1. Grammar

What is 'grammar'?

The English word *grammar* derives ultimately from the Greek *gramma*, meaning a 'letter'. In classical Greek and Latin the word *grammatica* referred to the general study of literature and language. When the word *grammar* came into English in the medieval period, it was used to refer only to Latin grammar; and it was not until the seventeenth century that the term took on a more general meaning and so had to be prefaced by 'Latin', 'English', 'French', etc.

Two meanings have competed with each other in English since the seventeenth century. In 1605 Francis Bacon wrote: "Concerning speech and words, the consideration of them hath produced the science of Grammar". While in 1637 Ben Jonson writes: "Grammar is the art of true and well speaking a Language". For Bacon, grammar is a science, a study of a set of phenomena; but for Jonson, grammar is an art, the skill or technique of speaking well. It is Jonson's definition, extended to include writing, that has predominated until recent times. Typical is the quote from an 1824 "English Grammar" by L Murray: "English grammar is the art of speaking and writing the English Language with propriety". Popularly, this is probably still what is understood by 'grammar'; but for linguists and students of language, it is Bacon's definition that is preferred.

'Grammar' vs 'a grammar'

Even having distinguished the 'science' from the 'art' definitions of 'grammar', the word is still ambiguous. On the one hand, the term 'grammar' can refer to certain features of a language, for example the "speech and words" of the Bacon quotation. That is to say that we may speak of the 'grammar of a language', being those characteristics of a language which we denote by the word 'grammar', rather than, say, the phonetic or semantic characteristics. On the other hand, the term 'grammar', or more accurately 'a grammar', is used to refer to a description either of the language as a whole (sounds, structures and meanings) or of the grammatical features of a language. A grammar in this sense is, then, a linguist's or grammarian's representation of those traits of a language designated by 'grammar' in the first sense.

Description vs prescription

Modern linguists, following Bacon's kind of definition of 'grammar', would claim to be describing the grammar of a language when they write a grammar. They would emphasise the word 'describe', in order to distinguish clearly what they are doing from the 'prescriptive' attitudes implied in the definitions of Jonson and Murray.

Descriptive linguists claim to be giving an account of what people actually do when they speak and write a language, rather than to be telling people what they ought to do when they speak and write their language, as the prescriptive grammarians tended to do. In the event, although different basic attitudes prevail, the distinction is probably not so clearcut as the terms 'descriptive' and 'prescriptive' imply. To be sure, prescriptive grammarians included rules in their grammars, such as "You should not end a sentence with a preposition"; but in so doing they still had to describe what a 'sentence' and a 'preposition' are. And a descriptive linguist producing a grammar of modern English, for example, has to make a choice of which English usage he is going to describe; and he would usually select the 'standard' variety, perhaps even 'standard educated usage', and by so doing he would have indulged in an implicit prescription.

'Good' and 'bad' grammar

There still remains a usage of the term 'grammar' that we have not mentioned, in instances such as: "She wrote me a nice letter, but her grammar was awful". Here the speaker is passing a judgement on someone's use of language. He is evaluating a piece of English usage against some supposed norm, called 'good English' or 'correct English'. The norm is usually what is called 'standard English' or 'educated English usage'. In other words, it is a particularly prestigious variety of English, which has achieved normative status because it is widely used in public life, and in its written form in most kinds of public communication (media, government, arts). Because it is a nationally used public variety of English does not automatically confer on it the epithet 'good', by which all alternative varieties are judged 'bad'. The use of double negation ("They didn't have none"), for example, is not intrinsically 'bad grammar'; it merely represents a 'non-standard' English usage. Linguists would prefer to replace the evaluative terms 'good' and 'bad' with the more neutral terms 'appropriate' and 'inappropriate'. That is to say, double negation may not be appropriate in the English of public communication, but in more casual or intimate circumstances among certain groups of English speakers it may well constitute a normal feature of English grammar.

Clearly, the extension of this argument is that no language can be regarded as a homogeneous entity. The term, 'the English language' subsumes a whole variety of 'Englishes', differentiated regionally and socially, one of which is the prestigious 'standard' variety, which is the variety of public communication, of the education system, and that is taught to foreign learners. When we talk of the grammar of a language, therefore, we need to specify which the variety of the language is whose grammar we are describing. Even then, we have to allow for variation within a variety; for example, in the previous sentence I used the relative pronoun *whose* to refer to the non-human noun *language*, which not all users of the 'standard' variety of English would do.

The scope of grammar

In his definition of the "science of Grammar", Bacon described the concerns of grammar as "speech and words". Certainly, words fall into the province of gram-

mar, but it is not clear what Bacon means by 'speech'. Although the term 'grammar' was at one time used to denote the study of all aspects of a language, linguists generally restrict the scope of the term these days. The linguistic study of a language, or variety of a language, is often said to comprise three components: phonetics/phonology, grammar, semantics. The three aspects of linguistic study are, however, not independent of each other.

Phonetics and phonology are both concerned with the pronunciation of language, how language sounds, the transmission of utterances through the medium of sound. The province of phonetics is the general study of speech sounds, their articulation, acoustic qualities and range; while phonology concentrates on the use particular languages make of the available human speech sounds. Phonology deals, for example, with the number and types of distinctive speech sounds employed by a language, their arrangement in syllables and words, besides the accompanying features of stress and intonation. One point where phonology intersects with grammar is in the consideration of the way in which grammatical categories are manifested in speech sound: note, for example, the different pronunciations of the English past tense ending (*-ed*) in the verbs *walked*, *filled*, *loaded*.

Semantics is the study of meaning. In a way, nearly all of language study is concerned with how language means. It is arguable, indeed, that there is no study of meaning separate from the study of grammar, phonology and the like. And some linguists would replace semantics with a branch of linguistics called lexicology, which comprises the study of words, their meanings and their interrelationships. Semantics is often conceived as being wider in scope than lexicology, concerned not only with the meaning of words, but also with the meaning of sentences and with meaning relations between words and between sentences.

Language is sometimes viewed as the means by which meanings are transmitted in sound via the organising principle of grammar. Grammar is thus concerned with the counters of meaning—words, parts of words, grammatical categories—and their combination into meaningful strings—sentences, texts—prior to vocalisation through the sound system of a language. The study of grammar is often subdivided into syntax and morphology, the former dealing with the structure of sentences, and the latter with the structure of words.

Syntax

The English word *syntax* derives from a similar Greek word meaning 'arrangement' and came to refer specifically to the arrangement of words. Syntax, then, is about the ways in which words may combine with one another, with the arrangements or patterns or orders of words that are possible in a language, and with the differences in meaning that different orders of words may bring about.

Let us illustrate these points from English. The string "Every good boy deserves favours" is a possible arrangement of words in English. The following are not permissible orderings (*Note*: '*' is used to mark an ungrammatical or unacceptable string, '?' to mark a doubtful string):

*"Deserves favours every good boy"
*"Good every boy favours deserves"
*"Boy good every deserves favours"
etc,

while "Favours deserves every good boy" could possibly be acceptable in an appropriate context, though this is perhaps doubtful. The distinction between possible and impossible, acceptable and unacceptable, is not always clearcut; native speakers of a language may vary in their judgements about the acceptability of an arrangement. (See further, Chapter 23.)

Consider now the string "Harry is tickling Susan". This is an acceptable arrangement in English, as is the following using the same words: "Susan is tickling Harry". But the two sentences do not mean the same: in the first, the person performing the action of tickling is Harry, while Susan is the one undergoing the action; in the second, their roles are reversed. Clearly, then, an alteration in the arrangement of words may bring about a change in meaning. In our investigation of syntax, therefore, we need to be able to identify elements in sentences that are movable and also determine the various functions they may perform in sentence structure.

Finally, look at the following string: "Very boy is because being atrocious naughty". This is obviously an unacceptable string in English, but it could not be made acceptable by a rearrangement of the words that it comprises. The problem lies in the incompatibility of a word like *very* with a word like *boy*, of a word like *because* with words like *is* or *being*, and of a word like *atrocious* with a word like *naughty*; although the last of these combinations could be made acceptable by converting *atrocious* into *atrociously*, giving the sequence "atrociously naughty". What these examples illustrate is that you cannot combine a word with just any other word in a language. In part, this arises from the lexical or semantic facts of a language; eg we talk about "naughty boys", but we do not usually talk about *"naughty buildings". But the incompatibilities we have been discussing do not arise particularly from the incongruities of meaning produced; rather from the grammatical, or more accurately, syntactic facts of the language. To put what we have been saying into the more precise linguistic terms we shall be using later: intensifying adverbs (like *very*) do not go together with nouns (like *boy*), conjunctions (like *because*) do not go together with verbs (like *is* and *being*), adjectives (like *atrocious*) do not go together with other adjectives (like *naughty*). Part of the task of a syntactic description is, then, to specify the possible combinations of words (more accurately, word classes—eg noun, adverb, conjunction, adjective) in a language.

Morphology

The word *morphology* derives from two Greek words which together mean 'the study of forms'. In botany, morphology is the study of plant forms. In linguistics, morphology is concerned with the grammatical structure of words; that is to say, not with the sounds or letters (or syllables) from which word forms are composed, but with constituents of words which have a grammatical function or meaning. For

example, the English word *redefinitions* may be analysed into four constituents: *re-define-tion-s*. Each of these constituents occurs as a constituent in the structure of other words in the language, with the same function or meaning (eg *define*, *re-inspect-*(*t*)*ion-s*). As in syntax, so in morphology, the permissible combinations of constituents are limited: the study of morphology must, therefore, identify the possible types of word constituent in a language, investigate the patterns and arrangements of the constituents in words, and determine the function(s) or meaning(s) realised by each constituent.

The study of the morphology of a language comprises two parts. On the one hand, morphology deals with the realisation of grammatical categories, systems, or meanings (such as number—singular/plural, tense—present/past) by means of inflections. The *-s* at the end of the English examples in the previous paragraph would be considered an inflection, realising the category of plural number in English. Sometimes an inflection will realise more than one grammatical category; for example, the *-s'* of *boys'* as in "the boys' bicycles" realises both plural number and possessive case, and it is pronounced in exactly the same way as *-s* realising only plural number in *boys* and *-'s* realising only possessive case in *boy's*.

On the other hand, morphology deals with the combining of word constituents as the means of deriving new words. One of the principal ways in which a language may add new words to its vocabulary is to make neologisms from existing word constituents by making combinations that have never been made before in the language. For example, the word *redefinition* is (was) made by adding the constituent *-tion* to the constituent (in fact a potentially independent word) *define*, and the constituent *re-* to the combination *definition*; or alternatively by adding *re-* to *define*, and then *-tion* to *redefine*. It is not clear in this case which order of combination might be regarded as more plausible. Sometimes, more than two constituents may be necessary in order to make a new combination; for example, *unparalleled* in English is a combination of *un-parallel-ed*. **Unparallel* is not a possible arrangement in English, and ?*paralleled* is probably at best marginal; so that we must regard *unparalleled* as a combination directly of three constituents.

Exercise 1

Say which of the following strings are grammatically unacceptable sentences of English, and explain why they are unacceptable:

1. All the birds have been migrating to warmer climates.
2. The dog big is intimidating the puppy small.
3. The baby eats not his dinner.
4. The violent wardrobe flew voraciously into the singing washing machine.
5. Every beautifully dancer receiving many into bouquet.
6. Some unhelpful decisions may not have been taken by wise men.

Exercise 2

Make as many acceptable English words as you can from the following word constituents:

1. re-
2. -ness
3. state
4. en-
5. -ment
6. large
7. -ly
8. treat

2. Word classes

Classification

From our earliest years we are taught to sort objects into groups on the basis of their similarity in one way or another. Indeed, there seems almost to be a human organisational urge to group similar things together in all kinds of areas of life. This is no less true in scientific study, where classification of phenomena forms an important part of description. In the study of language, then, we shall be grouping linguistic items into classes or types or categories on the basis of appropriate criteria of similarity.

One set of linguistic items that has traditionally been the subject of classification is that of words. In older grammars, word classes are often referred to as 'parts of speech', and the consideration of them forms the basis and main part of the grammar's content. In grammars such as these, deriving from a tradition based on Latin and Greek, eight parts of speech are recognised: Noun, Adjective, Pronoun, Verb, Adverb, Preposition, Conjunction, Interjection.

There are two problems with such a classification. One is that a classification worked out for one language may not be appropriate for another language; for instance, it is not clear where the English articles (*a*, *the*) fit into this Latin and Greek derived classification, since the classical languages did not have articles. The other problem concerns the criteria for the classification: on what basis are the words of a language grouped together into classes? One principle which emerges from this is that each language must be considered on its own terms.

Criteria of classification

One basis of classification that is applied uses semantic or 'notional' criteria, beginning from the question: What kinds of notions or concepts do the words refer to? It was this kind of criterion that gave rise to the traditional definitions of nouns as 'names of persons, places or things' and of verbs as 'doing words'. The problem with such definitions, even if they could be refined a great deal, is that they tend to remain vague and difficult to apply in any rigorous way. Such definitions are, however, useful as general characterising devices, and can serve as 'rules of thumb' in the recognition of words in an unfamiliar language.

A second basis of classification, that of morphology, is more easy to apply in a rigorous way, since it focuses on the form or structure of words; and this is not subject to the vagaries of intuitive or introspective judgements. A classification of words in English based on morphological criteria would, for example, note that one class of words (nouns) generally inflects for plural number and for possessive case (*girls*, *girl's*), while another class (verbs) inflects for past tense, present and past

participle (*showed*, *showing*, *shown*). Additionally, it might be noted that words containing the constituents *-tion*, *-ness*, *-ment*, *-ance* belong to the noun class (*celebration*, *keenness*, *retirement*, *remembrance*). However, derivational criteria of this kind would apply to only a small part of the membership of a class. Indeed, inflectional criteria are not necessarily universal in their application and are therefore not absolutely reliable: for example, nouns like *mud* or *information* in English do not form a plural, and the verbs *must* and *ought to* do not form a past tense or participles. Morphological criteria are not, moreover, universally applicable in the sense that some languages do not have words composed of more than one constituent, or alternatively do not have inflections regularly applying to all or most members of a class. The morphological basis of classification is, therefore, not totally reliable, though more reliable than the notional basis, and more reliably applied.

Exercise 3

In the following English data group the words into classes on the basis of their morphological characteristics:

1. big
2. bigger
3. biggest
4. read
5. reads
6. reader
7. sell
8. sells
9. seller
10. fast
11. faster
12. fastest
13. grow
14. grows
15. grower
16. tall
17. taller
18. tallest

Exercise 4

Do the same for the following Isthmus Zapotec (Mexico) data (SIL 1980: D1):

1. taburete	'chair'
2. ʃtaburetebe	'his chair'
3. ʃtaburetedu	'our chair'
4. pan	'bread'
5. ʃpanbe	'his bread'
6. ʃpandu	'our bread'
7. rukaadu	'we write'
8. rukaabe	'he writes'
9. zukaabe	'he will write'
10. bikaabe	'he wrote'
11. kukaabe	'he is writing'
12. ruʒooñedu	'we run'
13. ruʒooñebe	'he runs'
14. zuʒooñebe	'he will run'
15. biʒooñebe	'he ran'
16. kuʒooñebe	'he is running'

Syntactic criteria

A third basis for classification uses syntactic or 'functional' criteria. Words are grouped into classes according to their operation in syntactic structure: this covers

both position in relation to other words and function within linguistic units. For example, we could establish a class of words in English which fitted into the frame "the ____ man", and it would include items like *good*, *bad*, *big*, *small*, *fat*, *thin*, etc. Generalising, we would establish the class of words that could follow a definite article (*the*) and precede a noun. It would correspond to the traditional class of adjectives, but it would be established on the basis of a clear and more-or-less absolute criterion. In English, the same set of words would also be found in the frame "'Noun' is ____" (eg "the man is good/bad/big/small, etc"). In fact, this is not entirely correct, since there is a small set of adjectives in English that can function only in the pre-noun position (eg *mere* in "a mere boy"), and a small set that functions only after verbs like *be* (eg *asleep* in "the dog is asleep"). Apart from these few exceptions, the class of adjectives in English can be defined as the set of words that functions attributively (or as a modifier) before nouns and as a complement after *be* and similar verbs.

Consider the following Portuguese data (SIL 1980: D3)

1. A menina pequena chora pouco the girl little cries a little	'The little girl cries a little'
2. O menino pequeno chora the boy little cries	'The little boy cries'
3. A porca pequena come pouco the sow little eats a little	'The little sow eats a little'
4. O porco pequeno come muito the pig little eats a lot	'The little pig eats a lot'
5. O porco gordo come mais the pig fat eats more	'The fat pig eats more'
6. A menina gorda chora muito the girl fat cries a lot	'The fat girl cries a lot'

From this data it is possible to establish five word classes in Portuguese, on the basis of syntactic function:

1. A class of articles (*a*, *o*) functioning attributively before the noun.
2. A class of nouns functioning as head of the phrase functioning as subject of the sentence (*menina*, *menino*, *porca*, *porco*).
3. A class of adjectives functioning attributively after the noun (*pequena*, *pequeno*, *gorda*, *gordo*).
4. A class of verbs functioning as nucleus of the sentence (*chora*, *come*).
5. A class of adverbs functioning as attributive after the verb (*pouco*, *muito*, *mais*).

