

# Morphology

Francis Katamba



## 1 Introduction

### 1.1 THE EMERGENCE OF MORPHOLOGY

Although students of language have always been aware of the importance of words, **morphology**, the study of the internal structure of words did not emerge as a distinct sub-branch of linguistics until the nineteenth century.

Early in the nineteenth century, morphology played a pivotal role in the reconstruction of Indo-European. In 1816, Franz Bopp published the results of a study supporting the claim, originally made by Sir William Jones in 1786, that Sanskrit, Latin, Persian and the Germanic languages were descended from a common ancestor. Bopp's evidence was based on a comparison of the grammatical endings of words in these languages.

Between 1819 and 1837, Bopp's contemporary Jacob Grimm published his classic work, *Deutsche Grammatik*. By making a thorough analytical comparison of sound systems and word-formation patterns, Grimm showed the evolution of the grammar of Germanic languages and the relationships of Germanic to other Indo-European languages.

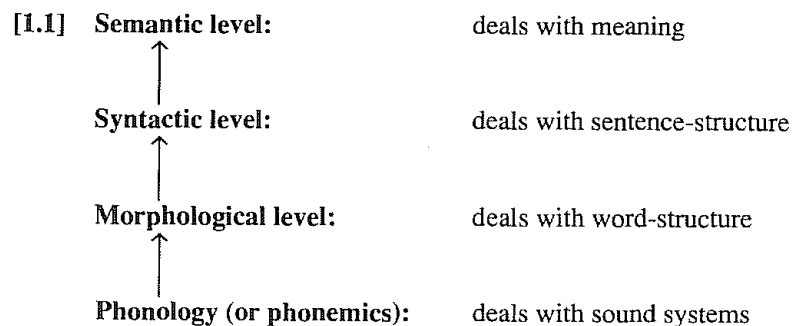
Later, under the influence of the Darwinian theory of evolution, the philologist Max Müller contended, in his Oxford lectures of 1899, that the study of the evolution of words would illuminate the evolution of language just as in biology morphology, the study of the forms of organisms, had thrown light on the evolution of species. His specific claim was that the study of the 400–500 basic roots of the Indo-European ancestor of many of the languages of Europe and Asia was the key to understanding the origin of human language (cf. Müller, 1899; cited by Matthews, 1974).

Such evolutionary pretensions were abandoned very early on in the history of morphology. In this century morphology has been regarded as an essentially **synchronic** discipline, that is to say, a discipline focusing on the study of word-structure at one stage in the life of a language rather than on the evolution of words. But, in spite of the unanimous agreement among linguists on this point, morphology has had a chequered career in twentieth-century linguistics, as we shall see.

### 1.2 MORPHOLOGY IN AMERICAN STRUCTURAL LINGUISTICS

Adherents to **American structural linguistics**, one of the dominant schools of linguistics in the first part of this century, typically viewed linguistics not so much as a 'theory' of the nature of language but rather as a body of

descriptive and analytical procedures. Ideally, linguistic analysis was expected to proceed by focusing selectively on one dimension of language structure at a time before tackling the next one. Each dimension was formally referred to as a **linguistic level**. The various levels are shown in [1.1].



The levels were assumed to be ordered in a hierarchy, with phonology at the bottom and semantics at the top. The task of the analyst producing a description of a language was seen as one of working out, in separate stages, first the pronunciation, then the word-structure, then the sentence-structure and finally the meaning of utterances. It was considered theoretically reprehensible to make use of information from a higher level, e.g. syntax, when analysing a lower level such as phonology. This was the doctrine of **separation of levels**.

In the early days, especially between 1920 and 1945, American structuralists grappled with the problem of how sounds are used to distinguish meaning in language. They developed and refined the theory of the **phoneme** (cf. Sapir, 1925; Swadesh, 1934; Twaddell, 1935; Harris, 1944).

As time went on, the focus gradually shifted to morphology. When structuralism was in its prime, especially between 1940 and 1960, the study of morphology occupied centre stage. Many major structuralists investigated issues in the theory of word-structure (cf. Bloomfield, 1933; Harris, 1942, 1946, 1951; Hockett, 1952, 1954, 1958). Nida's coursebook entitled *Morphology*, which was published in 1949, codified structuralist theory and practice. It introduced generations of linguists to the descriptive analysis of words.

The structuralists' methodological insistence on the separation of levels which we noted above was a mistake, as we shall see below in sections (1.3.2) and (1.3.3). But despite this flaw, there was much that was commendable in the structuralist approach to morphology. One of the structuralists' main contributions was the recognition of the fact that words may have intricate internal structures. Whereas traditionally linguistic analysis had treated the word as the basic unit of grammatical theory and lexicography,

the American structuralists showed that words are analysable in terms of **morphemes**. These are the smallest units of meaning and grammatical function. Previously, word-structure had been treated together with sentence-structure under grammar. The structuralists introduced morphology as a separate sub-branch of linguistics. Its purpose was 'the study of morphemes and their arrangements in forming words' (Nida, 1949:1). The contribution of the structuralists informs much of the discussion in the first part of this book.

### 1.3 THE CONCEPT OF CHOMSKYAN GENERATIVE GRAMMAR

The bulk of this book, however, presents morphological theory within the linguistic model of **generative grammar** initiated by Chomsky. Before we begin considering how this theory works, I will sketch the background assumptions made by generative grammarians so that we can place the theory of morphology in the wider theoretical context of generative linguistics.

The central objective of generative linguistics is to understand the nature of **linguistic knowledge** and how it is acquired by infants. In the light of this objective, a fundamental question that a theory of word-structure must address is, 'what kinds of information must speakers have about the words of their language in order to use them in utterances?' Attempts to answer this question have led to the development of sub-theories of the **lexicon** (i.e. dictionary) and of morphology.

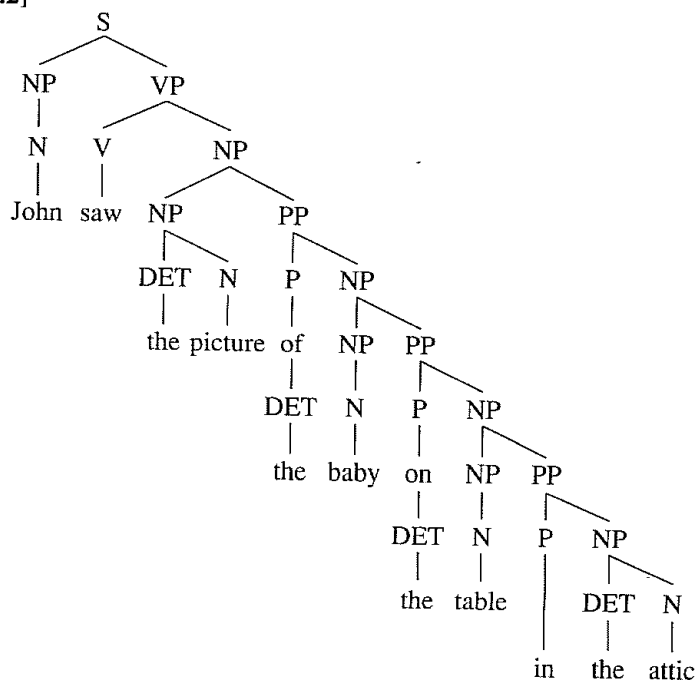
According to Chomsky (1980, 1981, 1986), the central goal of linguistic theory is to determine what it is people know if they *know* a particular language. Chomsky observes that knowing a language is not simply a matter of being able to manipulate a long list of sentences that have been memorised. Rather, knowing a language involves having the ability to produce and understand a vast (and indeed unlimited) number of utterances of that language that one may never have heard or produced before. In other words, **creativity** (also called productivity or open-endedness) is an aspect of linguistic knowledge that is of paramount importance.

Linguistic creativity is for the most part rule-governed. For instance, speakers of English know that it is possible to indicate that there is more than one entity referred to by a noun and that the standard way of doing this is to add *-s* at the end of a noun. Given the noun *book*, which we all have encountered before, we know that if there is more than one of these objects we refer to them as *books*. Likewise, given the nonsense word *smilts* as in the sentence *The smilts stink* which I have just made up, you know *smilts* would refer to more than one of these smelly things. Speakers

of English have tacit knowledge of the rule which says 'add -s for plural' and they can use it to produce the plural form of virtually any noun. I have emphasised the notion of rule, taking the existence of rules for granted.

I will now explain why a **generative grammar** is a system of explicit rules which may apply recursively to **generate** an indefinite number of sentences which can be as long as one wants them to be. **Recursiveness** has the consequence that, in principle, there is no upper limit to the length of sentences. A grammatical constituent like a noun phrase (NP) or a prepositional phrase (PP) can contain an indefinite number of further constituents of that category as in the sentence *John saw the picture of the baby on the table in the attic*. The recursion can be seen clearly in the tree diagram representing that sentence in [1.2]. As seen, NPs can contain NPs and PPs which in turn contain NPs which can contain NPs and PPs:

[1.2]



Notes: S – sentence; N – noun, NP – noun phrase; V – verb, VP – verb phrase; P – preposition, PP – prepositional phrase; DET – determiner.

One of our concerns will be to determine whether morphology should be recognised as a separate linguistic level (or **module**) that is independent of syntax and phonology (see [1.1] above and [1.3] below). Do morphological rules have certain properties which they do not share with rules in other parts of the grammar? Are recursive rules of the kind found in syntax

needed in morphology? This book will address these issues in depth. Here I will only attempt to give you a flavour of one of the issues that I will be exploring.

There are morphological processes which are similar to syntactic processes. For instance, certain adjectives which describe periods in history, such as *industrial*, can have the prefix *post-* before them as in *post-industrial*. And, given the adjective *post-industrial*, we can place another *post-* before it to yield *post-post-industrial*. Clearly, the word-formation process we witness here is recursive. We have the rule attaching *post-* to a word reapplying to its own output. This raises an interesting question: if morphological rules that build words are similar to syntactic rules that build sentences, what reason is there for assuming that morphology is essentially different from syntax?

Before we go any further we need to clarify the terms **grammar** and **rule of grammar**. These terms are used by linguists in four distinct senses. Firstly, in generative linguistics 'grammar' can refer to the implicit, totally unarticulated knowledge of rules and principles of their language that people have in their heads. This tacit knowledge enables them to distinguish between well-formed and ill-formed words and utterances in their language. For example, many English speakers may not be able to explain in an articulate manner why it is 'correct' to say *a grain* but 'incorrect' to say *a oat*. Nevertheless their knowledge of English grammatical structure enables them to determine that the former is correct and the latter is not.

Secondly, whereas in traditional approaches 'grammar' only includes morphology and syntax, in generative linguistics the term grammar is employed in a much wider sense. It covers not only morphology and syntax but also semantics, the lexicon and phonology. Hence, there are rules of grammar in every linguistic module. Phonological rules, morphological rules, syntactic rules and semantic rules are all regarded as rules of grammar.

Thirdly, grammar and rules of grammar may refer to a book containing a statement of the rules and principles inferred by linguists to lie behind the linguistic behaviour of speakers of a particular language. These rules simply describe regular patterns observed in the linguistic data.

Lastly, some grammars are books containing prescriptive statements. Such grammars contain rules that *prescribe* certain kinds of usage. Outside linguistics this view of grammar is still prevalent. The reason for this is clear. In everyday life rules are normally mechanisms for regulating behaviour – the behaviour of pupils in a school, members of a club, inmates of a prison, etc. In many traditional pedagogical grammars rules serve the same purpose. They are statements like 'A sentences must not end with a preposition.' They prescribe what the 'officially or socially approved' usage is – in the opinion of the grammarian.

In much of modern linguistics, however, rules have a different function.

4 They are not prescriptions of behaviour which the grammarian imposes on speakers, but rather they are statements of principles responsible for the observed regularities in the speech or writing or users of a particular language. The characterisation of regularities in observed patterns of usage is what the American structuralists regarded as the primary objective of linguistic investigations. Their grammatical rules were descriptive statements like 'The article precedes the noun in the English noun phrase.' This statement reflects the fact that *the book*, as in *I read the book*, is allowed whereas *\*book the*, as in *\*I read book the* is disallowed. (An asterisk indicates a disallowed form.)

Chomsky has shifted the focus of linguistic theory from the study of observed behaviour to the investigation of the knowledge that underlies that behaviour. In generative linguistics rules are intended to go beyond accounting for patterns in the data to a characterisation of speakers' linguistic knowledge. The primary objective of generative grammar is to model a speaker's linguistic knowledge.

Chomsky characterises linguistic knowledge using the concepts of competence and performance. **Competence** is a person's implicit knowledge of the rules of a language that makes the production and understanding of an indefinitely large number of new utterances possible while **performance** is the actual use of language in real situations. Chomsky proposes that competence, rather than performance, is the primary object of linguistic inquiry. Put simply, knowledge of a language entails mastery of an elaborate system of rules that enables a person to encode and decode a limitless number of utterances in that language. One sub-set of this rule system is the rules of word-formation which this book introduces you to. In section (4.1.3) of Chapters 4 and section (12.3.3) of Chapter 12 it will be shown that speakers of a language do not just commit to memory all the words they know. Their competence includes the ability to manipulate rules in order to create new words and to unscramble the meanings of novel or unfamiliar words which they encounter.

If knowing a language essentially involves mastering a system of rules, how do humans accomplish this task? Chomsky contends that the linguistic capacity of humans is **innate**. The general character of linguistic knowledge is determined by the nature of the mind which is endowed with a specialised **language faculty**. This faculty is determined in turn by the biology of the brain. The human child is born with a blue-print of language which is called **Universal Grammar**.

According to Chomsky, Universal Grammar is the faculty of the mind which determines the nature of language acquisition in the infant and of linguistic competence. The properties that lie behind the competence of speakers of various languages are governed by restricted and unified elementary principles rooted in Universal Grammar. This explains the striking underlying similarity between languages in their essential struc-

tural properties. Admittedly, languages differ from each other, but the structural differences between them occur within the fairly narrow range sanctioned by Universal Grammar. As we shall see (especially in Chapters 3, 8, 9 and 12) with regard to word-formation, very similar word-building principles recur in language after language. The language faculty of the mind is essentially the same in all humans. Hence languages can only differ from each other within the limits predetermined by the neurology and physiology of the human brain, which determine the nature of Universal Grammar. And Universal Grammar in turn determines the kinds of grammars of particular languages that can be acquired by infants.

The differences between the grammars acquired by individual speakers of, say, English and Arabic can be attributed to experience. An individual's experience serves to specify a particular grammar for the particular language which that individual is exposed to – within the range permitted by Universal Grammar.

How is Universal Grammar structured? It is **modular** in structure: it consists of various sub-systems of principles. Many of its principles consist of **parameters** which are fixed by experience on the basis of simple evidence of the kind available to the child. Chomsky compares Universal Grammar to an intricate electrical system that is all wired up, but not switched on. The system contains a finite set of switches, each one of which has a restricted number of positions. Exposure to a specific language experience is required to turn on these switches and give them the appropriate setting.

The basic idea of parameters is meant to capture the fact that many rules are interdependent. If one choice is made, it may either preclude some other choices or set in motion other related choices. This makes the task of language acquisition simpler than it would be if each rule had to be worked out independently of all other rules. The parametric approach assumes that the infant acquiring a language makes very clever guesses or hypotheses about the rules of the grammar being acquired on the basis of rules already acquired after experience of a particular language.

For a concrete example of a parameter, we will consider the **Right-hand Head/Left-hand Head Rule** which will be discussed in Chapter 12. This parameter is concerned with the position of the **head** of a grammatical constituent. Some languages, like English, normally place the head on the right, i.e. it is the last element of a constituent. For example, in the noun phrase *these big books* the right-handmost word, the noun *books*, is the head. It must come last. (Alternatives like *\*books big these* and *\*these books big* are forbidden.)

