

tion. If asked to fix it up, we would probably come up with something like “Disa put the baby to sleep,” but we also know that as it stands, (8a) is not a possible sentence of English. 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 impossible. 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 8.

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

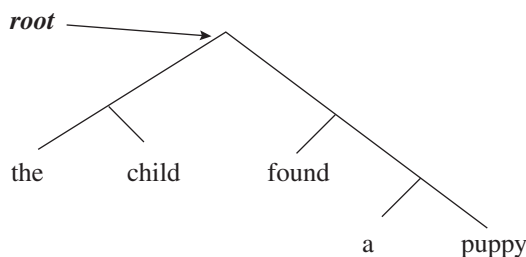
This template says that a determiner (an article) is followed by a noun, which is followed by a verb, and so on. It would describe English sentences such as the following:

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 a 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**:



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*, *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

Parts is parts.

WENDY’S COMMERCIAL, 2006

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, they form a constituent. For example, the set of words that can be used to answer a question is a constituent. So in answer 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.

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* 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*.

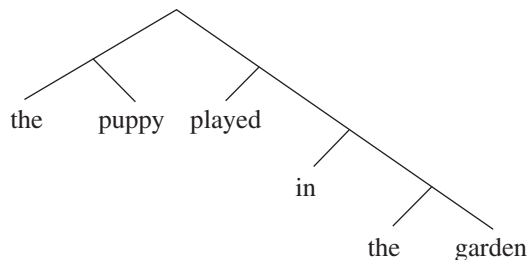
Some sentences have a prepositional phrase in the predicate. Consider

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.

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 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. For example, it could stand alone in answer to the question “Who did you buy an antique desk for?” In its second meaning, the phrase *with thick legs and large drawers* modifies the phrase *a desk for a lady*, and thus the structure is [[a desk for a lady][with thick legs and large drawers]].

Syntactic Categories



“ I MISS THE GOOD OLD DAYS WHEN ALL WE HAD TO WORRY ABOUT WAS NOUNS AND VERBS.”

© ScienceCartoonsPlus.com.

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 every other language in the world. NPs may function as the subject or as an object in a sentence. NPs often contain a *determiner* (like *a* or *the*) and a noun, but they 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: *Who found _____, _____ was seen by everyone,* and *What/who I heard was _____*. For example, **Who found with a balloon* is ungrammatical, as is **Have a nice day 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, AdjP (adjective phrase), PP (prepositional phrase), and AdvP (adverbial phrase), as well as lexical categories such as noun (N), verb (V), preposition (P), adjective (Adj), and adverb (Adv). Each lexical category has a corresponding phrasal category. Following is a list of lexical categories with some examples of each type:

Lexical categories

Noun (N)	<i>puppy, boy, soup, happiness, fork, kiss, pillow, cake, cupboard</i>
Verb (V)	<i>find, run, sleep, throw, realize, see, try, want, believe</i>
Preposition (P)	<i>up, down, across, into, from, by, with</i>
Adjective (Adj)	<i>red, big, candid, hopeless, fair, idiotic, lucky</i>
Adverb (Adv)	<i>again, carefully, luckily, 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 “counting words” such as *each* and *every*. Another less familiar category is **auxiliary (Aux)**, which includes the verbs *have*, *had*, *be*, *was*, and *were*, and the **modals** *may*, *might*, *can*, *could*, *must*, *shall*, *should*, *will*, and *would*. Aux and Det are **functional categories**, so called because their members have a grammatical function rather than a descriptive meaning. 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*). Auxiliaries provide the verb with a time frame, whether ongoing (*John is dancing*), completed in the past (*John has danced*), or occurring in the future (*John will dance*). Auxiliaries may also express notions such as possibility (*John may dance*), necessity (*John must dance*), ability (*John can dance*), and so on.

Lexical categories typically have particular kinds of meanings associated with them. 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* are not locational. 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 their syntactic distribution (where they occur in a sentence) and morphological characteristics. For example, we define a noun as a word that can occur with a determiner (*the boy*) and that can take a plural marker (*boys*), 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 the 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 and Rules

Who climbs the Grammar-Tree distinctly knows

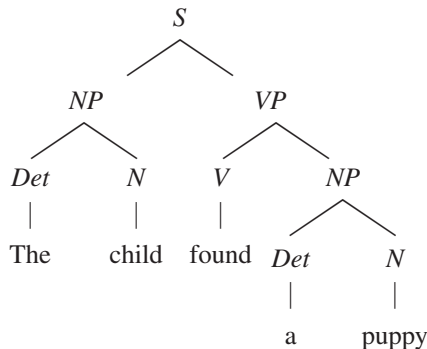
Where Noun and Verb and Participle grows.

