4° ATIN

07 sentence patterns

This chapter discusses the ways in which words can be linked together to form larger units. It explains how to analyze sentences into their 'constituents' (component parts), and shows ways of representing this type of analysis.

Words by themselves, or words strung together in a random way, are of relatively little use, a fact known by anyone who has visited a foreign country armed only with a dictionary, and no knowledge of the language. Does *me – bus* mean 'I am a bus', 'A bus hit me', 'I came by bus', or 'I want to go by bus'? So let us now look at how words may be combined together into longer utterances.

In this chapter, we shall consider, first, the ways in which words may be linked together to form larger units. Second, we shall discuss how to analyze sentences into their component parts, or **constituents** in linguistic terminology. Third, we shall suggest ways of representing this analysis.

Linking words together

Different languages use different devices for showing the relationship of one word to another. Most languages have one or two favourite devices. The following are especially common.

Word order

The device used most frequently in English is **word order**:

The large spider frightened Aunt Matilda.
Aunt Matilda frightened the large spider.

The words themselves in these two sentences are identical. It is the word order which indicates who frightened whom, and that it is the spider which is large, not Aunt Matilda. Languages which rely heavily on word order are known as **configurational** languages.

Inflections

In a language such as Latin, word endings or **inflections**, indicate the relationship between words. In the sentence:

Magna aranea perterruit Matildam amitam.
Large spider frightened Matilda aunt
'The large spider frightened Aunt Matilda'.

the word order is irrelevant. The sentence would still mean the same if the words were arranged quite differently as in:

Magna Matildam perterruit amitam aranea.
Large Matilda frightened aunt spider

The endings alone show that it was the spider which terrified Aunt Matilda, not the reverse, and that it is the spider, not Aunt Matilda, which is large. In linguistic terminology, Latin is a **non-configurational** language. Word order is not critical, though some word order preferences are found.

Function words

Another common device, used to some extent in both English and Latin, is the use of **function words**. These are words such as *of*, *by*, *that*, which indicate relationships between parts of the sentence:

Aunt Matilda was terrified by a spider.
The Queen of Sheba.
I know that Penelope will come.
Matilda amita ab aranea perterrita est.
Matilda aunt by spider frightened is[was]

There is some disagreement as to what counts as a function word in English. Part of the problem is that several English words, such as *to*, can be used both as a function word, and as a **content** word (one with intrinsic meaning):

Paul wants to go home. (function word)
Peter went to the river. (content word 'towards', 'as far as')

In addition, there are borderline cases, where *to* does not fit well into either type of usage:

Andrew's suit was made to order.
It seems to me a good idea.

Constituent analysis

Sentences are not simply random words strung together by means of various devices. We do not find English sentences such as:

**The large spider terrified Aunt Matilda swims of Sheba by a car.*

Instead, English (like every other language) has a limited number of recurring sentence patterns. A fundamental technique of syntactic analysis is to identify these patterns by a process of successive substitution. Take the sentence:

The duck bit the burglar.

In this sentence, *the* and *duck* can be replaced by a single word such as *Donald*, *it*, without altering the basic sentence pattern. This suggests that these two words are closely linked, and together constitute a single, larger component. Similarly, the words *the* and *burglar* go together, since they also could be replaced by a word such as *Albert*, *him*. So as a first stage, we have reduced a sentence with five original components down to three more basic ones (Figure 7.1):

figure 7.1

Of these three components, the final two could be replaced by a single word such as *slept*. We therefore conclude that they could be bracketed together as a single, larger component. We have therefore reduced a sentence with five components down to a basic two (Figure 7.2):

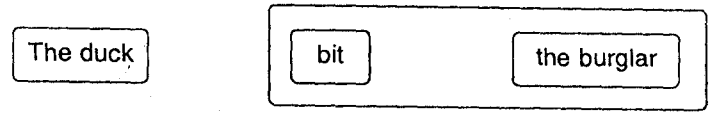


figure 7.2

The linguistic procedure which divides sentences into their component parts or constituents in this way is known as **constituent analysis**. The test of substitution is basic to such an analysis, though the process is not always as straightforward as the example above.

Tree diagrams

The successive layers of constituents which make up a sentence can be shown most clearly on a **tree diagram** – so called because its branches resemble the branches of a tree. In a tree diagram, a basic sentence type at the top branches downwards in ever-increasing complexity (Figure 7.3).

