CHAPTER 8
Syntax

Time flies like an arrow; fruit flies like a banana.

Oettinger (1966)

In an early observation on the difficulties of getting computers to process natural language, Anthony Oettinger used the example above to illustrate how we tend to interpret sentences based on an expected structure and when we arrive at a problematic interpretation, we are able to go back and try to use a different structure. This process brings to light the importance of recognizing the underlying structure of sentences in order to make sense of them. If we keep thinking that the structure of the second expression is the same as the first in the example, we'll miss something.

In Chapter 7, we moved from the general categories of traditional grammar to more specific methods of describing the structure of phrases and sentences. When we concentrate on the structure and ordering of components within a sentence, we are studying the **syntax** of a language. The word "syntax" comes originally from Greek and literally means "a putting together" or "arrangement." In earlier approaches, there was an attempt to produce an accurate description of the sequence or ordering "arrangement" of elements in the linear structure of the sentence. In more recent attempts to analyze structure, there has been a greater focus on the underlying rule system that we use to produce or "generate" sentences.

Syntactic rules

When we set out to provide an analysis of the syntax of a language, we try to adhere to the "all and only" criterion. This means that our analysis must account for *all* the grammatically correct phrases and sentences and *only* those grammatically correct phrases and sentences in whatever language we are analyzing. In other words, if we write rules for the creation of well-formed structures, we have to check that those rules, when applied logically, won't also lead to ill-formed structures.

For example, we might say informally that, in English, we put a preposition (*near*) before a noun (*London*) to form a prepositional phrase (*near London*). This will describe a large number of phrases, but does it describe all (and only) the prepositional phrases in English? Note that, if we use this as a rule of the grammar to create structures involving a preposition and a noun, we will end up producing phrases like **near tree* or **with dog*. These don't seem to be well-formed English structures, so we mark them with an asterisk *, indicating that they are ungrammatical.

We clearly need to be more careful in forming the rule that underlies the structure of prepositional phrases in English. We might have more success with a rule stating that we put a preposition before a noun phrase (not just a noun). In Chapter 7, we saw that a noun phrase can consist of a proper noun (*London*), a pronoun (*you*) or the combination of an article (*a*, *the*) with a noun (*tree*, *dog*), so that the revised rule can be used to produce these well-formed structures: *near London*, *with you*, *near a tree*, *with the dog*.

A generative grammar

When we have an effective rule such as "a prepositional phrase in English consists of a preposition followed by a noun phrase," we can imagine an extremely large number of English phrases that could be produced using this rule. In fact, the potential number is unlimited. This reflects another goal of syntactic analysis, which is to have a small and finite (i.e. limited) set of rules that will be capable of producing a large and potentially infinite (i.e. unlimited) number of well-formed structures. This small and finite set of rules is sometimes described as a **generative grammar** because it can be used to "generate" or produce sentence structures and not just describe them.

This type of grammar should also be capable of revealing the basis of two other phenomena: first, how some superficially different phrases and sentences are closely related and, second, how some superficially similar phrases and sentences are in fact different.

Deep and surface structure

Our intuitions tell us that there must be some underlying similarity involving these two superficially different sentences: *Charlie broke the window* and *The window was broken by Charlie*. In traditional grammar, the first is called an active sentence, focusing on what *Charlie* did, and the second is a passive sentence, focusing on *The window* and what happened to it. The distinction between them is a difference in their **surface structure**, that is, the different syntactic forms they have as individual English sentences. However, this superficial difference in form disguises the fact that the two sentences are very closely related, even identical, at some less superficial level.

This other "underlying" level, where the basic components (Noun Phrase + Verb + Noun Phrase) shared by the two sentences can be represented, is called their **deep structure**. The deep structure is an abstract level of structural organization in which all the elements determining structural interpretation are represented. That same deep structure can be the source of many other surface structures such as *It was Charlie who broke the window* and *Was the window broken by Charlie?*. In short, the grammar must be capable of showing how a single underlying abstract representation can become different surface structures.