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 sentences of a language are constructed. We will begin by building trees for simple sentences 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. In this section we will also introduce the syntactic rules that **generate** (a technical term for describe or specify) the different kinds of structures.

The following tree diagram provides labels for each of the constituents of the sentence “The child found a puppy.” These labels show that the entire sentence

belongs to the syntactic category of S (because the S-node encompasses all the words). It also reveals that *the child* and *a puppy* belong to the category NP, that is, they are noun phrases, and that *found a puppy* belongs to the category VP or is a verb phrase, consisting of a verb and an NP. It also reveals the syntactic category of each of the words in the sentence.



A tree diagram with syntactic category information is called a **phrase structure tree** or a **constituent structure tree**. This tree shows that a sentence is both a linear string of words and a hierarchical structure with phrases nested in phrases. Phrase structure trees (PS trees, for short) are explicit graphic representations of a speaker's knowledge of the structure of the sentences of his language.

PS trees represent three aspects of a speaker's syntactic knowledge:

1. The linear order of the words in the sentence
2. The identification of the syntactic categories of words and groups of words
3. The hierarchical structure of the syntactic categories (e.g., an S is composed of an NP followed by a VP, a VP is composed of a V that may be followed by an NP, and so on)

In chapter 3 we discussed the fact that the syntactic category of each word is listed in our mental dictionaries. We now see how this information is used by the syntax of the language. Words appear in trees under labels that correspond to their syntactic category. Nouns are under N, determiners under Det, verbs under V, and so on.

The larger syntactic categories, such as VP, consist of all the syntactic categories and words below that point, or **node**, in the tree. The VP in the PS tree above consists of syntactic category nodes V and NP and the words *found*, *a*, and *puppy*. Because *a puppy* can be traced up the tree to the node NP, this constituent is a noun phrase. Because *found* and *a puppy* can be traced up to the node VP, this constituent is a verb phrase. The PS tree reflects the speaker's intuitions about the natural groupings of words in a sentence. In discussing trees, every higher node is said to **dominate** all the categories beneath it. S dominates every node. A node is said to **immediately dominate** the categories one level below it. VP immediately dominates V and NP, the categories of which it is composed. Categories that are immediately dominated by the same node are **sisters**. V and NP are sisters in the phrase structure tree of "the child found a puppy."

A PS tree is a formal device for representing the speaker's knowledge of the structure of sentences in his language, as revealed by our linguistic intuitions. When we speak, we are not aware that we are producing sentences with such structures, but controlled experiments, such as the click experiments described earlier, show that we use them in speech production and comprehension. We will discuss these experiments further in chapter 9.

The information represented in a PS tree can also be represented by another formal device: phrase structure (PS) rules. PS rules capture the knowledge that speakers have about the possible structures of a language. Just as a speaker cannot have an infinite list of sentences in her head, so she cannot have an infinite set of PS trees in her head. Rather, a speaker's knowledge of the permissible and impermissible structures must exist as a finite set of rules that generate a tree for any sentence in the language. To express the structure given above, we need the following PS rules:

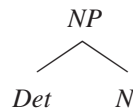
1. $S \rightarrow NP VP$
2. $NP \rightarrow Det N$
3. $VP \rightarrow V NP$

Phrase structure rules specify the well-formed structures of a language precisely and concisely. They express the regularities of the language and make explicit a speaker's knowledge of the order of words and the grouping of words into syntactic categories. For example, in English an NP may contain a determiner followed by a noun. This is represented by rule 2. This rule conveys two facts:

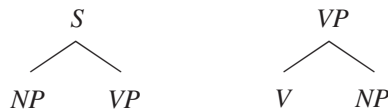
A noun phrase can contain a determiner followed by a noun in that order.

A determiner followed by a noun is a noun phrase.

You can think of PS rules as templates that a tree must match to be grammatical. To the left of the arrow is the dominating category, in this case NP, and the categories that it immediately dominates—that comprise it—appear on the right side, in this case Det and N. The right side of the arrow also shows the linear order of these components. Thus, one subtree for the English NP looks like this:



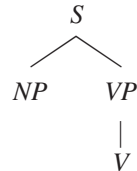
Rule 1 says that a sentence (S) contains (immediately dominates) an NP and a VP in that order. Rule 3 says that a verb phrase consists of a verb (V) followed by an NP. These rules are general statements and do not refer to any specific VP, V, or NP. The subtrees represented by rules 1 and 3 are as follows:



A VP need not contain an NP object, however. It may include a verb alone, as in the following sentences:

The woman laughed.
The man danced.
The horse galloped.