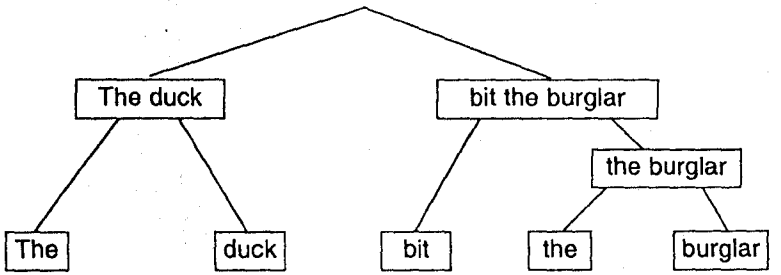


figure 7.3

The advantage of a tree diagram is that each join or node on the tree can be labelled, so that the whole construction becomes clearer (Figure 7.4).

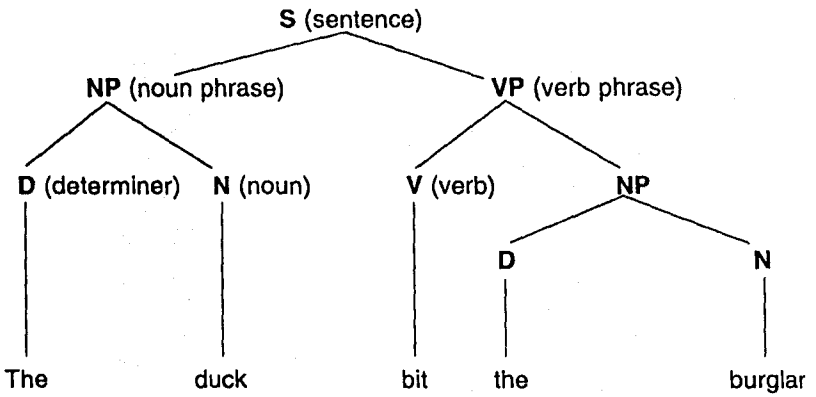


figure 7.4

A family metaphor is used to refer to the relationships on a tree (Figure 7.5). A higher node is a **mother**, and the nodes on the branches immediately beneath her are her **daughters**. Daughters of the same mother are known as **sisters**. A mother is said to **dominate** the nodes beneath her. She **immediately dominates** her daughters, but she also dominates her granddaughters, and great-granddaughters, as it were.

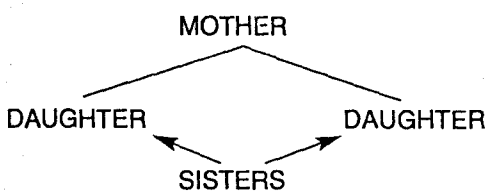


figure 7.5

Rewrite rules

An alternative way of expressing the information found on a tree diagram is by means of rewrite rules. A rewrite rule is a replacement rule, in which the symbol to the left of an arrow is replaced by an expanded form written to the right of the arrow.

S → NP VP

means 'Replace the symbol S by NP VP'.

VP → V NP

means 'Replace the symbol VP by V NP'.

NP → D N

means 'Replace the symbol NP by D N'.

The essential structure of *The duck bit the burglar* can therefore be summarized in just three rules:

S → NP VP
VP → V NP
NP → D N

On a tree diagram, these three rules would appear as in Figure 7.6.

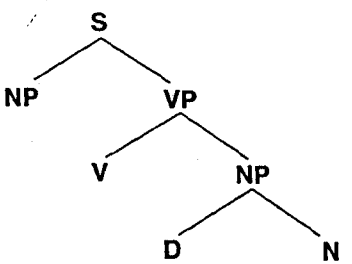


figure 7.6

These branching rules can then be supplemented by lexical substitution rules:

N → *duck, burglar*
V → *bit*
D → *the*

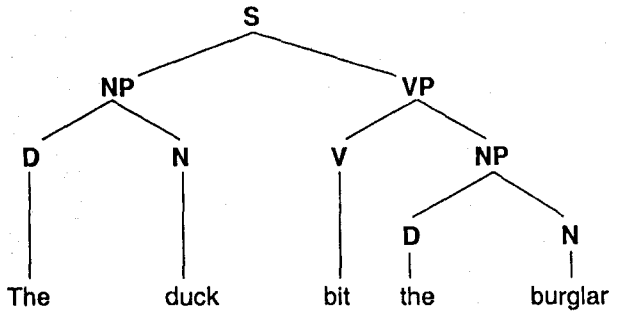


figure 7.7

The great advantage of rewrite rules is that they are perfectly **explicit**. They do not leave anything to the imagination. By following them, you could produce a perfect English sentence even if you did not know any English, since the rules are applied mechanically, step-by-step, one symbol at a time.

