This chapter has looked at how linguists handle sound structure. This is the first aspect which confronts anyone working on an unwritten language. In practice, sorting out the sound structure overlaps with the analysis of larger units, such as words. This is the topic of the next chapter.

Questions

1 What do you understand by the term **phoneme**?
2 What is a **minimal pair**? Suggest at least ten examples of minimal pairs in English.
3 What is an **allophone**? Give examples.
4 What are **distinctive features**?
5 What is a **natural class**?
6 What is **metrical phonology**?

06

words and pieces of words

This chapter looks at the problems encountered in identifying and defining the notion 'word'. It then discusses the identification and description of 'morphemes' (pieces of words). Finally, it looks at the way in which words can be assigned to 'word classes' (parts of speech).

The word appears to be a widespread concept. Even in primitive cultures, informants are often able to identify words. This is somewhat surprising, because nobody has yet proposed a satisfactory universal definition of the notion 'word', or provided a foolproof method of identification. People sometimes wrongly assume that a word is recognizable because it represents a 'single piece of meaning'. But it can easily be shown that this view is wrong by looking at the lack of correspondence between words from different languages. In English, the three words *cycle repair outfit* correspond to one in German, *Fahrradreparaturwerkzeuge*. Or the six words *He used to live in Rome* are translated by two in Latin, *Romae habitabat*. And even in English, a word such as *walked* includes at least two pieces of meaning, 'walk' and 'past tense'.

This chapter will deal with this matter. First, it will look at the problems of defining and identifying words. Second, it will consider pieces of words, or morphemes.

Defining words

The best-known definition of a word is that proposed by the American linguist Bloomfield, who defined it as a **minimum free form**, that is, the smallest form that can occur by itself. This is fairly unsatisfactory, because words do not normally occur by themselves in spoken speech. Even if you ask a simple question, a normal-sounding reply often requires more than one word:

Who did that?   John did.
What's that?   An oak tree.

Furthermore, some apparent words, such as *did*, *the*, *and*, are found alone only in exceptional circumstances, such as in answer to the question: 'What does *a-n-d* spell?'

Bloomfield's definition works best for written English, where we conventionally leave a space on either side. But linguists are concerned primarily with the spoken word, not the written, and the two do not necessarily coincide. For example, it seems to be purely accidental that the name of a certain type of snake, a *boa constrictor*, is written as two words rather than one, or that *seaside* appears as one word, but *sea shore* as two.

Why have linguists found it so hard to find a satisfactory definition of the notion 'word'? The answer seems to be that there are different types of word. Consider the rhyme:

A flea and a fly in a flue
Were imprisoned, so what could they do?
Said the flea: 'Let us fly'.
Said the fly: 'Let us flee'.
So they flew through a flaw in the flue.

At the simplest level, this rhyme contains thirty-six written words. But some of these are repeated. If we decide to leave out repeats and count the number of different words (in technical terms, count **word types** instead of **word tokens**), we come up against several problems. Should *fly* (noun) and *fly* (verb) be counted as the same, since they sound the same, or as different, because they have different meanings? Should *fly* and *flew* be regarded as the same, because they belong to the same verb, or as different because they have different forms? These problems can be solved only if we decide what kind of 'word' we are talking about. It is important to distinguish between **lexical items**, **syntactic words** and **phonological words**.

If by 'word' we mean **lexical item** (the technical term for 'dictionary entry'), then the sound sequence /flaɪ/ 'fly' represents two words, since most dictionaries have separate entries for *fly* (noun, N) and *fly* (verb, V):

fly N: an insect with two wings.
fly V: to move through the air in a controlled manner.

This is perhaps the most basic, and most abstract use of the word 'word'. However, both of these lexical items have various syntactic forms associated with them. The insect could occur as *fly* (singular) or *flies* (plural), and the verb could occur as *fly*, *flying*, *flies*, *flew*, *flown*. So if we counted the various syntactic forms as different words, the overall total would be much higher (Figure 6.1).

