5

Syntax: the analysis of sentence structure

… the game is to say something new with old words

RALPH WALDO EMERSON, *JOURNALS*, 1849

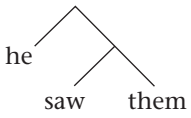
NOT MUCH can be said with a single word. If we are to use language to express complex thoughts and ideas, we must be able to combine and organize words into sentences. Not just any combination of words will do, however: the pattern in (1) is not permissible, even though the same words can be combined in a different way to form the acceptable sentence in (2).

(1) *House painted student a the.
(2) A student painted the house.

We say that an utterance is **grammatical** if native speakers judge it to be a possible sentence of their language. Hence, (2) is grammatical, but (1) is not.

This chapter will focus on **syntax**, the component of grammar that is concerned with the form of grammatical sentences. The starting point for work on syntax is the universally accepted idea that words belong to categories of different types (nouns, verbs, and so on) and that these categories can be combined in particular ways to form phrases—and ultimately, sentences. One widely accepted way to represent the internal structure of sentences makes use of 'tree diagrams', like the simplified one in figure 5.1.

FIGURE 5.1
A simplified syntactic structure (tree diagram)



As we will see a little later in this chapter, structures like this are built by two interacting operations. A **Merge** operation combines words to create larger phrases and sentences, and a **Move** operation can carry an element to a new position within the structure. Before examining these operations, however, it is necessary to have a look at words and the categories to which they belong.

Language Matters **What's the Longest Sentence in English?**

George Bernard Shaw wrote one that was 110 words long. William Faulkner's novel *Absalom, Absalom!* includes a 1300-word sentence. James Joyce managed to produce a 4391-word sentence (that goes on for forty pages) in *Ulysses*. But even that's not the longest known sentence—*The Rotter's Club* by Jonathon Coe contains a sentence that is 13 955 words long!

The bottom line is that there's no such thing as the world's longest sentence—any sentence can be made longer. That's because the operations that combine words can be used over and over again, without limit.

It's the right answer
I think it's the right answer
You know I think it's the right answer
Harry said you know I think it's the right answer.

….

a book
a book on the table
a book on the table near the bookcase
a book on the table near the bookcase in the office

The application of an operation to its own output to create an ever more complex structure is called **recursion**, and it's an essential part of our ability to build sentences.

5.1 Categories and structure

A fundamental fact about words in all human languages is that they can be grouped together into a relatively small number of classes called **syntactic categories** or **parts of speech**. This classification reflects a variety of factors, including the types of meaning that words express, the types of affixes that they take, and the types of structures in which they can occur.

5.1.1 Categories of words

Table 5.1 provides examples of the word-level categories that are most central to the study of syntax. The four most studied syntactic categories are **noun (N), verb (V), adjective (A)**, and **preposition (P)**. These elements, which are often called **lexical categories**, play a very important role in sentence formation, as we will soon see. A fifth and less studied lexical category consists of **adverbs (Adv)**, most of which are derived from adjectives.

Languages may also contain **non-lexical** or **functional categories**, including **determiner (Det), auxiliary verb (Aux), conjunction (Con)**, and **degree word (Deg)**. Such elements generally have meanings that are harder to define and paraphrase than those of lexical categories. For example, the meaning of a determiner such as *the* or an auxiliary such as *would* is more difficult to describe than the meaning of a noun such as *hill* or *vehicle*.

TABLE 5.1 Syntactic categories

Lexical categories ('content words')	**Examples**
Noun (N)	Harry, boy, wheat, policy, moisture, bravery
Verb (V)	arrive, discuss, melt, hear, remain, dislike
Adjective (A)	good, tall, old, intelligent, beautiful, fond
Preposition (P)	to, in, on, near, at, by
Adverb (Adv)	slowly, quietly, now, always, perhaps
Non-lexical categories ('functional categories')	**Examples**
Determiner (Det)	the, a, this, these, no (as in *no books*)
Degree word (Deg)	too, so, very, more, quite
Auxiliary (Aux)	
Modal	will, would, can, could, may, must, should
Non-modal	be, have, do
Conjunction (Con)	and, or, but

A potential source of confusion in the area of word classification stems from the fact that some items can belong to more than one category.

(3) *comb* used as a noun:
The woman found a comb.

comb used as a verb:
The boy should comb his hair.

(4) *near* used as a preposition:
The child stood near the fence.

near used as a verb:
The runners neared the finish line.

near used as an adjective:
The end is nearer than you might think.

How then can we determine a word's category?

Meaning

One criterion involves meaning. For instance, nouns typically name entities ('people and things'), including individuals (*Harry, Sue*) and objects (*book, desk*). Verbs characteristically designate actions (*run, jump*), sensations (*feel, hurt*), and states (*be, remain*). Consistent with these tendencies, *comb* in (3) refers to an object when used as a noun but to an action when used as a verb.

The typical function of an adjective is to designate a property or attribute of the entities denoted by nouns. Thus, when we say *that tall building*, we are attributing the property 'tall' to the building designated by the noun.

In a parallel way, adverbs typically denote properties and attributes of the actions, sensations, and states designated by verbs. In the following sentences, for example, the adverb *quickly* indicates the manner of Janet's leaving, while the adverb *early* specifies its time.

(5) Janet left quickly.
Janet left early.

A word's category membership does not always bear such a straightforward relationship to its meaning, however. For example, nouns such as *difficulty, truth*, and *likelihood* do not name entities in the strict sense. Moreover, even though words for actions tend to be verbs, some nouns also express this type of meaning (e.g., *push* in *give someone a push* and *run* in *have a run*).

Matters are further complicated by the fact that in some cases, words with very similar meanings belong to different categories. For instance, the words *like* and *fond* are very similar in meaning (as in *Mice like/are fond of cheese*), yet *like* is a verb and *fond* is an adjective.

