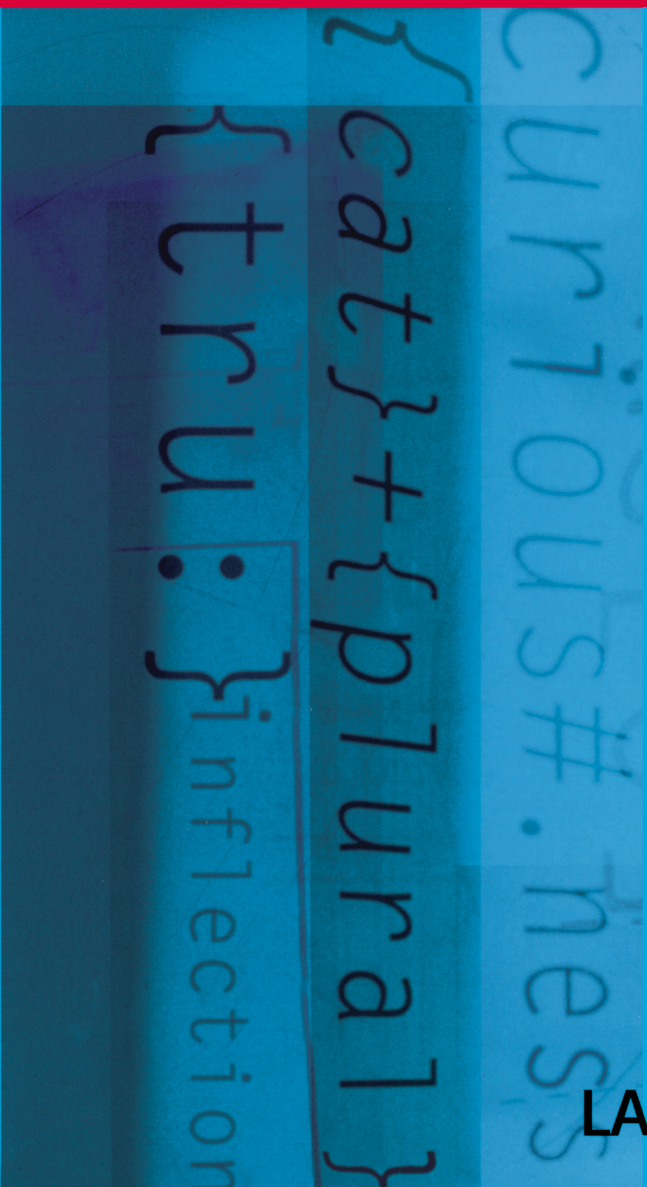
Introducing Linguistic Morphology

2ND EDITION



LAURIE BAUER

Introducing Linguistic Morphology

Introducing Linguistic Morphology

Second Edition

Laurie Bauer

Edinburgh University Press

Edinburgh University Press
22 George Square, Edinburgh

A CIP record for this book is available from the British Library

ISBN 0 7486 1704 3 (hardback)
ISBN 0 7486 1705 1 (paperback)

Typeset in 10/12 Palatino
Printed and bound in Great Britain by
CPI Antony Rowe, Chippenham and Eastbourne

Contents

Preface

The aim of this book is to provide an introduction to linguistic morphology. The study of morphology has been influenced by all major groups of linguists: by the philologists of the nineteenth century, by the structuralists in the twentieth century, by the transformational grammarians in the second half of the twentieth century and by linguists with other theoretical orientations as well. Part of the function of this book is thus to provide a coherent approach to the study of morphology, without gross distortion of the ideas that have come down to us from so many different sources.

In this book, Part One provides an introduction to the fundamental notions involved in the study of morphology. This Part provides a background in the morphological notions of the philologists and structuralists. Part Two provides an elaboration of this, going into considerably greater detail in a few areas of major importance and specifically raising questions which are glossed over in Part One. Part Three provides an introduction to some of the major issues in morphology today, in the transformational or post-transformational era. In a course on General Linguistics, it is assumed that each Part will be studied as part of the course work for a different year. But in some Universities there may be a specialised course in Morphology, which will consider all three Parts or sections of them. It is hoped that students who have worked their way through this book will have sufficient background to understand the current controversies in morphological theory and to be able to approach the original articles and books in which the developments are taking place at the moment.