Lexical items	Syntactic words
fly N	fly flies
fly V	fly flying flies flew flown

figure 6.1

A further complication occurs with a lexical item such as *flaw*. This has the two syntactic forms *flaw* (singular) and *flaws* (plural). But the singular form *flaw* then has two different sound sequences associated with it, /flɔː/ before a consonant, and /flɔːr/ before a vowel (Figure 6.2):

The flue had a flaw /flɔː/ which allowed the fly to escape.
There was a flaw /flɔːr/ in the flue.

Lexical item	Syntactic words	Phonological words
flaw N	flaw	/flɔː/
		/flɔːr/
	flaws	/flɔːz/

figure 6.2

These examples show that we must not expect an exact overlap between different types of word. And in some other languages, the situation is far more complex than in English. In Latin, for example, the lexical item *rosa* 'rose' has twelve different syntactic forms. In Welsh, the initial consonant of each word varies systematically, depending mainly on the preceding sound: the word for 'father' could be *tad, dad, thad,* or *nhad.* The last lines of the chorus in a well-known Welsh hymn have three different forms of the verb meaning 'sing': *canu, ganu* and *chanu* – and there is a fourth possibility, *nghanu*, which the hymn omits.

Identifying words

For anyone working on an unknown language, it is important to identify these various types of word. There are two main stages in the analysis. First, finding chunks such as *fly, flew,* which recur as self-contained units. Second, deciding how many lexical items are covered by each chunk (as with *fly*, which covers two lexical items), and conversely, deciding how many different chunks belong to the same lexical item (as with *fly, flew,* where different syntactic forms belong to one lexical item).

For the first stage, finding chunks which behave as self-contained units, we look for sequences which are uninter-

ruptible and mobile. These are useful guidelines in many languages. A sequence such as *chickens* cannot be interrupted. It is impossible to say **chick-little-ens*, or **chicken-little-s*. In addition, the sequence *chickens* can move about. It can occur next to different words, and in different parts of the sentence, as in: *Chickens lay eggs, foxes eat chickens, the chickens clucked loudly*, and so on.

To take another example, suppose we had come across the sequence *greentrousers*, and wanted to know whether this was one or more words. We would begin by looking for sentences which included any part of the sequence *greentrousers*. We might find:

Green leather trousers, Red trousers, Green shirts.

The fact that *greentrousers* can be interrupted by the word *leather* indicates that we are probably dealing with at least two words, *green* and *trousers*. This suspicion is confirmed by noting that both *green* and *trousers* occur with other words. But since *green* and *trousers* seem to be uninterruptible (we do not find **trous-green-ers*, for example), we surmise that each is a word.

At the end of this stage of the analysis, we have a rough list of 'words', though a list in which we are likely to have clumped together different lexical items which sound the same (homonyms), and to have separated different syntactic forms of the same lexical item.

For the second stage of the analysis, we need to consider the syntactic behaviour of these possible 'words', that is, their role in the overall sentence pattern. For example, *fly* N would show up as behaving differently from *fly* V, since each would fit into a different 'slot' in the sentences:

The fly buzzed.
Birds fly.

On the other hand, *fly* and *flew* would turn out to be somewhat similar, in that they would fit into the same general slot:

They fly home on Sunday.

They flew home on Sunday.

However, the syntactic behaviour of these different forms can be supplemented by an analysis of their make-up, or, in other words, the morphemes out of which they are constituted. Let us therefore go on to consider some basic facets of morphology.

Morphemes

The smallest syntactic unit is the morpheme. Morphemes vary in size. Neither syllables nor length are any guide to their identification. The essential criterion is that a morpheme cannot be cut up into smaller syntactic segments.