Inflection

Most linguists believe that meaning is only one of several criteria that enter into determining a word's category. As shown in table 5.2, inflection can also be very useful for distinguishing among different categories of words.

TABLE 5.2 Lexical categories and their inflectional suffixes in English

Category	Inflectional affix	Examples
Noun	plural *-s* possessive *-'s*	books, chairs, doctors John's, (the) man's
Verb	past tense *-ed* progressive *-ing* third person singular *-s*	arrived, melted, hopped arriving, melting, hopping arrives, melts, hops
Adjective	comparative *-er* superlative *-est*	taller, faster, smarter tallest, fastest, smartest

However, even inflection does not always provide the information needed to determine a word's category. In English, for example, not all adjectives can take the comparative and superlative suffixes (**intelligenter*, **beautifulest*) and some nouns cannot be pluralized (*moisture, bravery, knowledge*).

Distribution

A third and often more reliable criterion for determining a word's category involves the type of elements (especially functional categories) with which it can co-occur (its **distribution**).

For example, nouns can typically appear with a determiner, verbs with an auxiliary, and adjectives with a degree word in the patterns illustrated in table 5.3.

TABLE 5.3 Distributional properties of nouns, verbs, and adjectives

Category	Distributional property	Examples
Noun	occurrence with a determiner	a car, the wheat
Verb	occurrence with an auxiliary	has gone, will stay
Adjective	occurrence with a degree word	very rich, too big

In contrast, a noun cannot occur with an auxiliary, and a verb cannot occur with a determiner or degree word.

(6) a noun with an auxiliary:
*will destruction

a verb with a determiner:
*the destroy

a verb with a degree word:
*very arrive

Distributional tests for category membership are simple and highly reliable. They can be used with confidence when it is necessary to categorize unfamiliar words.

Language Matters **A Poem That Syntacticians Love**

Thanks to distributional and inflectional clues, it's often possible to identify a word's category without knowing its meaning. The poem "Jabberwocky" by Lewis Carroll illustrates this point in a particularly brilliant way—it's interpretable precisely because readers are able to figure out that *gyre* is a verb (note the auxiliary verb to its left), that *borogoves* is a noun (it's preceded by a determiner and takes the plural ending), and so on.

'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogoves,
And the mome raths outgrabe.

"Beware the Jabberwock, my son!
The jaws that bite, the claws that catch!
Beware the Jubjub bird, and shun
The frumious Bandersnatch!"

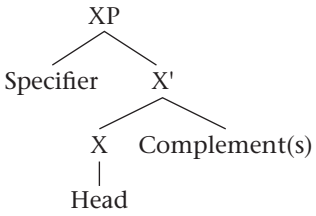
5.1.2 Phrase structure

Sentences are not formed by simply stringing words together like beads on a necklace. Rather, they have a hierarchical design in which words are grouped together into ever larger structural units called **phrases**—*the door, to the door, go to the door*, and so on.

The blueprint

As a first approximation, it is often suggested that the internal structure of phrases follows the design shown in figure 5.1 (X' is pronounced 'X-bar').

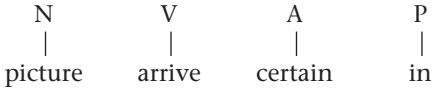
FIGURE 5.2
The X' Schema—a template for phrase structure



Heads

The **head** is the obligatory nucleus around which a phrase is built. For now we will focus on four categories that can function as the head of a phrase—nouns (N), verbs (V), adjectives (A), and prepositions (P).

FIGURE 5.3
Some examples of heads



Specifiers

Specifiers have no single semantic function. Structurally, though, they are alike in that they occur at the edge of a phrase. As illustrated in table 5.4, the specifier position in English is at a phrase's left margin (the beginning).

TABLE 5.4 Some specifiers

Head	Specifier	Examples
N	Determiner (Det) *the, a, some, this, those…*	***a*** *picture,* ***the*** *map,* ***those*** *people* ***some*** *guests*
V	Preverbal adverb (Adv) *never, perhaps, often, always, almost…*	***never*** *quit,* ***perhaps*** *go,* ***often*** *failed,* ***almost*** *forgot*
A or P	Degree word (Deg) *very, quite, more, almost…*	***very*** *smart,* ***quite*** *rich,* ***almost*** *in*

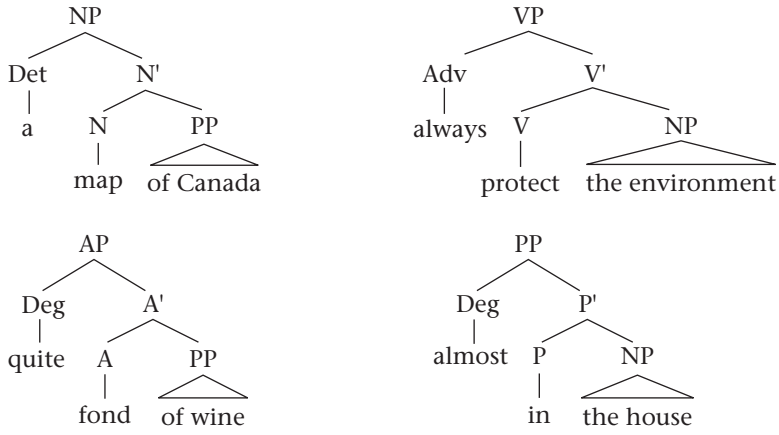
Note: *Almost* can be either an adverb or a degree word.

Complements

Complements, which are always phrases, provide information about entities and locations implied by the meaning of the head. For example, the meaning of *protect* implies something that is protected (*protect the environment*); the meaning of *in* implies a location (*in the house*); the meaning of *map* implies an area that is depicted, as in *a map of Canada;* and so on.

As illustrated in figure 5.4, the X' Schema ensures that when a phrase includes both a specifier and a complement in addition to the head, the specifier will occur higher than the complement. To simplify here, we don't show the internal structure of the complement phrases.

