

Language Files

Materials for An Introduction to Language

Compilers of the Second Edition:

CAROL JEAN GODBY
REX WALLACE
CATHERINE JOLLEY

Editors of Previous Versions:

DEBORAH B. SCHAFFER
JOHN W. PERKINS
F. CHRISTIAN LATTA
SHEILA GRAVES GEOGHEGAN

SYNTAX: WORD ORDER AND CONSTITUENT STRUCTURE

FILE
61

The sentence is one of the most fundamental units of linguistic organization. Though we use sentences all the time, we don't normally think about how they are structured; but a little consideration reveals that the principles by which words are organized into sentences are, in fact, quite complex. In this file, we will consider two basic principles of sentence organization; these are *linear order* and *hierarchical structure*.

I. LINEAR ORDER

The most obvious principle of sentence organization is linear order: the words in a sentence must occur in a particular sequence if the sentence is to convey the desired meaning. Consider, for example, the following sentence of English:

(1) John glanced at Mary.

If we rearrange the words in this sentence, we either come up with nonsense, as in (2)—

(2) Mary John at glanced.

or with a sentence whose meaning is distinctly different from that of (1)—

(3) Mary glanced at John.

Clearly, the ordering of the words in (1) is an essential aspect of its organization.

The ways in which words are ordered in a language may be described as the *word order rules* of that language. For instance, one of the many word order rules of English is that the grammatical subject of a sentence normally precedes the main verb, which in turn normally precedes its direct object; thus, *she resembles him* is English but *resembles she him* and *she him resembles* are not.

An important fact about word order rules is that they are language-specific—that is, languages vary in the ways in which they order words. An example will bring this out. Because of its characteristic ordering of subjects (S), main verbs (V), and objects (O), English may be categorized as an 'S-V-O language' (other languages, such as French and Swahili, may be similarly categorized). But there are also S-O-V languages (e.g., Turkish, Navaho), in which main verbs normally *follow* their direct object; V-S-O languages (e.g. Welsh, Hebrew), in which main verbs normally *precede* their subject; and a small number of V-O-S, O-V-S, and O-S-V languages as well. In addition, there are languages (such as Dyrbal, an Australian aboriginal language) in which the normal order of subject, verb, and object is remarkably free. Clearly, there is no set of word order rules which is valid for all languages.

II. HIERARCHICAL STRUCTURE

Although linear order is an important principle of sentence organization, sentences are more than just ordered sequences of words: they have internal hierarchical structure as well. That is, the individual words in a sentence are organized into natural, semantically coherent groupings, which are themselves organized into larger groupings, the largest grouping of all being the sentence itself. These groupings within a sentence are called *constituents* of that sentence. Consider sentence (4) as an example.

(4) Business executives eat at really fancy restaurants.

We can easily distinguish a number of meaningful groups of words in this sentence: *business executives* and *eat at really fancy restaurants*, for instance, have clear meanings of their own, and each makes a coherent contribution to the meaning of (4) as a whole. (Specifically, *business executives* is the subject of (4), and *eat at really fancy restaurants* is the predicate—see File 60, “Some Useful Syntactic Terms”.) For this reason, they are constituents of this sentence. On the other hand, some groups of words in sentence (4) do not naturally form meaningful units; *executives eat at* and *eat at really*, for example, don’t have clear meanings of their own. Thus, these groups of words are *not* constituents of (4).

There are a number of useful ways of distinguishing constituents from groups of words which aren’t constituents:

(a) Constituents can often be sensibly used alone, e.g. as exclamations or as answers to questions: —*What do business executives do?* —*Eat at really fancy restaurants.* This isn’t true of nonconstituents: if we were asked “Do fancy restaurants do much business?” we couldn’t sensibly answer “Well, executives eat at.”

(b) Parenthetical remarks like *of course*, *as you know*, *by the way*, etc. naturally appear between constituents: *Business executives, as you know, eat at really fancy restaurants.* But they sound rather unnatural if what precedes or follows is a non-constituent: *Business executives eat at really, as you know, fancy restaurants* (neither *business executives eat at really* nor *fancy restaurants* is a constituent of (4)).

(c) It is often possible to replace a constituent with a single word having the same meaning as that constituent. For example, if someone asked “What do business executives do?” we could answer either with sentence (4) or with sentence (5), in which the constituent *business executives* is replaced with the single word *they* (which, in this context, would mean the same thing as *business executives*).

(5) They eat at really fancy restaurants.

Similarly, if someone asked “Who eats at really fancy restaurants?” we could answer either with (4) or with sentence (6), in which the constituent *eat at really fancy restaurants* is replaced with the single word *do* (which would mean the same thing in this context).