The sentence in Figure 6.3 has eleven morphemes:

The	sleep	walk	ing	albatross	chant	ed	a	dream	y	lullaby
1	2	3	4	5	6	7	8	9	10	11

figure 6.3

The, albatross, a, lullaby, are all single morphemes because none of them can be syntactically split up further. *Alba-* and *-tross,* for example, do not have any other role to play in the syntax of English: they exist only as part of the single unit, *albatross*. *Chanted* and *dreamy,* on the other hand, each consist of two morphemes: *chant* is found in words such as *chanting, chants,* and is also a word by itself, while *-ed* is found in *wanted, batted* and so on. Similarly, *sleep-walking* consists of three morphemes, because *sleep, walk* and *-ing* are all found elsewhere. In theory there is no upper limit to the number of morphemes per word: *antidisestablishmentarianism,* for example, has at least six: *anti-dis-establish-ment-arian-ism.*

Recognition of morphemes

Linguists identify morphemes by comparing a wide variety of utterances. They look for utterances which are partially the same (Figure 6.4):

The	dinosaur	sniff-ed	arrogant-ly	and	plodd-ed	for-wards
The	dinosaur	grunt-ed	loud-ly	and	edg-ed	back-wards

figure 6.4

The partial similarity between *sniffed, grunted, plodded* and *edged* enables us to isolate the segment *-ed*. And the partial similarity between *arrogantly* and *loudly,* and between *backwards* and *forwards* makes it possible to isolate *-ly* and *-wards.*

In Turkish, the similarity between *adamlar,* 'men', and *kadınlar,* 'women', enables one to identify a plural suffix *-lar,* and the words for 'man', *adam,* and 'woman', *kadın.* In Swahili, the overlap between:

nitasoma	I will read
nilisoma	I read (*past*)
utasoma	you will read
ulisoma	you read (*past*)

allows us to identify *soma,* 'read'; *ni,* 'I'; *u,* 'you'; *ta,* future tense; *li,* past tense.

Not all morphemes are as easily segmentable as these examples. But the identification of morphemes is done wholly by means of this one basic technique – the comparison of partially similar utterances.

Types of morpheme

Morphemes such as *albatross, chant, lullaby,* which can occur by themselves as whole words are known as free morphemes. Those such as *anti-, -ed, -ly,* which must be attached to another, are bound morphemes. Bound morphemes are of two main types. Consider the sentence:

The owl look-ed up at the cloud-y sky.

Superficially, both *looked* and *cloudy* have a similar make-up, consisting of one free morpheme, followed by a bound one. Yet the bound morphemes differ in nature. *-ed* on the end of *looked* is an inflectional morpheme, since it provides further information about an existing lexical item *look,* in this case indicating that the looking occurred in the past. Other examples of inflectional morphemes are the plural, as in *owls,* and the possessive, as in *Peter's car.* However, *-y* on the end of *cloudy* behaves rather differently. It is a derivational morpheme, one which creates an entirely new word. *Cloud* and *cloudy* behave quite differently and fit into different slots in the sentence. Other examples of derivational morphemes are *-ness* as in *happiness, -ish* as in *greenish,* and *-ment* as in *establishment.*

In most cases, it is easy to tell the difference between inflection and derivation. Above all, inflectional endings do not alter the syntactic behaviour of an item in any major way. The word still fits into the same 'slot' in the sentence. Derivational endings create entirely new words. In addition, inflectional endings can

be added on to derivational ones, but not vice-versa. That is, we find words such as *establish-ment-s*, but not *establish-s-ment*.

English has relatively few inflectional morphemes. These are on the whole easy to identify, though they sometimes present problems of analysis, as discussed below.

Allomorphs

Sometimes a morpheme has only one phonological form. But frequently it has a number of variants known as allomorphs.

Allomorphs may vary considerably. Totally dissimilar forms may be allomorphs of the same morpheme. *Cats, dogs, horses, sheep, oxen, geese* all contain the English plural morpheme.

An allomorph is said to be phonologically conditioned when its form is dependent on the adjacent phonemes. An allomorph is said to be lexically conditioned when its form seems to be a purely accidental one, linked to a particular vocabulary item.

The English plural morpheme provides excellent examples of both phonologically and lexically conditioned allomorphs. Let us look at some of these.

Phonological conditioning

The study of the different phonemic shapes of allomorphs is known as morphophonology – sometimes abbreviated to morphonology.

/-z/ /-s/ /-ɪz/ are all phonologically conditioned allomorphs of the English plural morpheme. That is, each allomorph occurs in a predictable set of environments.