FIGURE 5.4
Phrases containing a head, a specifier, and a complement



However, it is common (and practical!) to represent tree structures in an abbreviated way, without the intermediate X', when there is no specifier and/or complement, as shown in figures 5.5 and 5.6.

FIGURE 5.5
Phrases consisting of just a specifier and a head

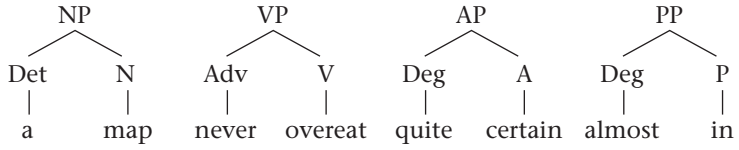
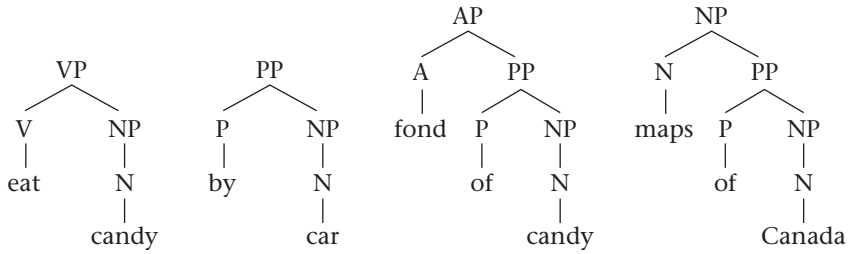


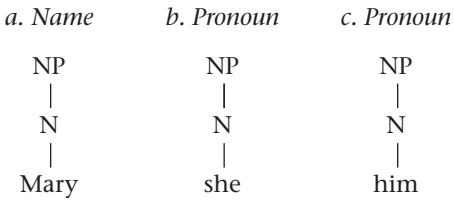
FIGURE 5.6
Phrases consisting of just a head and a complement



Only when the phrase contains both a specifier and a complement in addition to the head is it necessary to make use of the intermediate X' level.

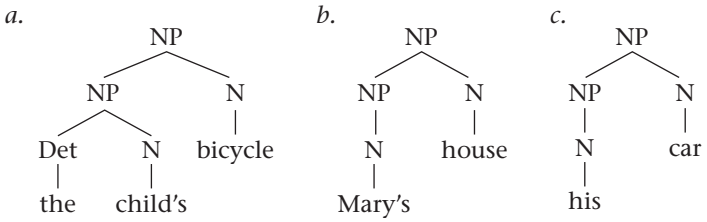
In the interests of being able to consider the largest number of patterns possible, we will adopt two common additional assumptions. First, we will treat both names (*Mary, Bob*, etc.) and pro*nouns* (*she, he, him, her*, etc.) as instances of the N category that do not normally take either specifiers or complements.

FIGURE 5.7
Names and pronouns



Second, we will assume that possessives (e.g., *the child's, Mary's, his*, etc.) are NPs that occur in the specifier position of a larger NP.

FIGURE 5.8
Possessives: NPs inside NPs



An appendix at the end of the chapter offers detailed instructions on how to draw tree structures; exercises 3 and 4 provide an opportunity to practice.

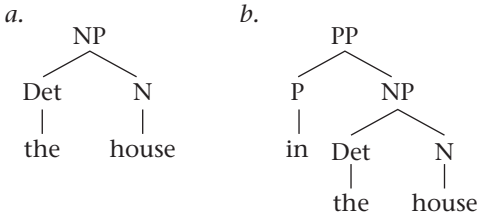
The Merge operation

We can now formulate the following operation for sentence building.

(7) *Merge*
Combine words in a manner compatible with the X' Schema.

As illustrated in figure 5.9, the Merge operation is able to take a determiner such as *the* and combine it with the N *house* to form the NP *the house*. It is then able to take a preposition such as *in* and combine it with the NP *the house* to form the PP *in the house*.

FIGURE 5.9
The Merge operation in action

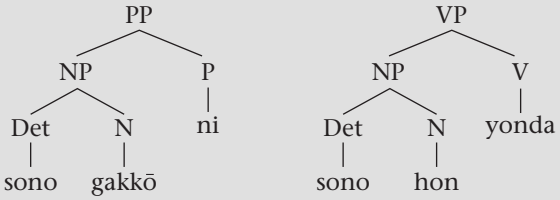


Further application of the Merge operation to additional words can lead to the formation of phrases and sentences of unlimited complexity.

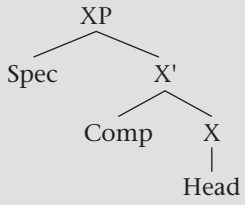
Language Matters **The Mirror Image**

Many languages have a head-complement order that is the mirror image of the one found in English—the complement occurs on the left side of the head rather than on the right side. (In both types of language, the specifier appears on the left side of the head.) Japanese works that way: the V occurs at the end of the VP, the P at the end of the PP, and so on.

[sono gakkō]-ni	[sono hon] yonda
that school at	that book read (Pst)
'at that school'	'read that book'



The version of the X' Schema needed for these languages looks like this—with the head to the right of its complement:



About half of the world's languages use this version of the X' Schema.

