

3

Syntax: The Sentence Patterns of Language

To grammar even kings bow.

J. B. MOLIÈRE, *Les Femmes Savantes*, II, 1672

It is an astonishing fact that any speaker of any human language can produce and understand an infinite number of sentences. We can show this quite easily through examples such as the following:

The kindhearted boy had many girlfriends.

The kindhearted, intelligent boy had many girlfriends.

The kindhearted, intelligent, handsome boy had many girlfriends.

.
. .
.

John found a book in the library.

John found a book in the library in the stacks.

John found a book in the library in the stacks on the fourth floor.

.
. .
.

The cat chased the mouse.

The cat chased the mouse that ate the cheese.

The cat chased the mouse that ate the cheese that came from the cow.

The cat chased the mouse that ate the cheese that came from the cow that grazed in the field.

In each case the speaker could continue creating sentences by adding another adjective, prepositional phrase, or relative clause. In principle, this could go on forever. All languages have mechanisms of this sort that make the number of sentences limitless. Given this fact, the sentences of a language cannot be stored in a dictionary format in our heads. Rather, sentences are composed of discrete units that are combined by rules. This system of rules explains how speakers can store infinite knowledge in a finite space—our brains.

The part of grammar that represents a speaker's knowledge of sentences and their structures is called **syntax**. The aim of this chapter is to show you what syntactic structures look like and to familiarize you with some of the rules that determine them. Most of the examples will be from the syntax of English, but the principles that account for syntactic structures are universal.

What the Syntax Rules Do

"Then you should say what you mean," the March Hare went on.

"I do," Alice hastily replied, "at least—I mean what I say—that's the same thing, you know."

"Not the same thing a bit!" said the Hatter. "You might just as well say that 'I see what I eat' is the same thing as 'I eat what I see!'"

"You might just as well say," added the March Hare, "that 'I like what I get' is the same thing as 'I get what I like!'"

"You might just as well say," added the Dormouse . . . "that 'I breathe when I sleep' is the same thing as 'I sleep when I breathe!'"

"It is the same thing with you," said the Hatter.

LEWIS CARROLL, *Alice's Adventures in Wonderland*, 1865

The **rules of syntax** combine words into phrases and phrases into sentences. Among other things, the rules determine the correct word order for a language. For example, English is a Subject–Verb–Object (SVO) language. The English sentence in (1) is grammatical because the words occur in the right order; the sentence in (2) is ungrammatical because the word order is incorrect for English. (Recall that the asterisk or star preceding a sentence is the linguistic convention for indicating that the sentence is ungrammatical or ill-formed according to the rules of the grammar.)

1. The President nominated a new Supreme Court justice.
2. *President the Supreme new justice Court a nominated.

A second important role of the syntax is to describe the relationship between the meaning of a particular group of words and the arrangement of those words. For example, Alice's companions show us that the word order of a sentence contributes crucially to its meaning. The sentences in (3) and (4) contain the same words, but the meanings are quite different, as the Mad Hatter points out.

3. I mean what I say.
4. I say what I mean.

The rules of the syntax also specify the **grammatical relations** of a sentence, such as **subject** and **direct object**. In other words, they provide information about who is doing what to whom. This information is crucial to understanding the meaning of a sentence. For example, the grammatical relations in (5) and (6) are reversed, so the otherwise identical sentences have very different meanings.

5. Your dog chased my cat.
6. My cat chased your dog.

In (7) we see that the phrase *ran up the hill* behaves differently from the phrase *ran up the bill*, even though the two phrases are superficially quite similar. For the expression *ran up the hill*, the rules of the syntax allow the word orders in (7a) and (7c), but not (7b). In *ran up the bill*, in contrast, the rules allow the order in (7d) and (7e), but not (7f).

7. (a) Jack and Jill ran up the hill.
- (b) *Jack and Jill ran the hill up.
- (c) Up the hill ran Jack and Jill.
- (d) Jack and Jill ran up the bill.
- (e) Jack and Jill ran the bill up.
- (f) *Up the bill ran Jack and Jill.