/-z/ occurs after most voiced phonemes as in *dogs, lambs, bees*. (A voiced phoneme is one in which the vocal cords vibrate, as in /b/, /d/, /g/, /v/, and vowels.)

/-s/ occurs after most voiceless phonemes, as in *cats, giraffes, skunks*. (A voiceless phoneme is one in which the vocal cords do not vibrate.)

/-ɪz/ occurs after sibilants (hissing and hushing sounds), as in *horses, cheeses, dishes*.

If we take /z/ as basic, then we can say first, that /-z/ turns into /ɪz/ after sibilants (Figure 6.5), and second, into /-s/ after voiceless sounds (Figure 6.6):

ø → ɪ/ [+ sibilant] —— z. e.g. /hɔːsz/ → /hɔːsɪz/

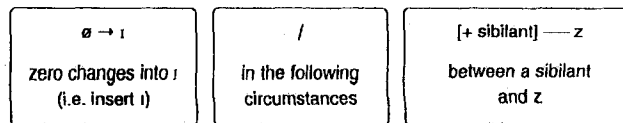| ø → ɪ	/	[+ sibilant] —— z
zero changes into ɪ (i.e. insert ɪ)	in the following circumstances	between a sibilant and z

figure 6.5

z → s/[-voice]—. e.g. /kætz/ → /kæts/

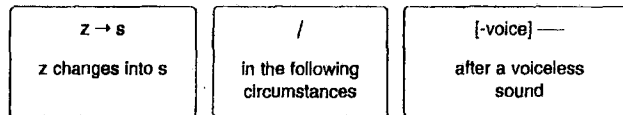| z → s	/	[-voice] ——
z changes into s	in the following circumstances	after a voiceless sound

figure 6.6

Note that these 'rules' must be applied in the order given above. If the order was reversed, we would get forms such as *[dɪʃs] instead of the correct [dɪʃɪz] for the plural of *dish*.

Lexical conditioning

Words such as *oxen, sheep, geese* present a problem. Although they function as plurals in the same way as *cats, dogs*, they are not marked as plurals in the same way. Such lexically conditioned plurals do not follow any specific rule. Each one has to be learnt separately.

Words such as *oxen, sheep, geese* can be identified as syntactically equivalent to the *cats* and *dogs* type of plural because they fit into the same 'slot' in a sentence (Figure 6.7):

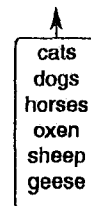The _____ are making a lot of noise

cats
dogs
horses
oxen
sheep
geese

figure 6.7

Oxen, sheep and *geese* each contain two morphemes:

> *ox* + plural
> *sheep* + plural
> *goose* + plural

But only *oxen* is easily divisible into two:

> *ox* + /-ən/ (*-en*)

Sheep can be divided into two if a zero suffix is assumed. A 'zero suffix' is a convenient linguistic fiction which is sometimes used in cases of this type. It is normally written /ø/:

> *sheep* + /ø/.

There is no obvious way to analyze *geese*. At one time, linguists suggested that the plural vowel /iː/ in /giːs/ (*geese*) which replaces the /uː/ in /guːs/ (*goose*) should be regarded as a special type of allomorph called a **replacive**. And they analyzed the plural as:

> /guːs/ + /iː/ ← (/uː/).

Here the formula /iː/ ← (/uː/) means '/iː/ replaces /uː/'.

But this is rather a strained explanation. These days, most linguists simply accept that the form /giːs/ (*geese*) represents two morphemes:

> *goose* + plural

and that these two cannot be separated. And a similar explanation is required for forms such as *went, took*, which represent:

> *go* + past tense
> *take* + past tense.

Word classes

In every language, there are a limited number of types of lexical item. These different kinds of word are traditionally known as 'parts of speech', though in linguistic terminology the label word class is more common. Word classes are conventionally given labels, such as noun, verb, adjective.