As a rule, the head is the only obligatory element of a constituent like an NP. *Books* is a well-formed NP but neither *these* nor *big* is a permissible NP on its own. Furthermore, in terms of meaning, the head *books* is the key

word in this NP. The function of *these* and *big* is merely to specify further the particular books referred to.

Likewise, at word level, in a compound like *farmhouse*, the head, *house*, is the last element and it is the pivotal element from a semantic point of view. (A *farmhouse* is a kind of *house*.) However, in some languages, such as Japanese, the reverse is the case. The head of a grammatical constituent is normally on the left. Once an infant has worked out the position of the head for one construction this can be generalised with a considerable degree of success to other constructions.

Universal Grammar consists of a number of modules which are inter-related. This is shown in [1.3] (which you should compare with [1.1] above):

- [1.3] (i) Lexicon and Morphology  
 (ii) Syntax  
 (iii) Phonetic Form (PF) (which deals with representation of utterances in speech)  
 (iv) Logical Form (LF) (which deals with meaning)

As seen, Universal Grammar includes the lexicon and morphology module. Knowledge of word-structure is a central aspect of linguistic competence. A case can be made for recognising morphology as a separate module of Universal Grammar. Yet at the same time, morphology (and the lexicon) are like a bridge that links the other modules of the grammar. It is therefore necessary to examine morphology not in isolation, but in relation to the other modules. Morphology interacts with both phonology and syntax as well as semantics. So, it can only be studied by considering the phonological, syntactic and semantic dimensions of words.

### 1.3.1 The Place of Morphology in Early Generative Grammar

Today the place of morphology in generative grammar is secure. But this is a recent development. After being in the limelight when structuralism peaked in the 1950s, morphology was at first eclipsed when generative grammar came on the scene. Generative grammarians initially rejected the validity of a separate morphological module.

From the point of view of advancing our understanding of word-structure, this stance was unfortunate. Since generative grammar has been the dominant school of linguistics in the second half of this century, it meant that the study of word-structure was in the shadows for more than a decade. Morphology did not re-emerge from oblivion until the mid-1970s. Fortunately, the eclipse was not total. A few isolated (for the most part non-generative) scholars such as Robins (1959) and Matthews (1972, 1974)

made important contributions to morphology during this time, as we shall see.

Part of the reason for the widespread neglect of morphology during the early years of generative grammar was the belief that word-formation could be adequately covered if it was partitioned between phonology and syntax. It was argued that no separate morphological level or component was needed in the grammar. Ways were found of describing the structure of words in a model of language that had a phonological component, a syntactic component and a semantic component but no morphological component. Those aspects of word-structure that relate to phonology (e.g. the alternation between *sane* [seɪn] and *sanity* [sæniːti]) would be dealt with using devices found in the phonological component. And those aspects of word-structure that are affected by syntax would be dealt with in the syntactic component.

The job of the syntactic component of the grammar was thought of as being to **generate** (i.e. to specify or enumerate explicitly) all the well-formed sentences of a language, without generating any ill-formed ones. Significantly, generating all the sentences of a language was seen as meaning generating all the permissible sequences of **morphemes** (not words), and showing which morpheme groupings formed syntactic constituents like noun phrases and verb phrases (also see p. 13 in this chapter). A specialised morphological component and a properly articulated lexicon were not part of the picture. Thus, Lees (1960), the first major descriptive study produced by a generative linguist, used syntactic rules to create derived words like the noun *appointment* from the verb *appoint*. As seen in [1.4a], Lees derived the sentence containing the noun *appointment* from a source sentence with the verb *appoint*. Likewise, he derived the abstract noun *priesthood* from a source sentence with the noun *priest*, as indicated in [1.4b].

- [1.4] a. The committee appoints John.  
 The committee's appointment of John.  
 (Source sentence: Lees, 1960: 67)
- b. John is a priest.  
 John's priesthood. (Source sentence: Lees, 1960: 110)

We will not examine the particulars of the syntactic rules which Lees uses. Our concern is that Lees saw this type of word-formation as taking place in the syntax and believed that he could dispense with morphology. We will revisit this issue in Chapter 12.

Let us now turn our attention to questions of phonological realisation. **Readjustment rules** (which were morphological rules in disguise) played a key role in this area. They operated on the final output of the syntactic component, making whatever modifications were necessary in order to

enable phonological rules to apply to the representation obtained after all syntactic rules had applied.

Unfortunately, there seems to have been no constraint on the power of readjustment rules. For instance, in *SPE (The Sound Pattern of English)* which appeared in 1968 and was the pivotal work in the development of generative phonological theory, Chomsky and Halle proposed (on p. 11) that the syntax should generate both the regular past tense form *mended* [<sub>v</sub>[<sub>v</sub>*mend*]<sub>v</sub> *past*]<sub>v</sub> and the irregular past tense form *sang* [<sub>v</sub>[<sub>v</sub>*sing*]<sub>v</sub> *past*]<sub>v</sub>. These bracketed strings, which were the output of the syntactic component, would form the input to the readjustment rules. Next, the readjustment rules would remove all the brackets associated with the past tense. In the case of *mend*, a general readjustment rule would replace *past* by *d*, while in the case of *sing* a special readjustment rule would delete the item *past*, together with the associated bracket labels, giving [<sub>v</sub>*sing*]<sub>v</sub>. The same readjustment rule would also attach the diacritic mark \* to the vowel /i/ indicating that eventually a phonological rule would change it into /æ/. The readjustment rules would give the forms [<sub>v</sub>[<sub>v</sub>*mend*]<sub>v</sub> *d*]<sub>v</sub> and [<sub>v</sub>*s\*ng*]<sub>v</sub>. These representations – and all other such representations yielded by readjustment rules – were referred to as **phonological representations**. Finally, phonological representations would be converted into the phonetic representations [mendɪd] and [sæŋ] by rules in the phonology module.

With the benefit of hindsight, we can see that readjustment rules were a mistake. They were rules with unbridled power. They could make whatever modifications were deemed necessary to enable phonological rules to apply to strings of morphemes produced by the syntax. It is very undesirable to have a batch of rules that empower us linguists to do whatever we like, whenever we like, so long as we come up with the answer we like. A theory becomes vacuous if it has rules that can insert all manner of elements, remove all manner of elements and make all manner of elements exchange places whenever we choose to, with no principles restricting our freedom. Effectively, this means that we are given *carte blanche* to start off with any arbitrary input, apply the rules, and come up with the 'correct' answer.

Furthermore, readjustment rules were a bad idea because they are evidence of a lack of interest in words *qua* words and in morphology as a linguistic level. Using rules of the syntax to specify permissible sequences of morphemes, regardless of whether they occurred in words or sentences, and using readjustment rules to turn strings generated by the syntax into strings that the phonology could process and assign a pronunciation to was merely skirting round the problem. Words are a central dimension of language. They have certain unique properties that they do not share with other elements of linguistic structure like sentences and speech sounds. A theory of language must include a properly developed model of word-formation that enables the linguist to describe words on their own terms –

without overlooking the ways in which word-formation rules interact with rules in other modules. As time went by, this became clear to generative linguists who, in increasing numbers, began to explore more satisfactory ways of dealing with word-structure.

### 1.3.2 The Morphology–Phonology Interaction

As regards the interaction with phonology, the selection of the form that manifests a given morpheme may be influenced by the sounds that realise neighbouring morphemes. Take the indefinite article in English. It has two manifestations. It is *a* before a word that begins with a consonant (e.g., *a pear*) and *an* before a word that begins with a vowel (e.g., *an orange*). We cannot describe the phonological shape of the indefinite article without referring to the sound at the beginning of the word that follows it.

### 1.3.3. The Morphology–Syntax Interaction

As regards the interaction with syntax, the form of a word may be affected by the syntactic construction in which the word is used. For instance, the verb *walk* has a number of forms including *walk*, *walks* and *walked*. The selection of a particular form of this verb on a given occasion is dependent on the syntactic construction in which it appears. Thus, in the present tense, the choice between the forms *walks* and *walk* depends on whether the subject of the verb is third person singular (in which case *walks* is selected as is *helshelit walks*) or not (in which case *walk* is selected as in *I/you/welthey walk*). In the past tense, *walk* is realised as *walked*.

Chomsky (1957: 39) deals with all these facts as uncontroversial syntactic phenomena, using the phrase structure rule below:

$$[1.5] \quad C \rightarrow \left\{ \begin{array}{l} S \quad \text{in the context NP}_{sing} \text{ —} \\ \emptyset \quad \text{in the context NP}_{pl} \text{ —} \\ past \end{array} \right\}$$

Notes: (i) '→' stands for 'expand' or 'rewrite as'. (ii) C stands for the various verbal suffixes that may be realised as *-s* (as in *walks*),  $\emptyset$  (i.e. zero) as in *walk* and *-ed* as in *walked*.

Chomsky's analysis does not separate phrase structure rules (e.g. Sentence → NP + VP; VP → Verb + NP) which enumerate permissible combinations of words in phrases and sentences from rules of word-structure like the one in [1.5] that gives *walks* from *walk*. All these rules are banded

together because they are concerned with enumerating permissible combinations of morphemes (see above).

Note, however, that this treatment of syntactically motivated alternation in the form of words is controversial. We have merely aired the problem for the present. We will postpone detailed discussion until Chapter 10.

Turning to semantics, the connection between morphology and the lexicon on the one hand with meaning on the other is obvious since a major role of the lexicon or dictionary is to list the meanings of words. This is because normally the relationship between a word and its meaning is arbitrary. There is no reason why a word has the particular meaning that it has. For instance, you just have to memorise the fact that the word *faillie* refers to a kind of head-dress worn in the seventeenth century. There is no way that you could discover this fact from the sounds or the structure of the word. We will come back to this topic in section (12.3.2).

It is less immediately obvious that, in addition to indicating the meaning of words and morphemes, the lexicon must also store other kinds of information relevant to the application of syntactic and phonological rules. Syntax needs to have access to **morphosyntactic properties** (i.e. properties that are partly morphological and partly syntactic) such as whether a noun is countable like *spades* or uncountable like *equipment*. This affects its behaviour in phrases and sentences. We may say *this spade* or *these spades* but we can only say *this equipment* (not *\*these equipments*).

Furthermore, some phonological rules apply to words differently depending on their morphosyntactic properties. For example, some phonological rules are sensitive to the difference between nouns and verbs. Thus, in the word *permit*, the main stress (shown here by underlining) falls on the first syllable if the word functions as a noun (*permit<sub>[noun]</sub>*). But if it functions as a verb (*permit<sub>[verb]</sub>*), main stress falls on the second syllable. Obviously, for phonological rules that assign stress to apply correctly, access to such morphosyntactic information is essential. This information must form part of the entry of the word in the lexicon.

The study of morphology, therefore, cannot be self-contained. The structuralist doctrine of the rigid separation of linguistic levels sketched in (1.2) is untenable. True, there are some issues that are the internal concerns of morphology. But many morphological problems involve the interaction between morphology and other modules of the grammar. For this reason, much of the space in the chapters that follow is devoted to the interaction between the lexicon and morphology with the mother modules.

## 1.4 ORGANISATION OF THE BOOK

The book is organised as follows:

**Part I** (Chapters 1–4) introduces basic concepts and traditional notions which are fundamental to all morphological discussions.

**Part II** (Chapters 5–9) explores the relationship between morphology, phonology and the lexicon in current generative theory.

**Part III** (Chapters 10–12) deals with the relationship between morphology and syntax in current generative theory.

Over the years, there have been several morphological theories that have been proposed by linguists. One way of introducing you to morphology would be to present a historical and comparative survey. I could have examined various theories in turn, and perhaps compared them. Or, alternatively, I could have been polemical and proselytising. I could have tried to persuade you that my preferred theory is the best theory. That is not what I shall do in this book.

Instead, I present you, sympathetically but at the same time critically, with one theoretically coherent approach to morphology, namely the theory of morphology in current mainstream generative grammar. This decision is sensible not only because this is the dominant model in the field today, but also because I think it offers the most promising solutions to the perennial problems in morphological analysis.

Even so, the book is inevitably selective. I have not attempted to represent every shade of opinion within the generative school. Rather I have focused on ideas and practices that seem to me to form part of the emerging 'canon' in mainstream generative morphology. Obviously, to some extent this is a matter of subjective judgement. In some cases my judgement may not be the same as that of some other linguists.

Of course, morphological theory in current mainstream generative grammar does not enjoy a monopoly of insight. The debt owed to other approaches will be evident, especially in the early chapters and in the bibliography.

A major feature of the book is that you will be asked to be an active investigator, not a passive reader. I have endeavoured to engage you actively and practically in *doing* morphology rather than in merely learning about its history and watching from the stalls how it is done. As you read each chapter, you are asked to pause at places and answer in-text questions and exercises before proceeding (the questions and exercises are signalled by lines across the page). Each chapter (after this one) ends with further exercises dealing with points raised in the body of the text. This insistence on getting you to analyse data is due to my firm conviction that the best initiation for anyone who wishes to become a linguist is to *do* linguistic analysis right from the start rather than to read about it.

In the text new morphological terms appear in bold type and they are

explained when they are first introduced. (They may also be in bold type when they appear subsequently in a context where they need to be highlighted.) Key terms from other branches of linguistics are explained in a glossary at the end. For any other linguistic terms that are unfamiliar, a good dictionary of linguistics, such as David Crystal's *A First Dictionary of Linguistics and Phonetics* (1980), should be consulted.

## 2 Introduction to Word-Structure

### 2.1 WHAT IS A WORD?

The assumption that languages contain words is taken for granted by most people. Even illiterate speakers know that there are words in their language. True, sometimes there are differences of opinion as to what units are to be treated as words. For instance, English speakers might not agree whether *all right* is one word or two and as a result disputes may arise as to whether *alright* is the correct way of writing *all right*. But, by and large, people can easily recognise a word of their language when they see or hear one. And normally their judgements as to what is or is not a word do coincide. English speakers agree, for example, that the form *splody* in the sentence *The splody cat sat on the mat* is not an English word – but all the other forms are.

#### 2.1.1 The Lexeme

However, closer examination of the nature of the 'word' reveals a somewhat more complex picture than I have painted above. What we mean by 'word' is not always clear. As we shall see in the next few paragraphs, difficulties in clarifying the nature of the word are largely due to the fact the term 'word' is used in a variety of senses which usually are not clearly distinguished. In taking the existence of words for granted, we tend to overlook the complexity of what it is we are taking for granted.

What would you do if you were reading a book and you encountered the 'word' *pockled* for the first time in this context?

[2.1] He went to the pub for a pint and then *pockled* off.

You would probably look up that unfamiliar word in a dictionary, not under *pockled*, but under *pockle*. This is because you know that *pockled* is not going to be listed in the dictionary. You also know, though nobody has told you, that the words *pockling* and *pockles* will also exist. Furthermore, you know that *pockling* and *pockle*, *pockles* and *pockled* are all in a sense different manifestations of the 'same' abstract vocabulary item.

We shall refer to the 'word' in this sense of abstract vocabulary item



using the term **lexeme**. The forms *pockling*, *pockle*, *pockles* and *pockled* are different **realisations** (or representations or manifestations) of the lexeme **POCKLE** (lexemes will be written in capital letters). They all share a core meaning although they are spelled and pronounced differently. Lexemes are the vocabulary items that are listed in the dictionary (cf. Di Sciullo and Williams, 1987).