Note, however, that the above rewrite rules could also have resulted in the sentence:

The burglar bit the duck.

This does not matter, as the sequence is a perfectly good sentence of English (though admittedly a somewhat unlikely one). The rewrite rules are there to tell us what is a well-formed English sentence, not to give us information about the probable behaviour of burglars.

Identifying constituents

As we have seen, every sentence can be broken down into successive layers of constituents. However, not all sentences can be analyzed with as little trouble as *The duck bit the burglar.* Consider the sentence:

The mouse ran up the clock.

How should this be analyzed? Should we bracket [*ran up*] together, on the assumption that these words could be replaced by a word such as *climbed*? Or should we bracket [*up the clock*] together, noting that the whole phrase could be replaced by a single word such as *upwards*? Problems of this type are solved by seeing whether the groups of words in question belong together as a constituent elsewhere, since words that are grouped together in one sentence are likely to recur as a single constituent in other sentences. One way of checking this is to construct sentences in which the original words occur in a different order:

Up the clock ran the mouse.
**The mouse ran the clock up.*

These sentences suggest that the words *up the clock* should be bracketed together, since they can be moved as a chunk to the front of the sentence. We may therefore analyze the sentence as:

[*The mouse*] [*ran*] [*up the clock.*]

and draw the tree diagram as in Figure 7.8.

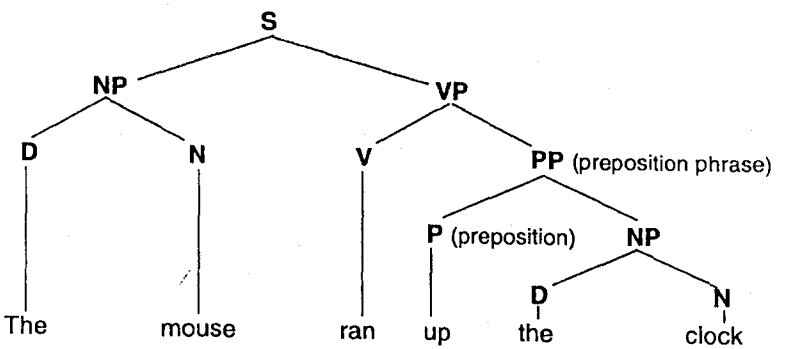


figure 7.8

The sentence discussed above must be analyzed differently from another, superficially similar sentence:

The mouse ate up the cheese.

We can show the difference by switching the sentence around:

**Up the cheese ate the mouse.*
(Compare: *Up the clock ran the mouse.*)
The mouse ate the cheese up.
(Compare: **The mouse ran the clock up.*)

We may therefore analyze the second sentence as:

[*The mouse*] [*ate up*] [*the cheese.*]

and draw the tree diagram as in Figure 7.9, using the extra node-labels VB for 'phrasal verb' and PRT for 'particle':

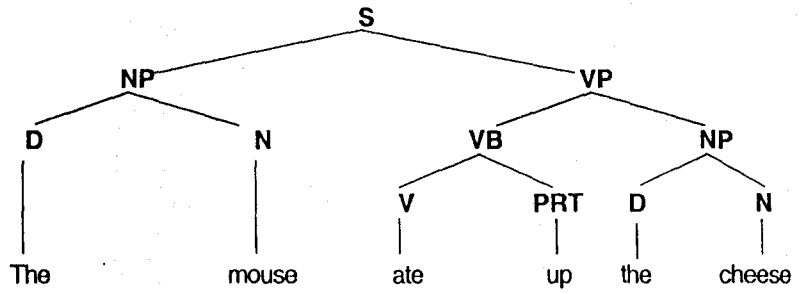


figure 7.9

Constituents behave in predictable ways, since languages ring the changes on a few recurring patterns. It is therefore possible to build up a store of specific 'tests' for the presence of a particular constituent in a given language. As *up the clock* suggests, one test for a PP (preposition phrase = phrase containing a preposition) is that a preposition cannot immediately follow its NP. Just as you cannot say:

**The mouse ran the clock up.*

so you cannot say:

**Fenella went the woods into.*
**Doris swam the bridge under.*

Let us now go on to consider this notion of 'tests' further, by considering 'NP tests'.