Words are classified into word classes partly on account of their syntactic behaviour, partly on the basis of their morphological form. That is, words from the same word class are likely to fit into the same slot in a sentence, and to be inflected in similar

ways. For example, the word class traditionally known as 'verb' can be recognized as a verb partly because it occurs after nouns (or phrases containing a noun), and partly because most verbs have an inflectional ending *-ed* to indicate the past:

> *Arabella detested snails.*
> *Marianna smiled.*

Careful analysis is needed, because in some cases, items which superficially appear to fit into the same slot in a sentence can turn out to be rather different in character. Consider the sentences:

> *Charlie ate caviare.*
> *Charlie ate well.*

At first sight, we might wrongly assume that *caviare* and *well* belong to the same word class. But a less superficial analysis reveals that they behave somewhat differently overall. If we tried altering the sentences around, we could say:

> *Caviare was eaten by Charlie.*
> *What Charlie ate was caviare.*

But we could not form the equivalent sentences with *well*:

> * *Well was eaten by Charlie.*
> * *What Charlie ate was well.*

These dissimilarities indicate that *caviare* and *well* are syntactically different, and belong to different word classes.

It is not always easy to tell how many word classes a language contains. Many traditional textbooks claim that English has eight 'parts of speech'. But this claim turns out to be based largely on old Latin grammars which were in turn translated from ancient Greek grammars, which mostly divided Greek words into eight word classes. If we look more closely, we find several discrepancies. For example, nouns and pronouns are traditionally classified as separate parts of speech, yet they have a large number of similarities:

> *Max laughed.*
> *He laughed.*

In fact, nouns and pronouns are more alike than the different types of word which are traditionally labelled adverbs. Words such as *quickly* and *very* are both usually classified as adverbs, but they behave quite differently:

> *He ran quickly.*
> * *He ran very.*

The number of word classes varies from language to language. Some word classes, such as noun and verb, may be universal. But others vary. Nouns, adjectives and verbs are on a continuum. At one end are nouns, words which maintain their identity over time, such as *tree, cat, river*. At the other end are verbs, words which signify rapid change, as in *walk, kick, push*. In the middle come properties, such as *large, beautiful, old*. In English, these form a separate word class, that of adjectives. But this is not inevitable. Some languages treat them as a type of verb, so-called stative verbs, ones which denote a state. Where English says:

> Petronella is happy.

a language such as Chinese might say, as it were:

> Petronella happies.

using a verb instead of an adjective. English also sometimes flips between verbs and adjectives. Compare the archaic *he ails* (stative verb) with the modern day *he is ill* (adjective).

Major word classes

English is sometimes considered to have four major word classes: noun (N), adjective (A), verb (V), preposition (P) (Figure 6.8).

Big	frogs	swim	under	water
A	N	V	P	N

figure 6.8

Of these four major classes, nouns, verbs and prepositions behave fairly differently from one another, though adjectives are somewhat strange, in that they have some noun-like qualities, and some verb-like ones. In *Blessed are the brave, brave* seems to have become a noun. And in *Mavis is asleep, asleep* seems fairly verb-like, since it fits into the same slot as *sleeping* in a sentence such as *Mavis is sleeping*.

It has been suggested that we should describe these four word classes in a manner parallel to the distinctive feature descriptions used for sounds, which can show shared similarities.

Noun	$[+ N, - V]$
Verb	$[- N, + V]$
Adjective	$[+ N, + V]$
Preposition	$[- N, - V]$

This seems to be a useful and economical way of capturing the similarities and differences between the major word classes.

The major word classes are known as **lexical categories**. Lexical categories contain content words, those with intrinsic meaning. They contrast with **functional categories**, which include 'little words' whose meaning is often difficult to specify, as *the, a*, which are determiners (D), or the **complementizer** *that* in *I know that Paul is ill*, often abbreviated to COMP or C. These function words are important for gluing pieces of sentences together into longer syntactic patterns.

This chapter has discussed words and morphemes. These fit into larger recurring patterns, which will be the topic of the next chapter.

Questions

1 Suggest three different ways in which the word **word** might be used.
2 What is a **morpheme**?
3 Distinguish between **inflection** and **derivation**.
4 What is the difference between **phonologically conditioned allomorphs** and **lexically conditioned allomorphs**?
5 How might one identify **word classes**?