Which ones of the words in [2.2] below belong to the same lexeme?

[2.2]	see	catches	taller	boy	catching	sees
	sleeps	woman	catch	saw	tallest	sleeping
	boys	sleep	seen	tall	jumped	caught
	seeing	jump	women	slept	jumps	jumping

We should all agree that:

<u>The physical word-forms</u> <i>see, sees, seeing, saw, seen</i> <i>sleeps, sleeping, slept</i> <i>catch, catches, catching, caught</i>	are realisations of	<u>the lexeme</u> SEE SLEEP CATCH
--	---------------------	--

<u>The physical word-forms</u> <i>jump, jumps, jumped, jumping</i> <i>tall, taller, tallest</i> <i>boy, boys</i> <i>woman, women</i>	are realisations of	<u>the lexeme</u> JUMP TALL BOY WOMAN
--	---------------------	---

### 2.1.2 Word-form

As we have just seen above, sometimes, when we use the term 'word', it is not the abstract vocabulary item with a common core of meaning, the lexeme, that we want to refer to. Rather, we may use the term 'word' to refer to a particular physical realisation of that lexeme in speech or writing, i.e. a particular **word-form**. Thus, we can refer to *see*, *sees*, *seeing*, *saw* and *seen* as five different words. In this sense, three different occurrences of any one of these word-forms would count as three words. We can also say that the word-form *see* has three letters and the word-form *seeing* has six. And, if we were counting the number of words in a passage, we would gladly count *see*, *sees*, *seeing*, *saw* and *seen* as five different word-forms (belonging to the same lexeme).

### 2.1.3 The Grammatical Word

The 'word' can also be seen as a representation of a lexeme that is associated with certain **morpho-syntactic properties** (i.e. partly morphological and partly syntactic properties) such as noun, adjective, verb, tense, gender, number, etc. We shall use the term **grammatical word** to refer to the 'word' in this sense.

Show why *cut* should be regarded as representing two distinct grammatical words in the following:

- [2.3] a. Usually I cut the bread on the table.  
b. Yesterday I cut the bread in the sink.

The same word-form *cut*, belonging to the verbal lexeme **CUT**, can represent two different grammatical words. In [2.3a], *cut* represents the grammatical word *cut*<sub>[verb, present, non 3rd person]</sub>, i.e. the present tense, non-third person form of the verb **CUT**. But in [2.3b] it represents the grammatical word *cut*<sub>[verb, past]</sub> which realises the past tense of **CUT**.

Besides the two grammatical words realised by the word-form *cut* which we have mentioned above, there is a third one which you can observe in *Jane has a cut on her finger*. This grammatical word is *cut*<sub>[noun, singular]</sub>. It belongs to a separate lexeme **CUT**, the noun. Obviously, **CUT**, the noun, is related in meaning to **CUT**, the verb. However, **CUT**, the noun, is a separate lexeme from **CUT**, the verb, because it belongs to a different word-class (see section 3.5 below).

The nature of the grammatical word is important in the discussion of the relationship between words and sentences and the boundary between morphology and syntax.

## 2.2 MORPHEMES: THE SMALLEST UNITS OF MEANING

Morphology is the study of word structure. The claim that words have structure might come as a surprise because normally speakers think of words as indivisible units of meaning. This is probably due to the fact that many words are morphologically simple. For example, *the*, *fierce*, *desk*, *eat*, *boot*, *at*, *fee*, *mosquito*, etc., cannot be segmented (i.e. divided up) into smaller units that are themselves meaningful. It is impossible to say what the *-quito* part of *mosquito* or the *-erce* part of *fierce* means.

But very many English words are morphologically complex. They can be broken down into smaller units that are meaningful. This is true of words like *desk-s* and *boot-s*, for instance, where *desk* refers to one piece of furniture and *boot* refers to one item of footwear, while in both cases the *-s* serves the grammatical function of indicating plurality.

The term **morpheme** is used to refer to the smallest, indivisible units of semantic content or grammatical function which words are made up of. By definition, a morpheme cannot be decomposed into smaller units which are either meaningful by themselves or mark a grammatical function like singular or plural number in the noun. If we divided up the word *fee* [fi:] (which contains just one morpheme) into, say, [f] and [i:], it would be impossible to say what each of the sounds [f] and [i:] means by itself since sounds in themselves do not have meaning.

How do we know when to recognise a single sound or a group of sounds as representing a morpheme? Whether a particular sound or string of sounds is to be regarded as a manifestation of a morpheme depends on the word in which it appears. So, while *un-* represents a negative morpheme and has a meaning that can roughly be glossed as 'not' in words such as *un-just* and *un-tidy*, it has no claim to morpheme status when it occurs in *uncle* or in *under*, since in these latter words it does not have any identifiable grammatical or semantic value, because *-cle* and *-der* on their own do not mean anything. (Morphemes will be separated with a hyphen in the examples.)

Lego provides a useful analogy. Morphemes can be compared to pieces of lego that can be used again and again as building blocks to form different words. Recurrent parts of words that have the same meaning are isolated and recognised as manifestations of the same morpheme. Thus, the negative morpheme *un-* occurs in an indefinitely large number of words, besides those listed above. We find it in *unwell*, *unsafe*, *unclean*, *unhappy*, *unfit*, *uneven*, etc.

However, recurrence in a large number of words is not an essential property of morphemes. Sometimes a morpheme may be restricted to relatively few words. This is true of the morpheme *-dom*, meaning 'condition, state, dignity', which is found in words like *martyrdom*, *kingdom*, *chiefdom*, etc. (My glosses, here and elsewhere in the book, are based on definitions in the *Oxford English Dictionary*.)

It has been argued that, in an extreme case, a morpheme may occur in a single word. Lightner (1975: 633) has claimed that the morpheme *-ric* meaning 'diocese' is only found in the word *bishopric*. But this claim is disputed by Bauer (1983: 93) who suggests instead that perhaps *-ric* is not a distinct morpheme and that *bishopric* should be listed in the dictionary as an unanalysable word. We will leave this controversy at that and instead see how morphemes are identified in less problematic cases.

List two other words which contain each morpheme represented below:

[2.4]	a.	-er	as in	play-er, call-er
		-ness	as in	kind-ness, good-ness
		-ette	as in	kitchen-ette, cigar-ette
	b.	ex-	as in	ex-wife, ex-minister
		pre-	as in	pre-war, pre-school
		mis-	as in	mis-kick, mis-judge

- a. Write down the meaning of each morpheme you identify. (If you are in doubt, consult a good etymological dictionary.)
- b. What is the syntactic category (noun, adjective, verb, etc.) of the form which this morpheme attaches to and what is the category of the resulting word?

I expect your answer to confirm that, in each example in [2.4], the elements recognised as belonging to a given morpheme contribute an identifiable meaning to the word of which they are a part. The form *-er* is attached to verbs to derive nouns with the general meaning 'someone who does X' (where X indicates whatever action the verb involves). When *-ness* is added to an adjective, it produces a noun meaning 'having the state or condition (e.g., of being *kind*)'. The addition of the diminutive morpheme *-ette* to a noun derives a new noun which has the meaning 'smaller in size' (e.g., a *kitchenette* is a small *kitchen* and a *cigarette* is smaller than a *cigar*). Finally, the morphemes *ex-* and *pre-* derive nouns from nouns while *mis-* derives verbs from verbs. We can gloss the morpheme *ex-* as 'former', *pre-* as 'before' and *mis-* as 'badly'.

So far we have described words with just one or two morphemes. In fact, it is possible to combine several morphemes together to form more complex words. This can be seen in long words like *unfaithfulness* and *reincarnation* which contain the morphemes *un-faith-ful-ness* and *re-in-carn-at-ion* respectively. But on what grounds do we divide up these words in this fashion? In the following sections we will examine the basis on which morphemes are identified.

### 2.2.1 Analysing Words

Up to now, we have used the criterion of meaning to identify morphemes. In many cases forms that share the same meaning may be safely assigned to the same morpheme. Where the meaning of a morpheme has been somewhat obscure, you have been encouraged to consult a good etymological

dictionary. Unfortunately, in practice, appealing to meanings listed in etymological dictionaries has its problems.

Consider the following words:

[2.5]	helicopter	pteropus	diptera
	bible	bibliography	bibliophile

Historically *pter* was borrowed from Greek, where it meant 'feather or wing'. The form *bibl-* also came from Greek where it meant 'papyrus, scroll, book'. Do you think *pter-* and *bibl-* should be recognised as morphemes in modern English?

I do not know what you decided. But I think it is questionable whether *pter-* is a morpheme of modern English. A *helicopter* is a kind of non-fixed wing aircraft which most speakers of English know about; *pteropus* are tropical bats with membranous wings popularly known as 'flying foxes' and *diptera* are two-winged flies (which few of us who are not entomologists know about). Obviously, *pter-* does occur in modern English words that have the meaning 'pertaining to wings'. What is doubtful is whether this fact is part of the tacit knowledge of speakers of English who are not versed in etymology. Most people probably go through life without seeing a semantic connection between 'wings' and 'helicopters'.

Similarly, as we have already noted, the words *bible*, *bibliography* and *bibliophile* have to do with books. Probably many English speakers can see the book connection in *bibliography* and *bibliophile*. But it is unlikely that anyone lacking a profound knowledge of English etymology (and a classical education) is aware that the word *bible* is not just the name of a scripture book and that it contains a morpheme which is found in a number of other words pertaining to books.

Clearly, we need to distinguish between etymological information, whose relevance is essentially historical, and synchronic information that is part of speakers' competence. Our primary task as morphologists is to investigate speakers' tacit knowledge of the rules of their language rather than to perform historical reconstruction. We shall discuss this further in Chapter 4. The point I am making is that over-reliance on meaning in isolating morphemes puts us in a quandary in cases where etymological meanings are shrouded in the mists of history and lose their synchronic relevance.

The common definition of the morpheme as the 'minimal meaningful unit' implies the claim that every morpheme has a readily identifiable meaning. But this is problematic. There are cases where we can justify

recognising a recurrent word-building unit as a morpheme although we cannot assign it a consistent meaning.

This is true of *-fer* in words like *pre-fer*, *in-fer*, *de-fer*, *con-fer*, *trans-fer* and *re-fer*. An etymological dictionary will tell us that *-fer* comes from the Latin word meaning 'bear, bring, send'. However, we would be hard-pressed to identify a consistent meaning like 'bring' attributable to *-fer-* in every instance above. For this reason some linguists, such as Aronoff (1976: 8–10), have argued that it is the word in its entirety rather than the morpheme *per se* that must be meaningful. Whereas all words must be meaningful when they occur on their own, morphemes need not be. Some morphemes, like *ex-* 'former' as in *ex-wife* and *pre-* 'before' as in *pre-war*, have a transparent, unambiguous meaning while others like *-fer* do not. Their interpretation varies depending on the other morphemes that occur together with them in a word.

In view of the above remarks, while semantic considerations must play a role in the identification of morphemes, given the pitfalls of a purely semantic approach, linguists tend to give a higher priority to more formal factors.

### 2.2.2 Morphemes, Morphs and Allomorphs

At one time, establishing mechanical procedures for the identification of morphemes was considered a realistic goal by structural linguists (cf. Harris, 1951). But it did not take long before most linguists realised that it was impossible to develop a set of **discovery procedures** that would lead automatically to a correct morphological analysis. No scientific discipline purports to equip its practitioners with infallible procedures for arriving at correct theories. Creative genius is needed to enable the scientist to make that leap into uncharted waters that results in a scientific discovery. What is true of science in general is also true of linguistics (cf. Chomsky, 1957: 49–60). Writing a grammar of a language entails constructing a theory of how that language works by making generalisations about its structure that go beyond the data that are observed.

Nevertheless, although there are no effective mechanical procedures for discovering the grammatical structure of a language in general or, in our case, the structure of its words, there exist reasonably reliable and widely accepted techniques that have been evolved by linguists working on morphology. These techniques are outlined in this section.

The main principle used in the analysis of words is the principle of **contrast**. We contrast forms that differ (i) in phonological shape due to the sounds used and (ii) in meaning, broadly defined to cover both lexical meaning and grammatical function. Thus, the phonological difference between /bɔɪ/ and /gɜ:1/ correlates with a semantic difference. The difference in meaning between the two sentences *The girl plays* and *The boy*

*plays* is attributable to the difference in lexical meaning between /bɔɪ/ and /gɜ:ɪ/. Likewise, the difference in grammatical function between *play-s* (present tense) and *play-ed* (past tense) is responsible for the difference in meaning between *The girl plays* and *The girl played*.

**DEFINITION:** The **morpheme** is the smallest difference in the shape of a word that correlates with the smallest difference in word or sentence meaning or in grammatical structure.

The analysis of words into morphemes begins with the isolation of morphs. A morph is a physical form representing some morpheme in a language. It is a recurrent distinctive sound (phoneme) or sequence of sounds (phonemes).

Study the data in [2.6] and identify the morphs:

- |                            |                        |
|----------------------------|------------------------|
| [2.6] a. I parked the car. | e. She parked the car. |
| b. We parked the car.      | f. She parks the car.  |
| c. I park the car.         | g. We park the car.    |
| d. He parks the car.       | h. He parked the car.  |

The morphs are:

Morph	Recurrs in
/aɪ/ 'I'	[2.6a] and [2.6c]
/ʃi:/ 'she'	[2.6e] and [2.6f]
/hi:/ 'he'	[2.6d] and [2.6h]
/ðə/ 'the'	in all the examples
/kɑ:/ 'car'	in all the examples
/pɑ:rk/ 'park'	<i>park</i> is found in all the examples, sometimes with an <i>-ed</i> suffix, sometimes with an <i>-s</i> suffix and sometimes on its own
/t/ '-ed'	suffixed to park in [2.6b, e, h]
/s/ '-s'	suffixed to park in [2.6d, f]

For our next example, we shall perform an analysis similar to the one we have just done for English on data from a less familiar language. Now study the data in [2.7] which are taken from Luganda and list all the morphs. (Although Luganda is a tone language, tone is omitted for simplicity's sake as it is not relevant here.)

[2.7]	tulilaba	kitabo	'we will see a book'
	tuligula	katabo	'we will buy a little book'
	baalaba	bitabo	'they saw books'
	tulilaba	butabo	'we will see little books'
	balilaba	kitabo	'they will see a book'
	tulilaba	bitabo	'we will see books'
	baatunda	butabo	'they sold little books'
	baligula	bitabo	'they will buy books'
	baagula	katabo	'they bought a little book'
	tutunda	bitabo	'we sell books'

- Hints:**
- The word meaning 'book' appears in all the sentences but in some it is singular and in others plural.
  - 'Book' sometimes refers to a normal size book, and in other cases to a *little* book.
  - We have three different verbs.
  - The verbs are in different tenses.
  - The verbs have different subjects.

The answer to [2.7] is given in [2.8].

- [2.8] -tabo 'book', tu- 'we', ki- 'singular' (normal size) noun prefix  
 -laba 'see', ba- 'they', bi- 'plural' (normal size) noun prefix  
 -gula 'buy', -li- 'future', ka- 'singular' (small size) noun prefix  
 -tunda 'sell', -a- 'past', bu- 'plural' (small size) noun prefix

In [2.8], each different morph represents a separate morpheme. But this is not always the case. Sometimes different morphs may represent the same morpheme. For instance, the past tense of regular verbs in English which is spelled *-ed* is realised in speech by /ɪd/, /d/ or /t/. The phonological properties of the last segment of the verb to which it is attached determine the choice:

- [2.9] It is realised as:
- /ɪd/ if the verb ends in /d/ or /t/  
 e.g. /mend/ ~ /mendɪd/ /peɪnt/ ~ /peɪntɪd/  
 'mend' 'mended' 'paint' 'painted'
  - /d/ after a verb ending in any voiced sound except /d/  
 e.g. /kli:n/ ~ /kli:nd/ /weɪ/ ~ /weɪd/  
 'clean' 'cleaned' 'weigh' 'weighed'
  - /t/ after a verb ending in any voiceless consonant other than /t/  
 e.g. /pɑ:k/ ~ /pɑ:kt/ /mɪs/ ~ /mɪst/  
 'park' 'parked' 'miss' 'missed'

Now compare the Luganda forms in [2.10] with those in [2.7] above.