NP tests

English NPs (noun phrases) recur in certain specifiable positions. Some of the main places in which they occur are:

- At the beginning of a sentence before the verb:
  The cat ate the canary.
- At the end of a sentence after the verb:
  The canary feared the cat.
- After *by* in a passive sentence:
  The canary was eaten by the cat.
- After an auxiliary verb in questions:
  Did the cat eat the canary?

Of course, other types of phrase can occur in some of these positions. But an NP such as *the cat* can occur in *all* of them. Consequently, if we find a phrase which we suspect might be an NP, we can apply these (and other) tests. For example, consider the sentences below:

Uncle Harry kicked the cat.
Suddenly Harry kicked the cat.

In order to find out whether the first two words in each sentence are an NP, we can apply the NP tests listed below:

- At the beginning of a sentence before a verb:
  Uncle Harry kicked the cat.
  Suddenly Harry kicked the cat.
- At the end of a sentence after a verb:
  The cat scratched Uncle Harry.
  **The cat scratched suddenly Harry.*
- After *by* in a passive sentence:
  The cat was kicked by Uncle Harry.
  **The cat was kicked by suddenly Harry.*
- After an auxiliary verb in questions:
  Did Uncle Harry kick the cat?
  **Did suddenly Harry kick the cat?*

The failure of *suddenly Harry* to pass most of these NP tests shows that it cannot be an NP, whereas the success of *Uncle Harry* indicates that it probably is an NP.

Adding in extra patterns

So far, our rewrite rules have dealt with only one structure, the pattern underlying *The duck bit the burglar.* Let us now add in some others. Consider the sentence:

The duck slept in the bath.

This has the same basic division into NP VP as *The duck bit the burglar.* But the structure of the VP differs. In *The duck slept in the bath,* the verb is followed by a preposition phrase (PP) (Figure 7.10).

The extra rewrite rules required for this are:

VP → V PP
PP → P NP

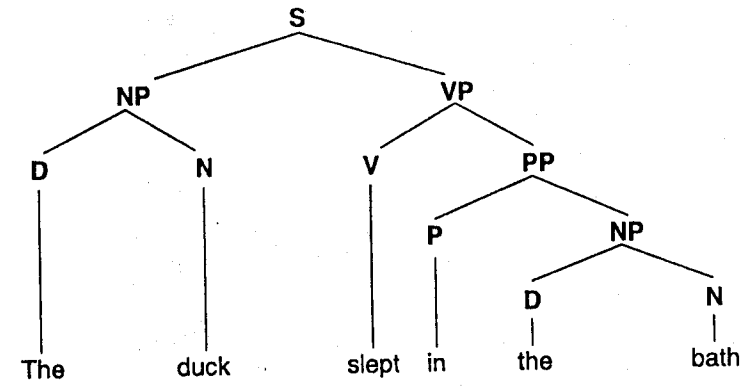


figure 7.10

However, the PP is not an essential part of the structure. It is an optional extra, since *The duck slept* is a well-formed sentence by itself. This can be shown by putting brackets round the PP in the rewrite rule, indicating that it is optional:

VP → V (PP)

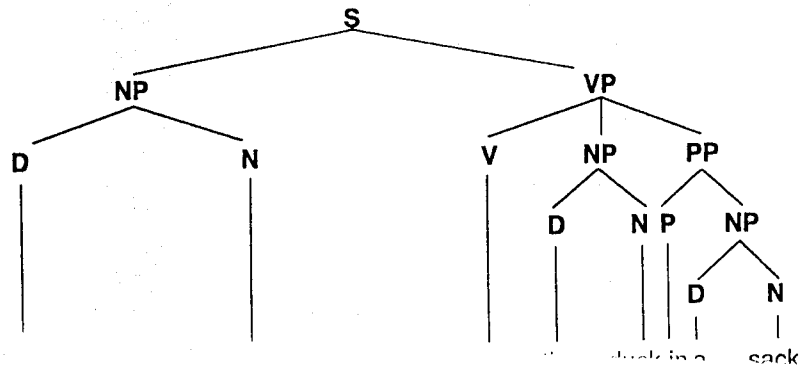
The rewrite rule above therefore underlies both *The duck slept in the bath*, and *The duck slept*. In the first, the optional PP has been selected. In the second, it has been omitted.