5.1.3 Sentences

The largest unit of syntactic analysis is the sentence. Sentences typically consist of an NP (often called 'the subject') and a VP that are linked together by an abstract category dubbed 'T' (for 'tense'). As illustrated in figure 5.10, T serves as the head of the sentence, taking the

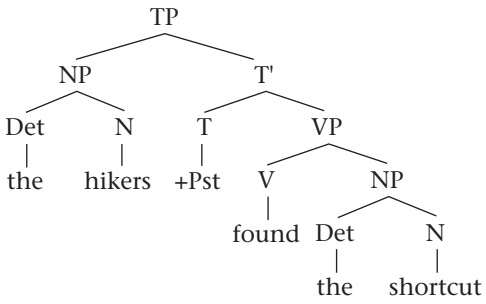


FIGURE 5.10
The structure of a typical sentence

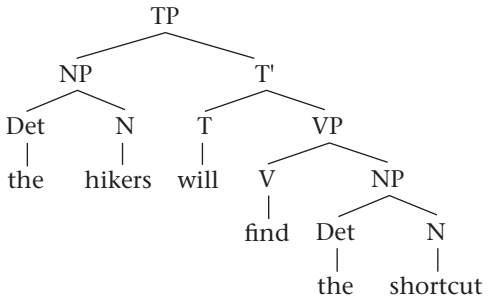
VP as its complement and the subject NP as its specifier (+Pst = past, –Pst = non-past). What we think of as a sentence or a sentential phrase, then, is really a TP.

The tense feature in T must be compatible with the form of the verb. So a sentence like the one above, whose head contains the feature +Pst, must contain a verb marked for the past tense (hence, *found* rather than *find*).

Although somewhat abstract, this analysis has the advantage of giving sentences the same internal structure as other phrases (with a specifier, a head, and a complement), making them consistent with the X' Schema. Moreover, because T, like all heads, is obligatory, we also account for the fact that all sentences have tense (i.e., they are all past or non-past).

The TP structure also provides us with a natural place to locate modal auxiliaries such as *can, may, will*, and *must*, most of which are inherently non-past, as shown by their incompatibility with time adverbs such as *yesterday*: ***He can/will/must work yesterday***. (The modals *could* and *would* can be either past or non-past: *He could swim when he was three/He could swim tomorrow*.) Although traditionally called auxiliary verbs, modals are treated as instances of the T category in contemporary syntactic analysis, as depicted in figure 5.11. (Because modals have inherent tense, we will assume that it is not necessary to have the feature ±Pst in the T position when they are used.)

FIGURE 5.11
A sentence with an auxiliary verb in the T position.



This structure neatly accounts not only for the fact that modals express an inherent tense but also for their positioning between the subject (the specifier) and the VP (the complement)—in the position reserved for the head of the sentence. (We will consider the non-modal auxiliary verbs *be* and *have* in section 5.4.1.)

Two types of auxiliary verbs in English

Modal auxiliaries: *will, would, can, could, shall, should, must, may, might*
Non-modal auxiliaries: *be, have, do*

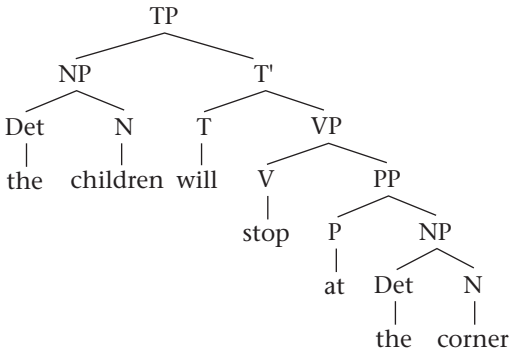
- Only non-modal auxiliaries can be inflected for past tense (*is~was, are~were, have~had, do~did*).
- When both types of auxiliaries appear in the same sentence, the modal always comes first (*They should have gone; They may be going*).
- Only modals are treated as instances of the T category.

The appendix at the end of the chapter outlines a procedure that will help you assign the right structure to sentences. Exercise 5 provides an opportunity to practice this procedure.

5.1.4 Tests for phrase structure

How can linguists be sure that they have grouped words together into phrases in the right way? The existence of the syntactic units, or **constituents**, found in tree structures can be independently verified with the help of special tests. Consider, for instance, the tree structure that the X' Schema requires for the sentence *The children will stop at the corner*.

FIGURE 5.12
The tree structure for *The children will stop at the corner*



The substitution test

One piece of evidence for syntactic units comes from the fact that they can often be replaced by an element such as *they, she, he, it, do so*, and so on. (This is called a **substitution test**.)

As illustrated in (8), *the children* can be replaced by *they*, and *stop at the corner* can be replaced by *do so*—confirming that each is a syntactic unit, just as the tree structure shows by grouping the component parts together under a phrasal label such as NP, VP, and so on.

(8) [$_{NP}$ The children] will [$_{VP}$ stop at the corner]. *They* always *do so*.
(*they = the children; do so = stop at the corner*)

A substitution test also confirms that *at the corner* is a unit, as it can be replaced by a single word in a sentence such as (9).

(9) The children stopped [$_{PP}$ at the corner] and we stopped *there* too.
(*there = at the corner*)

Elements that do not form a constituent cannot be replaced in this way. Thus, there is no word in English that we can use to replace *children stopped*, for example, or *at the*.

The movement test

A second indication that *at the corner* forms a constituent in figure 5.12 is that it can be moved as a single unit to a different position within the sentence. (This is called a **movement test**.) In (10), for instance, *at the corner* can be moved from a position after the verb to the beginning of the sentence.

(10) They stopped [$_{PP}$ at the corner]. → [$_{PP}$ At the corner], they stopped.

Of course, *at the*, which is not a syntactic unit, cannot be fronted in this manner (**At the, they stopped corner*).

The coordination test

Finally, we can conclude that a group of words forms a constituent if it can be joined to another group of words by a conjunction such as *and, or*, or *but*. (This is known as the **coordination test** since patterns built around a conjunction are called coordinate structures.) The sentence in (11) illustrates how coordination can be used to help establish that *stop at the corner* is a constituent.

(11) The children will [$_{VP}$ stop at the corner] *and* [$_{VP}$ look both ways].

5.2 Complement options

How can we be sure that individual words will occur with a complement of the right type in the syntactic structures that we have been building? Information about the complements permitted by a particular head is included in that head's entry in a speaker's lexicon. For instance, the lexicon for English includes an entry for *devour* that indicates that it requires an NP complement.

(12) *a.* *devour* with an NP complement:
The child devoured [$_{NP}$ the sandwich].

b. *devour* without an NP complement:
*The child devoured.