The pattern shown in (7) illustrates that sentences are not simply strings of words with no further organization. If they were, there would be no reason to expect *ran up the hill* to pattern differently from *ran up the bill*. These phrases act differently because they have different syntactic structures associated with them. In *ran up the hill*, the words *up the hill* form a unit, as follows:

He ran [up the hill].

The whole unit can be moved to the beginning of the sentence, as in (7c), but we cannot rearrange its subparts, as shown in (7b). On the other hand, in *ran up the bill*, the words *up the bill* do not form a natural unit, so they cannot be moved together, and (7f) is ungrammatical.

Our syntactic knowledge crucially includes rules that tell us how words form groups in a sentence, or how they are *hierarchically* arranged with respect to one another. Consider the following sentence:

The captain ordered all old men and women off the sinking ship.

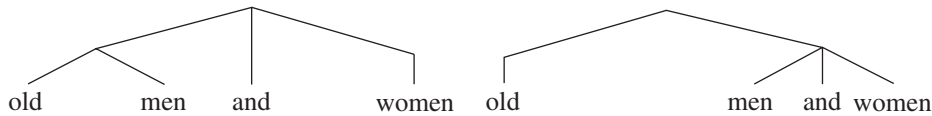
This phrase *old men and women* is ambiguous, referring to either old men and to women of any age or to old men and old women. The ambiguity arises because the words *old men and women* can be grouped in two ways. If the words are grouped as follows, *old* modifies only *men* and so the women can be of any age.

[old men] and [women]

When we group them like this, the adjective *old* modifies both *men* and *women*.

[old [men and women]]

The rules of syntax allow both of these groupings, which is why the expression is ambiguous. The following hierarchical diagrams, also called **tree diagrams**, illustrate the same point:



In the first structure *old* and *men* are under the same node and hence *old* modifies *men*. In the second structure *old* shares a node with the entire conjunction *men and women*, and so modifies both.

This is similar to what we find in morphology for ambiguous words such as *unlockable*, which have two structures, corresponding to two meanings, as discussed in chapter 2.

Many sentences exhibit such ambiguities, often leading to humorous results. Consider the following two sentences, which appeared in classified ads:

For sale: an antique desk suitable for lady with thick legs and large drawers.
We will oil your sewing machine and adjust tension in your home for \$10.00.

In the first ad, the humorous reading comes from the grouping [desk] [for lady with thick legs and large drawers] as opposed to the intended [desk for lady] [with thick legs and large drawers], where the legs and drawers belong to the desk. The second case is similar.

Because these ambiguities are a result of different structures, they are instances of **structural ambiguity**.

Contrast these sentences with:

This will make you smart.

The two interpretations of this sentence are due to the two meanings of *smart*—‘clever’ and ‘burning sensation.’ Such lexical or word-meaning ambiguities, as opposed to structural ambiguities, will be discussed in chapter 4.

Often a combination of differing structure and double word-meaning creates ambiguity (and humor) as in the cartoon:



Hilary B. Price. King Features Syndicate

Syntactic rules reveal the grammatical relations among the words of a sentence as well as their order and hierarchical organization. They also explain how the grouping of words relates to its meaning, such as when a sentence or phrase is ambiguous. In addition, the rules of syntax permit speakers to produce and understand a limitless number of sentences never produced or heard before—the *creative aspect of linguistic knowledge*. A major goal of linguistics is to show clearly and explicitly how syntactic rules account for this knowledge. A theory of grammar must provide a complete characterization of what speakers implicitly know about their language.

What Grammaticality Is Not Based On

Colorless green ideas sleep furiously. This is a very interesting sentence, because it shows that syntax can be separated from semantics—that form can be separated from meaning. The sentence doesn't seem to mean anything coherent, but it sounds like an English sentence.

HOWARD LASNIK, *The Human Language: Part One*, 1995

Importantly, a person's ability to make grammaticality judgments does not depend on having heard the sentence before. You may never have heard or read the sentence

Enormous crickets in pink socks danced at the prom.

but your syntactic knowledge tells you that it is grammatical. As we showed at the beginning of this chapter, people are able to understand, produce, and make judgments about an infinite range of sentences, most of which they have never heard before. This ability illustrates that our knowledge of language is creative—not creative in the sense that we are all accomplished poets, but creative in that none of us is limited to a fixed repertoire of expressions. Rather, we can exploit the resources of our language and grammar to produce and understand a limitless number of sentences embodying a limitless range of ideas and emotions.

