro k ramp ouezh ‘will eat the crepes’.*

\[
\begin{array}{c}
\text{e-tebro Yannig k rampouezh} \\
\text{Yannig}_{NP} \quad \text{e-tebro k rampouezh}_{VP}
\end{array}
\]

(28)

It should be clear that the nodes of the constituent structure trees in (27) and (28) need to be ordered only in terms of dominance, not in terms of precedence. All the information about the linear order of the constituents represented therein is encoded in the expressions mentioned at each node of the tree. For indeed, the nodes of these trees are *categorized expressions*, not just expressions (as the terminal nodes of the preceding trees) or categories (as the nonterminal nodes of the preceding trees). To acknowledge the differences between these trees and the preceding ones, the former have been given a special name—*analysis trees*.

It should also be clear that the analysis tree in (28) represents a truly discontinuous VP. Yet, it does not involve crossing branches. Again, this is because the linear order required by discontinuous constituency is not expressed by the linear order of the nodes of analysis trees; only by the expressions categorized at their nodes. See Huck (1984) for a more detailed discussion of discontinuity and wrapping operations.

Rules like (25) and (26) have been proposed to handle discontinuous constituents in morphology as well (McCarthy and Prince, 1986; Hammond 1990). Consider by way of example the rule in (29), where \( \tau_e \) is the category of trochee expletives that includes *friggin’* and *freakin’* (as well as others like them), and \( \tau_{pl} \) is the category of words like *California*, which consist of a plurality (two or more) of trochees.

\[
(29) \quad (\text{WRAP}, (\tau_e, \tau_{pl}), \tau_{pl})
\]

As might be expected, (29) is a rule that combines a trochee expletive with a multiple-trochee word by appealing to a wrapping operation that places the expletive immediately after the first trochee of the word it combines with (thus yielding another multiple-trochee word). The process whereby *Calif-friggin’-fornia* is generated by (29) is represented in (30).
§6 Current Approaches

Generally speaking, grammars induce linear order by ordering sisters. But truly discontinuous constituents call for wrapping sisters around nonsisters, and thus for ordering nonsisters. One could therefore ask whether the scope of linear order statements shouldn’t be widened so that they can apply both to sisters and to nonsisters. This is precisely the line of inquiry of a number of approaches to discontinuous constituency currently being explored.

Take for example Dowty (1996), who assumes that the default operation for combining two expressions syntactically is to merge their words into a single (unordered) multiset. Needless to say, this multiset needs to be ordered (or linearized) if it is to become a well-formed expression. Let us say then that such linearizations are admissible if and only if they abide by all the linear precedence statements made by the grammar. Now, since the multisets created by merging may contain nonsisters, the linear precedence statements of the grammar will apply, as desired, both to sisters and to nonsisters.

Suppose by way of example that four expressions $A$, $B$, $C$, $D$ were to combine syntactically (by default) as follows.

If there were no statements of linear precedence involving $A$, $B$, $C$, $D$ (i.e. if these expressions were freely ordered), then all of the twenty four possible linearizations of $\{A, B, C, D\}$ would be admissible. But suppose that there were two statements of linear precedence involving these expressions:

\begin{align*}
(32) & \quad A < B \\
& \quad C < D
\end{align*}

Now, of all the logically possible linearizations of $\{A, B, C, D\}$, only six are admissible:
Of these, all but the first and the last involve discontinuous constituents (we assume, of course, that constituency is determined by the rules that induced (31)).

But suppose that \{C, D\} was a bounded domain—say a sentence, whose constituents cannot generally mingle with constituents external to it. We could represent this fact by the marked form of syntactic composition represented in (34), where \{C, D\} is an element rather than a subset of \{A, B, \{C, D\}\}.

Now, given the statements of linear precedence in (32), only three linearizations of \{A, B, \{C, D\}\} are admissible. They are

Of these, the second involves a discontinuous constituent. Similar proposals have been developed, independently and in detail, by Reape (1996). Reape’s proposals have been extensively developed in the context of Head-drive Phrase Structure Grammar. See for example Kathol (2000).

References

McCarthy, John and Alan Prince (1986) “Prosodic Morphology”. Manuscript, University of Massachusetts at Amherst and Brandeis University.