The term **subcategorization** is used to refer to information about a word's complement options, such as the fact the verb *devour* belongs to a verb subcategory that requires an NP complement.

5.2.1 Complement options for verbs

Table 5.5 illustrates some of the more common complement options for verbs in English. The subscripted prepositions indicate subtypes of PP complements, where this is relevant. *Loc* stands for any preposition expressing a location (such as *near, on*, and *under*).

TABLE 5.5 Some examples of verb complements

Complement option	Sample heads	Example
Ø	vanish, arrive, die	The rabbit vanished.
NP	devour, cut, prove	The professor proved [$_{NP}$ *the theorem*].
AP	be, become	The man became [$_{AP}$ *very angry*].
PP_{to}	dash, talk, refer	The dog dashed [$_{PP}$ *to the door*].
NP NP	spare, hand, give	We handed [$_{NP}$ *the man*] [$_{NP}$ *a map*].
NP PP_{to}	hand, give, send	He gave [$_{NP}$ *a diploma*] [$_{PP}$ *to the student*].
NP PP_{for}	buy, cook, reserve	We bought [$_{NP}$ *a hat*] [$_{PP}$ *for Andy*].
NP PP_{loc}	put, place, stand	He put [$_{NP}$ *the muffler*] [$_{PP}$ *on the car*].
PP_{to} PP_{about}	talk, speak	I talked [$_{PP}$ *to a doctor*] [$_{PP}$ *about Sue*].
NP PP_{for} PP_{with}	open, fix	We opened [$_{NP}$ *the door*] [$_{PP}$ *for Andy*] [$_{PP}$ *with a crowbar*].

The verbs in the first line of table 5.5 (*vanish, arrive*, and *die*) don't take a complement, those in the second line take an NP complement, and so on.

When a verb's complement options include an NP, as in the case of *devour, give, buy*, and so on, it is said to be **transitive**, and its NP complement is often referred to as its **direct object**. Verbs like *vanish, arrive*, and *dash* that don't have an NP complement are called **intransitive**.

A word can belong to more than one subcategory. The verb *eat*, for example, can occur either with or without an NP complement and therefore belongs to both of the first two subcategories in table 5.5.

(13) After getting home, they ate (a snack).

Of course, not all verbs exhibit this flexibility. As we have already seen, *devour*—although similar in meaning to *eat*—requires an NP complement and therefore belongs only to the second subcategory in our table.

As the examples in table 5.5 also show, some heads can take more than one complement. The verb *put* is a case in point, since it requires both an NP complement and a PP complement (or a locative adverb such as *there*).

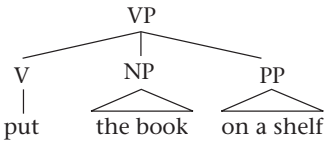
(14) *a. put* with an NP complement and a PP complement:
The librarian put [$_{NP}$ the book] [$_{PP}$ on the shelf].

b. put without an NP complement:
*The librarian put [$_{PP}$ on the shelf].

c. put without a PP complement:
*The librarian put [$_{NP}$ the book].

The VP *put the book on the shelf* has the structure in figure 5.13, in which the VP consists of the head *put* and two complements—the NP *the book* and the PP *on the shelf*.

FIGURE 5.13
A verb with two complements



5.2.2 Complement options for other categories

Various complement options are also available for Ns, As, and Ps. Tables 5.6, 5.7, and 5.8 provide examples of just some of the possibilities.

TABLE 5.6 Some examples of noun complements

Complement option	Sample heads	Example
Ø	car, boy, electricity	the car
PP_{of}	memory, failure, death	the memory [$_{PP}$ *of a friend*]
PP_{of} PP_{to}	presentation, description, donation	the presentation [$_{PP}$ *of a medal*] [$_{PP}$ *to the winner*]
PP_{with} PP_{about}	argument, discussion, conversation	an argument [$_{PP}$ *with Stella*] [$_{PP}$ *about politics*]

TABLE 5.7 Some examples of adjective complements

Complement option	Sample heads	Example
Ø	tall, green, smart	very tall
PP_{about}	curious, glad, angry	curious [$_{PP}$ *about China*]
PP_{to}	apparent, obvious	obvious [$_{PP}$ *to the student*]
PP_{of}	fond, full, sick	fond [$_{PP}$ *of chocolate*]

TABLE 5.8 Some examples of preposition complements

Complement option	Sample heads	Example
Ø	near, away, down	(he got) down
NP	in, on, by, near	in [$_{NP}$ *the house*]
PP	down, up, out	down [$_{PP}$ *into the cellar*]

Here again, subcategorization ensures that particular heads can appear in tree structures only if there is an appropriate type of complement. Thus, the adjective *sick* takes an *of*-PP as its complement, while the adjective *satisfied* takes a *with*-PP.

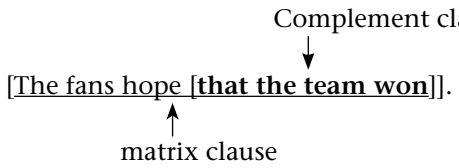
(15) *a.* sick [$_{PP}$ of cafeteria food] (compare: *sick with cafeteria food)
b. satisfied [$_{PP}$ with cafeteria food] (compare: *satisfied of cafeteria food)

A good deal of what we know about our language consists of information about words and the type of complements with which they can appear. Much of this information must be stored in the lexicon, since it cannot be predicted from a word's meaning.

5.2.3 Complement clauses

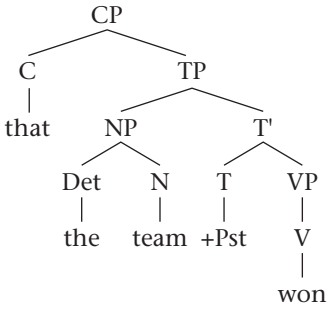
All human languages allow sentential phrases (or 'clauses', as they are often called) to function as complements. A simple example of this from English is given in (16).