Let us now consider another sentence:

The burglar put the duck in a sack.

This differs from the previous structures discussed in that it is essential to have both an NP and a PP after the V (Figure 7.11). If either were omitted, the sentence would be ill-formed:

**The burglar put the duck.*
**The burglar put in a sack.*



The rewrite rule in this case is:

VP → V NP PP

So far, then, we have three different rewrite rules for English VPs:

VP → V NP *The duck bit the burglar.*
VP → V (PP) *The duck slept. The duck slept in the bath.*
VP → V NP PP *The burglar put the duck in a sack.*

It would be useful to combine these three separate rules. As a first suggestion, one might simply number the types of verb (V1 for a verb such as *bit*, V2 for *slept*, V3 for *put*), and enclose them in another type of bracket { } which is used to denote alternative possibilities:

$$VP \rightarrow \left\{ \begin{array}{lll} V1 & NP & \\ V2 & (PP) & \\ V3 & NP & PP \end{array} \right\}$$

This means: 'Rewrite the VP as either V1 NP, or V2 (PP), or V3 NP PP'. However, if we wanted to include the full range of alternatives available in an English VP, the rewrite rules would become extremely long and complicated. A neater solution is to keep the rewrite rules fairly simple, and to use them in conjunction with a lexicon (dictionary) which specifies the structure associated with each V:

bit V [—NP]
slept V [—(PP)]
put V [—NP PP]

First, the item in question is listed, then the fact that it is a verb (V). In the square brackets come the structures associated with it. The long dash [—] indicates the place where the verb is inserted, so [—NP] says 'The verb in question must be followed by an NP'.

With these lexical entries, we need only one rewrite rule for the three types of verb:

VP → V (NP) (PP)

This rule says: 'A VP consists of a V optionally followed by an NP and/or a PP'. It accounts for all the possibilities discussed above, since one can slot in a verb only if it fits the structure chosen. For example, suppose we had chosen both the optional items, NP and PP, we must then slot in a verb followed by NP PP, in this case *put*. Similarly, if we had chosen V alone, the only

V which fits in this case is *sleep*.

With a detailed lexicon of this type, which can be expanded to include other word classes also, we no longer need substitution rules such as: V → *bit*, N → *burglar*.

Let us therefore summarize the rewrite rules and lexical entries for:

The duck bit the burglar.
The duck slept.
The duck slept in the bath.
The burglar put the duck in a sack.

A Rewrite rules

S → NP VP
VP → V (NP) (PP)
NP → D N
PP → P NP

B Lexicon

burglar	N	
duck	N	
sack	N	
bath	N	
bit	V	[—NP]
slept	V	[—(PP)]
put	V	[—NP PP]
the	D	
a	D	
in	P	

Of course, if more data had been considered, the rules and the lexicon would have to be complicated further. For example, if we had included a proper name such as *Donald*, the lexicon would have to specify which nouns are found with a determiner (D), as in *the duck*, *a sack*, and which not, as in *Donald* not **a Donald*. However, we set out to write rules for the sentence patterns in question, and we have done this as economically as possible.

Layers of branches

The tree diagrams we have considered so far have relatively few layers. But consider a sentence such as:

Maurice took a photograph of a platypus.