We showed that the structure of a sentence contributes to its meaning. However, grammaticality and meaningfulness are not the same thing, as shown by the following sentences:

Colorless green ideas sleep furiously.
A verb crumpled the milk.

Although these sentences do not make much sense, they are syntactically well formed. They sound funny, but their funniness is different from what we find in the following strings of words:

*Furiously sleep ideas green colorless.
*Milk the crumpled verb a.

There are also sentences that we understand even though they are not well-formed according to the rules of the syntax. For example, most English speakers could interpret

*The boy quickly in the house the ball found.

although they know that the word order is incorrect. To be a sentence, words must conform to specific patterns determined by the syntactic rules of the language.

Some sentences are grammatical even though they are difficult to interpret because they include nonsense words, that is, words with no agreed-on meaning. This is illustrated by the following lines from the poem “Jabberwocky” by Lewis Carroll:

’Twas brillig, and the slithy toves
Did gyre and gimble in the wabe

These lines are grammatical in the linguistic sense that they obey the word order and other constraints of English. Such nonsense poetry is amusing precisely because the sentences comply with syntactic rules and sound like good English. Ungrammatical strings of nonsense words are not entertaining:

*Toves slithy the and brillig ’twas
wabe the in gimble and gyre did

Grammaticality also does not depend on the truth of sentences. If it did, lying would be easy to detect. Nor does it depend on whether real objects are being discussed or whether something is possible in the real world. Untrue sentences can be grammatical, sentences discussing unicorns can be grammatical, and sentences referring to pregnant fathers can be grammatical.

The syntactic rules that permit us to produce, understand, and make grammaticality judgments are unconscious rules. The grammar is a mental grammar, different from the prescriptive grammar rules that we are taught in school. We develop the mental rules of grammar long before we attend school, as we shall see in chapter 9.

Sentence Structure

I really do not know that anything has ever been more exciting than diagramming sentences.

GERTRUDE STEIN, “Poetry and Grammar,” 1935

Suppose we wanted to write a template that described the structure of an English sentence, and more specifically, a template that gave the correct word order for English. We might come up with something like the following:

Det—N—V—Det—N

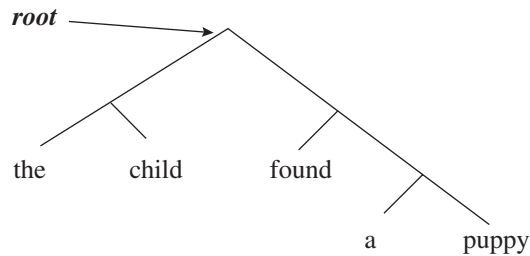
This template says that a determiner (e.g. an article like *the* or *a*) is followed by a noun, which is followed by a verb, and so on. It would describe English sentences such as the following:

The child found a puppy.
The professor wrote a book.
That runner won the race.

The implication of such a template would be that sentences are strings of words belonging to particular grammatical categories (“parts of speech”) with no internal organization. We know, however, that such “flat” structures are incorrect. As noted earlier, sentences have hierarchical organization; that is, the words are grouped into natural units. The words in the sentence

The child found a puppy.

may be grouped into [the child] and [found a puppy], corresponding to the subject and predicate of the sentence. A further division gives [the child] and then [[found] [a puppy]], and finally the individual words: [[the] [child]] [[found] [[a] [puppy]]]. It’s sometimes easier to see the parts and subparts of the sentence in a tree diagram, as we did earlier to illustrate ambiguity:



The “tree” is upside down with its “root” encompassing the entire sentence, *The child found a puppy*, and its “leaves” being the individual words *the*, *child*, *found*, *a*, and *puppy*. The tree conveys the same information as the nested square brackets. The hierarchical organization of the tree reflects the groupings and subgroupings of the words of the sentence.