(16)



The bold-faced bracketed phrase in (16) is called a **complement clause;** the larger underlined phrase in which it occurs is called the **matrix clause**. Words such as *that, whether*, and *if* are known as **complementizers** (Cs). Together with their TP complement, they form the CP (complementizer phrase) depicted in figure 5.14.

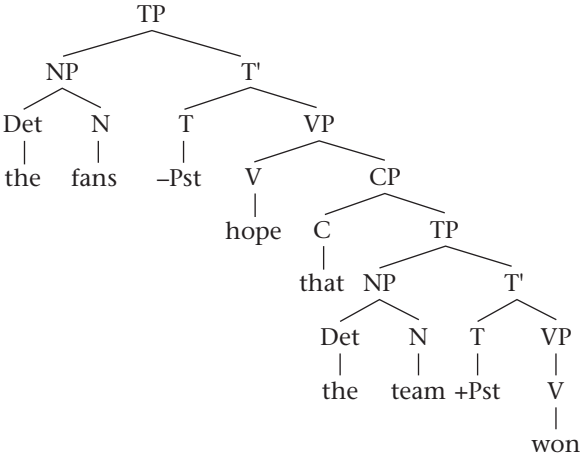
FIGURE 5.14
The internal structure of a CP



As we will see in section 5.3.2, there is even a type of element that can occur in the specifier position under CP.

When a CP occurs in a sentence such as (16), in which it serves as complement of the verb *hope*, the entire sentence has the structure in figure 5.15.

FIGURE 5.15
The structure of a sentence with an embedded CP



There is no limit on the number of embedded clauses that can occur in a sentence, as (17) shows.

(17) Harry said [$_{CP}$ that you know [$_{CP}$ that I think [$_{CP}$ that . . .

Table 5.9 provides examples of some verbs that are often found with a CP complement.

TABLE 5.9 Some verbs permitting CP complements

Complement(s)	Sample heads	Example
CP	believe, know, think, remember	They believe [$_{CP}$ *that Eric left*].
NP CP	persuade, tell, convince, promise	They told [$_{NP}$ *Mary*] [$_{CP}$ *that Eric had left*].
PP_{to} CP	concede, admit	They admitted [$_{PP}$ *to Mary*] [$_{CP}$ *that Eric had left*].

5.3 Move

As we have seen, it is possible to build a very large number of different sentences by allowing the Merge operation to combine words and phrases in accordance with the X' Schema and the subcategorization properties of individual words. Nonetheless, there are still many kinds of sentences that we cannot build. This section considers two such patterns and discusses the sentence-building operation needed to accommodate them.

5.3.1 *Yes-no* questions

The sentences in (18) are examples of ***yes-no* questions** (so called because the expected response is usually 'yes' or 'no').

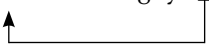
(18) *a.* **Should** those guys leave?
b. **Can** we meet at the library?

A defining feature of *yes-no* questions is that the auxiliary verb occurs at the beginning of the sentence rather than in its more usual position to the right of the subject, as illustrated in (19).

(19) *a.* Those guys **should** leave.
b. We **can** meet at the library.

How does the word order in (18) come about? The formation of question structures requires the use of an operation that we can call **Move**. Traditionally known as a **transformation** because it transforms an existing structure, Move transports the item in the T position to a new position to the left of the subject.

(20) Should those guys _ leave?



This analysis has at least two advantages. First, it allows us to avoid positing two types of modal auxiliary verbs in English: one that occurs between the subject and the VP and one that occurs to the left of the subject. Thanks to Move, all modal auxiliaries belong in the same place—in the T position, from which they can then be moved to the left of the subject in order to signal a question.

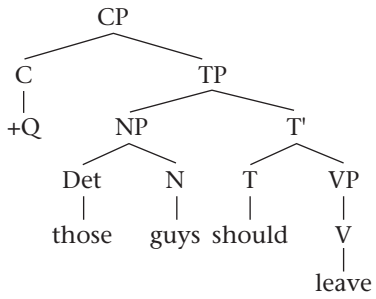
Second, the use of Move automatically captures the fact that the sentence *Should those guys leave?* is the question structure corresponding to *Those guys should leave*. According to the analysis presented here, both sentences initially have the same basic composition. They differ only in that the Move operation has applied to the T category in the question structure.

A landing site for T

In what position does the modal auxiliary 'land' when it is moved to the left of the subject? One promising idea assumes that TPs occur within a larger CP 'shell', in which the C position carries information about whether the sentence is a statement or a question. For the sake of

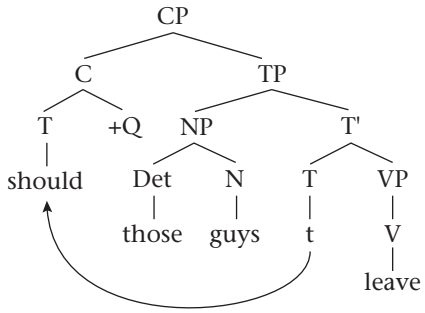
illustration, we use the symbol +Q to indicate a question; sentences with the feature −Q in their C position will be interpreted as statements.

FIGURE 5.16
A TP inside a CP shell, with the C carrying the +Q feature



In some languages, the Q feature is 'spelled out' as a separate morpheme (see the example from Yoruba in the box on page 156). In languages like English, where there is no such morpheme, the feature must attract another element to its position. The modal auxiliary in the T position is that element. As illustrated in figure 5.17, T is drawn to the C position, where it attaches right next to the +Q feature.

FIGURE 5.17
The T moves to the C position



A Move operation can do no more than change an element's position. It does not change the categories of any words and it cannot eliminate any part of the structure created by the Merge operation. Thus, *should* retains its T label even though it is moved into the C position (it changes its address, not its name). Moreover, the position that T formerly occupied remains in the tree structure. Called a **trace** and marked by the symbol *t*, it records the fact that the moved element comes from the head position within TP.