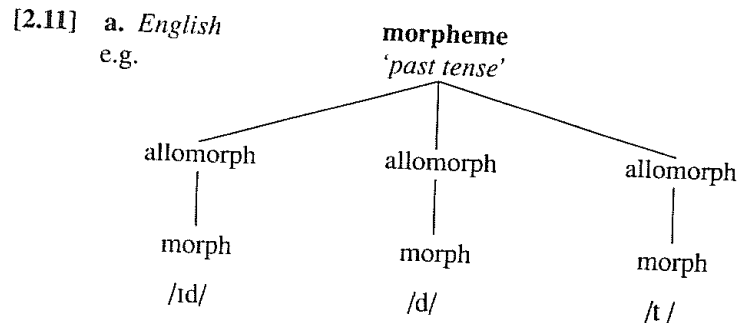
[2.10]	twaalaba	kitabo	'we saw a book'
	twaagula	bitabo	'we bought books'
	twaatunda	kitabo	'we sold a book'

The first person plural is represented by the form *tu-* in [2.7] and by *tw-* in [2.10]. What determines the selection of *tu* vs *tw*-?

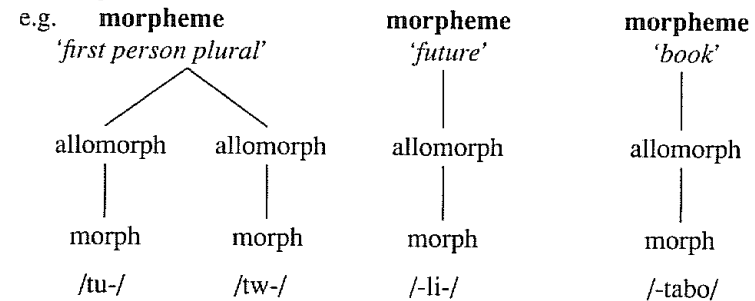
Observe that here again the difference in form is not associated with a difference in meaning. The morphs *tu-* and *tw-* both represent the first person plural in different contexts. *Tu-* is used if the next morpheme is realised by a form beginning with a consonant and *tw-* is selected if the next morpheme is realised by a form that begins with a vowel.

If different morphs represent the same morpheme, they are grouped together and they are called **allomorphs** of that morpheme. So, *tu-* and *tw-* are allomorphs of the 'first person plural' morpheme. (For simplicity's sake, for our present purposes, we are regarding 'first person plural' as a single unanalysable concept.) On the same grounds, /ɪd/, /d/ and /t/ are grouped together as allomorphs of the past tense morpheme in English.

The relationship between morphemes, allomorphs and morphs can be represented using a diagram in the following way:



b. Luganda



We can say that (i) /ɪd/, /d/ and /t/ are English morphs and (ii) we can group all these three morphs together as allomorphs of the past tense morpheme. Likewise, in Luganda we can say that (i) *tu-*, *tw-*, *-li-* and *-tabo* are morphs and furthermore (ii) *tu-* and *tw-* are allomorphs of the same morpheme since they represent the same superordinate concept, the morpheme 'first person plural'.

The central technique used in the identification of morphemes is based on the notion of **distribution**, i.e. the total set of contexts in which a particular linguistic form occurs. We classify a set of morphs as allomorphs of the same morpheme if they are in **complementary distribution**. Morphs are said to be in complementary distribution if (i) they represent the same meaning or serve the same grammatical function and (ii) they are never found in identical contexts. So, the three morphs /-ɪd/, /-d/ and /-t/ which represent the English regular past tense morpheme are in complementary distribution. Each morph is restricted to occurring in the contexts specified in [2.9]. Hence, they are allomorphs of the same morpheme. The same analysis applies also to Luganda *tu-* and *tw-*. Both morphs mean 'we' and they are in complementary distribution. *Tu-* occurs before consonants and *tw-* before vowels. They are therefore allomorphs of the first person plural morpheme. Morphemes realised by an invariant form (e.g., *future* and *book*) are said to have a single allomorph (cf. Matthews, 1974: 83).

Let us now examine some English words, focusing on the pronunciation of the underlined part of each word, which represents the negative morpheme *in-*. This morpheme can roughly be glossed as 'not':

[2.12] a.	<u>im</u> possible	[ɪmpɒstɪbl]
	<u>im</u> patient	[ɪmpetʃnt]
	<u>im</u> movable	[ɪmʊvəbl]
b.	<u>in</u> tolerable	[ɪntələrəbl]
	<u>in</u> decent	[ɪndi:sənt]



## 2.2.3 Grammatical Conditioning, Lexical Conditioning and Suppletion

We have seen in the last section that the distribution of allomorphs is usually subject to phonological conditioning. However, sometimes phonological factors play no role in the selection of allomorphs. Instead, the choice of allomorph may be **grammatically conditioned**, i.e. it may be dependent on the presence of a particular grammatical element. A special allomorph may be required in a given grammatical context although there might not be any good phonological reason for its selection. For example, in [2.15a], which is typical, in English the presence of the past tense morpheme in the majority of cases has no effect on the selection of the allomorph that represents the verb itself. But, as [2.15b] and [2.15c] show, in certain verbs the presence of the past tense morpheme requires the selection of a special allomorph of the verb:

[2.15]	Present tense		Past tense	
a.	walk	/wɔ:k/	Walked	/wɔ:kt/
	kiss	/kɪs/	kiss-ed	/kɪst/
	grasp	/grɑ:sp/	grasp-ed	/grɑ:spt/
b.	weep	/wi:p/	wep-t	/wept/
	sweep	/swi:p/	swep-t	/swept/
c.	shake	/ʃeɪk/	shook	/ʃʊk/
	take	/tu:k/	took	/tu:k/

In [2.15b], the choice of allomorph is grammatically conditioned. The presence of the past tense morpheme determines the choice of the /wep/ and /swep/ allomorphs in verbs that belong to this group. For the verbs in [2.15c] the past tense dictates the choice of the allomorphs *took* and *shook* of the verbs *take* and *shake* respectively.

In other cases, the choice of the allomorph may be **lexically conditioned**, i.e. use of a particular allomorph may be obligatory if a certain *word* is present. We can see this in the realisation of plural in English.

Normally the plural morpheme is realised by a phonologically conditioned allomorph whose distribution is stated in [2.16]:

- [2.16] a. select allomorph /-ɪz/ if a noun ends in an alveolar or alveopalatal sibilant (i.e. a consonant with a sharp, hissing sound such as /s z ʒ ʒ tʃ dʒ/).  
Examples: *asses* *mazes* *fishes* *badges* *beaches*  
          /æsɪz/ /meɪzɪz/ /fɪʃɪz/ /bædʒɪz/ /bi:tʃɪz/
- b. select allomorph /-s/ if a noun ends in a non-strident voiceless consonant (i.e. any one of the sounds /p t k f θ/).

Examples: *cups* *leeks* *carts* *moths*  
          /kʌps/ /li:ks/ /kɑ:ts/ /mʊθs/

- c. select allomorph /-z/ elsewhere (i.e. if the noun ends in a voiced nonstrident segment; this includes all vowels and the consonants /b d g d m n ŋ l r w j/).

Examples: *bards* *mugs* *rooms* *keys* *shoes*  
          /bɑ:dz/ /mʌgz/ /ru:mz/ /ki:z/ /ʃu:z/

Can you explain why the rule in [2.16] fails to account for the realisation of the plural morpheme in the word *oxen*?

I expect you to have failed to find a plausible explanation. There are cases where for no apparent reason the regular rule in [2.16] inexplicably fails to apply. The plural of *ox* is not \**oxes* but *oxen*, although words that rhyme with *ox* take the expected /ɪz/ plural allomorph (cf. /fɒksɪz/ *foxes* and /bɒksɪz/ *boxes*). The choice of the allomorph *-en* is **lexically conditioned**. It is dependent on the presence of the specific noun *ox*.

Finally, there exist a few morphemes whose allomorphs show no phonetic similarity. A classic example of this is provided by the forms *good*/*better* which both represent the lexeme GOOD despite the fact that they do not have even a single sound in common. Where allomorphs of a morpheme are phonetically unrelated we speak of **suppletion**.

The pair *good* and *better* is not unique in English. Find one other example of suppletion.

Other examples of suppletion in English include *bad* ~ *worse* (not \**badder*); *go* ~ *went* (not \**goed*)

## 2.2.4 Underlying Representations

Above we have distinguished between, on the one hand, regular, rule-governed **phonological alternation** (a situation where the choice between alternative allomorphs is regulated in quite predictable ways by the phonological properties of the different morphs that occur near each other (see section (2.2.2)) and cases of suppletion where there is phonologically arbitrary alternation in the realisation of a morpheme (see section (2.2.3)). This is standard in generative phonology (cf. Chomsky and Halle, 1968; Kenstowicz and Kisseberth, 1979; Anderson, 1974: 51–61).

Merely listing allomorphs does not allow us to distinguish between

eccentric alternations like *good ~ bett(-er)* and regular alternations like that shown by the negative prefix *in-* or by the regular *-s* plural suffix. The latter are general and will normally apply to any form with the relevant phonological properties, unless it is specifically exempted. Thus the regular plural rule in [2.16] above is used to attach */-s/, /-z/* or */-ɪz/* to virtually any noun that ends in the appropriate sound. By contrast, a rule of suppletion or lexical conditioning only applies if a form is expressly marked as being subject to it. Thus, for example, of all English adjectives, only *good* is subject to the suppletive rule that gives *bett-er* in the comparative; and only *ox* is subject to the lexically conditioned rule that suffixes *-en* to yield the plural *oxen*. Similarly, a grammatically conditioned rule will only be triggered if the appropriate grammatical conditioning factor is present. For example, the allomorph *sleep-* of the morpheme *sleep* only co-occurs with the past tense (or past participle) morpheme. It cannot be selected to co-occur with the present tense: *sleep* + [past] yields *slept* /slept/ (not /\*sli:pt/) while *sleep* + [present] gives *sleep* /sli:p/ (not /\*slep/).

To bring out the distinction between regular phonological alternation, which is phonetically motivated, and other kinds of morphological alternation that lack a phonetic basis, linguists posit a single **underlying representation** or **base form** from which the various allomorphs (or **alternants** i.e. alternative phonetic realisations) of a morpheme are derived by applying one or more **phonological rules**. The stages which a form goes through when it is being converted from an underlying representation to a **phonetic representation** constitute a **derivation**.

For a concrete example, let us consider again the representation of the *in-* morpheme in [2.13], which is repeated below as [2.17] for convenience:

- [2.17] a. select [ɪm] before a **labial** consonant (e.g. p, b, f, m) as in [ɪm]possible, [ɪm]patient, [ɪm]movable.  
 b. select [ɪŋ] before the **velar** consonants [k] (here spelt with 'c') and [g] as in [ɪŋ]compliance, [ɪŋ]compatible, [ɪŋ]gratitude.  
 c. select [ɪn] elsewhere, i.e. before an **alveolar** consonant like [t, d, s, z, n], as in [ɪn]tolerable, [ɪn]tangible and [ɪn]decent or before a vowel as in [ɪn]active, [ɪn]elegance.

The vital point to note is that the three parts of the rule in [2.17] are not independent of each other. By making three separate statements, we have missed a generalisation. A superior solution would be to restate [2.17] as [2.18]. The revised statement, in which we posit a single underlying representation from which the three allomorphs are derived, captures the fact that the alternation in the realisation of these allomorphs is due to a single factor, namely assimilation.

- [2.18] The nasal realising the morpheme *in-* /ɪn/ must appear in the phonetic representation as a nasal consonant that shares the place of articulation of the initial consonant of the form to which it is attached.

But how can we be certain that the base form is /ɪn/ rather than /ɪm/ or /ɪŋ/? We have seen that the nasal assimilates to the place of articulation of the consonant that follows it. The fact that when a vowel follows we still find [ɪn-] appearing as in [ɪn-ɔ:dɪbl] *inaudible*, and [ɪn-evɪtəbl] *inevitable* is very revealing. From a phonetic point of view, vowels do not have definite places of articulation, only consonants do. So, a consonant cannot assimilate to the place of articulation of a vowel. The occurrence of [ɪn-] before vowels is not due to place assimilation. Besides, the alveolar nasal is found regardless of whether the vowel that follows is made in the front of the mouth like [e], or in the back like [ɔ:]. So, the influence of the vowel cannot be responsible for the choice of [ɪn-].

A simple solution is to assume that [ɪn-] is the **default form**, i.e. the form selected unless there are explicit instructions to do otherwise. If we posit this form as the underlying representation, we do not need to change it before vowels or before alveolar consonants. We only need to change it before non-alveolar consonants. If, however we posited [ɪm-] or [ɪŋ-] as the underlying representation, we would need rules to modify them when they appeared not only before non-labial and non-velar consonants respectively but also before vowels. If two analyses can both account properly for the facts, the analysis that provides a simpler solution is preferred. Obviously, in this case it is the analysis in [2.18] (with /ɪn-/ as the base form) that wins.

Phonologically conditioned morphological alternations tend to be very general. Often allomorphs representing different morphemes will display the same phonological alternations if they occur in similar phonological environments. Thus, for example, the voice assimilation process displayed by the *-s* plural suffix in [2.16] is not unique to that morpheme. The *-s* third person singular present tense suffix in verbs shows exactly the same alternations, as you can see:

- [2.19] /-ɪz/ after sibilants e.g. /wɔ:ʃ/ *wash* ~ /wɔ:ʃɪz/ *washes*  
 /-z/ after voiced segments other than sibilants e.g. /ri:d/ *read* ~ /ri:dz/ *reads*  
 /-s/ after voiceless consonants other than sibilants e.g. /dʒʌmp/ *jump* ~ /dʒʌmps/ *jumps*

The same rule applies to genitives:

- [2.20] /-ɪz/ after sibilants e.g. /lɪz/ *Liz* ~ /lɪzɪz/ *Liz's*



- /-z/ after voiced segments other than sibilants e.g. /dʒeɪn/ *Jane* ~ /dʒeɪnz/ *Jane's*  
 /-s/ after voiceless consonants other than sibilants e.g. /mɪk/ *Mick* ~ /mɪks/ *Mick's*

If we make three separate statements, one for the plural, another for the third person singular and a third one for the genitive, we miss the generalisation that a sibilant suffix agrees in voicing with the last segment of the form to which it is attached, regardless of the morpheme the suffix represents.

However, this generalisation is captured if we posit just one underlying representation (or base form) for any sibilant suffix, and if that underlying representation is converted into different phonetic representations by the (informal) phonological rules below:

- [2.21] a. The underlying representation of any sibilant suffix is /z/.  
 b. It is realised as:  
 (i) /ɪz/ after alveolar and alveo-palatal sibilants (e.g. /s z ʃ ʒ tʃ dʒ/)  
 (ii) /z/ after voiced segments other than sibilants (e.g. vowels and voiced consonants like /b m d v/)  
 (iii) /s/ after voiceless consonants other than sibilants /p t k f θ/

The statement in [2.21] shows that the alternation in question is not the idiosyncratic property of any one morpheme but rather a general phonological process in the language. The terms **morphophonemics** (in American linguistics) and **morphophonology** (European linguistics) are used to refer to rules of this kind that account for the realisation of phonologically conditioned allomorphs of morphemes. The rule for the realisation /ɪn-/ in [2.18] is another example of a morphophonemic rule.