The tree diagram shows, among other things, that the phrase *found a puppy* divides naturally into two branches, one for the verb *found* and the other for the direct object *a puppy*. A different division, say, *found a* and *puppy*, is unnatural.

Constituents and Constituency Tests

The natural groupings or parts of a sentence are called **constituents**. Various linguistic tests reveal the constituents of a sentence. The first test is the “stand alone” test. If a group of words can stand alone, for example, as an answer to a question, they form a constituent. So in response to the question “What did you find?” a speaker might answer *a puppy*, but not *found a*. *A puppy* can stand alone while *found a* cannot. We have a clear intuition that one of these is a meaningful unit and the other is just a list of words.

The second test is “replacement by a pronoun.” Pronouns can substitute for natural groups. In answer to the question “Where did you find *a puppy*?” a speaker can say, “I found *him* in the park.” Words such as *do* (which is not a pronoun per se) can also take the place of the entire predicate *found a puppy*, as in “John found a puppy and Bill *did* too.” If a group of words can be replaced by a pronoun or a word like *do*, it forms a constituent.

A third test of constituency is the “move as a unit” test. If a group of words can be moved, they form a constituent. For example, if we compare the following sentences to the sentence “The child found a puppy,” we see that certain elements have moved:

It was *a puppy* that *the child* found.

A puppy was found by *the child*.

In the first example, the constituent *a puppy* has moved from its position following *found*; in the second example, the positions of *a puppy* and *the child* have been changed. In all such rearrangements the constituents *a puppy* and *the child* remain intact. *Found a* does not remain intact, because it is not a constituent.

In the sentence “The child found a puppy,” the natural groupings or constituents are the subject *the child*, the predicate *found a puppy*, and the direct object *a puppy*.

Some sentences have prepositional phrases in the predicate. Consider

The puppy played in the garden.

We can use our tests to show that *in the garden* is also a constituent, as follows:

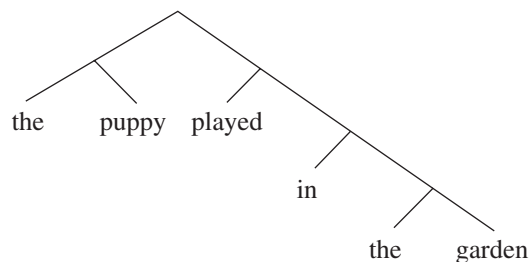
Where did the puppy play? *In the garden* (stand alone)

The puppy played *there*. (replacement by a pronoun-like word)

In the garden is where the puppy played. (move as a unit)

It was *in the garden* that the puppy played. (move as a unit)

As before, our knowledge of the **constituent structure** of a sentence may be graphically represented by a tree diagram. The tree diagram for the sentence “The puppy played in the garden” is as follows:



In addition to the syntactic tests just described, experimental evidence has shown that speakers do not mentally represent sentences as strings of words but rather in terms of constituents. In these experiments, subjects listen to sentences that have clicking noises inserted into them at random points. In some cases the click occurs at a constituent boundary, and in other sentences the click is inserted in the middle of a constituent. The subjects are then asked to report where the click occurred. There were two important results: (1) Subjects noticed the click and recalled its location best when it occurred at a major constituent boundary (e.g., between the subject and predicate); and (2) clicks that occurred inside the constituent were reported to have occurred between constituents. In other words, subjects displaced the clicks and put

them at constituent boundaries. These results show that speakers perceive sentences in chunks corresponding to grammatical constituents.

Every sentence in a language is associated with one or more constituent structures. If a sentence has more than one constituent structure, it is ambiguous, and each tree will correspond to one of the possible meanings. For example, the sentence *I bought an antique desk suitable for a lady with thick legs and large drawers* has two phrase structure trees associated with it. In one structure the phrase [a lady with thick legs and large drawers] forms a constituent; it could stand alone in answer to the question “Who did you buy an antique desk for?” In its second meaning, the phrase *thick legs and large drawers* modifies the phrase [desk for a lady]; it could stand alone in answer to the question “What did the desk have?”

Syntactic Categories



“Very traditional. He’s the noun. She’s the adjective.”