The Move operation used for *yes-no* questions is often informally called Inversion; it can be formulated as follows.

(21) *Inversion*
Move T to the C position.

Interesting evidence that T does in fact end up in the C position comes from patterns such as (22), which contain an embedded CP.

(22) I wonder [$_{CP}$ whether those guys should leave].

Here, the C position in the embedded clause is occupied by the complementizer *whether*.

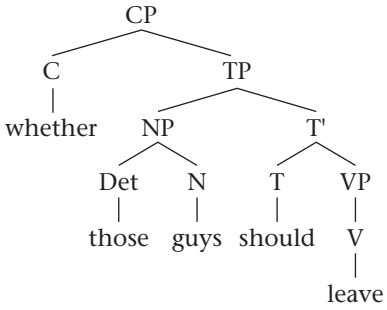


FIGURE 5.18
The C position in the embedded CP is filled by *whether*.

Assuming that no more than one word can occur in a head position, we predict that Inversion should not be able to apply in the embedded clause since there is nowhere for the moved auxiliary verb to land. The ungrammaticality of (23) shows that this is correct.

(23) Attempted inversion when there is a complementizer—the landing site is full:
*I wonder [$_{CP}$ **whether** those guys t leave].
**should*

Crucially, the acceptability of Inversion improves quite dramatically when there is no complementizer in the C position. (In fact, such sentences are perfectly acceptable in Appalachian English. For other English speakers, they may sound most natural when the embedded clause is interpreted as an indirect question.)

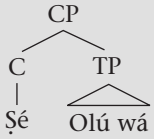
(24) Inversion in an embedded CP that does not have a complementizer:
I wonder [$_{CP}$ should those guys *t* leave].

To summarize before continuing, we have introduced two changes into our system of syntactic analysis. First, we assume that TPs occur inside CP shells even when there is no visible complementizer. Second, we assume that the Inversion transformation moves T (and its contents) to the C position in order to indicate a question. In addition to giving the correct word order for the question structure, this analysis helps explain why the result of applying Inversion sounds so unnatural when the C position is already filled by another element, as in (23).

Language Matters **Another Way to Ask a *Yes-No* Question**

Although Inversion is a widely used question-marking strategy around the world, many languages go about things in an entirely different way. Instead of moving something to the C position, they place a special question morpheme there to begin with. Yoruba (a Benue-Congo language spoken in Nigeria) works that way.

Yoruba
Ṣé Olú wá?
+Q Olu come
'Did Olu come?'



Source: Oluseye Adesola, *Yoruba: A Grammar Sketch*, Version 1.0, http://www.africananaphora.rutgers.edu/images/stories/downloads/casefiles/YorubaGS.pdf.

5.3.2 *Wh* questions

Consider now the question construction exemplified in (25). These sentences are called ***wh* questions** because of the presence of a question word beginning with *wh*.

(25) *a.* [$_{NP}$ Which languages] can Jerry speak?
b. [$_{NP}$ What] will they talk about?

TABLE 5.10 The syntactic category of ***wh*** words

***Wh* word**	**Syntactic category**	**Examples**
who	N	***Who** did you contact?*
what	N, when it occurs by itself	***What** did you see?*
	Det, when it occurs with a noun	***What** movie do you want to see?*
which	N, when in occurs by itself	***Which** do you prefer?*
	Det, when it occurs with a noun	***Which** car do you prefer?*
where	Adv	***Where** are you going?*
when	Adv	***When** did you move to Canada?*
why	Adv	***Why** did you leave the room?*
how	Adv, when it asks about a verb	***How** did they escape?*
	Deg, when it occurs with an adjective	***How** rich are they?*

There is reason to believe that the *wh* elements at the beginning of sentences such as those in (25) have been moved there from the positions indicated in (26).

(26) *a.* Jerry can speak [$_{NP}$ which languages]
b. They will talk about [$_{NP}$ what]

As illustrated here, *which languages* corresponds to the complement of *speak* (compare: *Jerry can speak two languages*) and *what* corresponds to the complement of *about* (compare: *They will talk about politics*).

How, then, do the *wh* phrases end up at the beginning of the sentence? The answer is that they are attracted there by the +Q feature, which triggers the application of a Move operation known as ***Wh* Movement**.

(27) *a.* [**Which languages**] can Jerry speak *t*?
Wh Movement
b. [**What**] will they talk about *t*?
Wh Movement

A landing site for *wh* words

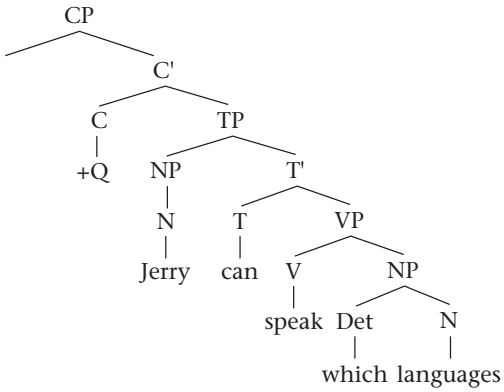
Because *wh* phrases end up to the left of the C position (filled in [27] by a moved modal), we can infer that they end up in the specifier of CP—the only available position in that region of the sentence. We can make this idea precise by formulating the *Wh* Movement operation as follows.

(28) Wh *Movement*
Move a *wh* phrase to the specifier position under CP.