These sentences have the structure:



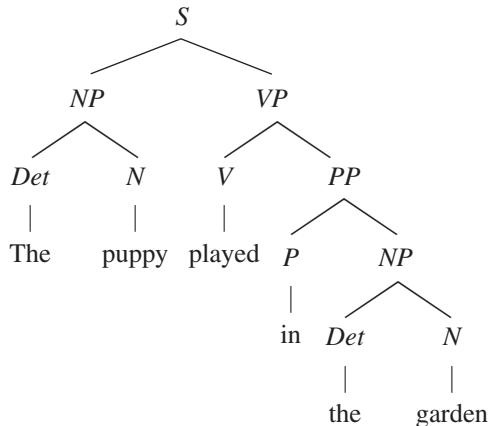
Thus a tree must have a VP that immediately dominates V, as specified by rule 4, which is therefore added to the grammar:

4. $VP \rightarrow V$

The following sentences contain prepositional phrases following the verb:

The puppy played in the garden.
The boat sailed up the river.
A girl laughed at the monkey.
The sheepdog rolled in the mud.

The PS tree for such sentences is



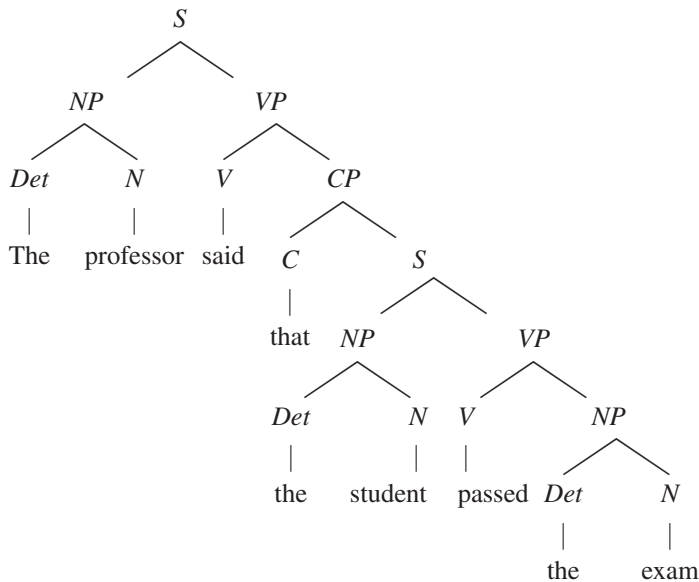
To permit structures of this type, we need two additional PS rules, as in 5 and 6.

5. $VP \rightarrow V PP$

6. $PP \rightarrow P NP$

Another option open to the VP is to contain or *embed* a sentence. For example, the sentence “The professor said that the student passed the exam” contains

the sentence “the student passed the exam.” Preceding the **embedded sentence** is the word *that*, which is a **complementizer (C)**. C is a functional category, like Aux and Det. Here is the structure of such sentence types:



To allow such embedded sentences, we need to add these two new rules to our set of phrase structure rules.

7. $VP \rightarrow V CP$
8. $CP \rightarrow C S$

CP stands for complementizer phrase. Rule 8 says that CP contains a complementizer such as *that* followed by the embedded sentence. Other complementizers are *if* and *whether* in sentences like

I don't know whether I should talk about this.

The teacher asked if the students understood the syntax lesson.

that have structures similar to the one above.

Here are the PS rules we have discussed so far. A few other rules will be considered later.

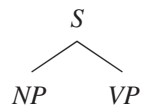
1. $S \rightarrow NP VP$
2. $NP \rightarrow Det N$
3. $VP \rightarrow V NP$
4. $VP \rightarrow V$
5. $VP \rightarrow V PP$
6. $PP \rightarrow P NP$
7. $VP \rightarrow V CP$
8. $CP \rightarrow C S$

Some Conventions for Building Phrase Structure Trees

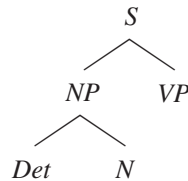
Everyone who is master of the language he speaks . . . may form new . . . phrases, provided they coincide with the genius of the language.