How many word classes?

The number of word classes that one identifies for a language will depend on the nature of the grammar of the language itself, as well as on the judgements of the linguistic analyst. Some languages are thought to have as few as three word classes. As we have already noted, a traditional 'notional' classification for English recog-

nises eight. A more rigorous 'functional' classification might also identify eight —Noun, Verb, Adjective, Adverb, Pronoun, Determiner, Preposition, Conjunction—with various subclasses. It would seem that all languages probably have a class approximating to that of Nouns and one approximating to that of Verbs in European languages, corresponding notionally to the 'things' of experience on the one hand, and to the 'events' of experience on the other.

If we establish word classes in a language on a syntactic, functional basis, it will often be the case that we find the membership of the classes overlapping. For example, the 'words' *cut*, *bottle*, *net*, *peel*, *screw*, etc, belong to both the Noun and the Verb class in English. In such a case, where we are dealing with a small set of overlapping members, we can say either that these words have multiple class membership, or that there are two 'words' *cut*, etc, in English, one a Noun and the other a Verb. You will find that dictionaries will sometimes deal with such a word under one headword and sometimes make two headwords (eg cut[1] N, cut[2] V). If, on the other hand, two functionally established classes are completely (or almost completely) overlapping, then it is usual to talk of a single word class with multiple functions. This is the case in English with the Adjective word class, which functions both attributively before Nouns and predicatively after 'copula' verbs (eg "the tall tree"—"the tree is tall"). Only a small number of adjectives is restricted in each case to attributive or predicative position (eg "a *mere* youth", "the baby is *asleep*").

Exercise 5

Set up word classes for the following Apinaye (Brazil) data (SIL 1980: D2):

1. kukrẽ kokoi eats monkey	'The monkey eats'
2. kukrẽ kra eats child	'The child eats'
3. ape kra works child	'The child works'
4. kukrẽ kokoi ratʃ eats monkey big	'The big monkey eats'
5. ape kra mɛtʃ works child good	'The good child works'
6. ape mɛtʃ kra works well child	'The child works well'
7. ape ratʃ mɨ mɛtʃ works a lot man good	'The good man works a lot'
8. kukrẽ ratʃ kokoi punui eats a lot monkey bad	'The bad monkey eats a lot'
9. ape punui mɨ piŋetʃ works badly man old	'The old man works badly'
10. ape piŋetʃ mɨ works a long time man	'The man works a long time'

Types of word class

It is often the case that some word classes in a language have a more readily extendable membership than others. The class membership is, in the first place, large and

continually fluctuates as new words are coined and some older ones fall out of fashion or otherwise become obsolete or archaic. Such word-classes in the European languages include Nouns, Verbs, Adjectives, and Adverbs. As new substances are invented or discovered, new thoughts and ideas developed, new sensations or qualities or ways of doing things and talking about things come into being, so new words are added to these word classes in the languages concerned.

These word classes are called 'open' or 'lexical' word classes: the term 'open' refers to the type of membership, while the term 'lexical' refers to the typical function of words in these classes. It is the members of these classes that carry the main burden of lexical meaning, referring to entities outside of language, in the structure of sentences. For example, note the open class words underlined in the following sentence:

"The motorist came round the corner and has knocked the child over".

It is predominantly the lexical words that survive in newspaper headlines and in telegrams.

The other type of word class is the 'closed' or 'grammatical' word class. Examples of these word-classes in the European languages are: Pronoun, Determiner, Preposition, Conjunction. Membership of these word classes is not readily expandable and changes only slowly over time; moreover, the membership is limited and can easily be enumerated. For example, the subclass of Personal Pronouns contains the following members in modern English: *I*, *me*, *we*, *us*, *you*, *he*, *him*, *she*, *her*, *it*, *they*, *them*.

The 'grammatical' label for these word classes refers to their mainly intra-linguistic function; they constitute the cement, holding the lexical building bricks together. Determiners, like *a*, *the* or *this*, have a function within a text; similarly many pronouns, like *he* or *they*. It would be incorrect, however, to assert that members of closed classes cannot refer at all; for example, prepositions like *into*, *below* or *along* clearly refer to specific kinds of describable spatial relationships. What we have here, as so often in language, is a continuum of more-or-less rather than the discrete classifications that scientists tend to make: at the 'most lexical' end of the continuum come nouns and verbs (though not all verbs), at the 'most grammatical' (ie 'least lexical') end come determiners and some pronouns; adjectives and adverbs are more lexical than grammatical, while prepositions and conjunctions are more grammatical than lexical. Only by examining the functions of the members of each word class can this categorisation be made for a particular language.

3. Noun

Reference

Nouns refer to 'things', in the broadest sense of that term. When we choose to refer to some entity by means of a noun, we choose to view it as a 'thing'. Some entities can be referred to only by nouns: *tree*, *window*, *van*. In other instances a choice can be made. Compare the following: "The bombers destroyed the village"—"the destruction of the village by the bombers". In the first, *destroy* is a verb and the action is viewed as an 'event'; in the second, *destruction* is a noun and the action is viewed as a 'thing', about which something further may then be predicated: "The destruction of the village by the bombers was unprovoked".

The kinds of 'things' that nouns refer to are varied. If a noun refers to a 'unique' thing, it is said to be a Proper Noun, such as personal names (Wolfgang Amadeus Mozart) or names of places or institutions (Berlin, Home Office); other nouns are 'common' nouns, which refer to classes of things or to instances of things. A traditional classification of common nouns is made into 'concrete' nouns and 'abstract' nouns. 'Concrete' nouns refer to tangible, observable things, that can be defined ostensively (ie by pointing to them and saying "That is a . . ."). In some dictionaries the definition of a concrete noun includes a photograph or a line drawing; such nouns are, arguably, best defined in this way. 'Abstract' nouns, on the other hand, cannot be defined ostensively or by means of illustrations; they refer to unobservable things like *imagination* or *truth* or *feminism*. They can be defined either by giving instances or by comparing their meaning with that of other related words in the vocabulary.

A further traditional classification of nouns is into 'animate' and 'inanimate'. Animate nouns refer to living 'things', ie people, animals, birds, fish, insects, etc. Sometimes, within this group, the distinction between 'person' (ie human) and non-person is grammatically important; for example, it underlies the choice between *who* and *which* as relative pronouns in English. Inanimate nouns refer to non-living things and thus include all the abstract nouns; alternatively, one might regard the animate/inanimate distinction as being a further subclassification of concrete nouns.

Gender

We turn now from semantic classifications of nouns to grammatical ones, which have been traditionally called 'gender', since in the European languages at least such classifications are based on natural gender or sex, ie the distinction between masculine, feminine and neuter. This is not the universal basis for a grammatical classification of nouns, however. In the Algonquian languages of North America,

for example, the gender classification is based on the animate/inanimate distinction. In Swahili there are six gender classes, the first containing mostly animate nouns, the second inanimate nouns, and the third plants. No gender system is completely consistent with its meaning base; there is always arbitrary assignment of words to genders from a semantic point of view, the overriding criterion has to be grammatical.

In German, for example, there are three genders: masculine, feminine and neuter. Virtually all nouns denoting male beings belong to the masculine gender, and those denoting female beings to the feminine gender, although *Weib* 'woman' and *Mädchen* 'girl' belong to the neuter gender. The neuter gender contains predominantly inanimate nouns, but these are also found in considerable numbers in the other two genders, eg *Tisch* 'table', *Baum* 'tree' are masculine, while *Blume* 'flower', *Wahrheit* 'truth' and many other abstract nouns are feminine.

There are no universal grammatical markers of gender; languages vary in the grammatical means used to signal the gender of a noun. Sometimes gender is marked in the noun itself, eg in Swahili by a set of prefixes. In French, nouns ending in *-tion* are usually feminine; in German, nouns ending in *-e* are usually feminine, as are abstract nouns in *-heit* or *-keit*. But such rules for the European languages do not account for a great number of words; here the markers of gender are outside the noun itself, usually in the forms of accompanying words like determiners or adjectives. In French, for example, the gender of a noun is identified by the form of the article accompanying it (*un*, *une* 'a'; *le*, *la* 'the' for masculine and feminine singular respectively), or by the form of a demonstrative determiner (*ce*, *cette* 'this'/'that'), or by the form of an accompanying adjective (eg "le garçon heureux", 'the happy boy'; "la fille heureuse", 'the happy girl').

A similar picture emerges for German. The singular definite article varies in form according to the gender of the noun: *der* masculine, *die* feminine, *das* neuter. There is also variation in the adjective, especially with the indefinite article: "ein grosser Mann" 'a big man', "eine grosse Frau" 'a big woman', "ein grosses Haus" 'a big house'. The pattern of agreement is, in fact, more complex than this, as there is a tendency to mark gender once only in any noun phrase.

Some languages do not make a gender classification of their nouns, for example Quechuan languages. It is arguable, moreover, that English does not exhibit grammatical gender: there is no mark of gender in the noun, apart from one or two endings that are restricted to nouns referring to females (eg *tigress*, *usherette*); and there is no mark in accompanying words (articles, adjectives, etc) either. The only argument for gender in English would be on the basis of the substitutability by a third person singular pronoun (ie *he*, *she* or *it*): this usage corresponds almost exactly with natural gender, except that *she* is sometimes extended in usage to refer to ships, cars, mountains, and there is variability of usage when referring to animals.

Exercise 6

Arrange the nouns in the following Chontal of Tabasco (Mexico) data (SIL 1980: C7) into genders, indicating how each gender is marked:

1. unkʔe hun	'one paper'	13. untu winik	'one man'
2. uʃkʔe hun	'three papers'	14. uʃtu winik	'three men'
3. wəkʔe hun	'six papers'	15. wʔtu winik	'six men'
4. untu tsitam	'one pig'	16. unkʔe wah	'one tortilla'
5. uʃtu tsitam	'three pigs'	17. uʃkʔe wah	'three tortillas'
6. wətu tsitam	'six pigs'	18. wəkʔe wah	'six tortillas'
7. untsʔit əhin	'one alligator'	19. untu mut	'one bird'
8. uʃtsʔit əhin	'three alligators'	20. uʃtu mut	'three birds'
9. wətsʔit əhin	'six alligators'	21. wətu mut	'six birds'
10. unkʔe jopo	'one leaf'	22. untsʔit tsan	'one snake'
11. uʃkʔe jopo	'three leaves'	23. uʃtsʔit tsan	'three snakes'
12. wəkʔe jopo	'six leaves'	24. wətsʔit tsan	'six snakes'

Countability

Many nouns refer to objects that are countable. The noun may be preceded or followed by a numeral indicating how many of the objects are being referred to, eg *six boxes*. Very often, the fact that more than one is being referred to is also marked in the noun (*boxes* rather than *box*). Besides numerals, quantity may also be indicated by 'indefinite' quantifiers, which do not specify an exact number, eg *several boxes*, *a lot of boxes*, *a few boxes*. Nouns, whose quantity may be specified in these ways, are called 'countable' nouns.

Some nouns refer to things that may not be counted. For example, it is not possible in English to speak of **six waters*. A plural form of the noun does not occur, and the quantifier *some* occurs before the singular form of the noun to indicate indefinite quantity: *some water* (compare *some boxes*—plural, but not **some box*). Such nouns are called 'uncountable' or 'mass' nouns. With some mass nouns in English a numeral may occur, but the meaning is not 'plural' but 'kinds of', eg *six wines*, *six cheeses*.

Number

Number is a grammatical category applicable to nouns when they are countable. Such nouns are often marked in languages for 'plural' number. That is to say, we may distinguish 'singular' number (referring to one) from 'plural' number (referring to more than one). Singular nouns do not usually have a positive mark for number: they are said to be 'unmarked'; it is plural number that is usually marked in nouns. Not all languages mark their nouns for number, and some only mark number when the noun is not accompanied by any other indication of quantity in the form of a numeral, indefinite quantifier, etc. Some languages, in addition to plural number, also mark a separate 'dual' number, when a noun refers to two objects.

Number is usually marked in nouns by means of suffixes (endings) or prefixes (added to the front). In European languages suffixation is the most common method. For example, in English, the regular mark of a plural is the (written) suffix *-(e)s*, eg *apples*, *apricots*, *peaches*. It might be noted that the pronunciation of the plural suffix is different for each of these three words, a point that will be taken up in Chapter 8. It might also be noted that, although suffixation by *-(e)s* is the regular

way of marking plural number in nouns in English, it is also marked in several other ways: compare, for example, *man—men*, *mouse—mice*, *ox—oxen*.

In German there are five different kinds of suffix to mark plural, three of which may in addition be accompanied by a change in the vowel of the noun (called 'umlaut', indicated in writing by '¨' above the vowel concerned). The first suffix is *-e*, with a noun like *Monat—Monate* 'months', and with umlaut in *Sohn—Söhne* 'sons'. The second suffix is *-(e)n*, with a noun like *Schmerz—Schmerzen* 'pains' or *Biene—Bienen* 'bees'; this suffix does not occur with umlaut. The third suffix is a 'zero'; that is, there is no ending, the plural is the same as the singular, as in a noun like *Daumen—Daumen* 'thumbs', or just with umlaut as in *Mantel—Mäntel* 'coats'. The fourth suffix is *-er*, with a noun like *Bild—Bilder* 'pictures', and with umlaut in *Rad—Räder* 'wheels'. The fifth and least frequent plural suffix, which is restricted mainly to words borrowed from other languages, especially English and French, is *-s*, as in a noun like *Park—Parks*; it does not occur with umlaut.

Number and gender

In some languages the marking of number and gender intersects. This is true to a limited extent for German: the *-e* suffix tends to be associated with masculine nouns, the *-(e)n* suffix with feminine nouns, and the *-er* suffix with neuter nouns. In Latin the intersection is more systematic: for example, *corvus—corvi* 'crow/s' represents a masculine plural, while *puella—puellae* 'girl/s' represents a feminine plural, and *bellum—bella* 'war/s' a neuter plural.

Consider the following data (SIL 1980: C8) from Konkomba (Ghana):

1. libil libaa	'one year'	7. kidʒuk kibaa	'one knife'
2. bubib bugmu	'five years'	8. ndʒum mulee	'two knives'
3. lilal libaa	'one axe'	9. kipipeek kibaa	'one place'
4. bulab bulee	'two axes'	10. npipeem mugmu	'five places'
5. lidʒool libaa	'one hill'	11. kigbalik kibaa	'one spear'
6. budʒoob buniin	'eight hills'	12. ngbalim muniin	'eight spears'

The nouns in this data are: *-bi-* 'year', *-la-* 'axe', *-dʒoo-* 'hill', *-dʒu-* 'knife', *-pipee-* 'place', *-gbali-* 'spear'. The first three of these (in Nos 1–6) belong to one gender, as marked by one set of prefixes and suffixes; the remaining three belong to a second gender, being marked by another set of prefixes and suffixes. The numeral (*-baa* 'one', *-lee* 'two', *-gmu* 'five', *-niin* 'eight') is also marked (by prefixes) according to the gender of the noun. The gender markers are, however, different in singular and plural. The markers of number thus intersect with the markers of gender, so that any prefix/suffix marks at the same time both a number (singular or plural) and a gender (1 or 2), according to the following scheme:

		Noun		Numeral
Gender 1	sing	li-	-l	li-
eg *-bi-* 'year'	pl	bu-	-b	bu-
Gender 2	sing	ki-	-k	ki-
eg *-dʒu-* 'knife'	pl	n-	-m	mu-

Case

Case is a further grammatical category that is marked in the noun and additionally (or alternatively) in words that accompany nouns (eg articles, adjectives). The function of case is to signal the grammatical/semantic relationships between nouns in a larger syntactic structure (eg clause or sentence). Languages vary in the extent to which case marking is used to signal these relationships. For example, in German the article (not the noun) is sometimes marked to signal whether the noun is functioning as Subject or Object in a sentence (see Chapter 11 for discussion of 'Subject', 'Object' and similar terms). In the following German sentence: "Der Hund hat den Mann gebissen"—'The dog bit the man', the noun *Hund* 'dog' is marked as Subject of the sentence by the form of the definite article *der* 'the', while *Mann* 'man' is marked as Object by the form of the definite article *den*. In English, on the other hand, the grammatical relationships 'Subject-of' and 'Object-of' are signalled merely by the relative ordering of the elements in the sentence. So, "The man bit the dog" has a different Subject and Object from "The dog bit the man"; while "Den Mann hat der Hund gebissen" has the same Subject and Object as "Der Hund hat den Mann gebissen".