To support the theoretical exposition of this book students should ideally, especially in the early stages, carry out exercises in morphological analysis, using data from a variety of languages, to see how the theoretical constructs apply in practice. Only a few such

exercises are provided in this book, alongside questions which involve more discussion of the notions which have been expounded in the chapters. Individual teachers may choose problems from the many available in specialised workbooks or invent their own.

In Appendix C there is a glossary of the technical terms of morphology, with definitions and, where appropriate, examples.

Technical terms from other areas of Linguistics such as 'phoneme' or 'direct object' are not listed, only those relevant to the study of morphology. Thus this appendix may be used as a quick reference source for the specialist terms in this book or as a means of revision. For technical terms from Linguistics that are not listed in the glossary, consult Crystal (1980) (or a later edition).

A brief note is also needed on my mode of glossing languages other than English. Morphs are separated out in the data by the symbol '·'. This symbol is also used in the English glosses to show which part of the English gloss corresponds to each morph in the original. Where a single morph in the original is glossed by more than one English word, the English glosses are hyphenated. Where necessary, a translation is also given between inverted commas. A simple example from French will illustrate this:

arriv·ons
arrive·1st-person-plural
'we are arriving'

In general I have tried to avoid abbreviations in the glosses, but *1st*, *2nd* and *3rd* are consistently used, and *sing*, *pl* are occasionally used for 'singular' and 'plural'. Other abbreviations are explained in the text or should be clear from the context.

Transcriptions of data from other languages are dependent on the system used by my sources; in many cases an orthographic form is given rather than a transcription. Where possible, phonetic symbols of the International Phonetic Association have been used. Transcriptions of English employ the system used by Gimson in the *English Pronouncing Dictionary* (Jones, 1977).

I should like to thank all those who have provided me with help, advice and suggestions during the writing of this book. In particular, the following people have commented on parts of various drafts and their aid has been invaluable: Winifred Bauer, Andrew Carstairs-McCarthy, Janet Holmes, Kate Kearns, Paul Warren. They are not to blame if I have not always followed all of their advice and they do not, needless to say, necessarily agree with what I have written. I should also like to thank reviewers of the first edition who pointed out errors and infelicities.

Part One: Fundamentals

CHAPTER 1

Introduction

How many words are there in this book? Can we always tell precisely what a word is? Do *motet, motion* and *motive* have anything to do with each other? What ways do we have of making new words in English? Are the same ways of forming new words found in all languages? Is it just coincidence that although you can have a word like *people* which means much the same as 'a lot of *persons*' and a word *peoples* which means, more or less, 'a lot of lots of *persons*', you cannot have a word *personss* meaning the same thing? Is it just coincidence that the ablative plural of the Latin word *re:x* 'king', *re:gibus*, meaning 'by/from/with the kings' is so much longer than the nominative singular *re:x*? (I use the phonetic length mark rather than the traditional macron to show long vowels in Latin.) All of these questions relate to **morphology**, the study of words and their structure.

It is a well-established observation that words occur in different forms. It is quite clear to anyone who has studied almost any of the Indo-European languages. Students of these languages learn **paradigms** like those below as models so that they can control the form-changes that are required. As illustrations, consider a verb paradigm from Latin and a noun paradigm from Icelandic. (The word 'paradigm' means 'pattern' or 'example'.)

(1)	amo:	'I love'
	ama:s	'you (singular) love'
	amat	'he/she/it loves'
	ama:mus	'we love'
	ama:tis	'you (plural) love'
	amant	'they love'

(2)

Singular	
nominative	hestur 'horse'
accusative	hest
dative	hesti
genitive	hests
Plural	
nominative	hestar
accusative	hesta
dative	hestum
genitive	hesta

In the nineteenth century, the term 'morphology' was given to the study of this change in the forms of words. The term is taken from the biological sciences and refers to the study of shapes. In linguistics this means the study of the shapes of words; not the phonological shape (which can be assumed to be fairly arbitrary) but rather the systematic changes in shape related to changes in meaning, such as those illustrated in the paradigms above or such as that relating the pairs of words below:

(3)

desert	deserter
design	designer
fight	fighter
kill	killer
paint	painter
twist	twister

By extension, the term 'morphology' is used not only for the study of the shapes of words but also for the collection of units which are used in changing the forms of words. In this sense, we might say that Latin has a more complex morphology than English. Again by extension, 'morphology' is also used for the sequence of rules which are postulated by the linguist to account for the changes in the shapes of words. In this sense we might contrast the morphology of language *L* with the syntax of language *L* (where the syntax is the sequence of rules postulated by the linguist to account for the ways in which words are strung together). In this sense we might also say that something is part of the job of 'the morphology of language *L*' or, more generally, of 'morphology', implying that this is true for all languages. We shall see later how all these senses fit together; such

extensions of meaning are common within linguistics and do not usually cause problems of interpretation.

Many traditional 'grammars' (in the sense 'grammar books') deal largely with such morphology as can be laid out in paradigms, like those presented above, and have little to say about syntax. This has led to the situation where many lay people today still believe that languages like Chinese or English do not have much grammar because they do not have extensive morphological paradigms. That is, for many people the term 'grammar' is equated with morphology. For most linguists today, however, 'grammar' includes both morphology and syntax (and, for most, phonology as well), and most of the linguistic study of 'grammar' in this sense has, since the middle of the last century, not been of morphology but of syntax. This is understandable. Syntax, especially from 1957 onwards, was a relatively new field of study, while morphology was considered well researched and well understood. It did not seem at that time as if there was a great deal that was new to say about morphology. Morphological descriptions of hundreds of languages were available but all the languages differed in what appeared to be essentially random ways. There did not seem to be any cross-linguistic generalisations to be made in morphology. Syntax, in the middle of the last century, was a far richer ground for linguistic discoveries. It was the excitement of the progress being made in the study of syntax which gave Linguistics such a boost in the 1960s. It was also progress in the study of syntax which eventually led to the realisation that there were still questions to be answered in morphology. As a result, there has, in recent years, been a resurgence of interest in morphology.

Since the 1970s the study of morphology has flourished in ways which could not have been imagined earlier in the twentieth century. This has come about partly through consideration of new data from a range of languages, partly through a consideration of the patterning of morphological data across languages, partly through innovations in the treatment of syntax and phonology (which have held implications for morphology), partly through a renewed interest in how the brain processes words and partly through the detailed study of morphological systems themselves. These various impulses have not led to a unified treatment of morphology but they have led to an increased sense of excitement among people who study morphology as it appears that morphology may provide a window on to wider linguistic behaviour.

This book provides an introduction to the study of morphology,

covering the input from these various sources and attempting some kind of synthesis in the light of the most recent research. It discusses both the general background to all morphological study and also some of the detail of recent theories of morphology.

REFERENCES AND FURTHER READING

At the end of each chapter there is a section of references and further reading. The book can be read without these sections although, in some later chapters, there are discussions of a few matters of import which are tangential to the main text. The references in these sections (and in the text of the chapters in Part Three) are given in terms of the surname of the author and the date of publication, and full details can be found in the list of references at the back of the book. It is sensible to get into the habit of reading some of the works referred to. No textbook gives an unbiased presentation of the facts and this one is certainly no exception. Only by reading other works can you make some attempt to counteract the in-built bias of the text. In other cases, it is probably just as well to assure yourself that I have not given an unfair account of other people's research. Sources of data can be consulted either for extra data for analysis or to see many of the extra complications which I have ignored for the sake of the example.

The data from Latin in this chapter, and in many other places in the book, is taken from Kennedy (1962). The data from Icelandic is from Einarsson (1945). Where references are not provided for language data, it is frequently because I am familiar with the language concerned. Much of the complex English data is gleaned from dictionaries and other reference books, only some of which are listed in the list of references at the end of the book.

Exercise

1. Consider the sentences (a)–(e):

(a) I bought a grammar of Maori.
(b) The grammar of English does not permit the ordering subject-object-verb.
(c) I tried to learn Icelandic but the grammar was too hard.
(d) The grammar of a language includes its phonology, its morphology and its syntax.
(e) I always hated grammar lessons in school.