### 2.3 THE NATURE OF MORPHEMES

Words can be divided into **segments of sound**. Thus, the word *book* can be divided into the segments /b, u, k/. Indeed, the division of words into phonemes forms the basis of alphabetic writing systems like that of English. But it is also possible, and natural to divide words into syllables. For instance, Japanese uses fifty distinct symbols to represent the fifty syllable types found in the language.

So, it is important to avoid confusing morphemes with syllables. **Syllables** are groupings of sounds for the purposes of articulation, while morphemes are the smallest units of meaning or grammatical function. A few examples should clarify the distinction. On the one hand, the words

*sofa* /səʊ fə/ and *balloon* /bə lu:n/ contain two syllables each while *camera* /kæ mə rə/ and *hooligan* /hu: lɪ gən/ contain three syllables each. (I have separated syllables with a space). But all these words have only one morpheme each. On the other hand, the word *books* /bʊks/ has one syllable, but two morphemes. They are the morpheme *book* /bʊk/ and the final *-s* /s/ which represents the plural morpheme although it is not a syllable in its own right.

When we divide a word into morphemes, we focus on strings of sound that are meaningful regardless of whether they constitute syllables at the phonological level. A question that lurks in the background concerns the precise nature of the relationship between strings of sounds and the meanings that they represent. This is the question to which we now turn. The discussion that follows draws on Matthews (1974).

At first, it might seem reasonable to assume that the relationship between morphemes and strings of phonemes, which are identified as morphs, is one of **composition**. In that case, we could say that the morpheme *book* /bʊk/ is *made up of* the phonemes /b/, /ʊ/ and /k/.

As we will see in a moment, an approach which assumes that morphemes are made up of phonemes leads to a theoretical cul-de-sac. It is preferable to view morphemes as being **represented** or **realised** or **manifested** by morphs. It is unsound to assume that morphemes are actually composed of (sequences of) phonemes because this would suggest that the meaning of a morpheme is a function of its phonemic composition. It is not, since phonemes in themselves cannot have meanings. The same phoneme /ə/ (spelled *-er*) can represent either the comparative degree of adjectives, as in *kind-er* and the noun-forming suffix *-er* as in *worker*, which is formed from the verb *work*, or it can be a part of a word without a discernible meaning of its own, as in *water*. Clearly, it is the morphs rather than morphemes that are made up of (sequences of) phonemes. Possible relationships between morphemes and (sequences of) phonemes can then be summarised in this fashion:

1. There may be a one-to-one correlation between morphemes and morphs (which are made up of individual phonemes or strings of phonemes). For instance, in French, the word *eau* /o/ (water) has one morpheme which is realised by a morph that is composed of just one phoneme, namely /o/. This is the simplest situation.

2. The relationship between sound and meaning in language is for the most part **arbitrary**, that is to say that is no good reason why a particular sound or string of sounds has a particular meaning. Given this, several different pairings of sounds with meaning are possible:

- (i) As we saw above a moment ago with regard to /ə/, a single phonological form may be used to represent different morphemes. Now we will

examine this point in more detail. Consider the phonological form /sart/ which happens to have three orthographic representations, each one of which represents a different morpheme. Also think about the phonological form /rart/ which has four spellings which represent four separate morphemes:

- [2.22] a. sight      site      cite  
 b. right      write      wright      rite

What we see here are **homophones**, i.e. forms which sound the same but differ in their meaning or grammatical function. From the point of view of the spoken language, there are only two morphs, namely the forms /sart/ and /rart/. The two morphs represent three and four morphemes respectively, but written English uses a different form to represent each morpheme in each case. Homophony is relatively common in language. Puns depend on it. And many a joke is due to the fact that morphs like *duck* and *bent* represent more than one morpheme.

- (ii) The converse is also common. A single morpheme may be represented by a variety of phonological representations. We have already seen this in the case of the plural morpheme, which has the three allomorphs [s], [z] and [ɪz] (see [2.16] above). The same applies to the negative prefix *in-* (which has the allomorphs [ɪm], [ɪn] and [ɪŋ] (see [2.17] above).
- (iii) The same string of sounds may cumulatively represent several morphemes. The *-s* ending in English verbs (e.g. *walk-s*) signals three morphemes simultaneously, namely, third person, present tense and singular number. If morphemes consisted of morphs this would not be possible. A separate morph would be needed to represent each morpheme. This shows just how abstract morphemes are, as opposed to morphs. Morphemes themselves are not composed of sounds but they are represented by morphs which are made up of sounds.

The term **portmanteau morph** is used to refer to cases like the above where a single morph simultaneously represents a bundle of several different grammatical elements.

Morphemes are to morphs what lexemes are to word-forms. Morphemes and lexemes are the abstract entities found in the lexicon while morphs and word-forms are the physical entities found in speech or writing.

In addition to different morphemes being represented by the same morphs, we can also have a situation where different grammatical words are represented by the same word-forms. This is called **syncretism**. It is a result of **neutralisation**. The same form is used to

represent distinct morphological concepts. Thus, in regular verbs, the same word-form represents two distinct grammatical words: e.g. *walk* + [past] *walked* (as in *I walked*) vs *walk* + [past participle] *walked* (as in *I have walked*). Irregular verbs like *see* and *take* exhibit no syncretism. They have distinct past tense and past participle forms: *see* + [past] (*saw*) and *take* + [past] (*took*); *see* + [past participle] (*seen*) and *take* + [past participle] (*taken*).

3. Finally, an approach that assumes a one-to-one correspondence between morphemes and morphs encounters difficulties when there is simply no match between morpheme and morph. There are two sets of circumstances in which this may happen:

- (i) The number of morphemes present may exceed the number of morphs available to represent them. This happens when a grammatical contrast which is marked overtly by a morph in some words is not overtly marked in others. Thus, for example, we know that in English, if the adverb *yesterday* or a phrase like *last week* is found in a sentence, the verb in that sentence must be in the past tense because that is the form of the verb that is required whenever a verb designates an event, action, state or process that happened prior to the moment of speaking or writing. As a rule, such a verb will end in *-ed*.

- [2.23] Last week the farmer sowed the corn.  
 Yesterday Jane painted the roof.

In the light of the last remark, explain how the past tense is marked in the following.

- [2.24] Last week I cut the grass.  
 I put those carnations in the vase yesterday.  
 Yesterday they shut the factory down.  
 The mob hit him last week.

We know that *cut*, *put*, *shut* and *hit* are every bit as past as *sowed* and *painted* in [2.23] because only verbs in the past tense can occur together with *yesterday* or *last week* in a sentence.

The past tense morpheme, which is represented by *-ed* in [2.23], is realised by a **zero allomorph** in [2.24]. In other words, we can infer from the structural patterns of the language that the verb is in the past tense although nothing about the shape of the word overtly shows this. If we allow ourselves to posit zero allomorphs, the assumption that morphemes 'consist' of phonemes must be rejected. Instead, we shall

regard morphemes as abstract entities which are *represented* by morphs. In speech, the morphs are composed of phonemes but the morphemes themselves are not. (See also the discussion of conversion in section (3.5).)

- (ii) The converse also occurs: the number of morphs that can be isolated may exceed the number of morphemes represented. In other words, there may be a surplus word-building element which does not realise any morpheme. Such an element is usually called an **empty morph**.

Describe in detail how the adjectives in [2.25a] and [2.25b] are derived from nouns.

[2.25]	Noun		Adjective	
a.	medicine	/medɪsɪn/	medicin-al	/medɪsɪnəl/
	person	/pɜːsən/	person-al	/pɜːsənəl/
	tribe	/traɪb/	trib-al	/traɪbəl/
b.	sense	/sens/	sens-u-al	/sensjʊəl/
	fact	/fækt/	fact-u-al	/fæktjʊəl/

In [2.25a] the adjectives are formed simply by adding the suffix *-al* to nouns. In [2.25b], however, there is an empty morph, *-u-* (/jʊ/) that does not represent any morpheme which is inserted immediately before *-al*.

'Empty morph' is an unfortunate choice of terminology – but we are stuck with it. If a morph is a morph by virtue of representing some morpheme, a surplus word-building element that does not represent any morpheme should not be regarded as a morph. Hence the appropriateness of the more neutral term **formative** for referring to any word-building element. Most formatives are morphs: they represent morphemes. And some are not. They are the so-called 'empty morphs'.

## 2.4 SUMMARY

The chapter opened with a discussion of the nature of the word (section (2.1)). We distinguished between lexemes, word-forms and grammatical words. Lexemes are abstract dictionary words like the verb SING. A lexeme is realised by one or more word-forms. Word-forms are concrete words that occur in speech and writing, e.g. *sing*, *sings*, *sang* and *sung*. We

also saw that the word can be viewed as a lexeme associated with a set of morpho-syntactic properties, e.g. *sing*<sub>[verb, present, 3rd person, singular]</sub>. In this case we are looking at a grammatical word.

The next section introduced the segmentation of words into the smallest abstract units of meaning or grammatical function. These units are called morphemes. We saw that the analysis of words into morphemes begins with the contrasting of pairs of utterances which are partially different in sound and meaning. Word-forms are segmented into morphs, which are recurrent physical word-forming chunks. Any morphs that represent the same meaning are grouped together as allomorphs of that morpheme. Meaning plays a role in this, but the main principle used is that of distribution. Morphs are listed as allomorphs of the same morpheme if they are in complementary distribution, i.e. if they are realisations of the same morpheme in different contexts. (Sometimes a morpheme has a single allomorph.)

Normally, the distribution of allomorphs is *phonologically conditioned*. The relationship between allomorphs has a phonetic motivation. A single underlying (base) form is postulated and the phonetic representation of the various allomorphs is derived from it using phonological rules.

But sometimes allomorphs may be grammatically conditioned or even lexically conditioned, i.e. a particular allomorph is selected if either a particular grammatical element or a particular word is present.

Occasionally there is suppletion, which means that an allomorph bears no phonetic similarity to other allomorphs of the same morpheme.

The last section dealt with the relationship between morphological and phonological representations. It was established that the relationship between morphemes on the one hand and morphs on the other, is one of representation (or realisation) rather than composition.

## EXERCISES

- Define and give one fresh example of each of the following:
  - lexeme; grammatical word; word-form.
  - morpheme; morph; allomorph; portmanteau morph; suppletion; empty morph; zero morph.
- (a) What is the allomorph of the plural morpheme that occurs in each group of words below?

- (b) Explain whether the choice of allomorph is phonologically, grammatically or lexically conditioned:
- (i) agenda data strata media desiderata  
(ii) stimuli radii fungi alumni
3. Study the following data and answer the questions that follow:
- dislike unwind report distrust uncover recover*  
*unable rewrite unlock landless disunited redraw*  
*ex-monk disallow penniless unhappy repel ex-coach*
- (a) What is the meaning of the morphemes represented in writing by *ex-*, *in-*, *dis-*, *un-*, *re-* and *-less*?
- (b) Are any of the meanings you recognise only of historical interest?
- (c) Comment on cases of homophony where a single morph represents more than one morpheme.
4. (a) Distinguish between phonological conditioning and grammatical conditioning of allomorphs.  
(b) Give one fresh example of each taken from any language that you know.
5. (a) Describe the tonal patterns found in the Rendille data below (from Oomen-van Schendel, 1977).  
(b) What are the functions of tone in these examples?
- |        |                                  |        |                                   |
|--------|----------------------------------|--------|-----------------------------------|
| nyíràx | 'young male camel'               | nyīrǎx | 'young female camel'              |
| hélèm  | 'ram'                            | hēlém  | 'sheep'                           |
| kélèh  | 'big male goat'                  | kēléh  | 'big male goat'                   |
| dúfàn  | 'camel kept for<br>slaughtering' | dūfán  | 'camels kept for<br>slaughtering' |
| náf    | 'a domestic animal'              | náf    | 'domestic animals'                |
- Note on tone marking: ´ marks High tone, ` marks Low tone, ˘ marks mid tone, ^ marks (High-Low) Falling tone.
6. Illustrating your answer with examples of your own, construct an argument for setting up underlying representations or base forms. (Read section (2.2.4) again before attempting this question.)

## 3 Types of Morphemes

### 3.1 ROOTS, AFFIXES, STEMS AND BASES

In the last chapter we saw that words have internal structure. This chapter introduces you to a wide range of word-building elements used to create that structure. We will start by considering roots and affixes.

#### 3.1.1 Roots

A root is the irreducible core of a word, with absolutely nothing else attached to it. It is the part that is always present, possibly with some modification, in the various manifestations of a lexeme. For example, *walk* is a root and it appears in the set of word-forms that instantiate the lexeme WALK such as *walk*, *walks*, *walking* and *walked*.

The only situation where this is not true is when suppletion takes place (see section (2.2.3)). In that case, word-forms that represent the same morpheme do not share a common root morpheme. Thus, although both the word-forms *good* and *better* realise the lexeme GOOD, only *good* is phonetically similar to GOOD.

Many words contain a root standing on its own. Roots which are capable of standing independently are called **free morphemes**, for example:

#### [3.1] Free morphemes

man	book	tea	sweet	cook
bet	very	aardvark	pain	walk

Single words like those in [3.1] are the smallest free morphemes capable of occurring in isolation.

The free morphemes in [3.1] are examples of **lexical morphemes**. They are nouns, adjectives, verbs, prepositions or adverbs. Such morphemes carry most of the 'semantic content' of utterances – loosely defined to cover notions like referring to individuals (e.g. the nouns *John*, *mother*), attributing properties (e.g. the adjectives *kind*, *clever*), describing actions, process or states (e.g. the verbs *hit*, *write*, *rest*) etc., expressing relations (e.g. the prepositions *in*, *on*, *under*) and describing circumstances like manner (e.g. *kindly*).

Many other free morphemes are **function words**. These differ from lexical morphemes in that while the lexical morphemes carry most of the 'semantic content', the function words mainly (but not exclusively) signal

- (b) Explain whether the choice of allomorph is phonologically, grammatically or lexically conditioned:
- (i) agenda data strata media desiderata  
(ii) stimuli radii fungi alumni
3. Study the following data and answer the questions that follow:
- dislike*    *unwind*    *report*    *distrust*    *uncover*    *recover*  
*unable*    *rewrite*    *unlock*    *landless*    *disunited*    *redraw*  
*ex-monk*    *disallow*    *penniless*    *unhappy*    *repel*    *ex-coach*
- (a) What is the meaning of the morphemes represented in writing by *ex-*, *in-*, *dis-*, *un-*, *re-* and *-less*?
- (b) Are any of the meanings you recognise only of historical interest?
- (c) Comment on cases of homophony where a single morph represents more than one morpheme.
4. (a) Distinguish between phonological conditioning and grammatical conditioning of allomorphs.  
(b) Give one fresh example of each taken from any language that you know.
5. (a) Describe the tonal patterns found in the Rendille data below (from Oomen-van Schendel, 1977).  
(b) What are the functions of tone in these examples?
- |        |                                  |        |                                   |
|--------|----------------------------------|--------|-----------------------------------|
| nyíràx | 'young male camel'               | nyīrǎx | 'young female camel'              |
| hélèm  | 'ram'                            | hēlém  | 'sheep'                           |
| kélèh  | 'big male goat'                  | kēléh  | 'big male goat'                   |
| dúfàn  | 'camel kept for<br>slaughtering' | dūfán  | 'camels kept for<br>slaughtering' |
| náf    | 'a domestic animal'              | náf    | 'domestic animals'                |
- Note on tone marking: ´ marks High tone, ` marks Low tone, ˘ marks mid tone, ^ marks (High-Low) Falling tone.
6. Illustrating your answer with examples of your own, construct an argument for setting up underlying representations or base forms. (Read section (2.2.4) again before attempting this question.)