(6) Business executives do.

But there is no word that could possibly replace the nonconstituent *eat at really* in (4) and mean the same thing, no matter what question was asked.

(d) Finally, sentences are always constituents, as are the individual words within a sentence: in sentence (4), for instance, the largest constituent is the sentence itself; the smallest constituents are the individual words *business*, *executives*, *eat*, *at*, *really*, *fancy*, and *restaurants*.

Two points must be kept in mind regarding constituent structure. First, given a group of words, we cannot say once and for all whether or not it is a constituent; rather, we can only say whether or not it is a constituent relative to a particular sentence. To see this, consider sentences (7) and (8).

(7) John and Bill raise weasels.

(8) Mary punched John and Bill kicked Jane.

In (7), *John and Bill* is a constituent: it functions as a coherent, meaningful unit within the sentence—in particular, as its subject. In (8), however, the very same sequence of words is *not* a constituent: because *John* is the direct object of the first clause and *Bill* is the subject of the second clause, the sequence *John and Bill* does not make a coherent contribution to the meaning of this sentence. (Notice that although *John and Bill* can be replaced with *they* in (7), this isn’t possible in sentence (8).) Thus, we can properly say that a group of words is a constituent only with respect to a particular sentence.

The second thing that must be kept in mind is that constituent structure is *hierarchical*—that is, one constituent may be part of another. Consider sentence (4) again.

(4) Business executives eat at really fancy restaurants.

Among the constituents in this sentence is the sequence *really fancy*. (To see this, note that *really fancy* can be used by itself:

—How fancy was it?

—Really fancy.

and that it can be replaced with the single word *such*:

—Who eats at really fancy restaurants?

—Business executives eat at such restaurants.)

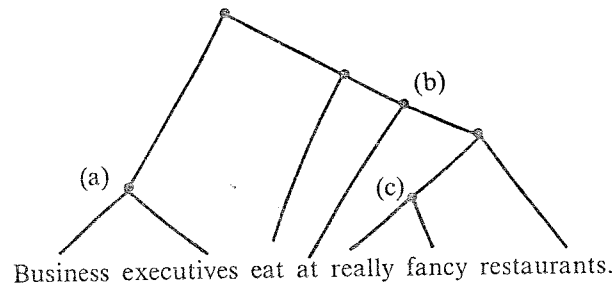
But *really fancy* is also part of a larger constituent, namely *really fancy restaurants*; this in turn is part of a larger constituent, *at really fancy restaurants*, which is itself part of the still larger constituent *eat at really fancy restaurants* and ultimately of the largest constituent in the sentence, namely the sentence itself. If we underline each of the constituents in (4), the hierarchical nature of its constituent structure becomes obvious:

(9) Business executives eat at really fancy restaurants.



Underlining is, as in (9), one way of representing the hierarchical nature of constituent structure. Another way is with *tree diagrams*: branching structures in which each constituent forms a 'branch'. For example, the tree diagram for sentence (4) is (10):

(10)



In this diagram, each of the constituents of sentence (4) forms a branch: for example, *business executives* corresponds to the branch labelled (a); *at really fancy restaurants*, to the branch labelled (b); and *really fancy*, to the branch labelled (c). Observe, in addition, that groups of words that are not constituents of sentence (4) do *not* form branches in this tree diagram: *executives eat at* and *eat at really*, for instance, clearly aren't constituents according to diagram (10). In principle, underlining is just as good as tree diagrams for representing constituent structure; but because tree diagrams are somewhat easier to read, they are usually preferred.

CONSTITUENT STRUCTURE AND AMBIGUITY

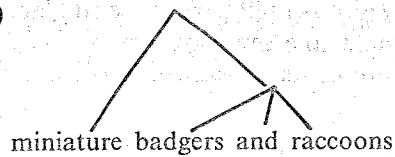
In every human language we can find individual expressions which have two or more distinct meanings. For example, the italicized portions of the following sentences of English can be interpreted in more than one way:

- (11) a. Larry raises *miniature badgers and raccoons*.
 b. We need *more intelligent leaders*.

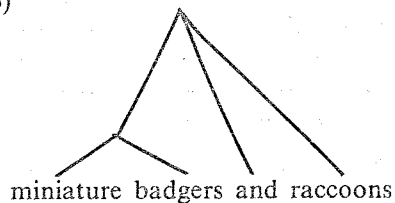
In (11a), *miniature badgers and raccoons* can mean either 'miniature badgers and miniature raccoons' or 'miniature badgers and raccoons (of any size)'; in (11b), *more intelligent leaders* can mean either 'a greater quantity of intelligent leaders' or 'leaders who are more intelligent'. This property of having two or more distinct meanings is called *ambiguity*; an expression with two or more distinct meanings is *ambiguous*.