There is a great variation in the number and kinds of case marking found in different languages. Some languages (eg Chinese) do not mark case at all. In English case marking in the noun is limited to a 'possessive' (or 'genitive') case, marked in writing by *-'s* in the singular noun and by *-s'* in the plural noun; it links two nouns in a semantic relationship of 'belonging' or 'possession' (eg "the cow's tail", "the girl's name") or in some kind of largely grammatical relationship (eg "the world's end", "the enquiry's conclusions"). In the pronoun in English, however, some distinctions are made between subject and object forms, eg *I—me*, *she—her*, *they—them*.

Four cases are distinguished in German, marked most consistently in the article: nominative, accusative, dative, genitive. The nominative and accusative are the Subject and Object cases respectively, mentioned earlier. The dative marks the Indirect Object in a sentence, eg *seiner Frau* '(to) his wife' in the following sentence: "Der Mann hat seiner Frau einen Brief geschickt", 'The man sent his wife a letter'. The genitive in German functions like the genitive/possessive in English. In addition, all the cases except the nominative are governed (see Chapter 16) by particular prepositions, eg *für* 'for' is followed by noun phrases in the accusative case, *von* 'by'/'from' by those in the dative, and *während* 'during' by those in the genitive. Latin has a further case, additional to the four found in German: the ablative, which usually translates into English by means of the prepositions *by*, *with* or *from*.

In Punjabi a 'direct' and an 'oblique' case are distinguished. The oblique case is used before postpositions (in common with other North-Indian languages, Punjabi has postpositions, coming after the noun referred to, rather than prepositions). The direct case is used for subjects and objects, although under certain circumstances these are realised by a noun in the oblique case followed by a particular postposition. In addition to the direct and oblique, Punjabi marks a 'vocative' case on nouns used as forms of address, an 'instrumental' case (in the plural only) to signal the

object 'with' which an action is performed, an 'ablative' case to signal movement 'from' a place, and a 'locative' case to signal 'at' a place.

In some languages, known as 'ergative' languages, the case of the Object of a transitive verb (for explanation of this term, see Chapter 10) is the same as that of the Subject of an intransitive verb. Such a language is Alawa, spoken in Australia. Among other cases, Alawa has a 'nominative', which is unmarked, and an 'operative', which is marked by a suffix. The nominative is used for the Subject of intransitive verbs and the Objects of transitive verbs; while the operative is used for the Subject of transitive verbs, as well as the instrument or location involved in an action. Consider the following examples:

lilmi na-wutala gutaru	'The man came to a hill'
man (nom) he-went hill (nom)	
lilmi-ri yaŋ ka-ŋatan-na da kiribu parakal-ta	
man (op) hit he-did-it kangaroo (nom) spear (op)	
	'The man killed the kangaroo with a spear'

Exercise 7

For the following Basari (Ghana) data (SIL 1980: C2), identify what the genders are and how they are marked for singular and plural in the noun and the demonstrative:

1. uni umbini	'this person'	biniib bimbini	'these persons'
2. ubɔti umbini	'this chief'	bibɔtiib bimbini	'these chiefs'
3. diyin dimbini	'this name'	ayin ŋimbini	'these names'
4. dibil dimbini	'this seed'	abil ŋimbini	'these seeds'
5. kusaau kumbini	'this farm'	tisaati timbini	'these farms'
6. kukabuu kumbini	'this basket'	tikabuti timbini	'these baskets'

4. Verb

Reference

Verbs refer to 'events'. Included under the heading of 'event' are: actions, eg *spill* in "The cat has spilled the milk"; 'process', eg *spill* in "The milk has spilled"; 'happenings', eg *occur* in "The accident occurred at lunch-time". One may also need to extend the notion of 'event' to include 'states', expressed in English by verbs like *seem*, *look* ("She seems/looks tired"), and above all *be* ("She is happy"). In the case of *be* in particular, however, two points might be noted. Firstly, many languages would not have the 'copula' verb *be* in sentences like "She is happy"; ie it would be equivalently "She happy" (cf Punjabi "ik pətla" = 'One (is) thin'). Secondly, it has been argued that for languages like English one should regard the 'verb' in sentences like "She is happy" as 'be happy'. The effect of this suggestion would be to vastly increase the number of verbs in a language like English, by including the majority of adjectives, and possibly many nouns as well (cf "She is a doctor"). One may well decide that the implementation of such a suggestion would not result in a very economical description of the language, since one would still need, for example, to account for the attributive use of adjectives in English (eg "the happy girl").

From the examples above ("The cat has spilled the milk"—"The milk has spilled") we can see that there may exist a relationship between actions and processes. Indeed, the use of *spill* in the first of these examples is sometimes referred to as an 'action-process', to distinguish it from action verbs that do not imply a simultaneous process, such as *laugh* in "Everybody laughed", or *kick* in "The referee has kicked the ball into the goal". Processes do not necessarily presuppose some action in order to trigger them off; all that is implied is some change taking place in the 'thing' undergoing the process, eg "The sky has cleared", "The leaves are turning yellow".

Tense

One piece of information that languages often encode with relation to events is the time when they take place, specifically either before the time of speaking, at the time of speaking, or after the time of speaking. These real-world time differences are encoded in language by means of the grammatical category of 'tense'. We thus often find verbs marked for 'past', 'present' and 'future' tenses, for example in French: "il parlait" 'he talked', "il parle" 'he talks', "il parlera" 'he will talk'.

In French, then, tense is marked by suffixes in the verb. Not all languages mark tense in this way; the English translations given for the French examples show that English marks present and past tense by means of suffixes (*-s* for present, and *-ed*

for past), but the future tense is marked periphrastically, by using another, 'auxiliary', verb. In fact, it is disputable whether a future tense can be clearly recognised in English, since it is possible to refer to future time in a number of ways, eg "he is going to talk", "he is to talk". Some languages do not mark tense in the verb by means of prefixes or suffixes at all, but by the use of auxiliary verbs or particles (see Chapter 9). In the Yurok language (North America), for example, tense is indicated by a series of particles placed before the verb, eg "ho s?egok" = 'past they-box', ie "they boxed". Other languages do not mark tense at all, but instead have an elaborate system of 'aspects' (see below); this seems to be the case in Classical Arabic.

We have suggested so far that tense, encoding real-world time, distinguishes past, present and future. These are not, however, the only time relations that may be encoded in language; they are time-relations viewed from the perspective of the time of speaking, the present. Some languages encode time relations that might be referred to as 'past-in-the-past' and 'past-in-the-future', eg in English "He had finished his breakfast, when . . .", "He will have completed the report by Thursday". The first of these sets one event in a past previous to another past event, while the second views an event as being past from a future perspective. These tenses are sometimes referred to as 'pluperfect' and 'future perfect' respectively.

Additionally, in some languages we find that there is no single past tense form, for example. Such is the case in French, where there are three past tenses, called the 'imperfect', 'preterite' and 'perfect'. A distinction between preterite and perfect is maintained in written French: the perfect implies that a past event has relationship to the present (eg "Jean m'a écrit deux fois cette semaine"—'John has written to me twice this week'), while the preterite places an event unequivocally in the past (eg "Jean lui écrivit à Noël"—'John wrote to him at Christmas'). In spoken French, however, the perfect has taken over the functions of the preterite, so that the latter is now restricted to the written form of the language. The imperfect should perhaps be considered under aspect (see below) rather than tense: it indicates 'past with duration' (eg "Il écrivait au moment où le téléphone sonna"—'He was writing when the telephone rang'), or 'past event repeated' (eg "Je lui écrivais chaque jour"—'I wrote to him every day'). But the imperfect in French is also used for past description ("Elle avait les cheuveux blonds"—'She had blond hair') and may be interchangeable with the preterite or perfect (eg "Il décidait soudain de se lever"—'He suddenly decided to get up").

Tense and time

Our discussion of tense up to now has centred on its relationship with time. But tense, although it clearly relates to real-world time, is nevertheless independent of it, in the sense that on occasions tense distinctions may have nothing to do with time. This independence of tense and time may be illustrated by the fact that they may not be in a one-to-one relationship in a language. Consider the 'present' tense in English: it may refer to present time, eg "I see no ships". But it may also be timeless in its reference, eg "The sun rises in the east", "Oil floats on water". It may be used with future reference, eg "Father comes home tomorrow"; or with

reference to habitual events, eg "We go shopping on Fridays". And in a few instances in English a present tense verb may refer to a past event, eg "I hear that you've changed your job", when the hearing was presumably before the time of speaking. A further illustration of this lack of one-to-one correspondence between tense and time is the perfect tense in German: on the one hand it is used like the perfect in French; on the other it may have a future-perfect reference, eg "Bald hat er es geschafft"—'He will have soon done it'.

The independence of tense and time is also evident in the fact that tenses may be used for other than time reference. The only difference in English between "May I use your telephone?" (present) and "Might I use your telephone?" (past) is one of tentativeness or politeness: the *might*-form is interpreted as a more polite request than the *may*-form. Similarly in conditional clauses in English, present tense implies a 'real' condition while past tense implies a 'hypothetical' condition, and both refer to future time: compare "If you visit the museum tomorrow . . .", "If you visited the museum tomorrow . . .". One might also point to the use of the present tense in past narratives (the so-called 'historic present') in order to mark a particular episode as climactic or to bring it into particular focus for some reason; spoken narratives in some styles of speaking are entirely in the present tense.

Tense and person

While dealing with tense we need to note that the marking of tense in the verb by means of suffixes (or prefixes) sometimes intersects with the marking of person/number. The category of person is dealt with in detail in the next chapter; let us note for the moment the distinction between the speaker (1st person singular/plural—*I*/*we*), the addressee (2nd person singular/plural—*you* for both in English), and the thing/person talked about (3rd person singular/plural—*he*, *she*, *it*/*they*). In the past tense in English, there is no mark of person along with that of tense: *-ed* signals past tense for all persons/numbers ("I talked", "you talked", "he talked", "we talked", "they talked"). In the present tense, there is a mark for the 3rd person singular, but for none of the other person/numbers ("he talks" but "I talk", "you talk", etc).

In French, on the other hand, each tense is marked by a different set of person-/number suffixes, so that each suffix is at the same time a mark of tense and person-/number, with very little overlap between forms. Consider the following table of forms for the verb *parler* 'speak', 'talk':

	Imperfect	Present	Future
1st sing *je*	parlais	parle	parlerai
2nd sing *tu*	parlais	parles	parleras
3rd sing *il*/*elle*	pårlait	parle	parlera
1st pl *nous*	parlions	parlons	parlerons
2nd pl *vous*	parliez	parlez	parlerez
3rd pl *ils*/*elles*	parlaient	parlent	parleront

The relative absence of overlap between forms applies primarily to the written form of the language; there is far more overlap in the spoken language, eg the 1st, 2nd

and 3rd person singular and the 3rd person plural forms of the imperfect are identical in pronunciation, as are the same forms of the present. Spoken French has diverged quite radically from the written form, which represents an older pronunciation.

Consider now the following data (SIL 1980: A3) from Mbembe (Eastern Nigeria):

1. nwa ogwo	'the child drinks'
2. nwa oci	'the child eats'
3. nwa oci eten	'the child eats fish'
4. aci eten	'you eat fish'
5. oci eten	'he eats fish'
6. nwa mogwo	'the child will drink'
7. maci eten	'you will eat fish'
8. kogwo	'he does not drink'
9. kagwo	'you do not drink'.

By comparing Nos 1 and 2 we can conclude that *nwa* means 'child', and *ogwo* and *oci* 'drinks' and 'eats' respectively. No 3 adds the further vocabulary item *eten* 'fish'. Nos 4 and 5 show us that the person is marked in the verb by means of a prefix, *o-* for 3rd person singular, *a-* for 2nd person: the verbs 'drink' and 'eat' are in fact then *-gwo*, *-ci*. Nos 6 and 7 show us that future tense is also marked in the verb with a prefix, independently of person and prior to it in order: *m-*. We conclude that present tense is *un*marked. Nos 8 and 9 show us additionally that the negative is marked in the verb by means of a prefix (*k-*), independently of person, though our data does not allow us to draw conclusions about its relationship to tense.

Exercise 8

Identify the constituents of the verbs in the following Congo Swahili data (SIL 1980: A2), and list them with their meanings:

1. ninasem	'I speak'
2. wunasem	'you speak'
3. anasem	'he speaks'
4. wanasem	'they speak'
5. ninaon	'I see'
6. nilion	'I saw'
7. ninawaon	'I see them'
8. niliwuon	'I saw you'
9. ananion	'he sees me'
10. wutakanion	'you will see me'

Now translate 'they saw him' into Congo Swahili.

Aspect

In addition to encoding in the verb the time at which an event takes place, many languages encode various ways of viewing the event, particularly from the point of view of the extension of the event in time. Ways of viewing the event are encoded in language by means of the grammatical category of 'aspect'. One aspect that is frequently found is the one we have already noted in the French imperfect: the

'durative', also termed 'continuous' and 'progressive'. The durative encodes a view of the event that regards it as having duration, continuing through a period of time, or being in progress. The English 'progressive' is an aspect with a similar meaning: compare "He slept all day" with "He was sleeping all day". The first views the sleeping as a once-and-for-all past event, the second views it as an event that has duration: the event is the same, the way of viewing it differs.

Another aspect that is often marked in languages is the 'completive' or 'perfective', which views an event as completed or finished rather than not yet completed. The difference in meaning may be illustrated from English with the following examples: "I was reading a book last night"—"I read a book last night". The first, in the past progressive form, implies an uncompleted event, while the second, in the simple past form, implies a completed event—the book was finished. A perfective aspect is found commonly in Slavonic languages among others. For example, in Polish one finds a set of perfective verbs contrastive with an equivalent set of imperfective verbs, eg *mrugnąć* (perfective) 'to blink once'—*mrugać* (imperfective) 'to blink several times', *zaboleć* (perfective) 'to begin aching'—*boleć* (imperfective) 'to ache', *domówić* (perfective) 'to finish speaking'—*mówić* (imperfective) 'to speak'.

As a counterpart to the continuous aspect, some languages positively mark a 'punctiliar' aspect: the event is viewed as taking place at a point in time. This is done in Alawa (Australia); compare the following forms: *ŋawiña* 'I will go' (future, no aspect marking), *ŋadiña* 'I can go' (present, punctiliar), *ŋadiři* 'I can go' (present, continuous).

As with tense, aspect may be marked in the verb by means of affixes (as in Alawa); or periphrastically with auxiliary verbs, as in the English progressive, which is formed with *be* as an auxiliary followed by a present participle (*is going*, *was talking*); or by means of particles.

Consider the following Ekpeye (Eastern Nigeria) data (SIL 1980: A6); there is no marker for 'future':

1.	edi	'he will eat'
2.	edikpo	'he will finish eating'
3.	edilɛ	'he has eaten'
4.	adikpolɛ	'we have finished eating'
5.	edikpohwɔ	'he will eventually finish eating'
6.	adigbalɛ	'we have eaten again'
7.	edikpohwɔlɛ	'he has eventually finished eating'
8.	emegba	'he will make again'
9.	amekpogbalɛ	'we have finished making again'
10.	amegbahwɔ	'we will eventually make again'

From the English glosses it is clear that a number of aspects are marked in the verb in Ekpeye, translated by English 'have', 'finish', 'eventually', 'again'. They are marked by the suffixes *-lɛ*, *-kpo*, *-hwɔ*, *-gba* respectively; and they occur in the order *-kpo-gba-hwɔ-lɛ*. The aspects might be termed 'perfect' (*-lɛ*), 'completive' (*-kpo*), 'culminative' (*-hwɔ*), 'iterative' (*-gba*). To complete the analysis, *-di-* 'eat' and *-me-* 'make' are the verbs, while *e-* 'he' and *a-* 'we' are person prefixes.

Exercise 9

Identify the constituents of the verb words in the following Inga (Colombia) data (SIL 1980: A7), and suggest terms for the aspects that are marked (*Note*: 3rd person singular is unmarked):

1. chayarca	'he arrived'
2. chayacurca	'he was arriving'
3. chayarcani	'I arrived'
4. chayacurcanimi	'I was definitely arriving'
5. chayacumi	'he is definitely arriving'
6. chayamucunimi	'I am definitely arriving here'
7. chayamurcachar	'he probably arrived here'
8. chayamucurcanguichar	'you were probably arriving here'
9. puñucuchar	'he is probably sleeping'

5. Pronoun

Types of pronoun

The term 'pronoun' implies an item which stands for or instead of a noun. In general, pronouns can be used in grammar where nouns would normally function, as Subject of a sentence, Object of a sentence, and so on. The class of pronouns, however, is not homogeneous; there are several types of pronoun.