JOHANN DAVID MICHAELIS, *Dissertation*, 1769

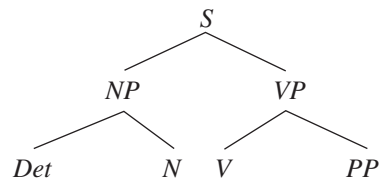
One can use the phrase structure rules as a guide for building trees that follow the structural constraints of the language. In so doing, certain conventions are followed. The *S* occurs at the top or “root” of the tree (it’s upside down). Another convention specifies how the rules are applied: First, find the rule with *S* on the left side of the arrow, and put the categories on the right side below the *S*, as shown here:



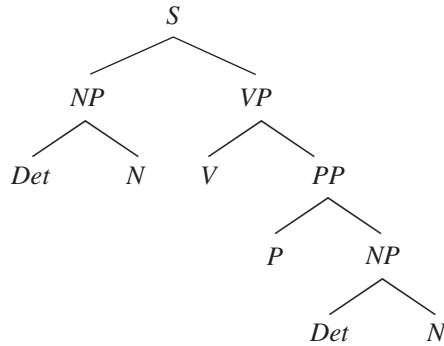
Continue by matching any syntactic category at the bottom of the partially constructed tree to a category on the left side of a rule, then expand the tree with the categories on the right side. For example, we may expand the tree by applying the NP rule to produce:



The categories at the bottom are *Det*, *N*, and *VP*, but only *VP* occurs to the left of an arrow in the set of rules and so needs to be expanded using one of the *VP* rules. Any one of the *VP* rules will work. The order in which the rules appear in the list of rules is irrelevant. (We could have begun by expanding the *VP* rather than the *NP*.) Suppose we use rule 5 next. Then the tree has grown to look like this:



Convention dictates that we continue in this way until none of the categories at the bottom of the tree appears on the left side of any rule (i.e., no phrasal categories may remain unexpanded). The *PP* must expand into a *P* and an *NP* (rule 6), and the *NP* into a *Det* and an *N*. We can use a rule as many times as it can apply. In this tree, we used the *NP* rule twice. After we have applied all the rules that can apply, the tree looks like this:



By following these conventions, we generate only trees specified by the PS rules, and hence only trees that conform to the syntax of the language. By implication, any tree not so specified will be ungrammatical, that is, not permitted by the syntax. At any point during the construction of a tree, any rule may be used as long as its left-side category occurs somewhere at the bottom of the tree. By choosing different VP rules, we could specify different structures corresponding to sentences such as:

The boys left. (VP \rightarrow V)

The wind blew the kite. (VP \rightarrow V NP)

The senator hopes that the bill passes. (VP \rightarrow V CP)

Because the number of possible sentences in every language is infinite, there are also an infinite number of trees. However, all trees are built out of the finite set of substructures allowed by the grammar of the language, and these substructures are specified by the finite set of phrase structure rules.

The Infinity of Language: Recursive Rules

So, naturalists observe, a flea
 Hath smaller fleas that on him prey;
 And these have smaller still to bite 'em,
 And so proceed ad infinitum.

JONATHAN SWIFT, "On Poetry, a Rhapsody," 1733

We noted at the beginning of the chapter that the number of sentences in a language is infinite and that languages have various means of creating longer and longer sentences, such as adding an adjective or a prepositional phrase. Even children know how to produce and understand very long sentences and know how to make them even longer, as illustrated by the children's rhyme about the house that Jack built.

This is the farmer sowing the corn,
 that kept the cock that crowed in the morn,
 that waked the priest all shaven and shorn,
 that married the man all tattered and torn,
 that kissed the maiden all forlorn,
 that milked the cow with the crumpled horn,
 that tossed the dog,

that worried the cat,
 that killed the rat,
 that ate the malt,
 that lay in the house that Jack built.

The child begins the rhyme with *This is the house that Jack built*, continues by lengthening it to *This is the malt that lay in the house that Jack built*, and so on.