Structural ambiguity

Let's say we have two distinct deep structures. One expresses the idea that "Annie had an umbrella and she bumped into a man with it." The other expresses the idea that "Annie bumped into a man and the man happened to be carrying an umbrella." Now, these two different versions of events can actually be expressed in the same surface structure form: *Annie bumped into a man with an umbrella*. This sentence provides an example of **structural ambiguity**. It has two distinct underlying interpretations that have to be represented differently in deep structure. Note that this is not the type of ambiguity that we experience in hearing *Their child has grown another foot*, which illustrates lexical ambiguity mainly because the word *foot* has more than one meaning (see Chapter 9).

The comedian Groucho Marx knew how to have fun with structural ambiguity. In the film *Animal Crackers*, he first says *I once shot an elephant in my pajamas*, then follows it with *How he got into my pajamas I'll never know*. In the non-funny interpretation, part of the underlying structure of the first sentence could be something like: "I shot an elephant (while I was) in my pajamas." In the other (ho, ho) interpretation, part of the underlying structure would be something like: "I shot an elephant (which was) in my pajamas." There are two different underlying structures with the same surface structure.

Tree diagrams

One of the best ways to create a visual representation of underlying syntactic structure is through **tree diagrams**. We can use the symbols introduced in Chapter 7 (Art = article, N = noun, NP = noun phrase) to label parts of the tree when we create a representation of how each part fits into the underlying hierarchical structure of phrases and sentences. The information in a labeled and bracketed phrase, on the left, can be expressed in a tree diagram, on the right, as shown in Figure 8.1.

Although this kind of "tree," with its "branches," on the right, seems to grow down rather than up, it functions rather well as a diagram representing all the grammatical information found in the other analysis on the left. It also shows very explicitly that there are different levels in the analysis. That is, there is a level of analysis at which a constituent such as NP is represented and a different, lower, level at which a constituent such as N is represented.

We can use a similar tree diagram to represent the structure of an English verb phrase (VP), as shown in Figure 8.2.

Tree diagram of an English sentence

We can now put together the structure of a whole sentence, hierarchically organized, as shown below in Figure 8.3. We start at the top of the tree diagram with (S) and divide it into two constituents (NP and VP). In turn, the NP constituent is

Figure 8.1

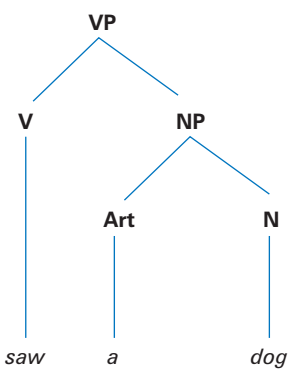

Figure 8.2

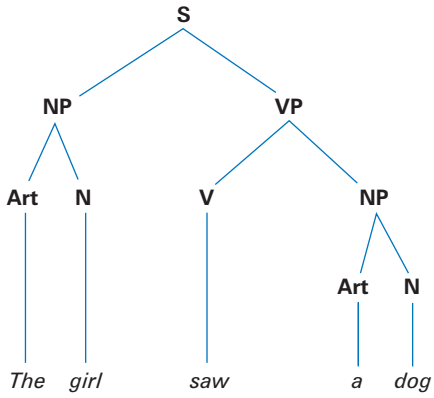

Figure 8.3

divided into two other constituents (Art and N). Finally, one word is selected that fits the label Art (*the*) and another that fits N (*girl*). You can go through the same procedure with the VP branches.

Symbols used in syntactic analysis

We have already encountered some symbols that are used as abbreviations for syntactic categories. Examples are "S" (= sentence), "NP" (= noun phrase), "N" (= noun), "Art" (= article), "V" (= verb) and "VP" (= verb phrase). Others, such as "PP" (= prepositional phrase), seem fairly transparent. There are three more symbols that are commonly used in syntactic description.

The first is in the form of an arrow →. It can be interpreted as "consists of" or "rewrites as." It is typically used in the following type of rule:

NP → Art N

This is simply a shorthand way of saying that a noun phrase (NP) such as *the dog* consists of or rewrites as (→) an article (Art) *the* and a noun (N) *dog*.

The second symbol is a pair of round brackets **( )**. Whatever occurs inside these round brackets will be treated as an optional constituent. For instance, we can describe something as *the dog* or *the small dog*, each of which is a noun phrase (NP). When we use a noun phrase in English, we can include an adjective (Adj) such as *small*, but we don't have to. It's an optional constituent in a grammatically well-formed noun phrase, as shown here:

NP → Art (Adj) N

This shorthand notation expresses the idea that a noun phrase (NP) rewrites as (→) an article (Art) and a noun (N), with the option of including an adjective (Adj) in a specific position between them. So, we can use this notation to generate *the dog*, *the small dog*, *a cat*, *a big cat*, *the book*, *a boring book* and an endless number of other similar noun phrases.