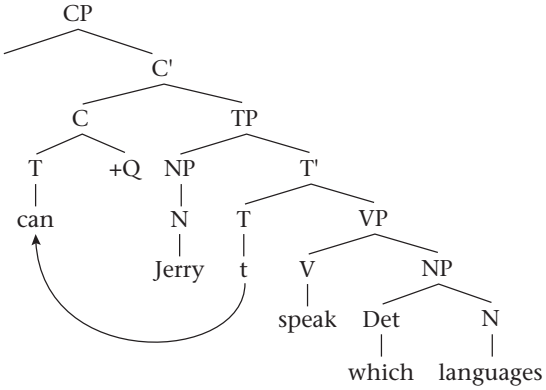
The sentence *Which languages can Jerry speak?* can now be analyzed in steps, the first of which involves formation of the structure in figure 5.19a, which includes an open specifier position under CP. *Wh* Movement and Inversion then apply, as depicted in figures 5.19b and 5.19c.

FIGURE 5.19
Steps for forming the sentence *Which languages can Jerry speak?*

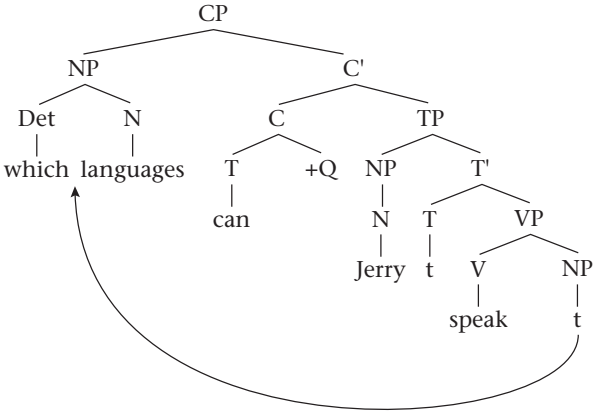
a. The structure produced by the Merge operation, with *which languages* functioning as complement of *speak*



b. Inversion: T moves to the C position



c. Wh Movement: the *wh* phrase moves to the specifier position in CP



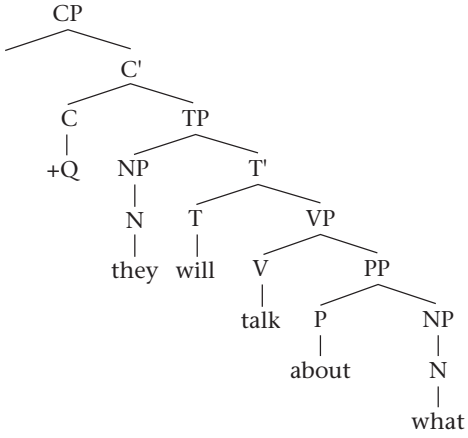
Like Inversion, *Wh* Movement cannot eliminate any part of the previously formed structure. The position initially occupied by the *wh* phrase is therefore not lost. That is because the

Move operation leaves behind an empty category (dubbed a **trace**) that marks the earlier position of the moved element. In the case at hand, the trace indicates that the NP *which languages* originates as the complement of the verb *speak*.

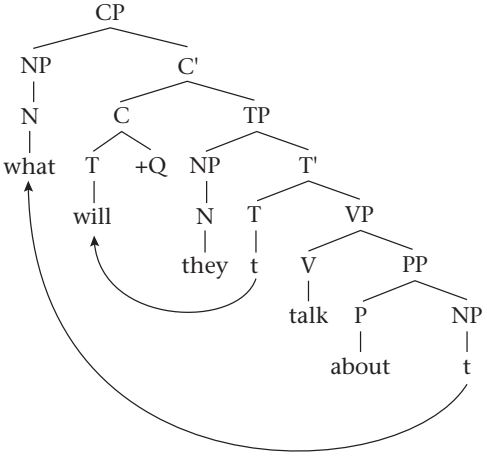
Here's a second example, involving sentence (27b).

FIGURE 5.20
Formation of the sentence *What will they talk about?*

a. The structure produced by the Merge operation



b. Inversion and *Wh* Movement (compressed here into a single step to save space)



Language Matters **Pied Piping**

In more formal varieties of English, there is a second possibility—the entire PP containing the *wh* word can undergo Wh Movement.

Movement of the PP *about what*:
[$_{PP}$ About what] will they *t* talk *t*?
Inversion
Wh Movement

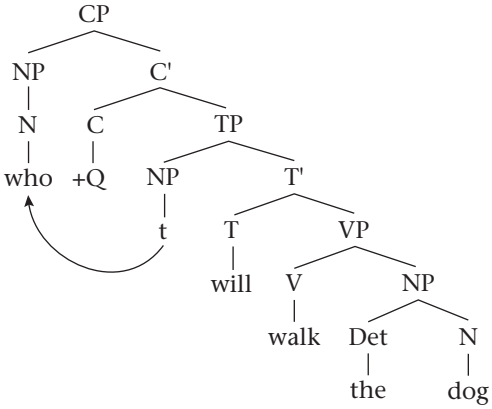
This phenomenon is known as 'pied-piping', a whimsical reference to the folk tale *The Pied Piper of Hamelin*, in which (in the words of Robert Browning) "the Piper advanced and the children followed."

In the examples considered so far, the *wh* word originates as the complement of a verb or preposition. In sentences such as the following, however, the *wh* word asks about the subject (the person who will walk the dog).

(29) Who will walk the dog?

The *wh* word in these patterns originates in the subject position. For the sake of generality, we assume that it subsequently moves to the specifier position in CP, even though the actual order of the words in the sentence does not change as a result of this movement (see figure 5.21). (We will assume that there is no Inversion in this type of question structure.)

FIGURE 5.21
Movement of a subject *wh* phrase



5.3.3 Deep structure and surface structure

The preceding examples show that two distinct types of mechanisms are involved in structure building. The first is the Merge operation, which creates tree structures by combining categories in a manner consistent with their subcategorization properties and the X' Schema. The second is the Move operation, which can modify these tree structures by moving an element from one position to another.

In the system sketched here, all instances of the Merge operation take place before any instances of the Move operation. This yields two distinct levels of syntactic structure, as shown in figure 5.22. The first, called **deep structure** (or **D-structure**), is formed by the Merge operation.

FIGURE 5.22
Structure-building operations