Often, an expression is ambiguous because it has more than one possible constituent structure. Consider, for example, the expression *miniature badgers and raccoons* in sentence (11a): it can have either of the following constituent structures.

(12)



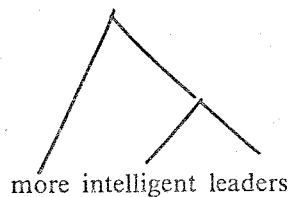
(13)



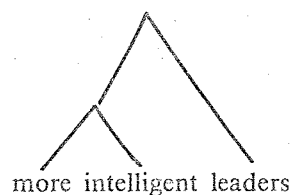
In (12), *badgers and raccoons* forms a constituent; (12) therefore represents the interpretation in which the adjective *miniature* applies to both the badgers and the raccoons. In (13), on the other hand, *miniature badgers* forms a constituent; (13) therefore represents the interpretation in which only the badgers are miniature. An expression which is ambiguous because it has more than one possible constituent structure is said to be *structurally ambiguous*.

The italicized portion of sentence (11b) is also structurally ambiguous: it can have either of the following constituent structures.

(14)



(15)

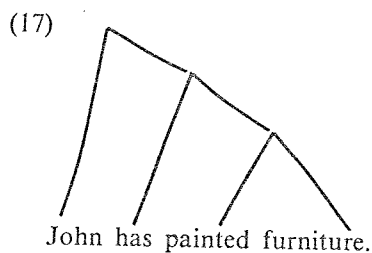


In (14), *intelligent leaders* is a constituent; for this reason, (14) represents the interpretation 'a greater quantity of intelligent leaders'. In (15), however, *more intelligent* forms a constituent; (15) therefore represents the meaning 'leaders who are more intelligent.'

Although structural ambiguity is a very common kind of ambiguity, it is not the only kind. Individual *words* are sometimes ambiguous (for example, *crane* can refer either to a kind of bird or to a large construction device); because words like *crane* have no internal constituent structure, they clearly can't be structurally ambiguous. (Instead, they are sometimes said to be *lexically ambiguous*.) A third sort of ambiguity is exemplified by sentence (16).

(16) John has painted furniture.

(16) can mean either 'John has applied paint to furniture' or 'John has furniture that is painted'. This ambiguity isn't structural: (17) is the correct constituent structure for (16), regardless of which meaning is intended.



(Under either interpretation of (16), *painted furniture* forms a constituent and can therefore be used alone:

- What has John done?
- Painted furniture.
- What does John have?
- Painted furniture.)

This sort of ambiguity will be discussed below.

SYNTACTIC CATEGORIES

A very important fact about constituent structure is that there are different types of constituents with very different uses. We refer to these different types of constituents as *syntactic categories*.

One of the most important syntactic categories is the category of *noun phrases*. Examples are the expressions in (18):

- (18) John
 mailmen
 most dogs
 many Americans
 a huge, loveable bear
 a student from Brazil
 the table in the corner
 the people that we interviewed
 John and his dog

A noun phrase can be used as the subject of a sentence, as in (19a); as the direct object, as in (19b); as the indirect object, as in (19c); and in many other ways as well.

- (19) a. *Most dogs* enjoy hamburger sauce.
 b. Harold likes *most dogs*.
 c. Lilian gave *most dogs* their rabies shots this morning.

In some cases, a single word can count as a noun phrase all by itself; this is true, for instance, of names (e.g. *John*), plural nouns (e.g. *mailmen*), nouns referring to substances (e.g. *water, dirt*), and pronouns (e.g. *I, she, them*). But many noun phrases begin with expressions like the following:

(20) the	most	at least five
a	all	my
every	few	Mary's
many	several	
some	three	

These are called *determiners*, and they combine with a single noun or with a noun modified by other sorts of expressions to produce a noun phrase: *the president, many Americans, two new sweaters, a student from Brazil*, etc.

Another extremely important syntactic category is the category of *verb phrases*. Some examples are the expressions in (21):

- (21) snore
 like Mary
 give a prize to John
 believe that dogs are smart
 want to leave
 sleep soundly
 can lift 100 pounds
 is wearing sunglasses
 go home and have a beer

A verb phrase can be used as the predicate of a sentence, as in the examples in (22).