How many different meanings of 'grammar' are there in these sentences and what precisely does 'grammar' mean in each sentence?

CHAPTER 2

The Basic Units

2.1 A CLOSE LOOK AT WORDS

How many words are there in (1)?

(1) The cook was a good cook as cooks go and, as cooks go, she went.

It probably didn't take you very long to come up with the answer '15'. Now think about how you arrived at that answer. What you did, in effect, was to count the items which appeared between spaces on the page. We could say that a word is a unit which, in print, is bounded by spaces on both sides. We will call this an **orthographic word** because it is linked to the spelling (orthography). In practice, we may not always feel certain just where words begin and end in the written language but strict application of the definition above will provide us with consistent answers. Consider the next example:

(2) I've been in hot water so often I feel like a tea-bag.

There are two places in (2) where we might raise questions about how many words are involved. *I've*, we know, has two parts and could have been written as *I have*. But an apostrophe is not a space and so, by applying the definition we formulated earlier, we can say that *I've* is one orthographic word. The same is true with *tea-bag* in (2). A hyphen is not a space and we are, thus, able to say that *tea-bag* is a single orthographic word, even though it is made up of two parts. Just what those parts are is a question to which we will return.

Now let us turn to a different question with respect to (1). How many different words does (1) contain? This question is more difficult to answer because the answer you give will depend on whether you

think *cook* and *cooks*, *go* and *went* are 'the same word' or not. They are clearly different orthographic words: they have different **forms** or shapes. So we might say that there are 11 different **word-forms** (or orthographic words) in (1): *the*, *cook*, *was*, *a*, *good*, *as*, *cooks*, *go*, *and*, *she*, *went*. On the other hand, there is another sense in which *cook* and *cooks* are forms 'of the same word'. Let us use the term **lexeme** for this sense and say that *cook* and *cooks* are different **word-forms** which belong to, or **realise**, the same **lexeme**. Similarly, *go* and *went* are different word-forms which realise a single lexeme. There are two fewer different lexemes in (1) than there are word-forms, so there are nine different lexemes in (1). The name by which we choose to refer to the lexeme is arbitrary and depends on the classification required in the general theory within which we are working. We could call the lexeme realised by *cook* and *cooks* '762' or '85/a/17-U5' or 'Samantha' but, to make things easier, we'll agree to call it COOK, using capital letters when we write it. We can now say, more succinctly than was possible before, that *cooks* is one of the word-forms which can realise the lexeme COOK. Similarly, we can say that *went* is one of the word-forms that can realise GO. The lexemes in (1) are THE, COOK, BE, A, GOOD, AS, GO, AND, SHE.

We can look at the same facts and the same terminology from a different angle. Suppose you were reading a book and you suddenly came across the following sentence, which you didn't understand.

(3) The posset was disembogued from the rehoboams.

Suppose you didn't understand the fourth orthographic word in (3). You might want to consult your dictionary to find out what it means. But you would not look up *disembogued* with a final *d* in your dictionary because your knowledge of English is sufficient to tell you that *disembogued* is just a form of the lexeme DISEMBOGUE. Similarly you wouldn't look up *rehoboams* but REHOBOAM, which is the name of the lexeme. Lexemes are dictionary words (not necessarily in the sense that they will be given a separate entry and act as a headword in the particular dictionary you happen to have on your shelf, but in the sense that you might expect their separate identity to be acknowledged in an ideal dictionary or the ideal dictionary that a speaker might have in their head). A lexeme comprises all the word-forms which can realise that lexeme. Thus the lexeme REHOBOAM is sometimes realised by the word-form *rehoboam* and sometimes by the word-form *rehoboams*.

The units which actually occur, which have a shape (whether that is an orthographic shape or a sound shape), are word-forms. Word-

forms realise (represent, belong to) lexemes. Lexemes, correspondingly, do not actually occur but are abstract dictionary words: abstract in that each lexeme may comprise a number of possible forms. The lexeme is, in a sense, what all the word-forms associated with it have in common.

For the time being, we shall assume that the word-forms of the spoken language are identical with the word-forms of the written language (that is, with orthographic words). This is actually an extremely awkward point and it will be taken up again in greater detail in Chapter 4.