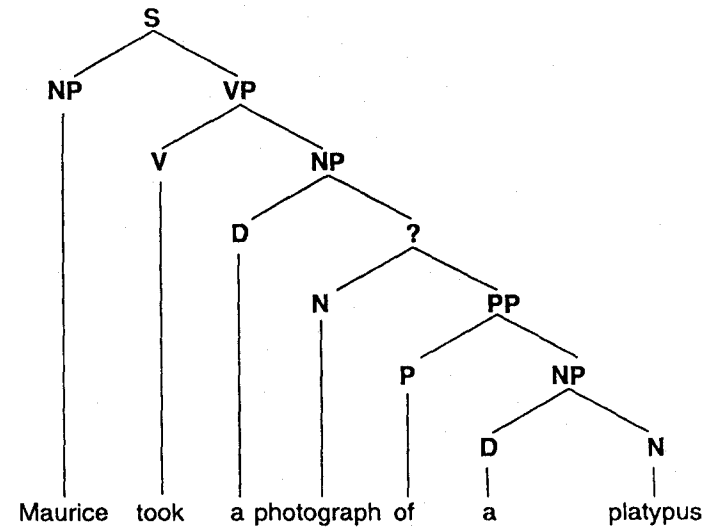


figure 7.12

The sequence *a photograph of a platypus* is clearly a noun phrase (NP) and the words *photograph* and *platypus* are nouns (N). But what of the intervening node, comprising *photograph of a platypus*? It seems to be something between an N and a full NP, so what is it? A useful solution is to give the label $\bar{N}$ (pronounced N-bar, since it has a bar along the top) to something that is neither a simple N, nor a whole NP. Some people also give the label $\bar{\bar{N}}$ (N-double bar) to a whole NP. (In Figure 7.13, a triangle has been drawn in place of the details of the PP. This is a standard procedure which avoids wasting time and space when the details are irrelevant to the point under discussion.)

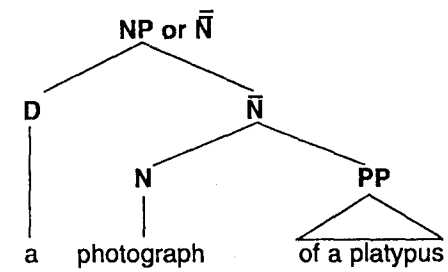


figure 7.13

The use of bars has one further major advantage: they can be used with adjectives (A), verbs (V), and prepositions (P), as well as with nouns (N). It is then easy to see similarities in structure between NPs, APs (adjective phrases), VPs and PPs which were not so evident before. It turns out that the **head** (main word) in one type of phrase is in a very similar position to the head in another. In other words, a noun in an NP is likely to be in a parallel location to an adjective in an AP, a verb in a VP and a preposition in a PP. For example, the AP *very proud of the platypus* has a structure that is similar in its branching pattern to the NP *a photograph of the platypus*. (In Figure 7.14, DEG stands for 'degree'.)

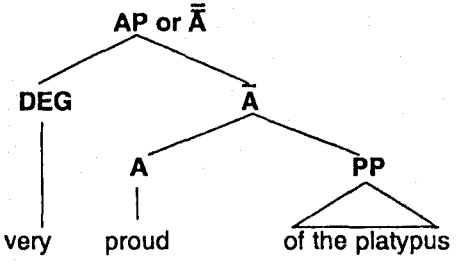


figure 7.14

A number of details still have to be worked out concerning **X-bar syntax** or **X-bar theory**, the name of this method of dealing with sentence patterns. For example, there is some controversy as to how many layers of bars it is useful to set up. But the theory appears to be here to stay, and it plays an important role in Chomsky's work.

Yet another way of handling layers has emerged in recent years, partly combined with X-bar syntax. **Functional phrases**, that is, phrases introduced by function words (Chapter 6) have a structure similar to lexical phrases, it has been claimed. For example, inflections (verb attachments) and the accompanying verb can be labelled an **inflectional phrase** (IP). So *to fish*, *will fish*, and *fished* are all IPs. In English, the inflection, I, sometimes known as INFL, mostly comes before the verb, as with *to*, *will*, though sometimes after it, as with *fished* (Figure 7.15).

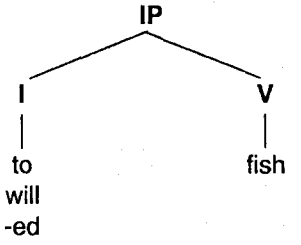


figure 7.15

Complex sentences

So far, we have assumed that all sentences are simple ones such as:

The duck bit the burglar.
The mouse ran up the clock.

In practice, however, many sentences have one or more sentence-like structures attached to them or inserted inside them. Consider:

Archibald played tennis, and Peter went fishing.

Here we have two sub-sentences of equal importance attached together to form a single one. This process is known as **conjoining**. In theory an indefinite number of sentences could be joined together:

Archibald played tennis, and Peter went fishing, and Pip played cricket, and Mary washed her hair, and Drusilla climbed the Eiffel tower...

However, conjoining is not the only process by which sentence-like structures are linked together. More often subsidiary sentences are inserted into one main sentence. This is known as **embedding** (Figure 7.16):

The rumour that the dinosaur had escaped worried the public.

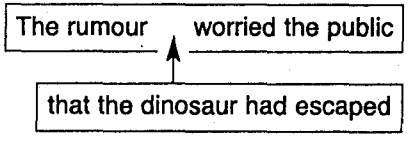


figure 7.16

In theory, a sentence can have an indefinite number of sentences embedded in it. In *The fact that the rumour that the dinosaur had escaped worried the public is not surprising*, the simple sentence has two others embedded in it (Figure 7.17).