© The New Yorker Collection 2003 William Haefeli from cartoonbank.com All Rights Reserved.

Each grouping in the tree diagrams of “The child found a puppy” is a member of a large family of similar expressions. For example, *the child* belongs to a family that includes *the police officer*, *your neighbor*, *this yellow cat*, *he*, *John*, and countless others. We can substitute any member of this family for the child without affecting the grammaticality of the sentence, although the meaning of course would change.

A police officer found a puppy.
Your neighbor found a puppy.
This yellow cat found a puppy.

A family of expressions that can substitute for one another without loss of grammaticality is called a **syntactic category**.

The child, *a police officer*, *John*, and so on belong to the syntactic category **noun phrase (NP)**, one of several syntactic categories in English and all languages. NPs may function as subjects or as objects in sentences. An NP often contains a *determiner* (like *a* or *the*) and a noun, but it may also consist of a proper name, a pronoun, a noun without a determiner, or even a clause or

a sentence. Even though a proper noun like *John* and pronouns such as *he* and *him* are single words, they are technically NPs, because they pattern like NPs in being able to fill a subject or object or other NP slots.

John found the puppy.
He found the puppy.
Boys love puppies.
The puppy loved him.
The puppy loved John.

NPs can be more complex, as illustrated by the sentence:

The girl that Professor Snape loved married the man of her dreams.

The NP subject of this sentence is *the girl that Professor Snape loved*, and the NP object is *the man of her dreams*.

Syntactic categories are part of a speaker's knowledge of syntax. That is, speakers of English know that only items (a), (b), (e), (f), and (g) in the following list are NPs even if they have never heard the term *noun phrase* before.

1. (a) a bird
(b) the red banjo
(c) have a nice day
(d) with a balloon
(e) the woman who was laughing
(f) it
(g) John
(h) went

You can test this claim by inserting each expression into three contexts: *What/who I heard was _____*, *Who found _____?* and *_____ was seen by everyone*. For example, **Who found with a balloon?* is ungrammatical, as is **Went was seen by everyone*, as opposed to *Who found it?* or *John was seen by everyone*. Only NPs fit into these contexts because only NPs can function as subjects and objects.

There are other syntactic categories. The expression *found a puppy* is a **verb phrase (VP)**. A verb phrase always contains a **verb (V)**, and it may contain other categories, such as a noun phrase or **prepositional phrase (PP)**, which is a preposition followed by an NP, such as *in the park*, *on the roof*, *with a balloon*. In (2) the VPs are those phrases that can complete the sentence "The child _____."

2. (a) saw a clown
(b) a bird
(c) slept
(d) smart
(e) ate the cake
(f) found the cake in the cupboard
(g) realized that the Earth was round

Inserting (a), (c), (e), (f), and (g) will produce grammatical sentences, whereas the insertion of (b) or (d) would result in an ungrammatical sentence. Thus, (a), (c), (e), (f), and (g) are verb phrases.

Lexical and Functional Categories

There are ten parts of speech, and they are all troublesome.

MARK TWAIN, "The Awful German Language," in *A Tramp Abroad*, 1880

Syntactic categories include both phrasal categories such as NP, VP, AP (adjective phrase), PP (prepositional phrase), and AdvP (adverbial phrase), as well as lexical categories such as noun (N), verb (V), preposition (P), adjective (A), and adverb (Adv). Each lexical category has a corresponding phrasal category. Following is a list of phrasal categories and lexical categories with some examples of each type:

Phrasal categories

Noun Phrase (NP)	<i>men, the man, the man with a telescope</i>
Verb Phrase (VP)	<i>sees, always sees, rarely sees the man, often sees the man with a telescope</i>
Adjective Phrase (AP)	<i>happy, very happy, very happy about winning</i>
Prepositional Phrase (PP)	<i>over, nearly over, nearly over the hill</i>
Adverbial Phrase (AdvP)	<i>brightly, more brightly, more brightly than the Sun</i>

Lexical categories

Noun (N)	<i>puppy, boy, man, soup, happiness, fork, kiss, pillow</i>
Verb (V)	<i>find, run, sleep, throw, realize, see, try, want, believe</i>
Preposition (P)	<i>up, down, across, into, from, by, with, over</i>
Adjective (A)	<i>red, big, happy, candid, hopeless, fair, idiotic, lucky</i>
Adverb (Adv)	<i>again, always, brightly, often, never, very, fairly</i>

Many of these categories may already be familiar to you. As mentioned earlier, some of them are traditionally referred to as *parts of speech*. Other categories may be less familiar, for example, the category **determiner (Det)**, which includes the articles *a* and *the*, as well as **demonstratives** such as *this*, *that*, *these*, and *those*, and "quantifiers" such as *each* and *every*.