The third symbol is in the form of curly brackets { }.These indicate that only one of the elements enclosed within the curly brackets must be selected. For example, we have already seen that a noun phrase can consist of an expression such as *the dog* (article plus noun), or *it* (pronoun), or *Cathy* (proper noun). Using the abbreviations "Pro" (for pronoun) and "PN" (for proper noun), we can try to capture this observation about English with three separate rules, as shown on the left. However, it is more succinct to write one rule, as shown on the right, using curly brackets.

NP → Art (Adj) N
NP → Pro **NP → {Art (Adj) N, Pro, PN}**
NP → PN

It is important to remember that, although there are three constituents inside these curly brackets, only one of them can be selected on any occasion.

The list of common symbols and abbreviations is summarized here.

S sentence	**NP** noun phrase	**PN** proper noun
N noun	**VP** verb phrase	**Adv** adverb
V verb	**Adj** adjective	**Prep** preposition
Art article	**Pro** pronoun	**PP** prepositional phrase

* ungrammatical sentence
→ consists of / rewrites as
( ) optional constituent
{ }one and only one of these constituents must be selected

Phrase structure rules

When we use a tree diagram format, we can think of it in two different ways. In one way, we can simply treat it as a static representation of the structure of the sentence shown at the bottom of the diagram. We could then propose that, for every single sentence in English, a tree diagram of this type could be drawn. An alternative view is to treat the tree diagram as a dynamic format, in the sense that it represents a way of generating not only that one sentence, but also a very large number of other sentences with similar structures.

This second approach is very appealing because it would enable us to generate a very large number of sentences with what look like a very small number of rules. These rules are called **phrase structure rules**. As the name suggests, these rules state that the structure of a phrase of a specific type will consist of one or more constituents in a particular order. We can use phrase structure rules to present the information of the tree diagram in another format. That is, the information shown in

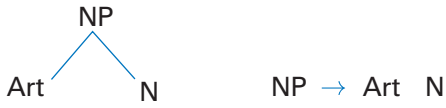

Figure 8.4

the tree diagram on the left in Figure 8.4 can be expressed in the phrase structure rule on the right.

According to this basic rule, "a noun phrase rewrites as an article followed by a noun." Using this format, we can create a more detailed set of rules.

The first rule in the following set of simple (and necessarily incomplete) phrase structure rules states that "a sentence rewrites as a noun phrase and a verb phrase." The second rule states that "a noun phrase rewrites as either an article plus an optional adjective plus a noun, or a pronoun, or a proper noun." The other rules follow a similar pattern.

S → NP VP

NP → {Art (Adj) N, Pro, PN}

VP → V NP (PP) (Adv)

PP → Prep NP

Lexical rules

Phrase structure rules generate structures. In order to turn those structures into recognizable English, we also need **lexical rules** that specify which words can be used when we rewrite constituents such as PN. The first rule in the following set states that "a proper noun rewrites as *Mary* or *George*." (It's a very small world.)

PN → {*Mary, George*}
N → {*girl, dog, boy*}
V → {*followed, helped, saw*}
Art → {*a, the*}
Pro → {*it, you*}

We can rely on these rules to generate the grammatical sentences shown below in (1)–(6), but not the ungrammatical sentences shown in (7)–(12).

(1) *A dog followed the boy.*
(2) *Mary helped George.*
(3) *George saw the dog.*
(4) *The boy helped you.*
(5) *It followed Mary.*
(6) *You saw it.*
(7) **Dog followed boy.*
(8) **The helped you boy.*
(9) **George Mary dog.*
(10) **Helped George the dog.*
(11) **You it saw.*
(12) **Mary George helped.*

As a way of visualizing how the phrase structure rules form the basis of these sentences, we can draw the tree diagrams for sentences (1) and (6), as in Figure 8.5.