Perhaps the words most commonly thought of as pronouns are those represented by the English set *I/me*, *you*, *he/him*, *she/her*, *it*, *we/us*, *they/them*. These are the 'personal' pronouns. It is to be expected that a similar set of pronouns would be more or less universal, although not necessarily with all the distinctions found in the English set.

Allied to the personal pronouns one may also find a set of 'reflexive' pronouns, eg in English *myself*, *yourself*, *himself*, *herself*, *itself*, *ourselves*, *yourselves*, *themselves*. In English these have two uses: when the Object of a sentence is the same as the Subject (eg "she has hurt herself"); and as an emphatic pronoun (eg "I did it myself"). Many languages do not have such an elaborate system of reflexive or emphatic pronouns; one form may well serve for all persons, as in North-Indian languages. In German the reflexive pronoun is the same as the accusative case form of the personal pronoun for 1st and 2nd persons, with a special form for 3rd person (singular and plural) – *sich*, eg "Ich habe mich verletzt" 'I have hurt myself', "Sie hat sich verlatzt" 'she has hurt herself'. The emphatic pronoun in German is an invariant form—*selbst*, eg "Ich habe es selbst gemacht" 'I did it myself'.

Also related to the personal pronouns one may find a set of 'possessive' pronouns, eg in English *mine*, *yours*, *his*, *hers*, *its*, *ours*, *theirs*, used in sentences like "This coat is yours" (ie '. . . belongs to you'), or "Mine is the brown one" (ie '(the coat) belonging to me . . .').

Other types of pronoun found in languages include: 'relative' pronouns, used to introduce relative clauses (see Chapter 13), eg *who*, *which*, *whose*, *that* in English; 'interrogative' pronouns, used in questions, eg *who?*, *what?* in English; 'indefinite' pronouns, used to refer to a non-specified person or thing, eg *someone*, *anything*, *everybody* in English; 'demonstrative' pronouns, with a distinguishing or pointing reference, *this/these*, *that/those* in English, eg in "That is a good idea", "This looks like the one we want".

Person

We noted in the previous chapter that probably all languages have a grammatical category of person, distinguishing a 1st, 2nd and 3rd person, referring to the speaker, the addressee, and the person/thing being talked about, respectively.

Unless, unusually, the reference of a 3rd person pronoun is clear in the situational context, a 3rd person pronoun will have a textual function: it will be standing for a noun occurring previously in the text; eg "Once upon a time there was an *emperor*. *He* ruled the country harshly". The reference of 1st and 2nd person pronouns is clear in the situational context, ie the person talking and the person being talked to, respectively.

Many languages have more than one 2nd person pronoun; indeed, English appears to be unusual in having only one. In French, for example, we find *tu* and *vous*. Although *vous* is the plural counterpart to singular *tu*, it is also used to speak to a singular addressee. A distinction is made in terms of familiarity or politeness: *tu* is used when talking to family or friends, *vous* with strangers. But the distinction is also connected with relative social status: a social superior may speak to a social inferior (eg adult to child, master to servant) with *tu*, expecting to be addressed with *vous* in return. This use of 2nd person pronouns to signal social asymmetry has, however, been gradually breaking down in the increasing egalitarianism of post-war society. Similarly in German: *du*, which has its own plural form *Ihr*, is the 'familiar' 2nd person pronoun, and *Sie* the 'polite' one, used for both singular and plural. A further level of politeness or deference may be added by using a 3rd person form of address; this may also be found in English, in deferential butler-talk, eg "My lord has requested the car. Would he like a rug?"

In Javanese there are three 2nd person pronouns, whose use depends on whether one is addressing a superior, an equal, or an inferior in social status; although how many of these an individual may use will depend on his own social status, and in Javanese it is not only 2nd person pronouns that are marked for social status but other vocabulary items as well (eg there are three forms for 'now', two for 'rice', etc). Similarly, in Hindi and Urdu there are three 2nd person pronouns, one for addressing inferiors, one for addressing intimates, and one for expressing politeness or respect. In Bengali there are also 'ordinary' and 'honorific' forms of the 3rd person pronoun.

Person and number/gender/case

We have already mentioned that number intersects with person in the pronoun system; that is to say, pronouns occur in singular and plural forms for each person. In the case of 2nd and 3rd person pronouns the reference is to a plurality of addressees or things under discussion, eg *Ihr*, *they*. But with 1st person pronouns we do not have a straightforward case of plurality; unless the speaker is a chorus, 1st person plural must refer to 1st person (*I*) plus others. Now 'others' could include either the addressee (*you*) or persons other than the addressee (*they*) or both together. Some languages encode these possible distinctions by having more than one 1st person plural pronoun: an 'exclusive' 1st person plural pronoun (ie *I* + *they*, but not *you*), and an 'inclusive' 1st person plural pronoun (ie *I* + *you*, and also possibly *they*). Alternatively, a distinction may be made, as in Alawa (see below), between *I* + *you* but not *they*, and *I* + *you*/*they*.

Usually only in the 3rd person, person also intersects with gender, often reflecting the gender distinctions made in the noun class. In French, for example, a mas-

culine/feminine distinction is made in both singular and plural 3rd person pronouns: *il/elle*, *ils/elles*. In German the masculine/feminine/neuter gender distinction is found only in the 3rd person singular pronoun (*er*, *sie*, *es*), the plural pronoun being common for all genders (*sie*). A similar pattern is found in English, where the 3rd person singular pronoun is the only place in the language that a gender distinction is maintained (*he*, *she*, *it*). In Punjabi we find the reverse case: the masculine/feminine gender distinction of nouns is not replicated in the 3rd person pronoun.

A third intersection of grammatical categories may occur in the pronoun between person and case, affecting potentially all person/numbers. English makes a case distinction (subject/object) in the personal pronoun system, though no longer anywhere else in the language, eg *I/me*, *she/her*. In fact, the 'subject'/'object' labels represent a simplification: the *I*-forms are restricted to subject position, though native speakers argue about whether one says "It is I" or "It's me" (cf the less likely "It's I", "It is me"); but the *me*-forms are used everywhere else in the structure of sentences, as object ("They saw me"), after a preposition ("Don't wait for me"), as an emphatic ("Who's coming?"—"Me!").

Pronoun systems

We will now diagram two personal pronoun systems, to show how the categories we have been discussing intersect to produce an array of pronoun forms.

German:

		Nominative	Accusative	Dative
sing	1	ich	mich	mir
	2 familiar	du	dich	dir
	3 m	er	ihn	ihm
	3 f	sie	sie	ihr
	3 n	es	es	ihm
pl	1	wir	uns	uns
	2 familiar	ihr	euch	euch
	2 sing/pl polite	Sie	Sie	Ihnen
	3	sie	sie	ihnen

It will be noted that the 2nd person 'polite' forms are identical in pronunciation with the 3rd person plural, although situational context will always disambiguate them. In writing, it is conventional to capitalize the initial letter of the 2nd person forms.

Alawa (Australia):

	Direct	Indirect
sing 12	ñanu	ñaka
1	ŋina	ŋapa
2	ñagana	ñaba
3 m	nula	nipa
3 f	ŋadula	ŋatu
non-sing 12	ñalu	ñalaŋa

Alawa (Australia) (cont.)

	Direct	Indirect
dual 1	ŋařu	ŋařaŋa
2	wuřu	wuřuŋa
3	yiřula	yiřuŋa
pl 1	ŋalu	ŋalaŋa
2	wulu	wuluŋa
3	yilula	yiluŋa

In Alawa a case distinction is made between 'direct' and 'indirect': the direct pronoun is used for the Subject of most clauses and wherever a pronoun is placed at the beginning of a clause for emphasis; the indirect pronoun is used elsewhere in the clause and for possession. The usual three persons are distinguished, together with an additional 1st person, marked '12': in the singular, this person refers to the speaker plus one hearer, in the non-singular to the speaker plus two or more hearers. Three numbers are distinguished for 1st, 2nd and 3rd persons singular, dual and plural. In the 3rd person singular a gender distinction (masculine/feminine) operates. There are no overlaps among the twenty-four forms of the personal pronoun.

Bound and free pronouns

The pronouns that we have discussed so far in this chapter all have the status of free and independent words. In the previous chapter we noted that person/number is also frequently marked in the verb by means of a prefix or suffix, eg in French or Congo Swahili. However, these two languages differ in the way in which they mark person in the verb: in the case of French the category of person intersects with that of tense, so that a particular suffix signals both person and tense simultaneously, eg "je parle", "je parlais", "je parlerai"—where *-e*, *-ais* and *-erai* mark both 1st person and present, imperfect and future tenses respectively. In the case of Swahili (Exercise 8) the category of person is marked independently of the category of tense: *ni-*, *wu-*, *a-*, and *wa-* are person prefixes, while *na-*, *li-* and *taka-* are tense prefixes. When person is marked in this way in the verb, independently of tense and aspect in particular (as categories that do not intersect with person in free pronouns), then we are justified in talking about a set of bound pronoun forms. Many languages exhibit parallel sets of free and bound pronouns.

Consider the system of personal pronouns in Bimoba (Ghana):

	Emphatic free	Non-emphatic free	Bound subject	Bound object
sing 1	min	n	m-/ma-	-n
2	fin	a	f-/fa-	-a
3 animate	ngɔɔ	u	w-/wu-	-u/-ɔ
pl 1	timm	te	t-	-et
2	yimm	i	y-/yi-	-i
3 animate	ngamm	be	b-	-eb
sing/pl 3 inanimate	ngann	le	l-	-er

Bimoba has two sets of free pronouns, one used in emphatic contexts, the other when the pronoun is not being emphasised. There are also two sets of bound pronouns, distinguished for case: a subject set, and an object set. The pronouns are distinguished for the usual three persons, and for singular and plural number. In the 3rd person a gender distinction is made between 'animate' and 'inanimate', and the 3rd person inanimate pronoun shows no distinction of number.

Exercise 10

In the following Zulu data (SIL 1980: E11), identify the personal pronouns and suggest what functions the different pronouns have (*Note*: *-ya-* in the verb word is an aspect marker which appears in the present tense when no complement follows the verb):

1.	Umfana u-fana ukudla boy he-want food	'The boy wants food'
2.	U-fana ukudla he-want food	'He wants food'
3.	uSipho yena u-fana ukudla Sipho he he-want food	'_____ but Sipho (himself) wants food'
4.	Yena u-fana ukudla he he-want food	'_____ but he (himself) wants food'
5.	Yena uSipho u-fana ukudla he Sipho he-want food	'_____ but he, Sipho (himself), wants food'
6.	U-shaya imbongolo he-hit donkey	'He hits a donkey'
7.	U-shaya yona he-hit it	'He hits it'
8.	U-ya-yi-shaya he- -it-hit	'He hits it'
9.	U-shaya yona imbongolo he-hit it donkey	'He hits it, the donkey'
10.	U-ya-yi-shaya yona he- -it-hit it	'_____ but he hits *it*'
11.	U-ya-yi-shaya imbongolo he- -it-hit donkey	'_____ but he hits it, the donkey'
12.	U-ya-yi-shaya yona imbongolo he- -it-hit it donkey	'_____ but he hits *it*, the *donkey*'

6. Word structure

Morpheme

In considering data in previous chapters we have on a number of occasions been involved in identifying constituents of words (eg in Exercises 7–9). The implication of this procedure is that words may have an identifiable structure and be composed of smaller recurrent parts. There is a sense in which words are composed of sounds or letters, eg *cat* is composed of *c* + *a* + *t*; but such structure is not grammatical structure. We are concerned with the analysis of words into recurrent parts with an identifiable constant grammatical function or meaning. Such constituents of words are called 'morphemes'. We may, then, define a morpheme as "the minimal unit of grammar", and impose the condition that a morpheme must have a discernible meaning or function, which is more or less constant for all its occurrences in the structure of the words of a language.

Morphemes can be identified by comparing the words of a language which have a similar form. Recurrent sequences (of sounds or letters) with the same function or meaning are recognised as morphemes. Consider the following English examples:

refuse	dismiss	revive	refer
refus-al	dismiss-al	reviv-al	refer-al
refus-al-s	dismiss-al-s	reviv-al-s	refer-al-s

A comparison of these forms enables us to identify the following morphemes: *refuse*, *dismiss*, *revive*, *refer* with their usual dictionary meanings; *-al*, which has the function of deriving a noun from a verb; *-s*, which signals the grammatical category 'plural number'.

Now consider the following data (SIL 1980: A1) from Jebero (Peru):

1. nuŋʃa	'a little canoe'
2. nuŋʃawək	'my little canoe'
3. wilaʃa	'a little child'
4. tulapəŋ	'your leg'
5. piðəknəŋ	'his house'
6. piðəklusaʔ	'houses'
7. piðəkpəŋlusaʔ	'your houses'

By comparing Nos 1 and 3, we can recognise morphemes *nuŋ-* meaning 'canoe', *wila-* 'child', and *-ʃa* 'little' or 'diminutive' (cf *-let* in English *piglet*, *notelet*). Returning to No 2, we conclude that *-wək* is a morpheme meaning 'my', ie '1st person singular possessive'. Looking now at Nos 5 and 6, we can identify a morpheme *piðək-* meaning 'house', and conclude that *-nəŋ* is a morpheme meaning 'his' or '3rd person singular possessive', and *-lusaʔ* is a morpheme meaning 'plural'.

From No 7 we recognise a further morpheme *-pəŋ* meaning 'your' or '2nd person possessive'. Returning finally to No 4 we conclude that it is composed of the morpheme *tula-* meaning 'leg' and the '2nd person possessive' morpheme *-pəŋ*. A comparison of *-nəŋ* and *-pəŋ* might suggest a further subdivision into *-əŋ* 'possessive' and *-n/-p* '3rd person'/'2nd person'; but this analysis is contradicted by *-wək*, which cannot be analysed in this way: on the basis of the data given we are not justified in proposing a further analysis of *-nəŋ* and *-pəŋ*.

Summarising, the Jebero data contains the following morphemes:

Nouns: *nuŋ-* 'canoe', *piðək-* 'house', *tula-* 'leg', *wila-* 'child'.
Possessives: *-wək* '1st person singular', *-pəŋ* '2nd person', *-nəŋ* '3rd person singular'.
Number: *-lusaʔ* 'plural'.
Derivative: *-ʃa* 'diminutive'.

Root and affix

If you look back at the examples we have discussed, it is evident that a word is composed of a central morpheme, to which other morphemes are added. Semantically, the central morpheme carries the main lexical reference of the word, and the other morphemes modify that reference in various ways. So, for example, the English word *refusals* has *refuse* as its central morpheme, and the Jebero word *nuŋʃawək* has *nuŋ* as its central morpheme. The central morpheme is called the 'root' of the word, and the peripheral morphemes that are attached to the root are called 'affixes'.

Affixes are always 'bound' morphemes: they cannot occur alone as words, they occur only in combination with other morphemes. Thus in the English examples, *-al* and *-s* cannot be independent words, they are always bound to another morpheme. You will notice that in writing these morphemes, as indeed in writing many of the other morphemes we have identified in this and previous chapters, in fact all the bound morphemes, we have written them with a hyphen on the side or sides where they are bound. In the case of *-al* and *-s*, they are bound on the left, similarly all the bound affixes in the Jebero data; in Exercise 8, however, the Congo Swahili verb word has affix morphemes bound on the right.

Root morphemes may be bound or 'free', ie able to stand alone as an independent word. In the English data above, the roots *refuse*, *dismiss*, etc, are free morphemes, since they constitute words by themselves. Most roots in English are free. In the Jebero data it appears that the roots are bound; there is no example where a root stands alone, but it is not inconceivable, for example, that *nuŋ* might constitute an independent word in Jebero, with the meaning 'canoe', similarly *wila* 'child', *tula* 'leg', *piðək* 'house'. The limited data that we have does not allow us to definitely conclude this, although it could represent a reasonable hypothesis awaiting confirmation from further data. In the Congo Swahili data of Exercise 8, it seems fairly certain that the verb roots are bound, since they are preceded by at least a tense morpheme and a person morpheme functioning as subject.