## 3 Types of Morphemes

### 3.1 ROOTS, AFFIXES, STEMS AND BASES

In the last chapter we saw that words have internal structure. This chapter introduces you to a wide range of word-building elements used to create that structure. We will start by considering roots and affixes.

#### 3.1.1 Roots

A root is the irreducible core of a word, with absolutely nothing else attached to it. It is the part that is always present, possibly with some modification, in the various manifestations of a lexeme. For example, *walk* is a root and it appears in the set of word-forms that instantiate the lexeme WALK such as *walk*, *walks*, *walking* and *walked*.

The only situation where this is not true is when suppletion takes place (see section (2.2.3)). In that case, word-forms that represent the same morpheme do not share a common root morpheme. Thus, although both the word-forms *good* and *better* realise the lexeme GOOD, only *good* is phonetically similar to GOOD.

Many words contain a root standing on its own. Roots which are capable of standing independently are called **free morphemes**, for example:

#### [3.1] Free morphemes

man	book	tea	sweet	cook
bet	very	aardvark	pain	walk

Single words like those in [3.1] are the smallest free morphemes capable of occurring in isolation.

The free morphemes in [3.1] are examples of **lexical morphemes**. They are nouns, adjectives, verbs, prepositions or adverbs. Such morphemes carry most of the 'semantic content' of utterances – loosely defined to cover notions like referring to individuals (e.g. the nouns *John*, *mother*), attributing properties (e.g. the adjectives *kind*, *clever*), describing actions, process or states (e.g. the verbs *hit*, *write*, *rest*) etc., expressing relations (e.g. the prepositions *in*, *on*, *under*) and describing circumstances like manner (e.g. *kindly*).

Many other free morphemes are **function words**. These differ from lexical morphemes in that while the lexical morphemes carry most of the 'semantic content', the function words mainly (but not exclusively) signal

grammatical information or logical relations in a sentence. Typical function words include the following:

[3.2] Function words

articles:	a the
demonstratives:	this that these those
pronouns:	I you we they them; my your his hers; who whom which whose, etc.
conjunctions:	and yet if but however or, etc.

Distinguishing between lexical and grammatical morphemes is normally both useful and straightforward. However, there are cases where this distinction is blurred. This is because there are free morphemes (i.e. simple words) which do not fit neatly into either category. For example, a conjunction like *though* signals a logical relationship and at the same time appears to have considerably more 'descriptive semantic content' than, say, the article *the*.

While only roots can be free morphemes, not all roots are free. Many roots are incapable of occurring in isolation. They always occur with some other word-building element attached to them. Such roots are called **bound morphemes**. Examples of bound morphemes are given below:

[3.3] a.	-mit	as in permit, remit, commit, admit
b.	-ceive	as in perceive, receive, conceive
c.	pred-	as in predator, predatory, predation, depredate
d.	sed-	as in sedan, sedate, sedent, sedentary, sediment

The bound roots *-mit*, *-ceive*, *-pred* and *sed-* co-occur with forms like *de-*, *re-*, *-ate*, *-ment* which recur in numerous other words as prefixes or suffixes. None of these roots could occur as an independent word.

Roots tend to have a core meaning which is in some way modified by the affix. But determining meaning is sometimes tricky. Perhaps you are able to recognise the meaning 'prey' that runs through the root *pred-* in the various words in [3.3c] and perhaps you are also able to identify the meaning 'sit' in all the forms in [3.3d] which contain *sed-*.

These roots are **linate**, i.e. they came into English from Latin (normally via French). I suspect that, unless you have studied Latin, you are unable to say that *-mit* means 'send, do' and *-ceive* means 'take' without looking up *-mit* and *-ceive* in an etymological dictionary. In present-day English none of these meanings is recognisable. These formatives cannot be assigned a clear, constant meaning on their own.

In the last chapter the morpheme was defined as the smallest unit of meaning or grammatical function. In the light of the foregoing discussion,

the insistence on the requirement that every morpheme must have a clear, constant meaning (or grammatical function) seems too strong to some linguists. There are morphemes that lack a clear meaning. Instead, they suggest, it is the word rather than the morpheme that must always be independently meaningful whenever it is used. As we saw in section (2.2.1) above, the crucial thing about morphemes is not that they are independently meaningful, but that they are recognisable **distributional units** (Harris, 1951). As Aronoff (1976: 15) puts it, we can recognise a morpheme when we see a morph 'which can be connected to a linguistic entity outside that string. What is important is not its meaning, but its arbitrariness.'

The reason for treating those recurring portions of words that appear to lack a clear, constant meaning as morphemes representing some morpheme is that they behave in a phonologically consistent way in the language which is different from the behaviour of morphologically unrelated but phonologically similar sequences. Take *-mit*, for example. Aronoff (1976: 12-13) points out that, notwithstanding the tenuous semantic link between instances of all the linate root *-mit*, they nevertheless share a common feature which is not predictable from any properties of the phonetic sequence [mit]. All instances of linate *-mit* have the allomorph [mɪf] or [mɪs] before the suffixes *-ion*, *-ory*, *-or*, *-ive*, and *-able* / *-ible*, as you can see:

[3.4] linate root [mit]	[mɪf] before <i>-ion</i>	[mɪs] before <i>-ive</i> , <i>-ory</i>
permit	permission	permissive
submit	submission	submissive
admit	admission	admissive
remit	remission	remissory

By contrast, any other phonetic form [mit] (*-mit*) does not undergo the same phonological modification before such suffixes. Thus, although forms like *dormitory* and *vomitory* have a [mit] phonetic shape preceding the suffix *-ory*, they fail to undergo the rule that changes /t/ to [s]. If that rule applied, it would incorrectly deliver *\*dormissory* or *\*vomissory*, since the same phonetic sequence [mit] as that in [3.4a] precedes the suffix *-ory*. Clearly, the [mit] sequence in *vomitory* and *dormitory* is not a morph representing the linate *-mit* morpheme. The rule that supplies the allomorph [mɪs] of verbs that contain [mit] is only activated where [mit] represents the linate root *-mit*.

What this discussion shows is that even where the semantic basis for recognising a morpheme is shaky, there may well be distributional considerations that may save the day. Only linate *-mit* has the allomorph

[-mis-]. Any word-form that displays the [mit] ~ [mis] alternation in the contexts in [3.4] contains the latinate root morpheme *-mit*.

### 3.1.2 Affixes

An **affix** is a morpheme which only occurs when attached to some other morpheme or morphemes such as a root or stem or base. (The latter two terms are explained in (3.1.3) below.) Obviously, by definition **affixes** are bound morphemes. No word may contain only an affix standing on its own, like *\*-s* or *\*-ed* or *\*-al* or even a number of affixes strung together like *\*-al-s*.

There are three types of affixes. We will consider them in turn.

#### (i) Prefixes

A **prefix** is an affix attached *before* a root or stem or base like *re-*, *un-* and *in-*:

[3.5] re-make un-kind in-decent  
re-read un-tidy in-accurate

#### (ii) Suffixes

A **suffix** is an affix attached *after* a root (or stem or base) like *-ly*, *-er*, *-ist*, *-s*, *-ing* and *-ed*.

[3.6] kind-ly wait-er book-s walk-ed  
quick-ly play-er mat-s jump-ed

#### (iii) Infixes

An **infix** is an affix inserted into the root itself. Infixes are very common in Semitic languages like Arabic and Hebrew as we will see in section (3.6) below and in more detail in Chapter 9. But infixing is somewhat rare in English. Sloat and Taylor (1978) suggest that the only infix that occurs in English morphology is /-n-/ which is inserted before the last consonant of the root in a few words of Latin origin, on what appears to be an arbitrary basis. This infix undergoes place of articulation assimilation. Thus, the root *-cub-* meaning 'lie in, on or upon' occurs without [m] before the [b] in some words containing that root, e.g. *incubate*, *incubus*, *concupine* and *succubus*. But [m] is infixes before that same root in some other words like *incumbent*, *succumb*, and *decumbent*. This infix is a frozen historical relic from Latin.

In fact, infixation of sorts still happens in contemporary English. Consider the examples in [3.7a] which are gleaned from Zwicky and Pullum (1987) and those in [3.7b] taken from Bauer (1983):

[3.7] a. Kalamazoo (place name) → Kalama-goddam-zoo  
instantiate (verb) → in-fuckin-stantiate

b. kangaroo → kanga-bloody-roo  
impossible → in-fuckin-possible  
guarantee → guaran-friggin-tee

(Recall that the arrow → means 'becomes' or is 're-written as'.)

As you can see, in present-day English infixation, not of an affix morpheme but of an entire word (which may have more than one morpheme, e.g. *bloody-y*, *fuck-ing*) is actively used to form words. Curiously, this infixation is virtually restricted to inserting expletives into words in expressive language that one would probably not use in polite company.

### 3.1.3 Roots, Stems and Bases

The stem is that part of a word that is in existence before any *inflectional* affixes (i.e. those affixes whose presence is required by the syntax such as markers of singular and plural number in nouns, tense in verbs etc.) have been added. Inflection is discussed in section (3.2). For the moment a few examples should suffice:

[3.8]	<u>Noun stem</u>	<u>Plural</u>
	cat	-s
	worker	-s

In the word-form *cats*, the plural inflectional suffix *-s* is attached to the simple stem *cat*, which is a bare root, i.e. the irreducible core of the word. In *workers* the same inflectional *-s* suffix comes after a slightly more complex stem consisting of the root *work* plus the suffix *-er* which is used to form nouns from verbs (with the meaning 'someone who does the action designated by the verb (e.g. *worker*)'). Here *work* is the root, but *worker* is the stem to which *-s* is attached.

Finally, a **base** is any unit whatsoever to which affixes of any kind can be added. The affixes attached to a base may be **inflectional affixes** selected for syntactic reasons or **derivational affixes** which alter the meaning or grammatical category of the base (see sections (3.2) and (10.2)). An unadorned root like *boy* can be a base since it can have attached to it inflectional affixes like *-s* to form the plural *boys* or derivational affixes like *-ish* to turn the noun *boy* into the adjective *boyish*. In other words, all **roots** are **bases**. Bases are called **stems** only in the context of inflectional morphology.

Identify the inflectional affixes, derivational affixes, roots, bases, and stems in the following:

[3.9]	faiths	frogmarched
	faithfully	bookshops
	unfaithful	window-cleaners
	faithfulness	hardships

I hope your solution is like this:

[3.10]	<u>Inflectional</u> <u>Affixes</u>	<u>Derivational</u> <u>Affixes</u>	<u>Roots</u>	<u>Stems</u>	<u>Bases</u>
	-ed	un-	faith	faith	faith
	-s	-ful	frog	frogmarch	faithful
		-ly	march	bookshop	frogmarch
		-er	clean	windowcleaner	bookshop
		-ness	hard	hardship	window-clean
		-ship	window		window-cleaner
					hardship

It is clear from [3.10] that it is possible to form a complex word by adding affixes to a form containing more than one root. For instance, the independent words *frog* and *march* can be joined together to form the base (a stem, to be precise) *frog-march* to which the suffix *-ed* may be added to yield *[[frog]-[march]-ed]*. Similarly, *window* and *clean* can be joined to form the base *[[window]-[clean]]* to which the derivational suffix *-er* can be added to produce *[[[window]-[clean]]er]*. And *[[[Window]-[cleaner]]]* can serve as a stem to which the inflectional plural ending *-s* is attached to give *[[[[Window]-[cleaner]]]s]*. A word like this which contains more than one root is called **compound word** (see section (3.4) below and Chapter 12).

### 3.1.4 Stem Extenders

In section (2.3) of the last chapter we saw that languages sometimes have word-building elements that are devoid of content. Such empty formatives are often referred to, somewhat inappropriately, as empty morphs.

In English, empty formatives are interposed between the root, base or stem and an affix. For instance, while the irregular plural allomorph *-en* is attached directly to the stem *ox* to form *ox-en*, in the formation of *child-r-nen* and *breth-r-en* it can only be added after the stem has been extended by

attaching *-r-* to *child-* and *breth-*. Hence, the name **stem extender** for this type of formative.

The use of stem extenders may not be entirely arbitrary. There may be a good historical reason for the use of particular stem extenders before certain affixes. To some extent, current word-formation rules reflect the history of the language.

The history of stem extender *-r-* is instructive. A small number of nouns in Old English formed their plural by adding *-er*. The word 'child' was *cild* in the singular and *cilder* in the plural (a form that has survived in some conservative North of England dialects, and is spelled *childer*). But later, *-en* was added as an additional plural ending. Eventually *-er* lost its value as a marker of plural and it simply became a stem extender:

[3.11]	a. <u>Singular</u>	<u>Plural</u>	b. <u>New singular</u>	<u>Plural</u>
	b. cild 'child'	cild-er	cilder	cilder-en

## 3.2 INFLECTIONAL AND DERIVATIONAL MORPHEMES

As we have already hinted, affix morphemes can be divided into two major functional categories, namely **derivational morphemes** and **inflectional morphemes**. This reflects a recognition of two principal word building processes: **inflection** and **derivation**. While all morphologists accept this distinction in some form, it is nevertheless one of the most contentious issues in morphological theory. I will briefly introduce you here to the essentials of this distinction but postpone detailed discussion until Chapter 10.

Inflectional and derivational morphemes form words in different ways. Derivational morphemes form new words either:

- (i) by changing the meaning of the base to which they are attached, e.g. *kind* vs *un-kind* (both are adjectives but with opposite meanings); *obey* vs *dis-obey* (both are verbs but with opposite meanings). Or
- (ii) by changing the word-class that a base belongs to, e.g. the addition of *-ly* to the adjectives *kind* and *simple* produces the adverbs *kind-ly* and *simp-ly*. As a rule, it is possible to derive an adverb by adding the suffix *-ly* to an adjectival base.



Study the following data and answer the questions below:

- [3.12] I ducked                      He was sheepish  
           two ducks                    three ducklings  
           He is humourless        You are ducking the issue  
           He ducks

- a. Identify the suffixes in the underlined words. To what word-class do the words to which the suffixes are added belong, and what word-class results?  
 b. For each suffix determine whether it is inflectional or derivational. Briefly justify your decision.

I hope your answer is very close to the following:

[3.13]	Suffix	Input	Output	Remarks
a.	-ed	V	V	inflectional:- it marks past tense in <i>duck-ed</i>
	-s	N	N	inflectional:- it marks plural number (in <i>(two) duck-s</i> and <i>duckling-s</i> )
	-s	V	V	inflectional:- a portmanteau morph marking 3rd person, present tense and singular in <i>(he) ducks</i>
	-ing	V	V	inflectional:- it marks progressive aspect (i.e. incomplete action in <i>ducking</i> )
b.	-ling	N	N	derivational:- it changes meaning to 'small duck'
	-ish	N	Adj	derivational:- it changes word-class and meaning to 'like a sheep'
	-less	N	Adj	derivational:- it turns a noun into an adjective and adds the meaning 'lacking (e.g. humour)'

Sometimes the presence of a derivational affix causes a major grammatical change, involving moving the base from one word-class into another as in the case of *-less* which turns a noun into an adjective. In other cases, the change caused by a derivational suffix may be minor. It may merely shift a base to a different sub-class within the same broader word-class. That is what happens when the suffix *-ling* is attached to *duck* above.