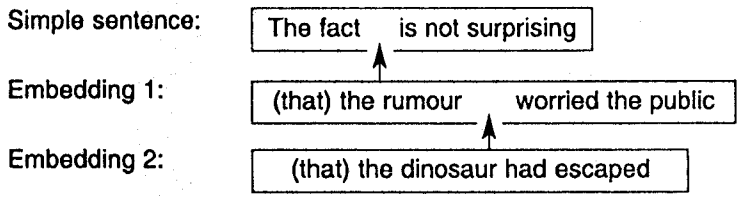


figure 7.17

Another example of embedding is the old nursery rhyme (Figure 7.18).

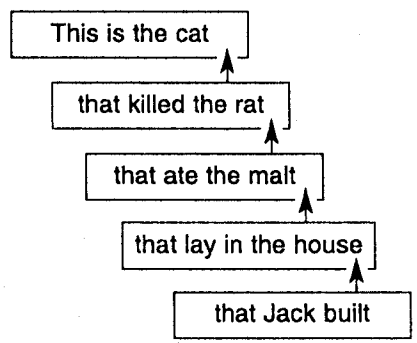


figure 7.18

Both embedding and conjoining illustrate an important property of language – that of recursion.

Recursion is the possibility of repeatedly re-using the same construction, so that there is no fixed limit to the length of sentences. This has important implications. It means that we can never make a complete list of all the possible sentences of any language. Instead, we must work out the system of rules which underlie the sentences.

It is quite easy to incorporate recursion into the rewrite rules, if one allows a symbol such as VP to be rewritten to include an S:

$$VP \rightarrow V \quad S$$

This rule (which would need to be combined with the other VP rule discussed earlier) allows one to generate a sentence such as:

Mavis believes the burglar took the duck. (Figure 7.19).

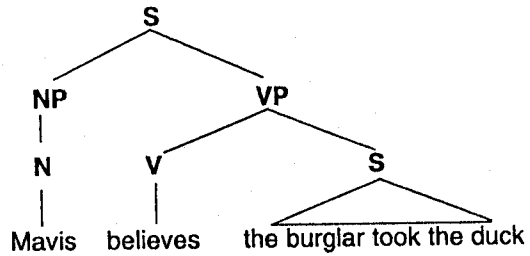


figure 7.19

So far, this chapter has shown how linguists analyze sentence patterns, with particular attention to configurational languages (those which rely on word order). There are extra problems involved in the investigation of non-configurational languages, but the notion of checking whether one constituent (component part of sentence) can be substituted for another is basic to all syntactic analyses.

Verbs: the syntax–meaning overlap

Verbs straddle the gap between syntax and semantics: the structure surrounding them provides clues to their meaning. Take the nonsense sentence:

The wickwock jipped.

Here, the wickwock has clearly done something alone, maybe jumped, or hiccuped. Or consider:

The wickwock grunched the mobe.

In this case, the wickwock has done something or other to something else.

The nouns accompanying verbs display different semantic roles, or, in more recent terminology, **thematic relations** – from the word *theme* – the label sometimes given to a noun involved in an action, though not initiating it, as in:

The snowball (theme) *rolled down the hill.*

But how many different roles are there? Some are obvious. An **agent** initiates action, and a **patient** receives it, as in:

The dog (agent) *chewed a bone* (patient).

A **recipient** receives something, as in:

Paul sent a letter to Patsy.
Patsy received a letter from Paul.

But problems arise. Consider:

Veronica leapt into the water.

Is *the water* a **recipient**? Or is it a **goal** Veronica is aiming at? And supposing Veronica had *fallen* into the water, what then? This example shows the difficulty of deciding how many roles there are, and which one is which. The overall aim is to specify a set of relations which can be used to describe any human language, and discussions continue.

This chapter has looked at syntactic patterns, and also drawn attention to the overlap between syntax and meaning. The next chapter will discuss how linguists handle meaning.

Questions

1 Suggest three ways in which languages show the relationship of one word to another.
2 What is a **tree diagram**, and why is it useful?
3 What are **rewrite rules**?
4 Draw a tree diagram for each of the following sentences:
The bus ran down the dog.
The boy ran down the street.
5 Distinguish between **conjoining** and **embedding**. Give examples.
6 What are **thematic relations**?