Exercise 11

Identify the morphemes in the following Isthmus Zapotec (Mexico) data (SIL 1980: A5); list the morphemes with their meanings, and mark the bound morphemes with hyphens as appropriate:

1.	ñee	'foot'	10.	ʒigilu?	'your chin'
2.	kañee	'feet'	11.	kaʒigitu	'your(pl) chins'
3.	ñeebe	'his foot'	12.	kaʒigidu	'our chins'
4.	ñeelu?	'your foot'	13.	ʒike	'shoulder'
5.	kañeetu	'your(pl) feet'	14.	ʒikebe	'his shoulder'
6.	kañeedu	'our feet'	15.	kaʒikelu?	'your shoulders'
7.	ʒigi	'chin'	16.	diaga	'ear'
8.	kaʒigi	'chins'	17.	kadiagatu	'your(pl) ears'
9.	ʒigibe	'his chin'	18.	kadiagadu	'our ears'

Types of affix

In the examples we have been concerned with so far, we have come across two types of affix: those bound to the right (eg *ka-* 'plural' in the previous exercise); and those bound to the left (eg *-be* '3rd person singular masculine possessive' in the previous exercise). An affix which is bound to the right is called a 'prefix', and an affix which is bound to the left is called a 'suffix'. We introduced these terms informally in Chapter 3. Prefixes and suffixes are the most commonly occurring types of affix found in the world's languages, and in European languages they are probably the only regularly occurring affixes. But there are two further types of affix that are found less commonly: the 'infix' and the 'suprafix'.

An infix is an affix which interrupts the root; it is inserted into the root, instead of being joined at the side as with a prefix or suffix. Consider the following data (SIL 1980: A15) from Chontal of Oaxaca (Mexico):

1.	tsetse	'squirrel'	tseɬtse	'squirrels'
2.	tuwa	'foreigner'	tuɬwa	'foreigners'
3.	te?a	'elder'	teɬ?a	'elders'
4.	meko?	'spoon'	meɬko?	'spoons'

The 'plural' morpheme is an infix *-ɬ-*, placed between the two syllables of the root morpheme.

Look now at the following Kamhmu? (Laos) data (SIL 1980: A16):

1.	kap	'to grasp with tongs'	krnap	'tongs'
2.	poot	'to walk on'	prnoot	'platform round a house'
3.	sal	'to place in ear-lobe'	srnal	'ear ornament'
4.	hiip	'tọ eat with a spoon'	hrniip	'spoon'

The morpheme *-rn-* is an infix, placed after the initial consonant of the root, with the function of deriving a noun from a verb; it might be called a 'nominaliser' morpheme.

The final affix type, the suprafix, is so called because it accompanies the root, and is usually written above the root. Suprafixes are, therefore, features of tone or stress

or nasalisation, which have an identifiable meaning or grammatical function. This may be illustrated from English: the word *import* has two stress patternings associated with a difference of word-class membership. If *ímport* receives its main or primary stress on the first syllable, it belongs to the noun class; but if it receives its main stress on the second syllable—*impórt*—it belongs to the verb class. Arguably, therefore, the shift in stress patterning may be regarded as a morpheme (suprafix), since it has the same function as, for example, the suffix *-ment* in *statement* (verb → noun) or the prefix *be-* in *befriend* (noun → verb), depending on whether one takes *ímport* (noun) or *impórt* (verb) as the basic (root) morpheme.

Consider the following data (SIL 1980: A17) from Bekwarra (Nigeria). The language is tonal (ie pitch change on a word may signal a lexical/meaning difference), having a system of three tones: high tone is marked ´; low tone is marked `; mid tone is unmarked. Compare: *abe éfàà* 'they grind', *abe éfaà* 'they learn', *abe éfaa* 'they roast'.

1. abe éfàà	'they grind'
2. abe efàà	'they ground'
3. abe èfàà	'they should grind'
4. abe éhàrà	'they answer'
5. abe ehàrà	'they answered'
6. abe èhàrà	'they should answer'

The tone on the initial syllable of the root is a suprafix with a number of distinct meanings: high tone signals 'present tense'; mid tone signals 'past tense'; low tone signals 'obligation'.

Word structure

We have described the structure of words in terms of the constituent morphemes that compose words. For each word there is a root morpheme, to which may be added one or more affixes. A root morpheme is the minimal form of a word, and some words consist only of a root morpheme, eg *art*, *circle*, *difficult* in English. In some languages (eg Chinese or Vietnamese) the vast majority of words are of this kind; ie there are few if any affixes in these languages. In languages like English, words consisting of a single root morpheme occur freely, as well as words composed of a root and one or more affixes, eg *give-s*, *un-decide-d*, *stigmat-is(e)-ing*, *general-is(e)-ation-s*, *re-institut-ion-al-is(e)*-ation. In a language like Swahili probably the majority of words contain affixes attached to a root.

When multiple affixation occurs, it is usually the case that the affixes occur in a specific order, so that it is important in the description of word structure to state affix orders if this is appropriate. We have done this, for example, in the solutions to Exercises 9 and 11. Consider now the following data (SIL 1980: A11) from Sierra Popoluca (Mexico):

1. anʌkpa	'I go'
2. tanʌkgakpa	'You and I go again'
3. minʌktaʔmgakum	'You all went again'

4.	nʌkjahpa	'They go'
5.	nʌknejahum	'They have gone'
6.	anʌkneta?mgakum	'We (exclusive) have gone again'
7.	nʌkjahgakpa	'They go again'
8.	nʌkpa	'He goes'

Let us first of all identify the morphemes, and then establish the order of affixes. From its recurrence in all the eight examples, it would seem clear that *nʌk-* is the root morpheme meaning 'go'; it appears to be bound on the right (cf Nos 4, 5, 7, 8). Comparing Nos 1 and 4, which differ only in person, we conclude that *a-* is the '1st person singular' morpheme, and *-jah* the '3rd person plural' morpheme. A comparison of Nos 4 and 8 reveals that *-gak* means 'again' or 'iterative'. From No 2 we can now conclude that *ta-* means 'you and I' or '1st/2nd person singular'. What Nos 1, 2, 4, 7 and 8 have in common by contrast with the remainder is that they are present tense, which appears to be marked by *-pa*. If this is so, we conclude from No 8 that '3rd person singular' is unmarked. Nos 3, 5 and 6 contain the suffix *-um*, which appears to mean 'past tense'. Comparing Nos 3 and 6, we can conclude that *-ta?m* means 'plural (for 1st/2nd person)', while *mi-* is the morpheme for '2nd person'; we must revise our assessment of *a-* to '1st person', being singular in No 1 without *-ta?m*, but becoming plural in No 6 only by the presence of *-ta?m*. The remaining morpheme to be identified is *-ne* in Nos 5 and 6, which corresponds to *have* in the English gloss, ie it has the meaning 'perfective'.

Summarising, we have identified the following morphemes in the verb word in Sierra Popoluca:

Verb Root: *nʌk-* 'go'.
Person Affix: *a-* '1st person', *mi-* '2nd person', *ta-* '1st/2nd person', *-jah* '3rd person plural'.
Tense Affix: *-pa* 'present', *-um* 'past'.
Aspect Affix: *-gak* 'iterative', *-ne* 'perfective'.
Number Affix: *-ta?m* 'plural'.

The order of morphemes is as follows:

Person (1st–2nd)—Root—Perfective—Plural/3rd person plural—Iterative—Tense.

Root and stem

We have identified the root as the part of a word to which affixes are added. Alternatively, the root may be viewed as the part of a word that remains after all the affixes have been removed. What we need in addition to the terms 'root' and 'affix' is a term to designate a structure (word) composed of root plus affix(es), but which may itself be further affixed. The term used is 'stem'. A stem is a word which may still have additional affixes attached to it. The term 'stem' thus includes 'root': a root is a minimal stem. Look at the following English example:

re-align-ment-s
ROOT
-STEM-
——STEM—

Align is the root. It is prefixed by *re-*, and *realign* is a stem, since it may be further affixed, eg by *-ment*. *Realignment* is also a stem, since it may be affixed by the 'plural' morpheme *-s*. However, *realignments* is no longer a stem, since it may not take any further affixes.

Exercise 12

Identify the morphemes in the following data (SIL 1980: A19) from Chatino (Mexico), and list them with their meanings (*Notes*: treat *ngu-* as a single morpheme; ~ marks nasalisation):

1.	ngudaō	'we(inclusive) gave'
2.	ndaō	'we(inclusive) give'
3.	ngudaba	'we(exclusive) gave'
4.	ngudā	'I gave'
5.	nguda	'You/he gave'
6.	nda	'You/he give'
7.	ndja	'You/he pay' (ie are caused to give)
8.	ngudja	'You/he paid'
9.	kuda	'You/he will give'
10.	kusiʔju	'You/he will cut'
11.	nsiʔjū	'I cut (present)'
12.	nsjiʔju	'You/he are caused to cut'
13.	ngusiʔjuō	'We(inclusive) cut (past)'
14.	ngusiʔjuwō	'You(plural) cut (past)'
15.	ngusjiʔju	'You/he were caused to cut'

7. Combining morphemes

Derivation and inflection

In the previous chapter we considered the decomposition of words into their constituent parts, ie morphemes. That is to say, we took an analytical aproach to the consideration of word structure. In this chapter we are going to take a 'synthetic' approach, attempting to answer the questions: "How do morphemes combine together to form words?", "What kinds of processes are at work?"

We need first of all to notice that two general processes are at work in combining morphemes into words: 'derivation' and 'inflection'. Derivation is a lexical process which forms a new word out of an existing word by the addition of a morpheme. For example, in English the suffix *-ment* may be added to the verb *state* to derive the noun *statement*, which is a different word, an additional vocabulary item in English. Similarly in the Kamhmuʔ examples cited in the previous chapter, the addition of the infix *-rn-* to, say, the verb *poot* 'to walk on' derives a noun *prnoot* 'platform around a house', an additional lexical item in this language. Strictly speaking, the term 'derivation' refers to the creation of a new word by means of the addition of an affix to a stem; but we shall also be considering other word-formation processes in this chapter, which have a similar, lexical, function.

The other general process at work in combining morphemes, inflection, is a grammatical process. For example, the plural morpheme in English is an inflectional morpheme. Inflections represent alternative grammatical forms of a word. The plural form *boys*, for instance, does not represent a different lexical item from the singular form *boy*, but rather merely the addition of the morphemic realisation of the grammatical category of plural number. An inflectional morpheme, then, does not derive a different vocabulary item; it realises alternative grammatical forms of the same word.

The distinction between derivation and inflection is a morphological one. It should not be concluded that a particular meaning (eg 'plural') is universally represented either inflectionally or derivationally. Plural number, for example, may be marked equally well by quantifiers (eg *many*, *several*) or numerals, ie lexically, as by an inflectional affix. Similarly, 'person' may be realised by an affix in the verb word (ie inflectionally) or alternatively by a separate pronoun word (ie lexically); compare the distinction made in Chapter 5 between bound and free pronouns. Or consider the Sierra Popolucá data in the previous chapter: here the meaning 'iterative' is realised grammatically by the inflectional morpheme *-gak*, whereas in English this meaning is realised lexically by the separate word *again*. The term 'derivational' and 'inflectional' are, then, a means of distinguishing a morphological process that creates new vocabulary items on the one hand, and on the other a morphological process that realises alternative grammatical forms of the same word.

Compounding

We look first under the broad heading of derivation at a word-formation process known as 'compounding'. Compounding involves the combination of two roots (Root + Root), to form a new word or stem. For example, in English the word *driftwood* is made up of two roots, the verb *drift* and the noun *wood*, both of which may be independent words in their own right. Similarly in German, the word *Lebensabend* 'old age' is a compound of the nouns *Leben* 'life' and *Abend* 'evening'.

In compounding, two roots combine in such a way that each loses its independent word status and the new combination takes on the characteristics of a single word. These single-word characteristics may include the semantic dimension, eg a compound may have a unitary meaning not immediately derivable from its elements, as in German *Augenblick* 'moment', made up of *Auge* 'eye' and *Blick* 'look/glance'. Or they may include the grammatical dimension, eg a compound will inflect like a single word, as English *babysit*. This has all the usual verb inflections of English, eg *babysits*, *babysitting*, *babysat*; but the *baby* part of the compound does not inflect, eg for 'plural', however many babies may be involved. Additionally, single-word characteristics may include the phonological dimension, eg a compound will take on the stress features of single words, or regular phonological changes may occur when roots are compounded. In English, for example, single words have characteristically one primary (or main), stress, so that in a compound one of the roots normally loses its primary stress; compare "a bláck boárd" and "a bláckboard". In German, a number of roots forming the first element of a compound regularly take an additional *-s*, eg *Universitätsprofessor* 'university professor', *Universitätsgebäude* 'university building', *Universitätsstudium* 'study at a university'.

Consider now the following data (SIL 1980: I8) from Igede (Nigeria):

1. iyo	'meat'	enyi	'water'	iyenyi	'fish'
2. imi	'hunger'	enyi	'water'	imenyi	'thirst'
3. ube	'room'	utoji	'medicine'	ubutoji	'clinic'.

It will be noticed that all the nouns in this data begin with a vowel and end with a vowel. However, it appears that two vowels cannot occur together in a word, so that on compounding the final vowel of the first element is omitted. Note that in the compounds in this data it is the first element that is semantically primary, ie *iyenyi* 'fish' is a kind of meat (*iyo*), *imenyi* 'thirst' is a kind of 'hunger' (*imi*), and *ubutoji* 'clinic' is a kind of 'room' (*ube*). In our English and German examples, representative of these languages generally, the order is reversed: *driftwood* is a kind of 'wood', *Universitätsgebäude* is a kind of *Gebäude* 'building'.

Compounding is not always a matter of the combination of two roots. Take the compound *watchmaker* in English, representative of a common type of English compound. This is a combination of the noun root *watch* and the derived noun *maker*, which has a verb root *make*. But there is no **watchmake* in English, so that the second element of *watchmaker* is not, strictly speaking, a root, but a stem. Compounds themselves should also be regarded as stems, since they may be

affixed; eg *watchmakers* by the 'plural' *-s* suffix, *overtaker* by the 'nominaliser' *-er* suffix. Indeed, compounds may be elements in further compounds, eg in German *Mitunterschrift* 'joint signature', made up of *mit* 'with' plus *Unterschrift* 'signature', which in turn is a compound made up of *unter* 'under' plus *Schrift* 'writing'.

Exercise 13

Examine the following data (SIL 1980: I9) from Mbembe (Nigeria) and say what phonological processes occur as a result of compounding:

1. ana	'oil'	ɛꝑya	'market'	ɛnaꝑya	'selling of oil'
2. ɔbɔːk	'arm'	ɛdɔŋɔ	'throat'	ɛbɔːdɔŋɔ	'wrist'
3. ɔsɔːm	'house'	ɛmma	'mouth'	ɛsɔːmma	'threshold'
4. eci	'tree'	ɛraŋa	'root'	ɛciraŋa	'medicine'
5. ɛciraŋa	'medicine'	ɔsɔːm	'house'	ɔciraŋasɔːm	'hospital'
6. eci	'tree'	ikwuma	'half'	icikwuma	'log of wood'
7. ɔda	'sleep'	icen	'eye'	idacen	'dream'
8. ɛba	'breast'	asi	'liquid'	abasi	'milk'

Derivation

We look secondly at derivation proper, the coining of new items of vocabulary by means of affixation. All four types of affix may be used derivationally. We mentioned a derivational suffix and a derivational infix at the beginning of the chapter. A derivational prefix can be illustrated from English, eg *be-* in *befriend*, deriving a verb from the noun *friend*. The stress change on the English word *import*, deriving a noun from a verb or vice versa, could be regarded as a derivational suprafix. Derivation involves the addition of an affix to a root or stem to form a different lexical item (word or new stem).

The majority of derivational affixes effect a change in the word class membership of the stem to which they are added, ie the derived word (or stem) has a different word class membership from that of the stem being affixed. For example, in German the addition of the suffix *-heit* to the adjective *wahr* 'true' derives the noun *Wahrheit* 'truth', in English the addition of the prefix *en-* to the adjective *large* derives the verb *enlarge*. A derivational affix may not in some cases effect a change in the word class membership of the stem to which it is added, eg in English the negative prefix *un-* when added to the adjective *true* derives another adjective *untrue*, and when added to the verb *tie* derives another verb *untie*. Intermediate between these two types of derivational affix is a third type which does not effect a change of word class, but of sub-class within a class, eg in English there are several derivational suffixes forming abstract nouns from concrete nouns: *machine-ery*, *boy-hood*, *dictator-ship*.

Consider now the following data (SIL 1980: I2) from Mbembe (Nigeria):

1. bèŋa	'to grow'	àbèŋíjí	'growth'
2. ɓɔr	'to be pale'	àɓɔríjí	'paleness'
3. dòŋ	'to be deep'	àdòŋíjí	'depth'

4. gòrɔ	'to be proud'	àgòríjí	'pride'
5. báŋ	'to be dirty'	ábáŋíjí	'dirtiness'
6. cámɛ	'to be small'	ácámíjí	'smallness'

In this data, the derivation of a noun from a verb involves a simultaneous prefixation and suffixation. Prefixation is by *a-* which 'agrees' in tone with the root, low tone before low tone (Nos 1–4), high tone before high tone (Nos 5 and 6). Suffixation is by *-íjí*. If the verb root ends in a vowel (Nos 1, 4 and 6), this is deleted when *-íjí* is affixed.