Further examples are given below. In [3.14a] the diminutive suffix *-let* is attached to nouns to form diminutive nouns (meaning a small something).

In [3.14b] the derivational suffix *-ship* is used to change a concrete noun base into an abstract noun (meaning 'state, condition'):

- [3.14] a. pig ~ pig-let    b. friend ~ friend-ship  
           book ~ book-let    leader ~ leader-ship

The tables in [3.15] and [3.16] list some common derivational prefixes and suffixes, the classes of the bases to which they can be attached and the words that are thereby formed. It will be obvious that in order to determine which morpheme a particular affix morph belongs to, it is often essential to know the base to which it attaches because the same phonological form may represent different morphemes depending on the base with which it co-occurs.

Note: These abbreviations are used in the tables below: N for noun, N (abs) for abstract noun, N (conc) for concrete noun. V for verb, Adj for adjective, and Adv for adverb.

[3.15]	Prefix	Word-class of input base	Meaning	Word-class of output word	Example
	in-	Adj	'not'	Adj	in-accurate
	un-	Adj	'not'	Adj	un-kind
	un-	V	'reversive'	V	un-tie
	dis-	V	'reversive'	V	dis-continue
	dis-	N (abs)	'not'	N (abs)	dis-order
	dis-	Adj	'not'	Adj	dis-honest
	dis-	V	'not'	V	dis-approve
	re-	V	'again'	V	re-write
	ex-	N	'former'	N	ex-mayor
	en-	N	'put in'	V	en-cage

[3.16]	Suffix	Word-class of input base	Meaning	Word-class of output word	Example
	-hood	N	'status'	N (abs)	child-hood
	-ship	N	'state or condition'	N (abs)	king-ship
	-ness	Adj	'quality, state or condition'	N (abs)	kind-ness
	-ity	Adj	'state or condition'	N (abs)	sincer-ity
			etc.		
	-ment	V	'result or product of N doing the action indicated by the verb'	N	govern-ment

-less	N	'without'	Adj	power-less
-ful	N	'having'	Adj	power-ful
-ic	N	'pertaining to'	Adj	democrat-ic
-al	N	'pertaining to, of the kind'	Adj	medicin-al
-al	V	'pertaining to or act of'	N (abs)	refus-al
-er	V	'agent who does whatever the verb indicates'	N	read-er
-ly	Adj	'manner'	Adv	kind-ly

To sum up the discussion so far, we have observed that derivational affixes are used to create new lexemes by either: (i) modifying significantly the meaning of the base to which they are attached, without necessarily changing its grammatical category (see *kind* and *unkind* above); or (ii) they bring about a shift in the grammatical class of a base as well as a possible change in meaning (as in the case of *hard* (Adj) and *hardship* (N (abs))); or (iii) they may cause a shift in the grammatical sub-class of a word without moving it into a new word-class (as in the case of *friend* (N (conc)) and *friend-ship* (N (abs))).

With that in mind, study the data below which contain the derivational prefix *en-*.

- (i) State the word-classes (e.g. noun, adjective, verb, etc.) of the bases to which *en-* is prefixed.
- (ii) What is the word-class of the new word resulting from the prefixation of *en-* in each case?
- (iii) What is the meaning (or meanings) of *en-* in these words?

Consult a good dictionary, if you are not sure. Is there reason to regard *en-* as a homophonous morph?

[3.17]	Base	New word	Base	New word
	cage	en-cage	noble	en-noble
	large	en-large	rich	en-rich
	robe	en-robe	rage	en-rage
	danger	en-danger	able	en-able

You will have established that the new word resulting from the prefixation of *en-* in [3.17] is a verb. But there is a difference in the input bases.

Sometimes *en-* is attached to adjectives as seen in [3.18a], and sometimes to nouns, as in [3.18b]:

[3.18]	a.	Adj base	New word	Verb	b.	Noun base	New word	verb
		able	en-able			robe	en-robe	
		large	en-large			danger	en-danger	
		noble	en-noble			rage	en-rage	
		rich	en-rich			cage	en-cage	

Interestingly, this formal difference correlates with a semantic distinction. So, we conclude that there are two different prefixes here which happen to be homophonous. The *en-* in [3.18a] has a causative meaning (similar to 'make'). To *enable* is to 'make able', to *enlarge* is to 'make large', etc., while in [3.18b] *en-* can be paraphrased as 'put in or into'. To *encage* is to 'put in a cage' and to *endanger* is to 'put in danger' etc.

Let us now turn to inflectional morphemes. Unlike derivational morphemes, inflectional morphemes do not change referential or cognitive meaning. We have already seen that a derivational affix like *un-* can change kind into *un-kind*. In this case, the derived word has a meaning which is opposite to that of the input. The addition of an inflectional affix will not do such a thing. Furthermore, while a derivational affix may move a base into a new word-class (e.g., *kind* (adjective) but *kind-ly* (adverb)), an inflectional morpheme does not alter the word-class of the base to which it is attached. Inflectional morphemes are only able to modify the form of a word so that it can fit into a particular syntactic slot. Thus, *book* and *books* are both nouns referring to the same kind of entity. The *-s* ending merely carries information about the number of those entities. The grammar dictates that a form marked as plural (normally by suffixing *-s*) must be used when more than one entity is referred to. We must say *ten books*; *\*ten book* is ruled out, although the numeral ten makes it clear that more than one item is being referred to.

See the table in [3.19] for a sample of frequently used inflectional suffixes. English has no inflectional prefixes but some other languages do (see Luganda in [2.7]).

[3.19]	Suffix	Stem	Function	Example
	-s	N	plural	book-s
	-s	V	3rd person, singular, present tense	sleep-s
	-ed	V	past tense	walk-ed
	-ing	V	progressive (incomplete action)	walk-ing
	-er	Adj	comparative degree	tall-er
	-est	Adj	superlative degree	tall-est

Below I have presented an additional inflectional suffix. What is this suffix called and what is its function in each example?

- [3.20] a. Janet's book  
 b. The Winter's Tale  
 c. in two days' time

The *-s* suffix in [3.20] is usually called the **genitive suffix**. Quirk and Greenbaum (1973) list these, among others, as the uses of the genitive suffix in English:

- (i) marking the noun referring to the possessor of something (as in *Janet's book*),  
 (ii) marking a noun that describes something (as in *The Winter's Tale*),  
 (iii) marking a noun used as a measure (*in two days' time*).

We will return this and refine our analysis of genitive *-s* in (section 10.5).

### 3.3 MULTIPLE AFFIXATION

What we are now going to explore are some of the ways in which complex words are formed by creating bases which contain several derivational morphemes. Let us take the Latin root *-dict-* meaning 'speak, say' which is found in *diction, dictate, dictatorial, contradict, benediction, etc.* Starting with *-dict-*, we can form complex words such as *contradictory* and *contradictoriness* by attaching several affixes to the root, i.e. we can have **multiple affixation**. This process can take place in a number of rounds, with the output created by one round of affixation serving as the input to a later round:

[3.21]	Root	<i>-dict<sub>v</sub></i>	Output
	base:	<i>-dict<sub>v</sub></i>	(round one: prefixation: → <i>contradict<sub>v</sub></i> )
			add <i>contra</i> -Preposition)
	base:	<i>contradict<sub>v</sub></i>	(round two: first → <i>contradict-ory<sub>Adj</sub></i> )
			suffixation: add <i>-ory<sub>Adj</sub></i> )
	base:	<i>contradictory<sub>Adj</sub></i>	(round three: second → <i>contradictoriness<sub>N</sub></i> )
			suffixation: add <i>-ness<sub>N</sub></i> )

Words may have multiple affixes either with different suffixes appearing

in a sequence as in [3.21] or with the same prefix recurring as below in [3.22].

- [3.22] a. the latest re-re-re-make of *Beau Geste*.  
 b. the great-great-great-great grandson of the last Tsar of Russia.

What [3.22] shows is that, with a limited number of morphemes, morphological prefixation rules can apply recursively in English (see section 1.3). However, **performance** difficulties in working out what exactly *great-great-great-great grandson* or *re-re-re-make* means do severely restrict the chances of such words being used. But the point is that the grammar cannot exclude them as ill-formed. Recursive rules are one of the devices that make morphology open-ended. They make possible the creation of new words with the same morphemes being used again and again (cf. section 4.1).

Re-attaching the same morpheme again and again is permitted, but unusual. What is common is multiple affixation of different affixes. It is such affixation that we will concentrate on. We have already seen an example of it in *contradict-ori-ness* in [3.21].

Take the free morpheme *nation* and add to it as many prefixes and suffixes as you can. Attempt to go through at least four rounds of affixation.

I hope you have come up with something like this:

- [3.23] nation  
 nation-al  
 national-ise  
 denationalis-at-ion  
 anti-denationalisation  
 pre-antidenationalisation  
 de-nationalise  
 (but there is no \*denationalisate)

Observe that where several prefixes or suffixes occur in a word, their place in the sequence is normally rigidly fixed. Whereas there is usually some scope for rearranging words in different orders in sentences, as you can see:

- [3.24] a. You can play badminton.    b. What I need is a nice cup of tea.  
 Can you play badminton?    A nice cup of tea is what I need.

there is virtually no possibility of re-arranging morphemes within a word.

So, for example, the morphemes in *de-nation-al-ise* must appear in that order. Rearranging the affixes produces ill-formed strings like \**ise-nation-de-al-* or \**al-ise-nation-de*. The main problem and interest, as we will see in section (6.2.1), is determining the order of derivational affixes where several of them occur in a word.

### 3.4 COMPOUNDING

As we briefly saw in (3.1.3), a **compound word** contains at least two bases which are both words, or at any rate, root morphemes.

Analyse the following compounds into their constituent elements: *teapot*, *week-end*, *hairdresser*, *kind-hearted*.

I expect you to have worked out an answer close to the following:

- [3.25] a. [tea]<sub>N</sub> [pot]<sub>N</sub> → [teapot]<sub>N</sub>  
           [week]<sub>N</sub> [end]<sub>N</sub> → [week-end]<sub>N</sub>  
       b. [hair]<sub>N</sub> [[dress]<sub>V</sub>-er]<sub>N</sub> → [hairdresser]<sub>N</sub>  
           [kind]<sub>A</sub> [[heart]<sub>N</sub>-ed]<sub>A</sub> → [open-ended]<sub>A</sub>

Compounding is a very important way of adding to the word stock of English as we will see. Sometimes it is bare roots that are combined in compounds as in [3.25a], and sometimes an input base contains an affixed form as in [3.25b]. We will discuss compounds again in a preliminary way in the next chapter and return to them in more detail in Chapter 12.

### 3.5 CONVERSION

We have seen that complex words may be formed either by compounding or by affixation, or by a combination of the two. We are going to see now that there is an alternative word-formation strategy which is commonly used in English. Words may be formed without modifying the form of the input word that serves as the base. Thus *head* can be a noun or verb. This is called **conversion**.

How do you know whether *head* is a noun or verb in the following?

- [3.26] a. The head of the village school has arrived.  
           The heads of the village schools have arrived.  
       b. She will head the village school.  
           She headed that school.

It is partly the morphological structure, and partly the syntactic position that the word occupies that tells you whether it is a noun or a verb. From a syntactic point of view, we know that in [3.26a] *the head* is a noun phrase. The key word in a noun phrase must be a noun. As *head* occurs following *the* and is the key word in this construction, *head* must be a noun. But from a morphological point of view, we cannot tell whether *head*, is a noun or verb when it occurs with no affixes. However, in the case of *heads*, the presence of the *-s* morph which here realises the plural in nouns gives us a useful clue.

By contrast, in [3.26b] *head* must be a verb. It comes after the auxiliary verb *will* in a slot that is typically filled by verbs. In the second example, *head* has attached to it the *-ed* morph representing the past tense morpheme which is only found in verbs. Furthermore, from a syntactic point of view, we know that *she* is the subject and *that school* is the object. The sentence must also have a verb. The verb occurs between the subject and the object. (The order of sentence constituents in English is Subject Verb Object.) So, *headed* must be the verb, since it occurs between the subject and the object.

Conversion is also referred to as **zero derivation** in the literature (cf. Marchand, 1969; Adams, 1973) and is subsumed under affixation, by analogy to zero affixation in inflectional morphology (cf. section (2.3)). It is claimed that zero morphs (i.e. ones lacking any overt marking) are used as suffixes in derivational morphology as well. For instance, the verb *head* is derived by suffixing a zero morph to the noun *head*. This is done by analogy to the derivation of a verb like *victim-ise<sub>V</sub>* (from the noun *victim<sub>N</sub>* where the overt verb-forming suffix *-ise* is used).

The use of zero in derivational morphology is controversial. Since neither the original noun *head*, nor the derived verb *head*, has an overt suffix, if we assume that zero suffixation takes place here, we end up with a somewhat absurd situation where a zero suffix on the noun is said to contrast with a zero suffix on the derived verb. It is more prudent to recognise conversion as a distinct word-forming mechanism and to restrict zero morphs to inflectional morphology where it is supported by the evidence. See section (6.2.3) for further discussion.

## 3.6 MORPHOLOGICAL TYPOLOGY

We suggested in the opening chapter that although languages vary enormously in their structure they nonetheless show surprising similarities. The study of the significant shared structural properties which languages have in common is the domain of **language universals**. Many of the universals are abstract principles of **Universal Grammar** which determine the properties of rules that grammars of individual languages may have (e.g. the Strict Cycle Condition discussed in section (6.2.4)).

An integral part of the study of universals in language is the study of differences between languages. This might look odd to begin with. But it turns out that differences between the structural patterns found in different languages appear to occur within a fairly restricted range. There are **parameters** within which most differences between languages occur. Just as tram lines determine where trams can go in a city (while leaving them plenty of options), pre-set parameters determine the structural patterns from which different languages may select.

Structural patterns are not randomly distributed. There are a number of strongly preferred patterns which recur in language after language, while other patterns are rare, or non-existent (Greenberg, 1963; Comrie, 1981; and especially Chomsky, 1986). The study of the range of patterns within which languages may vary is the domain of **language typology**.

Our concern in this book is with both the similarities and differences between languages in the ways in which they form words. On the basis of typical patterns of word-formation linguists recognise five broad morphological types:

- (i) **analytic** (also called isolating) languages;
- (ii) **agglutinating** (also called agglutinative) languages;
- (iii) **inflecting** (also called synthetic or fusional) languages;
- (iv) **incorporating** (also called polysynthetic) languages;
- (v) **infixing** languages.

We will now consider the morphological types in turn, starting with examples of analytic morphology from Chinese:

- [3.27] a. Tā bǎ shū mǎi le.  
he OM book buy Asp.  
'He bought the book.'
- b. Tā chǎo le yíge cài hěn xiāng.  
he cook Asp. a dish very delicious.  
'He cooked a dish that was very delicious.'

Note: Asp. is short for 'perfective aspect'. It indicates that an action is

completed. OM is short for 'object marker', i.e. the morpheme that indicates the object of the verb.

(data from Li and Thompson, 1978)

As you can see from the morpheme by morpheme translation, in Chinese bound morphemes are infrequent. Usually the words are bare, unaffixed root morphemes.