Another less familiar category is T(ense), which includes the **modal** auxiliaries *may*, *might*, *can*, *could*, *must*, *shall*, *should*, *will*, and *would*, and abstract tense morphemes that we discuss below. T and Det are **functional categories**, so called because their members have grammatical functions rather than descriptive meanings. For example, determiners specify whether a noun is indefinite or definite (*a boy* versus *the boy*), or the proximity of the person or object to the context (*this boy* versus *that boy*). Tense provides the verb with a time frame, whether present (*John knows Mary*), or past (*John danced*). In English, T is expressed as a (sometimes silent) morpheme on the verb, except in the future tense, which is expressed with the modal *will*. Modals also express notions such as possibility (*John may dance*); necessity (*John must dance*); ability (*John can dance*); and so on. The modals belong to a larger class of verbal elements traditionally referred to as **auxiliaries** or helping verbs, which also include *have* and *be* in sentences such as *John is dancing* or *John has danced*.

Each lexical category typically has a particular kind of meaning associated with it. For example, verbs usually refer to actions, events, and states (*kick*, *marry*, *love*); adjectives to qualities or properties (*lucky*, *old*); common nouns to

general entities (*dog, elephant, house*); and proper nouns to particular individuals (*Noam Chomsky*) or places (*Dodger Stadium*) or other things that people give names to, such as commercial products (*Coca-Cola, Viagra*).

But the relationship between grammatical categories and meaning is more complex than these few examples suggest. For example, some nouns refer to events (*marriage* and *destruction*) and others to states (*happiness, loneliness*). We can use abstract nouns such as *honor* and *beauty*, rather than adjectives, to refer to properties and qualities. In the sentence “Seeing is believing,” *seeing* and *believing* are nouns but are not entities. Prepositions are usually used to express relationships between two entities involving a location (e.g., *the boy is in the room, the cat is under the bed*), but this is not always the case; the prepositions *of, by, about, and with* often have other than locational meanings.

Because of the difficulties involved in specifying the precise meaning of lexical categories, we do not usually define categories in terms of their meanings, but rather on the basis of where they occur in a sentence, what categories co-occur with them, and what their morphological characteristics are. For example, we define a noun as a word that can occur with a determiner (*the boy*) and that can (ordinarily) take a plural marker (*boys*); a verb as a word that can occur with an adverb (*run fast*) or modal (*may go, will dance*); an adjective as a word that can occur with a degree word (*very hungry*) or a morphological marker (*hungrier*), among other properties.

All languages have syntactic categories such as N, V, and NP. Speakers know the syntactic categories of their language even if they do not know the technical terms. Our knowledge of syntactic classes is revealed when we substitute equivalent phrases, as we just did in examples (1) and (2), and when we use the various syntactic tests that we have discussed.

Phrase Structure Trees

Who climbs the Grammar-Tree distinctly knows

Where Noun and Verb and Participle grows.

JOHN DRYDEN, “The Sixth Satyr of Juvenal,” 1693

Now that you know something about constituent structure and grammatical categories, you are ready to learn how the phrases and sentences of a language are constructed. We will begin by illustrating trees for simple phrases and then proceed to more complex structures. The trees that we will build here are more detailed than those we saw in the previous sections, because the branches of the tree will have category labels identifying each constituent.

One of the striking things we observe when we consider the various phrasal categories discussed above is that they have a similar organization. Consider the following examples of each of the phrasal categories

NP: *the mother of James Whistler*

VP: *sing an aria*

AP: *wary of snakes*

PP: *over the hill*