- (22) a. John and Bill *like Mary*.
 b. Henry *wants to leave*.

Certain verbs, such as *snore, swim, talk, and die*, can form a verb phrase all by themselves; verbs of this sort are called *intransitive verbs*. Certain other verbs form a verb phrase by combining with a noun phrase. Such verbs are called *transitive verbs*; the noun phrase with which a transitive verb combines is its direct object (see File 60), as in the following examples:

- (23) like Mary
 chase cars
 annoy three burly sergeants
 develop every roll of film

Other verbs combine with other sorts of expressions to form a verb phrase: verbs like *give* and *owe* combine with two noun phrases (a direct object and an indirect object), as in *give a prize to John* and *owe Larry's brother several hundred dollars*; verbs like *try* and *manage* combine with a verb phrase marked with *to*, as in *want to leave* and *manage to finish the pizza*; and so on.

Certain verb phrases consist of a *helping verb* (e.g. *can, should, might, will, be, have*) plus a smaller verb phrase: *can lift 100 pounds, should wear sunglasses, might want to leave*. When the helping verb is *have*, the verb which follows it is in its past participle form: *John has chopped the onions, Henry has found the wallet*. When the helping verb is *be*, the verb which follows it may be in its past participle form (if the sentence has a *passive* meaning—see File 60) or in its present participle form:

- (24) a. The onions were chopped by John.
 b. The wallet was found by Henry. (passive sentences)
 c. John is chopping the onions.
 d. Henry is finding the wallet.

Another important syntactic category is the category of *adjective phrases*, of which the following are examples:

- (25) smart
 very fat
 as crazy as John
 more intelligent than Mary
 certain to win

Adjective phrases are often used to modify nouns and thus often appear as constituents of noun phrases: *a very fat individual; someone as crazy as John*.

Adverbial phrases, such as those in (26), are often used to modify verbs and adjectives, and thus appear as constituents of verb phrases and adjective phrases, as in (27).

- (26) soundly
fiercely
as fluently as John
almost certainly
- (27) sleep soundly
speak French as fluently as John (verb phrases)
fiercely loyal
almost certainly able to walk (adjective phrases)

Another important syntactic category is that of *prepositional phrases*. Prepositional phrases always consist of a *preposition* (e.g. *to, from, with, for, in, at, on, under, about, through*) plus a noun phrase:

- (28) to the movies
from Brazil
with John and Bill
for nothing

A prepositional phrase can be a constituent of a wide range of expressions:

- (29) go to the movies (verb phrase)
a student from Brazil (noun phrase)
angry with John and Bill (adjective phrase)
separately from the others (adverbial phrase)

Perhaps the most important syntactic category of all is the category of *sentences*. Sentences are, of course, often used by themselves:

- (30) a. It is raining.
b. I like hamburger sauce.

But a sentence may also appear as a constituent of another expression; for example, each of the following expressions has a sentence as a constituent.

- (31) the fact that *it is raining*
a student *who met Susan last Thursday* (noun phrases)
discover that *it is raining*
know *who met Susan last Thursday* (verb phrases)
glad that *it is raining* (adjective phrase)

In addition, certain adverbial phrases consist of a sentence preceded by a *subordinating conjunction* (e.g. *if, though, when, after, because*): *if it is raining, though it is raining*, and so on. Any sentence which is a constituent of an expression of another category is an *embedded sentence*.

One other syntactic category that should be mentioned is the category of *coordinating conjunctions*: these are words like *and, or, and but*, which are used to connect two or more expressions of the same category. As the following examples suggest, *and* can be used to connect expressions of virtually any category:

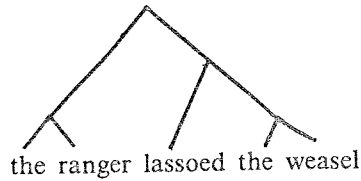
- (32) John and his dog (noun phrases)
go home and have a beer (verb phrases)
faster than a speeding bullet and more powerful than a locomotive (adjective phrases)
quickly and very easily (adverbial phrases)
over the river and through the woods (prepositional phrases)
It is raining and it may sleet. (sentences)

Note that any expression resulting from the connection of two or more smaller expressions belongs to the same category as they do. That is, because *John* and *his dog* are noun phrases, so is *John and his dog*; because *go home* and *have a beer* are verb phrases, so is *go home and have a beer*; and so on.