Even more true of derivation than of compounding is the possibility of a word resulting from the application of several derivational processes. We noted this point in the previous chapter. For example, the English word *intermarriage* may be viewed as the prefixation of the verb root *marry* by *inter-* (which does not change the word class), and then suffixation by *-age* to derive a noun. Similarly, the German word *misstönig* is first of all the prefixation of the noun root *Ton* 'sound' by *miss-* to derive *Misston* 'dissonance' (still a noun), and the suffixation of this stem by *-ig* accompanied by umlaut of the root vowel to derive the adjective *misstönig* 'dissonant'.

Exercise 14

Examine the following Lisu (North Thailand) data (SIL 1980: I3) and identify the derivational affixes, giving each a 'meaning'.

1. syɨ	'to sweep'	9. tshɨ	'to wash'
2. syɨdwu	'broom'	10. tshɨswu	'washer woman'
3. syɨswu	'sweeper'	11. tshɨgwu	'washing place'
4. tshye	'to hide'	12. tshɨdwu	'thing for washing clothes'
5. tshyegwu	'hiding place'	13. pū	'to write'
6. tya	'to live'	14. pūdwu	'pencil'
7. tyaswu	'inhabitant'	15. pūswu	'writer'
8. tyagwu	'house'		

Other word-formation processes

Apart from the non-morphological ways of adding new vocabulary to a language such as borrowing words from other languages (cf *detente* in English from French, or *le weekend* in French from English) or creating words from roots in classical Latin and Greek (eg many scientific and medical terms), there are a few more strictly morphological ways of forming new words, but of a minor nature by comparison with compounding and derivation proper.

Perhaps the most widespread of these minor word-formation processes is one known as 'conversion', which could be described as derivation without affixation. A word (or stem) belonging to one word class is transferred to another word class without any change of form by means of affixation. For example, the word *bottle* in English is a noun, but it is also used as a verb without any change of form, except that as a verb it is subject to the usual inflectional affixation of verbs (*he/she bottles*, *bottling*, *bottled*). Compare also: *peel* (noun and verb), *race* (verb and

noun), *dirty* (adjective and verb), *escape* (verb and noun), *picture* (noun and verb). Similarly in other languages, eg French *travail/travailler* 'work', *lutte/lutter* 'fight' (noun and verb, the *-er* being the verb inflectional suffix for the infinitive, ie citation form), *innocent* 'innocent' (adjective and noun) together with the verb *innocenter* 'to declare innocent'; also German *Blick/blicken* 'look, glance', *Arbeit/arbeiten* 'work', *Rauch/rauchen* 'smoke' (all noun and verb, with *-en* the inflectional suffix for the infinitive form of the verb).

A second minor word-formation process is one known as 'back-formation', which is in a way the converse of derivation: a word is coined by the removal of a (supposed) affix. For example, the verb *babysit* in English was derived from the noun *babysitter* by the removal of the *-er*, which was associated with the 'agentive' suffix of words like *walker*, *seller*, *flyer*, etc. The verb *edit* was derived from the noun *editor* in a similar way. Unlike derivation proper, however, the process of back-formation is not obvious to the modern native-speaker without access to historical linguistic information. Nevertheless, back-formation is a process by which new words may yet be coined in a contemporary language.

A last minor word-formation process that we might mention is the coining of new words by forming acronyms, ie taking the initial letters of the words in a phrase and forming them into a new word. For example, the word *radar* was formed in this way, standing for 'radio detecting and ranging'; also *laser*, 'lightwave amplification by stimulated emission of radiation'. But this word-formation process is popular with international organisations, cf *NATO* 'North Atlantic Treaty Organisation', *UNESCO* 'United Nations Educational, Scientific and Cultural Organisation'.

Inflection

We come finally in this chapter to inflectional morphemes, those which add grammatical meaning to a word. Inflectional morphemes cannot change the word class membership of a stem to which they are affixed. They realise grammatical categories and produce different grammatical forms of the same word. Take the verb in English: as a rule, it has potentially five different forms—base form/present tense (eg *show*), 3rd person singular present tense (eg *shows*), past tense (eg *showed*), present participle (eg *showing*), past participle (eg *shown*). In fact, for the majority of verbs the past tense and the past participle have the same form (eg *walked*). The inflectional morphemes here realise categories like person/tense (*-s*), past tense (*-ed*), and participles (*-ing*, *-n*); and the resulting verb word forms are used appropriately in specific lexical and syntactic contexts, eg the present participle is used with *be* to realise 'progressive aspect' ("they are *showing*") or after certain other verbs ("they like *reading* stories").

In English, inflectional morphemes are always suffixes (cf also 'plural' and 'possessive' in the noun—*girls*, *girl's*); but any of the four types of affix may be inflectional. In the Congo Swahili data of Exercise 8, the inflections for person and tense are both prefixes, eg *ninasem* 'I speak' (*ni-* '1st person singular', *na-* 'present tense'). In the previous chapter, one of the illustrations of an infix was the inflectional morpheme 'plural' in Chontal of Oaxaca, eg *tuɬwa* 'foreigners' (*-ɬ-* 'plural'). And in the

Chatino data of Exercise 12, the inflectional morpheme '1st person singular' is a suprafix of nasalisation, eg *ngudā* 'I gave' (*ngu-* 'past tense', *-da* 'give', ~ '1st person singular').

We have noted on a number of occasions that grammatical categories intersect and find a realisation by a common morpheme; for example, number and person intersect in pronouns; tense, person and number intersect in verbs. That is to say, one inflectional affix may 'carry' a number of grammatical meanings, and it is not possible to identify a separate segment of the word for each meaning. For example, the *-s* inflection in the English verb (eg *shows*) 'carries' the meanings '3rd person', 'singular number', and 'present tense'. This point can perhaps be illustrated best from the conjugational and declensional paradigms of Latin or Greek grammar. Consider, for instance, the verb form *amo* 'I love' in Latin. It is composed of the root *am-* 'to love' and the inflectional suffix *-o*, meaning '1st person singular present active indicative'; ie it 'carries' five meanings. It is '1st person' by contrast with, say, '2nd person', eg *amas* 'you love'. It is 'singular' in number by contrast with 'plural', eg *amamus* 'we love'. It is 'present' in tense by contrast with, say, 'past', eg *amabam* 'I loved'. It is 'active' in voice by contrast with 'passive', eg *amor* 'I am loved'. And it is 'indicative' in mood by contrast with, say, 'subjunctive', eg *amarem* 'I should love'. Thus, there is no necessary one-to-one relationship between inflectional affix and grammatical category or meaning.

Exercise 15

Identify the morphemes in the following Sierra Aztec (Mexico) data (SIL 1980: A9) and list them with their meanings. Say if the affixes are inflectional or derivational. Chart the order of morphemes.

1.	nimitsita	'I see you'
2.	nikita	'I see him'
3.	nikmaka	'I give it to him'
4.	tikmaka	'You give it to him'
5.	tinetʃita	'You see me'
6.	nannetʃmaka	'You(pl) give it to me'
7.	tikonmaka	'You give it to him, Sir'
8.	tikonitatihtsinoh	'You see him, most honoured Sir'
9.	tinetʃonita	'You see me, Sir'
10.	tinetʃonmakatihtsinoh	'You give it to me, most honoured Sir'
11.	nannetʃonmakatsikah	'You(pl) give it to me, honoured Sirs'
12.	nannetʃonitatihtsinoh	'You(pl) see me, most honoured Sirs'
13.	tinetʃonitatsikah	'You see me, honoured Sir'
14.	nannetʃonmakatihtsinoh	'You(pl) give it to me, most honoured Sirs'

8. Morpheme variants

Some knowledge of phonetics is assumed in this chapter.

Allomorphs

In Chapter 3 the reader was requested to note the different pronunciations of the 'plural' suffix in the English words *apples*, *apricots* and *peaches*. We come now to consider the nature of such differences. What differences like this mean is that a particular morpheme (eg 'plural number') may not be always pronounced (or spelt) in the same way wherever it occurs as a constituent of a word. Morphemes, that is to say, may have variant realisations (pronunciations or spellings), depending on their context of occurrence. These variant realisations are known as 'allomorphs'. From our examples above, the English 'plural' morpheme has the allomorphs /z/, /s/ and /ɪz/ respectively. Strictly speaking, then, the 'morpheme' is an abstract unit, whose concrete realisation in sound or spelling is by means of a 'morph' or—if there are variant realisations—'allomorphs'. We can equate the term 'morpheme' with the 'meaning' or 'function', and the term 'morph' with sounds or letters.

Some morphemes have only a single realisation; they always appear with the same pronunciation/spelling in a particular variety of the language (eg dialect or accent). For example, the 'present participle' morpheme in standard English has the single realisation *-ing*/ɪŋ/in all its contexts of occurrence, ie whichever verb root it is affixed to (cf *sowing*, *sleeping*, *finding*). Most inflectional morphemes in English, however, do have alternative realisations (allomorphs), such as the 'plural' inflection. But it is not only inflectional morphemes that may have allomorphs; derivational morphemes and root morphemes may also have variant forms. For example, the 'negative' prefix *un-* has variant forms in *unpleasant*/ʌm/, *unkind*/ʌŋ/ and *unchanged*/ʌn/; the root *resign* has variant forms in *resigns*/rīzaīn/and *resignation*/rezīgn/.

There are two kinds of descriptive explanation for morpheme variants. On the one hand, the pronunciation of a morpheme may vary because of its phonological context, ie the nature of the sounds in the accompanying morpheme(s): such allomorphs are said to be 'phonologically conditioned'. On the other hand, the realisation of a morpheme may vary quite arbitrarily when combined with certain other morphemes: such variation is said to be 'morphologically conditioned', ie the variation depends purely on which morpheme accompanies that with variant forms, but not on its phonological form.

Exercise 16

Identify the variant forms (allomorphs) of the 'past tense' morpheme in the following English data. Be sure to consider the pronunciation:

1. believe	believed	6. miss	missed
2. float	floated	7. redeem	redeemed
3. snatch	snatched	8. give	gave
4. allow	allowed	9. catch	caught
5. need	needed	10. put	put

Phonological conditioning

The allomorphs of a morpheme are said to be phonologically conditioned when the variation in pronunciation can be explained from the phonological context, ie the sounds composing the adjoining morph(s). Take the allomorphs of the 'plural' morpheme in English, with which we began this chapter: /z/, /s/, /ɪz/. These allomorphs are phonologically conditioned: which one of them occurs depends on the nature of the final sound of the noun stem to which the 'plural' morpheme is affixed. If the final sound of the noun stem is a sibilant (ie/s, z, ʃ, ʒ, tʃ, dʒ/), then the 'plural' morpheme is realised by the allomorph /ɪz/; if the stem-final sound is a voiceless consonant (other than a sibilant), the allomorph is /s/; if the stem-final sound is a voiced sound, ie a vowel or a voiced consonant (other than a sibilant), then the allomorph of the 'plural' is /z/.

This conditioning just described may be expressed in the following formula:

	/ɪz/	after stem-final sibilants
{plural} →	/s/	after-stem-final voiceless consonants
	/z/	elsewhere

The morpheme is conventionally written in curly brackets (braces), and the allomorphs in the slashed brackets of phonemic transcription. The arrow represents the statement "is realised by the allomorphs", and the contexts of conditioning are put after each allomorph. Note that the allomorph with the narrowest distribution (ie occurring after the least number of sounds) is put first, and that with the widest distribution last. Each successive conditioning statement excludes the one(s) that have preceded, so that the last one is appropriately expressed by an "elsewhere" statement, since it refers to every phonological environment not previously mentioned. A similar statement can be made for the 'past tense' suffix identified in Exercise 16:

	/ɪd/	after stem-final alveolar plosives
{past tense} →	/t/	after stem-final voiceless consonants
	/d/	elsewhere

Phonological conditioning, thus, implies providing a plausible phonological explanation from surrounding context for allomorphic variation. Consider now the following data (SIL 1980: H10) from Ilocano (Philippines):

1. tugawko	'my chair'	7. bagasko	'my rice'
2. tugawmo	'your chair'	8. bagasmo	'your rice'
3. tugawna	'his chair'	9. bagasna	'his rice'
4. sabak	'my banana'	10. sidak	'my food'
5. sabam	'your banana'	11. sidam	'your food'
6. sabana	'his banana'	12. sidana	'his food'

The nouns in this data are composed of a root followed by a 'possessive' suffix. Three possessive morphemes are represented in this data: '1st person singular', '2nd person', '3rd person singular masculine'. The '3rd person singular masculine possessive' morpheme is realised by a single morph /-na/ (Nos 3, 6, 9, 12). The other two possessive morphemes are each realised by two allomorphs, the '1st person singular possessive' by /-ko/ and /-k/, and the '2nd person possessive' by /-mo/ and /-m/. The allomorphs of both morphemes are phonologically conditioned in the same way: /-ko/ and /-mo/ occur after a root-final consonant, while /-k/ and /-m/ occur after a root-final vowel. A statement for the '1st person singular possessive' would be as follows:

$$\left\{\begin{array}{c}\text{1st person} \\ \text{singular possessive}\end{array}\right\} \rightarrow \begin{array}{ll} /k/ & \text{after root-final vowel} \\ /ko/ & \text{elsewhere}\end{array}$$

And similarly for the '2nd person possessive':

$$\left\{\begin{array}{c}\text{2nd person} \\ \text{possessive}\end{array}\right\} \rightarrow \begin{array}{ll} /m/ & \text{after root-final vowel} \\ /mo/ & \text{elsewhere}\end{array}$$

Exercise 17

Identify the allomorphs of the 'plural' and '2nd person possessive' morphemes in the following Turkish data (SIL 1980: H12), and state the phonological conditioning. (*Note*: the answer lies in the vowels.)

		plural	'your' + sing noun	'your' + pl noun
1. mum	'candle'	mumlar	mumun	mumlarɨn
2. kibrit	'match'	kibritler	kibritin	kibritlerin
3. yzym	'grape'	yzymler	yzymyn	yzymlerin
4. sɨnɨf	'class'	sɨnɨflar	sɨnɨfɨn	sɨnɨflarɨn
5. ders	'lesson'	dersler	dersin	derslerin
6. satʃ	'hair'	satʃlar	satʃɨn	satʃlarɨn
7. gøz	'eye'	gøzler	gøzyn	gøzlerin
8. top	'gun'	toplar	topun	toplarɨn
9. kuʃ	'bird'	kuʃlar	kuʃun	kuʃlarɨn
10. diʃ	'tooth'	diʃler	diʃin	diʃlerin
11. ok	'arrow'	oklar	okun	oklarɨn
12. køk	'root'	køkler	køkyn	køklerin

Morphological conditioning

Allomorphs of a morpheme, whose variation cannot be accounted for phonologically, are said to be 'morphologically conditioned'. That is to say, a particular allomorph of a morpheme occurs with an arbitrary set of other morphemes; the presence of one of that set determines or conditions the occurrence of the particular allomorph. For example, the allomorphs of the 'plural' morpheme in German, which we discussed in Chapter 3, are morphologically conditioned allomorphs: *-e*/ə/ occurs with noun morphemes like *Tag* 'day', *Mond* 'moon', *Schuh* 'shoe', etc, *-er* /ɐ/ occurs with noun morphemes like *Ei* 'egg', *Kind* 'child', *Licht*

'light', etc, and so on. Learners of German know that for each noun learnt as a vocabulary item they must also learn which plural allomorph it is associated with; apart from a few rules of thumb (eg feminine nouns ending in *e* usually take plural in *-n*), there is no general predictability, and certainly not of a phonological nature.

A similar kind of statement of allomorphic variation can be made for morphologically conditioned allomorphs as for phonologically conditioned ones, except that in the case of morphological conditioning a list of possible accompanying morphemes must be provided for each allomorph, rather than a general statement of conditioning. A statement for the 'plural' morpheme in German would have the following kind of form:

{plural} →

/s/ with the following noun stems:
Detail 'detail', *Hotel* 'hotel', *Team* 'team', etc

/ɐ/ with the following noun stems:
Brett 'board', *Kleid* 'dress', *Leib* 'body', etc

/ø/ with the following noun stems:
Haken 'hook', *Koffer* 'case', *Fenster* 'window', etc

/(ə)n/ with the following noun stems:
Biene 'bee', *Frau* 'woman', *Vetter* 'cousin', etc

/ə/ elsewhere, ie with the following noun stems:
Erfolg 'success', *Hund* 'dog', *Dieb* 'thief', etc

To avoid over-complicating this statement, no account has been taken of the operation of 'umlaut' on stem vowels with some of these suffixes in respect of specific noun stems.