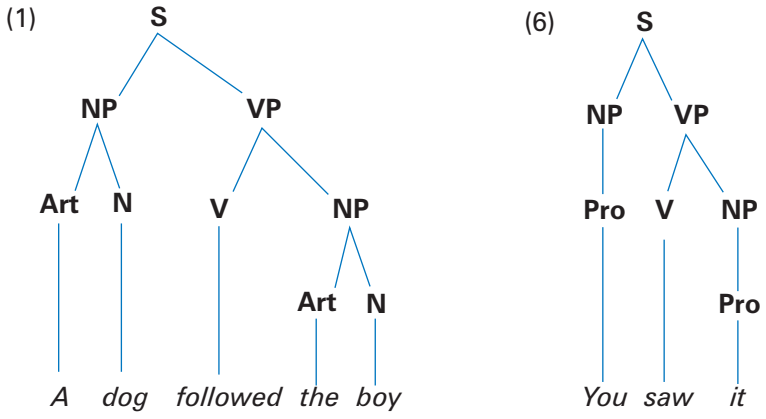

Figure 8.5

Movement rules

The very small set of phrase structure rules just described is a sample of what a more complex phrase structure grammar of English, with many more parts, would look like. These rules can be treated as a representation of the underlying or deep structures of sentences in English. One feature of these underlying structures is that they will generate sentences with a fixed word order. That is convenient for creating declarative forms (*You can see it*), but not for making interrogative forms, as used in questions (*Can you see it?*). In making the question, we move one part of the structure to a different position. This process is based on a **movement rule**.

In order to talk about this process, we need to expand our phrase structure rules to include an **auxiliary verb (Aux)** as part of the sentence. This is illustrated in the first rewrite rule below. Auxiliary verbs (sometimes described as "helping" verbs) take different forms in English, but one well-known set can be included in the rudimentary lexical rule for Aux below. The examples listed here for Aux, such as *can* and *will*, are called "modal verbs" and they are always used with the basic form of the main verb. The basic forms of some verbs are included in the third rewrite rule here.

S → NP Aux VP

Aux → {*can, could, should, will, would*}

V → {*follow, help, see*}

With these components, we can specify a simple movement rule that is involved in the creation of one basic type of question in English.

NP Aux VP ⇒ Aux NP VP

This rule states that if we have one structure of the type *You* (NP) + *can* (Aux) + *see it* (VP), then we can turn it into a different structure by moving the Aux component to the first position in the sequence in order to create *Can you see it?*. Similarly, if we start with *You will help Mary*, we can use the Aux-movement rule to produce *Will*

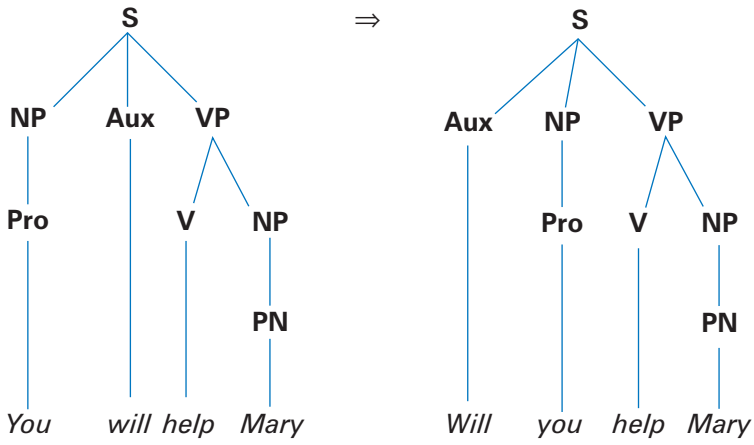

Figure 8.6

you help Mary?. Note that this type of rule has a special symbol ⇒ and can be illustrated in the process of one tree, on the right, being derived from the tree on the left, as in Figure 8.6.

Using this simple rule, we can also generate these other questions:

Can you see the dog? *Should Mary follow you?*
Could the boy see it? *Would George help Mary?*

These are all surface structure variations of a single underlying structure.

As we try to capture more aspects of the structure of complex English sentences, we inevitably need to identify more rules and concepts involved in the analysis of syntax. (We've barely scratched the surface structures.) However, having explored some of the basic issues, terminology, and methods of syntactic analysis in order to talk about structure in language, we need to move on to consider how we might incorporate the analysis of meaning in the study of language.