LABELLED TREE DIAGRAMS

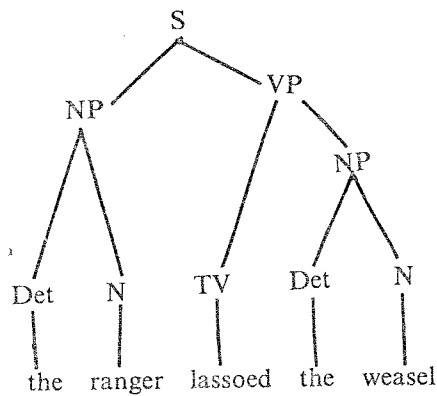
As we saw above, constituent structure can be perspicuously represented by means of tree diagrams such as the following:

(33)



(33) indicates what groups of words are constituents in the sentence *the ranger lassoed the weasel*; it does not, however, indicate the syntactic categories to which the constituents belong. This limitation can be overcome if we mark each constituent in a tree structure like (33) with the appropriate category label; the result of this procedure is a *labelled tree diagram* such as (34).

(34)



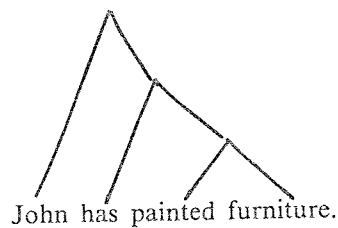
(The following abbreviations are used here: S = Sentence; NP = Noun Phrase; Det = Determiner; N = Noun; VP = Verb Phrase; TV = Transitive Verb.)

By using labelled trees, we can account for certain nonstructural sorts of ambiguity. Consider again sentence (16):

(16) John has painted furniture.

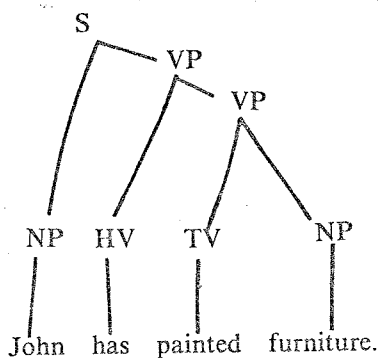
As we saw above, this sentence is ambiguous in a nonstructural way: it has (17) as its structure whether it means 'John has applied paint to furniture' or 'John has furniture that is painted'.

(17)



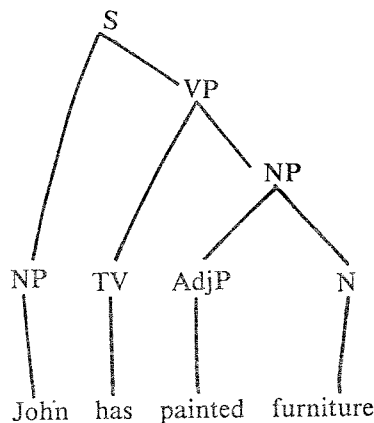
Notice, however, that the constituents in (17) can be plausibly labelled in two different ways. On the one hand, we can label *has* as a helping verb (HV) and *painted furniture* as a verb phrase (in which *painted* is a transitive verb):

(35)



(35) corresponds to the meaning 'John has applied paint to furniture'. On the other hand, we can label *has* as a transitive verb and *painted furniture* as a noun phrase (in which *painted* is an adjective phrase):

(36)



(36) represents the meaning 'John has furniture that is painted'. Observe that (35) and (36) are identical if the category labels are ignored; thus, sentence (16) is ambiguous because its constituents can be categorized in two different ways, but not because of any structural ambiguity.

SUMMARY

In this file, a number of important syntactic concepts have been discussed. We have seen that *linear order* is an important principle of sentence organization, but that sentences have *hierarchical organization* as well; that the fundamental unit of hierarchical organization is the *constituent*; and that there are several different *categories* of constituents. In addition, we have developed a graphic means of representing sentence structure: *labelled tree diagrams*, which reflect the constituent structure of a sentence and indicate the syntactic category to which each constituent belongs.

EXERCISES

1. Which of the italicized expressions in the following sentences are constituents? Which are not? Why? (Use the tests (a)-(c) for constituency mentioned above to show this in each case.)
 - a. John *ate the stale* candy.
 - b. John *ate the stale* candy.
 - c. My little *brother snores*.
 - d. My little *brother snores*.

2. Using labelled tree diagrams, give the constituent structure of the following sentences.
- Mary snores.
 - Two elephants noticed the mice.
 - The train slowly departed.
 - Ed knows you like cheese.

3. Discuss the ambiguity of the following sentences. Which is lexically ambiguous? Which is structurally ambiguous? Which is ambiguous because its constituents can be categorized in two different ways?
- They are baking apples.
 - John sat on Jumbo's trunk.
 - Marian knows a Chinese art expert.