Chinese is an example of an **analytic language**, i.e. a language where each morpheme tends to occur as a word in isolation. Words virtually never have inflectional affixes. Thus, the object marker *bǎ* is an independent word. By contrast, in other language types normally object markers are inflectional affixes that are part of a noun or pronoun. In English the subject pronoun *he* contrasts with the object pronoun *him* in *He saw Lauren vs Lauren saw him*. The change from *he* to *him* in the pronoun marks the change in grammatical function. Similarly, in English markers of aspect and tense are usually inflectional affixes of the verb such as *-ed*, as in *cook-ed* (vs *cook*). By contrast, in Chinese, in [3.27b] the aspectual morpheme is realised, not by an affix, but by the independent word *le*. (Note in passing that there are some Chinese words containing more than one morpheme. Usually they are compounds like *jue-she* (literally 'chew-tongue'), 'gossip' (noun) and *zhen-tou* ('pillow', literally 'rest-head').

Let us now turn to another language, Turkish.

- a. Divide the following words into morphs and assign each morph to a morpheme.
- b. How do the morphs match up with morphemes?

[3.28]	el	'the hand'	elimde	'in my hand'
	elim	'my hand'	ellerim	'my hands'
	eller	'the hands'	ellerimde	'in my hands'

Your answer to the first question should be: *el* 'hand', *-im* 'my' (genitive), *-ler* 'plural' and *-de* 'in'.

Turkish is a classic example of an **agglutinating** language. In this kind of language there tends to be a more or less one-to-one matching of morphemes with morphs:

## 3.6 MORPHOLOGICAL TYPOLOGY

We suggested in the opening chapter that although languages vary enormously in their structure they nonetheless show surprising similarities. The study of the significant shared structural properties which languages have in common is the domain of **language universals**. Many of the universals are abstract principles of **Universal Grammar** which determine the properties of rules that grammars of individual languages may have (e.g. the Strict Cycle Condition discussed in section (6.2.4)).

An integral part of the study of universals in language is the study of differences between languages. This might look odd to begin with. But it turns out that differences between the structural patterns found in different languages appear to occur within a fairly restricted range. There are **parameters** within which most differences between languages occur. Just as tram lines determine where trams can go in a city (while leaving them plenty of options), pre-set parameters determine the structural patterns from which different languages may select.

Structural patterns are not randomly distributed. There are a number of strongly preferred patterns which recur in language after language, while other patterns are rare, or non-existent (Greenberg, 1963; Comrie, 1981; and especially Chomsky, 1986). The study of the range of patterns within which languages may vary is the domain of **language typology**.

Our concern in this book is with both the similarities and differences between languages in the ways in which they form words. On the basis of typical patterns of word-formation linguists recognise five broad morphological types:

- (i) **analytic** (also called isolating) languages;
- (ii) **agglutinating** (also called agglutinative) languages;
- (iii) **inflecting** (also called synthetic or fusional) languages;
- (iv) **incorporating** (also called polysynthetic) languages;
- (v) **infixing** languages.

We will now consider the morphological types in turn, starting with examples of analytic morphology from Chinese:

- [3.27] a. Tā bǎ shū mǎi le.  
he OM book buy Asp.  
'He bought the book.'
- b. Tā chǎo le yíge cài hěn xiāng.  
he cook Asp. a dish very delicious.  
'He cooked a dish that was very delicious.'

Note: Asp. is short for 'perfective aspect'. It indicates that an action is

completed. OM is short for 'object marker', i.e. the morpheme that indicates the object of the verb.

(data from Li and Thompson, 1978)

As you can see from the morpheme by morpheme translation, in Chinese bound morphemes are infrequent. Usually the words are bare, unaffixed root morphemes.

Chinese is an example of an **analytic language**, i.e. a language where each morpheme tends to occur as a word in isolation. Words virtually never have inflectional affixes. Thus, the object marker *bǎ* is an independent word. By contrast, in other language types normally object markers are inflectional affixes that are part of a noun or pronoun. In English the subject pronoun *he* contrasts with the object pronoun *him* in *He saw Lauren vs Lauren saw him*. The change from *he* to *him* in the pronoun marks the change in grammatical function. Similarly, in English markers of aspect and tense are usually inflectional affixes of the verb such as *-ed*, as in *cook-ed (vs cook)*. By contrast, in Chinese, in [3.27b] the aspectual morpheme is realised, not by an affix, but by the independent word *le*. (Note in passing that there are some Chinese words containing more than one morpheme. Usually they are compounds like *jue-she* (literally 'chew-tongue'), 'gossip' (noun) and *zhen-tou* ('pillow', literally 'rest-head').

Let us now turn to another language, Turkish.

- a. Divide the following words into morphs and assign each morph to a morpheme.
- b. How do the morphs match up with morphemes?

[3.28]	el	'the hand'	elimde	'in my hand'
	elim	'my hand'	ellerim	'my hands'
	eller	'the hands'	ellerimde	'in my hands'

Your answer to the first question should be: *el* 'hand', *-im* 'my' (genitive), *-ler* 'plural' and *-de* 'in'.

Turkish is a classic example of an **agglutinating language**. In this kind of language there tends to be a more or less one-to-one matching of morphemes with morphs:

English is predominantly isolating. The vast majority of the 45 words in this sentence, which is typical of modern English, are simple. They contain just one morpheme. But English is not a thoroughbred isolating language. Five of the words, namely *year-s*, *precise-ly*, *hav-ing*, *no-thing*, and *water-y* contain two morphs representing two distinct morphemes. These words exemplify a degree of agglutination. In addition, there are also several words containing one morph which represents several morphemes concurrently, e.g. *me* (1st person, singular, accusative pronoun); *my* (1st person, singular, possessive pronoun), *I* (1st person, singular, nominative pronoun), *thought* (THINK, past) and *would* (WILL, past). In words like this, trying to designate a portion of the word as a morph representing one of the morphemes would be futile. Such words show that, to a certain extent, English is a synthetic language. Even infixation (which is not exemplified by [3.33]) is found occasionally in English, as in *incumbent*, *succumb*, and *decumbent*, where *-m-* is infixed in the root *-cub-* (see p. 44 above).

Greenberg (1954) made a proposal regarding *typology* that is widely accepted. He suggested that the number of morphemes in a representative sample of sentences should be divided by the number of words to work out the ratio of morphemes to words in a language. The result should form the basis of our typological classification.

- (i) If a language has between 1.00 and 1.99 morphemes per word it is **analytic** (isolating). With 1.68 morphemes per word in Greenberg's sample of sentences, English falls in the essentially isolating category. (It is similar to Chinese – see [3.27].)
- (ii) A language averaging between 2.00 and 2.99 morphemes per word, is **synthetic** (inflecting) if the realisation of the different morphemes tends to be simultaneous (as in Latin – see [3.30]).
- (iii) A language averaging between 2.00 and 2.99 morphemes per word is **agglutinative** if each morpheme tends to be realised by a separate morph (as in Turkish in [3.28]).
- (iv) A language is incorporating if it averages 3.00 morphemes per word or more (e.g. Eskimo – see [3.31]).

It is important to realise that probably no language has an unalloyed analytic, agglutinating, inflecting or incorporating morphological system. All that the classification attempts to do is reflect the dominant tendencies found in a particular language.

### 3.7 WP AND THE CENTRALITY OF THE WORD

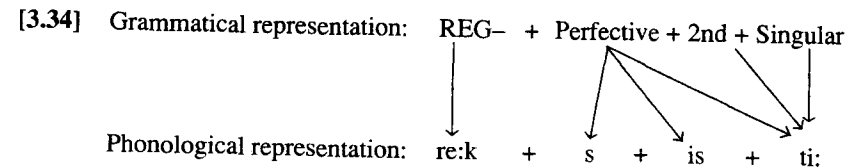
A central question which morphological theory needs to address is 'what is the key unit which morphological theory deals with?' In structuralist

morphology the answer was unequivocally, 'the morpheme'. However, in recent years, various scholars have proposed that it is not the morpheme but rather the **word** that should be regarded as the central unit of morphological analysis. This debate has important repercussions for how we formulate our theory of morphology and the lexicon.

**Word-and-paradigm morphology** (WP) is one theory that puts the word at the centre. It was first mentioned in modern linguistics by Hockett (1954) who identified it as the approach assumed in traditional grammars based on Latin. This model was articulated in Robins (1959) and extensively revised by Matthews (1972). It has since been elaborated by S. R. Anderson (1977, 1982, 1984, 1988a). Unfortunately, in spite of its inherent merits, this approach has not been adopted by many linguists.

But although there are not many WP morphologists, the critique of morpheme-based approaches to morphology which this theory embodies has contributed to a healthy re-examination of the nature of morphological representations in recent years. WP is critical of the somewhat naïve view of the relationship between morphological representations and morphs found in some structuralist models of morphology. Matthews (1972) has shown that a theory of the morpheme that relies on the assumption that morphemes are always typified by a one-to-one pairing of morphemes with morphs is misguided. True, in straightforward cases of agglutination like the Turkish example in [3.29], a bit of the phonological representation may directly correspond to a bit of the morphological representation. But the phenomenon of **portmanteau morphs** that is found frequently in inflecting languages illustrates the difficulties that arise if morphemes are assumed to be always matched in a straightforward way with morphs.

Matthews (1972: 132) suggests that the Latin word /re:ksisti:/ 'you (sg.) ruled (or I have ruled)' could be analysed as in [3.34]:



The morphemes second person and singular are both realised by the portmanteau morph *ti:* while the perfective is multiply signalled, partly in the selection of *rek-* (see below) and by the suffixes *-s-*, *-is-* and *-ti:*. The justification for this analysis will be clear if you compare parts of the perfect and imperfect forms of the verb *regere* 'rule':

<u>Imperfect</u>		<u>Perfect</u>	
rege:bam	'I was ruling'	re:ksi:	'I have ruled'
rege:ba:s	'you (sing.) were ruling'	re:ksisti:	'you (sing.) have ruled'
rege:bat	'he was ruling'	re:ksit	'he has ruled'

If you examine the second person singular forms, for example, you observe that the root REG- has the phonological realisation /reg-/ in the imperfect but /rek-/ in the perfective. So, the distinction between perfective and imperfective is in part realised in the root itself. (See the diagram in [3.34].) The ending /-ti:/ marks second person singular if the grammatical representation also includes the perfective. If the verb is in the imperfective, the second person singular is marked instead by /a:s/. The crucial point is that these various morphs do not have a clear identifiable meaning on their own. They can only be interpreted in the wider context of the word as a whole of which they form a part. To know how second person singular is going to be realised we need to take into account the rest of the grammatical representation manifested in a particular word. A sensible solution, and one that WP morphology advocates, is one that recognises a combination of morphs as simultaneously signalling a particular meaning if they co-occur in a word that has a certain combination of grammatical properties.

I will not introduce you to the formalism of WP because that formalism is not important for the generative theory of morphology that I am outlining. If you wish to see WP rules, turn to Matthews (1972) and S. R. Anderson (1982). My aim has been to show that while morphemes are important theoretical entities, the word is the key unit of morphological representation. While still recognising the relevance of morphemes, present-day morphological theory in generative grammar is word-based. The pivotal role of the word will become especially obvious in Part II of the book.

## EXERCISES

1. Examine carefully the following sentence:

Mr Nickleby shook his head, and motioning them all out of the room, embraced his wife and children, and having pressed them by turns to his languidly beating heart, sunk exhausted on his pillow.

(Charles Dickens, *Nicholas Nickleby*)

- (a) List five free and three bound morphemes that occur in this sentence.  
 (b) List three functional morphemes in the sentence.
2. (a) Identify the morphemes in the Swahili words below, distinguishing between roots and affixes.  
 (b) State the meaning of each morpheme.  
 (c) State whether the affix morphemes are: (i) prefixes or suffixes, and (ii) inflectional or derivational.  
 (d) On the basis of these data, would you classify Swahili as an isolating, agglutinating, synthetic or incorporating language?

nilipata	'I got'	niliwapiga	'I hit them'
walipata	'they got'	walitupiga	'they hit us'
nilipiga	'I hit'	walikipiga	'they hit it'
nilikipata	'I got it'	utatupiga	'you will hit us'
ulikipata	'you got it'	ulipata	'you got'
nitakipata	'I will get it'	watakupiga	'they will hit you'
ulipiga	'you hit'	ulitupiga	'you hit us'
watakupiga	'they will hit it'	nitakupata	'I will get you'

Note: Here the form 'hit' as in 'you hit' represents the past tense form of the verb *hit* and 'you' stands for 'second person singular'.

3. (a) Make a morphological analysis of the following Latin data:

<u>Present tense</u>		<u>Pluperfect</u>	
regō	'I rule'	rēkseram	'I had ruled'
regis	'you (sing.) rule'	rēkserās	'you (sing.) had ruled'
regit	's/he rules'	rēkserat	's/he had ruled'
regimus	'we rule'	rēkserāmus	'we had ruled'
regitis	'you (pl.) rule'	rēkserātis	'you (pl.) had ruled'
regunt	'they rule'	rēkserant	'they had ruled'

### Future simple

regam	'I shall rule'
regēs	'you (sing.) shall rule'
reget	's/he will rule'
regēmus	'we will rule'
regētis	'you (pl.) will rule'
regent	'they will rule'

- (b) Referring to your analysis, highlight the pitfalls of a theory of word-structure that assumes that there is always a one-to-one matching of morphs with morphemes.



4. (a) What is the morphological function of tone in the Lulubo words below?  
 (b) State exactly how tone is used to perform this function.  
 (c) Explain whether or not Lulubo fits in the morphological typology given in this chapter.

àz̄	'long'	àz̄	'to become long'
inḍá	'good'	inḍà	'to become good'
òsú	'good'	òsù	'to become good'
álí	'deep'	àlì	'to become deep'
àkèlí	'red'	àkèlì	'to become red'
áfórò	'yellow'	àfòrò	'to become yellow'

(data from Andersen, 1987)

## 4 Productivity in Word-Formation

### 4.1 THE OPEN-ENDEDNESS OF THE LEXICON

One of the goals of morphological theorising is to account for the ways in which speakers both understand and form not only 'real' words that occur in their language, but also potential words which are not instantiated in use in utterances. While it is true that a large percentage of 'real' words listed in dictionaries (such as *pear* and *pair*) are memorised, it is equally true, and of great theoretical interest, that countless words used in conversation (and to a lesser extent in writing) are new, made up on the spur of the moment. So, morphology has to throw light not only on the structure of established words like *pair*, but also on that of freshly coined **neologisms** like *snail-mail* (meaning the postal service, as opposed to modern electronic mail).

The consensus appears to be that the words of a language are **listable** in a way in which sentences are not (see 12.3.2). The meanings of many words (e.g. *pear* and *pair*) must be listed in the **lexicon** because there is nothing about their sounds or morphological structure that would enable one to work out their meaning. In this respect morphology differs from syntax. Syntax cannot be restricted to cataloguing only those sentences that occur in some **corpus** (i.e. a body of texts), since language is vast and no list of sentences, no matter how long, could exhaust the set of possible well-formed sentences. Typically, speakers do not merely recycle sentences memorised from previous conversations. Rather, they tend to construct fresh sentences to suit the occasion.

However, by and large, people do not routinely make up new words each time they speak. Nonetheless, the lexicon cannot be seen as a *static* list. No dictionary, however large, (not even the complete Oxford English Dictionary, including all its supplements) can list every word in the English language. Why is this so?

Until recently, word-formation rules have tended to be seen as being largely passive in the sense that they are basically used to analyse existing words rather than to create new ones. It is significant in this connection that, whereas reasonably comprehensive dictionaries and wordlists for dozens of languages exist, there are no equivalent, all-encompassing sentence lists for any language. Lists of sentences such as those found in phrase-books for foreigners make no pretence of being exhaustive.

The verdict on whether or not morphology and the lexicon deal with what is effectively a closed list of words will hinge, to some extent, on our attitude to nonce-words (like *uncomplicatedness*), created by an individual,