Consider now the following data (SIL 1980: H2) from Hixkaryana (Brazil):

	Verb stem	Adjective	
1.	tawa-	tawanje	'dark'
2.	tak-	takdje	'warm'
3.	bata-	bataʃe	'rotten'
4.	tak-	takrje	'wet'
5.	tawas-	tawasdje	'light'
6.	tutʃu-	tutʃurje	'red'
7.	tatʃenot-	tatʃenotdje	'cold'

In this data, the 'adjectiviser' morpheme has four allomorphs, which are morphologically conditioned as follows:

{adjectiviser} →

/nje/ with the following verb stem:
tawa-

/ʃe/ with the following verb stem:
bata-

/rje/ with the following verb stems:
tak-, *tutʃu-*

/dje/ elsewhere, ie with the following verb stems: *tak-*, *tatʃenot-*, *tawas-*

Exercise 18

Identify the morphemes in the following Isthmus Zapotec (Mexico) data (SIL 1980: H5), and describe the allomorphs of the tense morphemes.

1.	rukaadu	'we write'	14.	ruyubibe	'he looks for'
2.	rukaabe	'he writes'	15.	zuyubibe	'he will look for'
3.	zukaabe	'he will write'	16.	biyubibe	'he looked for'
4.	bikaabe	'he wrote'	17.	kuyubibe	'he is looking for'
5.	kukaabe	'he is writing'	18.	rireebe	'he goes out'
6.	ruʒooñebe	'he runs'	19.	zareebe	'he will go out'
7.	zuʒooñebe	'he will run'	20.	bireebe	'he went out'
8.	biʒooñebe	'he ran'	21.	kareebe	'he is going out'
9.	kuʒooñebe	'he is running'	22.	ribanibe	'he wakes up'
10.	ridʒelabe	'he finds'	23.	zabanibe	'he will wake up'
11.	zadʒelabe	'he will find'	24.	bibanibe	'he woke up'
12.	bidʒelabe	'he found'	25.	kabanibe	'he is waking up'
13.	kadʒelabe	'he is finding'	26.	kabanidu	'we are waking up'

Mixed conditioning

Sometimes the allomorphs of a morpheme cannot be accounted for completely by either phonological conditioning or morphological conditioning alone; some of the allomorphs are found to be phonologically conditioned and some morphologically conditioned. This is, in fact, the case with the allomorphs of the 'plural' morpheme and of the 'past tense' morpheme in English. In our previous discussions of these morphemes we have considered only the 'regular' forms, ie /s/, /z/, /ɪz/ for 'plural' and /t/, /d/, /ɪd/ for 'past tense' (but cf Exercise 16). Consider, however, the following plurals of English nouns: *deer*, *mice*, *oxen*, *teeth*, *criteria*. The allomorph of the 'plural' in *deer* is zero /ø/, ie there is no visible or audible mark of plurality; in *mice* the allomorph is the stem vowel change /aʊ → aɪ/; in *oxen* it is the suffix /ən/; in *teeth* it is the stem vowel change /u → i) and in *criteria* it is the change taken over from Latin from /ən/ (*criterion* /kraɪtɪərɪən/) to /ə/ (/kraɪtɪərɪə/), ie in effect the loss of final /n/. These allomorphs are all morphologically conditioned; they apply to only a small number of noun stems, and there are more of a similar kind. Clearly, then, the allomorphs of the 'plural' morpheme in English represent an instance of mixed conditioning.

In making a statement of mixed conditioning, since the general principle is that allomorphs with the most restricted distribution are listed first and in mixed conditioning the phonologically conditioned allomorphs generally have the wider distribution, those that are morphologically conditioned are stated first. For example, a statement of the variants of the 'plural' morpheme in English would have the following kind of form:

{plural} →	/ən/	with the following noun stems: *ox*, *child*
	/øn/	with the following noun stems: *deer*, *sheep*, etc
	/aʊ→aɪ/	with the following noun stems: *mouse*, *louse*, etc
	/u→i/	with the following noun stems: *tooth*, *goose*, etc
	/ən→ə/	with the following noun stems: *criterion*, *phenomenon*, etc

/ɪz/ after stem-final sibilants
/s/ after stem-final voiceless consonants
/z/ elsewhere

Now consider the following Northern Tepehuan (Mexico) data (SIL 1980: H19):

	Singular	Plural
1. 'rabbit'	toʃi	totoʃi
2. 'man'	kʌli	kʌkʌli
3. 'foreigner'	obai	obai
4. 'tree'	uʃi	uʃi
5. 'son'	mara	mamara
6. 'stone'	obai	oxodai
7. 'friend'	aduni	aaduni
8. 'arrow'	uyi	uxuyi
9. 'turkey'	tova	totova
10. 'older brother'	ʃiʌgi	ʃiʃiʌgi
11. 'species of bird'	adatomali	aadatomali
12. 'needle'	oyi	oxoyi
13. 'younger brother'	sukuli	susukuli
14. 'species of fish'	aaʃi	aaʃi
15. 'rat'	dʌgi	dʌdʌgi
16. 'water jar'	ayi	axayi

There are three allomorphs of the 'plural' morpheme in this data. The allomorph with the widest distribution is the reduplication of the initial syllable (consisting of a Consonant + Vowel or just of a Vowel, ie/(C)V/), Nos 1, 2, 5, 7, 9, 10, 11, 13, 15: this allomorph can be regarded as phonologically conditioned. The one with the next widest distribution is the reduplication of the initial syllable consisting of just a Vowel with the addition of the consonant /x/, ie /Vx/: this must be regarded as morphologically conditioned, since the same initial vowel occurs as in the regular allomorph, viz /a/, cf Nos 7/14 with 16. The third allomorph is /ø/, Nos 3, 4 and 14, which must also be regarded as morphologically conditioned. Our statement for the 'plural' morpheme in this data might, then, be as follows:

{plural} → /ø/ with the following noun stems: *obai* 'foreigner', *uʃi* 'tree', *aaʃi* 'species of fish'
/Vx/ with the following noun stems: *odai* 'stone', *uyi* 'arrow', *oyi* 'needle', *ayi* 'water jar'
/(C)V)/ elsewhere

Exercise 19

Make a statement for the '3rd person singular masculine possessive' (*his*) and the '3rd person singular masculine reflexive possessive' (*his own*) morphemes in the following data (SIL 1980: H20) based on *Kaiwa* (Brazil).

1. iñāka	'his head'	oāka	'his own head'
2. inambi	'his ear'	onambi	'his own ear'
3. ipo	'his hand'	opo	'his own hand'
4. ijavati	'his corn'	oavati	'his own corn'
5. ijɨbɨ	'his land'	oɨbɨ	'his own land'

6.	iñāpeku	'his tongue'	oāpeku	'his own tongue'
7.	hiʔi	'his water'	oʔi	'his own water'
8.	hiʔabɨ	'his hair'	oʔabɨ	'his own hair'
9.	iñāmba	'his place'	oāmba	'his own place'
10.	ijape	'his shell'	oape	'his own shell'
11.	ijagwa	'his dog'	ojagwa	'his own dog'
12.	ijape	'his track'	gwape	'his own track'
13.	ijupa	'his lying place'	gwupa	'his own lying place'
14.	iñāta	'his home'	ŋwāta	'his own home'
15.	iñēmbiʔu	'his food'	ŋwēmbiʔu	'his own food'
16.	ijɨviri	'his younger brother'	gwɨviri	'his own younger brother'
17.	hiʔaŋgwe	'his shadow'	gwʔaŋgwe	'his own shadow'

9. What is a word?

'Word' an ambiguous term

Within linguistics, as indeed within ordinary language, the term 'word' is multiply ambiguous. We use it to refer to different kinds of unit. For example, I might say that this sentence contains eleven words. By 'word' here I would mean blocks of letters separated by spaces. A dictionary will claim to contain "15,000 words", and the term has changed its reference: for example, the items *sing*, *sings*, *singing*, *sang* and *sung* will count as only one of the 15,000 'words' in the dictionary. Or consider the case of 'homonyms', either spelt the same (ie 'homographs') or pronounced the same (ie 'homophones'); for example, *long* (opposite of *short*) and *long* ('desire strongly') are homonyms (both homographs and homophones), *bow* (as in "bow and arrow") and *bow* ('inclination of the head') are homographs but pronounced differently, while *bow* ('inclination of the head') and *bough* are homophones but not homographs. In each of these cases, we may ask whether they are the same word, and the answer will depend on the point of view that we take.

If we take the point of view of the written language, regarding words as combinations of letters (the 'orthographic' point of view), then *row* (ie $r + o + w$) represents one word, however many 'meanings' it might have. Similarly, from the point of view of pronunciation ('phonological'), regarding words as combinations of sounds (phonemes), then /baʊ/ (ie /b/ + /ɑ/ + /ʊ/) represents one word, however many spellings or meanings it might have. From the point of view of the dictionary, however, every item with a 'different meaning' (though that is open to interpretation) is a different word; so that *bow* (as in "bow and arrow"), *bow* ('inclination of the head') and *bough* are all separate words, as indeed are *bow* ('inclination of the head'—noun) and *bow* ('incline the head'—verb). However, as we noted earlier, *bows*, *bowing* and *bowed* would not be separate words in the dictionary, but regarded as 'forms' of the word *bow* (verb). Clearly, then, we need some new terminology to enable us to make the necessary distinctions that we have been discussing.

Word-forms

Following P H Matthews (*Morphology*, CUP 1979), we will refer to orthographic and phonological words as 'word-forms'. These are words viewed as combinations of letters or combinations of sounds, respectively. In languages that are well-established in the written medium, the question of what constitutes an orthographic word-form poses few problems: the convention of writing 'words' with spaces either side provides the answer, except in the case of emergent compounds. For example, is *bird bath* spelt thus as two orthographic word-forms, or hyphenated *bird-bath*, or together *birdbath*, the last two counting each as one word-form. Decisions of this

nature at the orthographic level are generally dependent on prior lexical or phonological decisions about word status.

Phonologically, the question of what is a word is considerably more fraught. In some languages, there are clear phonological indicators of word boundaries. In languages with a fixed stress, where the main (or primary) stress falls on the same syllable in every word, a clear word pattern emerges. Another such indicator, which may combine with that of fixed stress, is the syllable structure in a language. Some languages have a very simple syllable structure, which does not allow combinations of consonants within a syllable; in those which do allow consonant clusters, there are specifiable constraints on combinations. For example, in the sequence /bɔldtri/ "bald tree" in English, the boundary would have to be established between /d/ and /t/, because the only permissible consonant clusters are /ld/ syllable-finally and /tr/ syllable-initially. A few languages exhibit a further phonological feature that contributes to the establishing of phonological word-forms, a feature termed 'vowel harmony'. This feature, found for example in Turkish (see Exercise 17), means that within a given word all the vowels will be of a given type (eg back rounded), so that a new syllable with a vowel not of this type will indicate the beginning of a new word.

In spite of these kinds of phonological indicators, it is extremely difficult in some languages to establish word forms phonologically. French is a notorious example, with stress distributed at utterance rather than word level and with extensive contraction of 'grammatical' words. Consider the following sequence: /ʒənleʒamɛvy/ "je ne l'ai jamais vu" ('I have never seen him/her'), where it is difficult to apply any of the phonological criteria we have discussed.

Lexemes

The term that has been coined to refer to 'words' from the point of view of the dictionary (or lexicon) is 'lexeme', on the pattern of 'phoneme' and 'morpheme'. 'Words' in this sense correspond to the headwords of dictionary entries. So far we have assumed that a lexeme comprises a single word-form, but this is by no means always the case. One problematical area is that of compound words (cf Chapter 7). Many compounds are identifiable as a single word-form because they adopt phonological characteristics of ordinary words, eg a single primary stress in English or German: *dríftwood*, *Dańpfschifffahrt* 'stéamboat joúrney'. But other root + or combinations, that one would want to recognise as compounds semantically root grammatically, do not adopt these phonological characteristics, eg *sélf-séeker*, whose compound status is recognised orthographically by the hyphen. Other compounds, whose single-word status is confirmed phonologically are, nevertheless, written as two word-forms, eg *cáttle grid*, *cárrier pigeon*. Another group of lexemes in English that has single-word status phonologically but not orthographically are the prepositional and phrasal verbs, eg *look after*, *look up* (ie 'visit'), *look up to*—to which we shall return later.

Leaving compounds aside, there are other combinations of word-forms that constitute a single item lexically: these are 'fixed phrases', such as metaphors or idioms,

eg *on the straight and narrow*, *hammer and tongs*, *get the wrong end of the stick*. Characteristic of such expressions, which occur as items in many dictionaries and are thus to be treated as single lexemes, is that their meaning cannot be read off from the meaning of their constituents and that they have become institutionalised as 'fixed phrases' with idiosyncratic meanings. There is, however, in English at least, no clear boundary between sequences of words, whose meaning derives from the amalgamation of the meaning of each constituent word, and those fixed phrases whose meaning is not derivable in this way. For example, there are sequences of words that regularly occur together (collocations), of which one member only does not have its usual meaning, eg "a happy chance". *Happy* is normally predicated of people, and here the meaning is slightly different from its normal usage: do we then count "happy chance" as one lexeme or two? For any particular language there may well not be any clear-cut answers; lexicographers of English have varied in their judgements of what are to be counted as lexemes in the language. Clearly, though, there may be no direct one-to-one correspondence between lexeme and word-form in a language.

Words

We reserve the term 'word' for the grammatical level of analysis. Words are units of grammar, whose structure is described in terms of morphemes, and which enter into the structure of units at higher levels of grammatical patterning, eg phrases and clauses. The different forms of lexemes (eg *sing*, *sings*, *singing*, *sang*, *sung* as forms of SING) thus constitute separate words grammatically. Equally, homonyms like *skins* (plural of noun *skin*) and *skins* (3rd person singular present tense of verb *to skin*) are different grammatical words. So are the plural, the possessive, and the plural possessive, all represented by the pronunciation /gɛlz/ (but spelt *girls*, *girl's*, *girls'* respectively).

One classic definition of the 'word' regards it as "the minimal free form", ie the smallest unit that may occur on its own, eg as a one-word response to a question. Clearly, under this definition an inflectional morpheme would be excluded from word status, since it may only occur bound to a root morpheme or stem. However, such a definition also calls into question the status of many so-called 'grammatical' or 'form'-words, for example the articles (*a* and *the* in English). It would be unusual to hear *the* as a one-word response, except as a reply to a request for clarification ("Did you say *a* tiger or *the* tiger?"—"The"); but such replies may also comprise bound morphemes ("Did you say *dis*belief or *un*belief?"—"Dis").

The status of *the* as a word, as of 'form'-words generally, depends not on their ability to function as minimal responses, but on their detachability from other words and their relative freedom of occurrence. For example, *the* may occur directly before nouns ("the tiger"), before adjectives ("the ferocious tiger"), before quantifiers and numerals ("the many/seven ferocious tigers"), etc. If *the* is not to be regarded as free, and thus grammatically a word, it would be necessary to find some other class of word or morpheme to which it would be affixed: from our examples it is clear that this is not the case. An alternative solution would be to say

that *the* is neither a word nor a bound morpheme, but has some indeterminate status, eg as a 'particle'—of which more below.

We consider first of all whether the 'word' or the 'morpheme' is to be considered the smallest unit of grammar. The traditional division of grammar into morphology and syntax regards the word as pivotal in both: in morphology it is the unit whose structure is described in terms of morphemes; and in syntax it is the smallest unit in the analysis of sentences. However, different types of language will lead to alternative evaluations of the relative status of word and morpheme. In an 'isolating' language like Vietnamese, where there are no inflectional or derivational affixes, it may be questionable whether a distinction between word and morpheme is strictly necessary. In an 'agglutinating' language like Turkish, where each morph realises only one morpheme, so that every word can be segmented into units, each of which has a single 'meaning', it would seem sensible not to insist on a division between morphology and syntax, but to regard the morpheme as the smallest unit in the syntactic hierarchy. In a 'synthetic' language like Latin, however, where a segment of a word (a morph) may realise a number of 'meanings' (eg the *-a* of *puella* 'girl' realises 'nominative' case, 'singular' number, 'feminine' gender), the distinction between morphology (as the description of the grammar of the word) and syntax (the description of the grammar of the sentence) is essential. We have described the three classic types of language: few languages conform exclusively in their structure to one only of these types. Linguistic analysts, therefore, need to make judgements on the relative status of word and morpheme in the description of a particular language on the basis of the nature of that individual language.

Consider now the following data (SIL 1980: K1) from Kaiwa (Brazil). Word boundaries are not marked!