You can add any of the following to the beginning of the rhyme and still have a grammatical sentence:

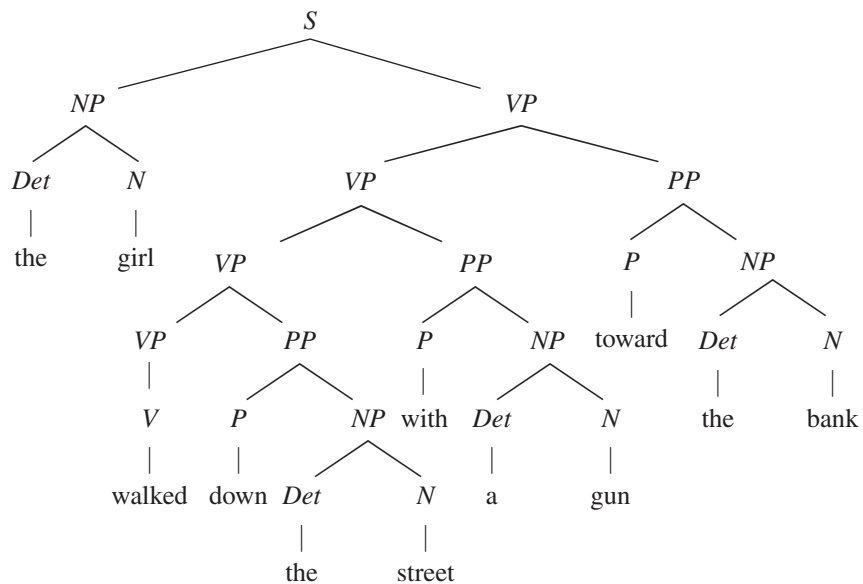
I think that . . .
 What is the name of the unicorn that noticed that . . .
 Ask someone if . . .
 Do you know whether . . .

Once we acknowledge the unboundedness of sentences, we need a formal device to capture that crucial aspect of speakers' syntactic knowledge. It is no longer possible to specify each legal structure; there are infinitely many.

To see how this works, let us first look at the case of multiple prepositional phrases such as [The girl walked [down the street] [over the hill] [through the woods] . . .]. VP substructures currently allow only one PP per sentence (VP → V PP—rule 5). We can rectify this problem by revising rule 5:

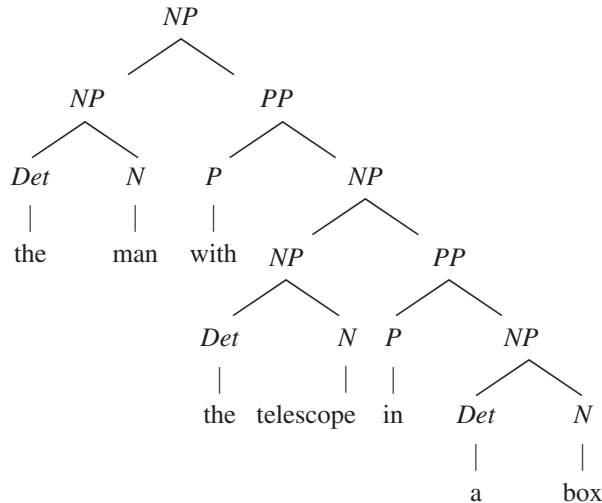
5. VP → VP PP

Rule 5 is different from the previous rules because it repeats its own category (VP) inside itself. This is an instance of a **recursive rule**. Recursive rules are of critical importance because they allow the grammar to generate an infinite set of sentences. Reapplying rule 5 shows how the syntax permits structures with multiple PPs, such as in the sentence “The girl walked down the street with a gun toward the bank.”



In this structure the VP rule 5 has applied three times and so there are three PPs: [down the street] [with a gun] [toward the bank]. It is easy to see that the rule could have applied four or more times, for example by adding a PP like *for no good purpose*.

NPs can also contain PPs recursively. An example of this is shown by the phrase *the man with the telescope in a box*.



To show that speakers permit recursive NP structures of this sort, we need to include the following PS rule, which is like the recursive VP rule 5.

9. NP → NP PP

The PS rules define the allowable structures of the language, and in so doing make predictions about structures that we may not have considered when formulating each rule individually. These predictions can be tested, and if they are not validated, the rules must be reformulated because they must generate all and only the allowable structures. For example, rule 7 (VP → V CP) in combination with rules 8 (CP → C S) and 1 (S → NP VP) form a recursive set. (The recursiveness comes from the fact that S and VP occur on both the left and right side of the rules.) Those rules allow S to contain VP, which in turn contains CP, which in turn contains S, which in turn again contains VP, and so on, potentially without end. These rules, formulated for different purposes, correctly predict the limitlessness of language in which sentences are embedded inside larger sentences, such as *The children hope that the teacher knows that the principal said that the school closes for the day* as illustrated on the following page.