1. ɨwɨ'tuma	'The wind is already blowing (*lit.* 'It's winding')
2. ɨwɨ'tupo'rã	'lovely wind'
3. ipo'rãɨwɨ'tu	'The wind is lovely'
4. ipɨ'tãha'gwe	'Its feathers are red'
5. ha'gwepɨ'tãpo'rã	'Its feathers are a beautiful red'
6. iho'vɨha'gwe	'Its feathers are green'
7. ha'gweho'vɨpo'râ	'Its feathers are a beautiful green'
8. ha'gwema	'It already has feathers' (*lit.* 'It's feathering')
9. ipo'rãha'gwe	'Its feathers are beautiful'

In this data it is clear that some items occur in the same position with respect to other (kinds of) items, while for others their relative position is variable. In particular: *-ma*, which is a 'verbaliser' (deriving a verb from a noun), and *i-*, also a 'verbaliser' (but deriving a verb from an adjective), must be regarded as bound morphemes. The other items, however, which we establish as items on the basis of their non-interruptibility, must be regarded as words, since they are mobile items: *ha'gwe* 'feathers', *ɨwɨ'tu* 'wind'; *ho'vɨ* 'green', *pɨ'tã* 'red', *po'rã* 'lovely/beautiful'.

Particles

We now take up the question of items which are clearly not morphemes, because of their relative freedom of occurrence, but which do not seem to attain to full word

status. We earlier cited the question of *the* in English. A further example from English would be the so-called 'adverb particles', which combine with lexical verbs to form 'phrasal verbs', eg *look up* (= 'visit'), *slow down*, *give in*, *knock out*, *bring on* (= 'induce'), *take off* (= 'imitate'). These are regarded as single lexemes (eg in the *Longman Dictionary of Contemporary English*), but they are not compounds, because the adverb particle may be separated from the lexical verb (eg "the boxer knocked him out"), neither is the adverb particle an affix morpheme, for the same reason. But it is not a free word, since it is clearly part of the verb lexeme; and like the article, an adverb particle may not normally be the locus of intonational prominence in speech.

Another interesting case in English is that of the items *more* and *most*, which along with the suffixes *-er* and *-est* are 'allomorphs' of the 'comparative' and 'superlative' morphemes respectively. They are not affix morph(eme)s, however, since they may occur before words other than adjectives, eg nouns (cf "more difficult", "more money"; "most decisive", "most respect"). Should we, then, perhaps regard them as particles, since they are like affix morph(eme)s in respect of the inflections of the adjective, but like words in their general behaviour in the sentence?

In German, a set of word-forms called 'modal particles' occurs, including *doch*, *ja*, *wohl*, *(ein)mal*. They never receive any intonational prominence, have a generally 'emphatic' function, and are frequently not translatable (into English, at least!). For example: "Sie wissen *doch*, wie das so ist" ('You know how it is, don't you?'), "Das kann *wohl mal* vorkommen" ('That might well happen').

If we define particles as items that do not have bound morpheme status, but are not clearly words because they are generally unable to function as minimal responses, we ought probably to include prepositions and conjunctions within the category of particle. Prepositions, after all, are often functionally equivalent to inflectional affixes of synthetic languages, cf Latin *urbe* 'from the city', Old High German *swertu* 'with a sword'.

Clitics

Particles are items that have a marginal status as words. Clitics are items that have a marginal status as morphemes. That is to say, they clearly form part of another word, eg on phonological grounds, but they are moveable and do not regularly form part of one particular class of words. A candidate for the status of clitic in English is the 'possessive' suffix, which we generally regard as being an inflectional suffix of the noun. It does not, however, always occur on the noun: if the noun is postmodified (see Chapter 13), the 'possessive' clitic is attached to the last item in the noun phrase; cf "the president*'s* proposals", "the president of the United States of America*'s* proposals", "the president who addressed the nation yesterday*'s* proposals". The last of these examples is perhaps less likely to occur; in such cases the alternative formulation with *of* is probably preferred ("the proposals of the president, who addressed the nation yesterday").

Also a candidate for clitic status is the Latin item *-que* 'and', which only occurs as part of another word, but which may be affixed to almost any word, eg "pueri puellaeque" 'boys and girls', "*vir mulierque*" 'a man and a woman'.

Consider now the following data (SIL 1980: K11) from Choctaw (USA):

1. alatik-mat talowa — 'the girl sings'
   girl sings
2. alatik osi-mat talowa — 'the little girl sings'
   girl little sings
3. alatik osi taθapi-mat talowa — 'the four little girls sing'
   girl little four sing
4. hatak-mat abilini-ma pilatok — 'the man threw the chair'
   man chair threw
5. hatak-mat abilini chite-ma pilatok — 'the man threw the big chair'
   man chair big threw
6. hatak-mat abilini-ma ayimpa-pito pilatok
   man chair table-on threw
   'the man threw the chair on the table'
7. hatak-mat abilini-ma alatik osi-pito pilatok
   man chair girl little-to threw
   'the man threw the chair at the little girl'

Here the items *-mat*, *-ma* and *-pito* are to be regarded as clitics; they are affixed to stems, but not always to the same stem. The item *-mat* is a subject-marker and it occurs on the last item of the noun phrase functioning as subject, whether that is a noun (No 1), an adjective (No 2), or a numeral (No 3). The item *-ma* is an object marker and occurs similarly on a noun (No 4), or an adjective (No 5). The item *-pito* is a postposition, occurring on the last item of the noun phrase governed by it, a noun (No 6) or an adjective (No 7).

We must conclude, therefore, that the descriptive status of words, morphemes and any other intermediate unit (particles and clitics) must depend on the nature of the individual language and on the considered judgement of the linguistic analyst describing that language.

Exercise 20

Consider which items in the following Tlingit (Alaska) data (SIL 1980: K9) might be regarded as particles or clitics.

1. ax hidee-t oowagut — 'He went to (and arrived at) my house'
   my house-poss-to he-went
2. ax hidee tlen-t oowagut — 'He went to my big house'
   my house-poss big-to he went
3. ax hidee-dei ʼwoogoot — 'He went to (towards) my house'
   my house-poss-to he-went
4. ax hidee tlen-dei woogoot — 'He went to my big house'
   my house-poss big-to he-went
5. doo aat-gaa woogoot — 'He went for (to fetch) his aunt'
   his aunt-for he-went
6. doo aat has-gaa woogoot — 'He went to fetch his aunts'
   his aunt pl.-for he-went

7.	ax aat nak ax hide-dei woogoot my aunt without my house-to he-went	'He went to my house without my aunt'
8.	doo aat has nak woogoot his aunt pl. without he-went	'He went without his aunts'
9.	tleil doo aat teen woogoot neg. his aunt with he-went	'He didn't go with his aunt'
10.	doo aat has teen woo.aat his aunt pl. with he-went	'He went with his aunts'

10. Sentences (1)

Nouns and verbs

Sentences may be viewed essentially as the combinations of nouns and verbs. More precisely, a sentence may be viewed as the combination of a verb with one or more nouns. We mentioned earlier that nouns refer to the 'things' of our experience (persons, objects, ideas, feelings, etc), and that verbs refer to 'events' (including actions, processes and states). Sentences, then, are concerned with the myriad ways in which the 'things' of experience are involved in the 'events' of experience, or rather with the myriad ways in which we are able to talk about them.

Consider the following sentences of English:

1. Horses eat hay.
2. The committee has awarded the novelist first prize.
3. The heckler argued with the speaker about his ideas.
4. The children played in the garden after tea.

The first of these sentences contains just one verb and two nouns: the verb refers to an 'action', and the nouns to the 'things' involved in the action. In the second sentence, the structure still essentially comprises a verb and nouns (viz *award*, *committee*, *novelist*, *prize*): the additional words are dependent on or modify these main items. Similarly in the third sentence, where the verb and nouns are *argue*, *heckler*, *speaker*, *ideas*; except that here, besides 'modifying' words like *the* and *his*, there are also 'relational' words (prepositions) like *with* and *about*. The fourth sentence (in which the verb and nouns are *play*, *children*, *garden*, *tea*) also has prepositions (*in*, *after*) in addition to modifiers, but here the prepositions have a spatial and temporal reference respectively, and are substitutable. The function of the prepositions in No 4 is thus different from that of the prepositions in No 3: you can only argue *with* somebody *about* something; but you can play *in*, *on*, *outside*, etc somewhere, *after*, *before*, *at*, etc, some point in time. But the basic point remains: sentences are essentially structures comprising a verb and a number of nouns.

The verb is central

A useful way of approaching the analysis of sentence structure is to regard the verb as the central element which determines how many (and which kind of) nouns may be present in a sentence, including any non-substitutable relational items (eg prepositions and case inflections). This function of the verb as the determining element in a sentence derives from a 'dependency' view of sentence structure (see Chapter 22). Which other elements a verb determines is attributable to the 'meaning' of the verb. For example, we implied earlier that the verb *argue*

determines the presence of three nouns: *someone* argue with *someone* about *something*. In fact, our English specification allows us to indicate additionally that two of the nouns refer to persons and one to a non-person. Clearly, this kind of statement has a good measure of generality, since other kinds of sentence structure occur with *argue* that make some refinements of this statement necessary; eg "The boys are arguing", "He is arguing with his sister", "They are arguing about the rules of the game". Two further points need to be noticed about *argue*: firstly, it may be used to refer generally to the action, when only the participants in the argument are mentioned, but not the 'something' argued about; secondly, the participants in the argument may be collected together as one (plural) noun (*the boys*, *they*), with or without mention of the 'something' argued about.

Consider another example: *buy*. This verb could be said to determine the following nouns: *someone* buy *something* from *someone/somewhere* for *some amount*, eg "Bill bought the radio from the corner shop for £10". We might add: at *some time*, eg "the day before yesterday"; though we might regard this as not being strictly determined by the verb *buy*, since almost any event may be specified for point of time. A refinement of this general statement for *buy* would have to indicate that not all of the nouns need to be specified, ie that the someone/somewhere from which the thing is bought and the amount for which it is bought may be omitted (eg "Jack has bought a new car", "Bill bought a radio for £10", "Harry buys his newspapers from the corner shop").

English does not have case marking for sentence elements. We will now consider a couple of examples from German, which does mark case (though as we have seen—Chapter 3—not in the noun itself). Our first example is *geben* 'give', which determines as follows: *someone* (nominative case) geben *someone* (dative case) *something* (accusative case), eg "Der Mann hat seiner Frau einen Diamanten gegeben" 'The man gave his wife a diamond'. Our second example from German is *anklagen* 'accuse', which determines as follows (*wegen* is a preposition often translated by 'because of' in English): *someone* (nominative case) anklagen *someone* (accusative case) wegen *something* (genitive case), eg "Die Polizei hat den Gefangenen wegen des Diebstahls angeklagt" 'The police accused the prisoner of theft'.

Exercise 21

Make 'determining' statements for the following English verbs, adding any 'refining' details you think necessary:

1. write
2. win
3. spend
4. wait
5. invite
6. clean
7. fight
8. report

Sentence types

From the limited number of examples that we have looked at so far in this chapter, it has probably not been apparent that for any language a limited number of patterns of sentence structure occurs. That is to say, it is possible to make generalis-

ations about the types of sentence patterns that occur in a language, many of which are applicable across languages.

Traditionally, a distinction has been made between 'transitive' and 'intransitive' sentences. Intransitive sentences have only one noun directly involved in the event: the 'doer' of an action, or the 'undergoer' of a process; eg "The children are laughing", "The tree is falling". In transitive sentences, more than one noun is directly involved in the event, though usually not more than three. The second noun usually refers to the 'undergoer' in an action, eg "The goalkeeper caught the ball". The third noun, in what is sometimes called a 'ditransitive' sentence, usually refers to the 'recipient' or 'beneficiary' involved in the action, eg "The shopkeeper sent the customer a bill". In German, and other languages with case systems, these different nouns are identifiable by their case marking: in German, the first noun (ie the one occurring alone in intransitive sentences, and the 'doer' in transitive ones) is in the nominative case, the second noun is in the accusative case, and the third noun is in the dative case.

Two further sentence types commonly found are 'stative' and 'equative' sentences, both of which are characterised by the presence of the verb *be* in English (though this is not the only verb which may occur), eg

Stative:	"The leaves are brown".
Equative:	"Harry is the ship's doctor".

In the case of stative sentences, the second (non-verbal) element is often—as in our example—an adjective rather than a noun; though we could equally well say "The leaves are a brown colour", where the second element is now a noun (*colour*), modified by an adjective (*brown*). In stative sentences the adjective or second noun describes (the state of) the first noun. In equative sentences, as the term suggests, there is an equational relationship between the two nouns; their order is often reversible, eg we could have said for our example "The ship's doctor is Harry", though this would have a different contextual meaning. In languages that have case inflections (eg Latin, German), the second element is often in the same case as the first noun (eg nominative case), cf German "Dieser Mann ist der Schiffsarzt" 'This man is the ship's doctor'.

Our list of sentence types is by no means exhaustive. Each language must be considered on its merits, on the basis of the possible generalisations from occurring sentence patterns. Consider now the following data from Kolami (India):

1. amd mangten he slept	'he slept'
2. add ellang seddin she house went	'she went to the house'
3. amd aambal tinden he rice ate	'he ate rice'
4. amd anung pustok siyten he me book gave	'he gave me a book'
5. ramak poḍam aṇḍan Ram tall is	'Ram is tall'

Here we can clearly recognise an intransitive sentence (No 1), with just one noun and an action verb; a transitive sentence (No 3), with two nouns and an action verb; and a ditransitive sentence (No 4), with three nouns. No 5 falls into the stative (or 'descriptive') type, with a noun and an adjective describing the noun. No 2 is somewhat of a problem, since it looks like a transitive sentence, with two nouns; but the action verb *go* does not involve the second noun (*ellang* 'house') in the same way that *tinden* 'ate' involves *aambal* 'rice', ie the house does not have something done to it. Rather, the house is the place or the location, towards which the action is directed. Such sentence types, with a location as the second noun but structured like a transitive sentence, are often called 'semi-transitive' sentences; they occur typically in North-Indian languages.

Exercise 22

Identify the sentence types in the following Engenni (Nigeria) data (SIL 1980: E7):

1. edei nɛ adidɛ man the rich-man	'The man is a rich man'
2. Ade misɛ n'eki slept in-market	'Ade slept in the market'
3. Ade doriya is-tall	'Ade is tall'
4. edei nɛ mɔni ozhi avu n'ukwo man the saw thief one at-farm	'The man saw a thief at the farm'
5. Ade do eseni stole fish	'Ade stole a fish'
6. eseni kuya fish is-scarce	'Fish is scarce'
7. edei nɛ fina n'ɔkɔ man the enter in-canoe	'The man entered the canoe'
8. adidɛ nɛ wu rich-man the died	'The rich man died'
9. edei dori nɛ Ade man tall the	'The tall man is Ade'

Sentences without verbs

In the previous exercise, two of the examples were sentences without a verb (Nos 1 and 9), which would seem to contradict our earlier statement that a sentence was constituted by a verb and one or more nouns. In fact, that statement does not have general validity because of one other kind of sentence: one that contains only a verb, eg Latin "Pluit" 'It is raining'. But the more interesting cases of the statement not being valid are those where there is no verb. In the Engenni data (Exercise 22), the sentences without verbs are equative sentences. And we find that it is not unusual for equative and stative/descriptive sentences to be verbless (where in English the verb *be* would be used).

Consider the following examples from Western Desert (Australia):

1. wati mirpanpa man angry	'The man is angry'
2. wati ngalyayala doctor	'The man is a doctor'

Similarly in Hixkaryana (Brazil) for equative sentences, eg

1. Hoñko woto wild-pig game	'The game is/was wild pig'
2. Towankaɾ̃hemɨ anaro komo courageous-one other group	'Others were courageous'

And in Tlingit (Alaska) there is a locative sentence that is verbless, eg

1. ee xòonee nèilu your friend inside-LM	'Your friend is at home'
2. aatlèin dáanaa doo jèewu lots money his hand-LM	'He has lots of money'

(LM = locative marker, ie an inflection signalling location).

Exercise 23

Identify the sentence types of the examples in the following data (SIL 1980: E8) based on Pocomchi (Guatemala):

1. huan ʃokik pan ripat John entered in his-house	'John entered his house'
2. re huan ʃilwatʃ rasbes recognised older-brother	'John recognised his older brother'
3. re ribahil maria huan her-husband Mary John	'John is Mary's husband'
4. riʃq'un na rik'uʃum tulul her-daughter will eat bananas	'Her daughter will eat bananas'
5. re tulul ʃirix pat the bananas behind house	'The bananas are behind the house'
6. re kinaq' wiki beans there	'The beans are there'
7. re maria re riʃq'un Mary her-daughter	'Her daughter is Mary'
8. re be inox pan tinamit road goes in town	'The road goes into town'
9. ripat ʃirix k'aybal her-house behind market	'Her house is behind the market'
10. maria na ri'ayem tulu pan tinamit Mary will sell bananas in town	'Mary will sell bananas in town'