We are now in a position to distinguish between the lexemes and the word-forms in a sentence like (1) but this will not always be sufficient for all purposes. Consider the next four sentences.

(4) (a) Lee walked home.
    (b) Lee went home.
(5) (a) Lee has walked home.
    (b) Lee has gone home.

The word-form *walked* occurs in both (4a) and in (5a) and, in both cases, it realises the lexeme WALK. Yet *walked* in these two sentences is not precisely the same element, as we can see when we compare with (4b) and (5b), which contain equivalent forms of the lexeme GO. In (4a) *walked* realises WALK + past tense, while in (5a) it realises WALK + past participle. We might want to say that *walked* in (4a) and (5a) are different words, even though they are homophonous word-forms and realise the same lexeme. We will say that they are different **grammatical words**. Grammatical words are defined in terms of their place in the paradigm and named by descriptions such as 'the past participle of WALK', which spell out that place.

We now have three different kinds of 'words': word-forms (including orthographic words), lexemes and grammatical words. The word 'word' might seem to be rather a liability under these circumstances. However, it turns out that the amount of specificity in the three other terms is not always needed and it is useful to have a superordinate term for word-form, lexeme and grammatical word. This allows us to avoid, when appropriate, the precision implicit in these other terms. In this book, **word** will be used with this less specific meaning.

In the printed text, word-forms are already separated out for us. But, even if they were not, we would still be able to discover the beginnings and ends of words fairly simply. Indeed, Ancient Greek was frequently written without gaps between the words and this

corresponds to the way in which we talk. There are no gaps between the words in normal conversation. If there were, schoolboy jokes such as asking someone to say 'I chased the bug around the tree' or 'I'll have his blood, he knows I will' would not work, and neither would elaborate jokes like the following French couplet:

(6) Gal, amant de la reine, alla, tour magnanime,
Galamment de l'arène à la Tour Magne, à Nîmes.

'Gal, the queen's lover, went (which was generous of him) gallantly from the arena to the Magne Tower at Nimes.'

The two lines in this couplet are (or can be) absolutely homophonous, even though they mean different things because the word-breaks come in different places. So there are no gaps between words in speech and what we hear could be represented by running all the words together in print, as in

(7) Menbecomeoldbuttheyneverbecomegood.

Even here it is a relatively straightforward task for a speaker of English to sort out where the word-forms begin and end. We do this because we can recognise patterns and carry out substitutions. Firstly, we recognise certain strings of letters (or strings of sounds in the spoken language) which we know are found in other sentences with a meaning which would also make sense here. So we can recognise the sequence *men* in (7) because it also comes up in sentences like:

(8) Menareconservativeafterdinner.

(9) Menlosetheirtempersindefendingtheirtaste.

(10) Afterfortymenhavemarriedtheirhabits.

In a similar way we could isolate each of the word-forms in (7). Notice that we might try to isolate the first two letters of (7) and we would find *me* in other places, but then we would be unable to find a word beginning *nb*. So firstly we can recognise recurrent sequences. But, even if we found a sequence which we did not recognise, we could substitute other things which we know are word-forms in the place of these strings. So for *men* in (7) we could substitute *people, lecturers, sopranos* and so on, in place of *become* we could have *get, are, turn,* and so on, but we could not replace the *m* in *men* or *menb* in (7) with a word-form like *weeds* and still end up with a sentence. So substitution allows us to determine word-forms.

Before we leave words, we need to consider constructions like *kick*

the bucket 'die' or *red herring* 'irrelevant distraction'. These look like ordinary syntactic constructions but their meanings are not derivable from the meanings of the component elements, as would be the case with *kick the ball* or *red paint*. Such phrases are called **idioms** and we would expect to find them listed in a dictionary (whether an actual book or an ideal mental facility). Idioms thus share with lexemes the quality of having dictionary entries. At the same time, an idiom like *red herring* seems to be made up of two lexemes, each of which has its own dictionary entry. So, while lexemes are, roughly defined, dictionary words, they are not the only things that are found in dictionaries. This means we need a different label for things listed in dictionaries. Several such labels are found in the literature, including **lexical item** and **listeme**. Lexical items in this sense include lexemes, idioms, phrasal verbs in English (*root out*, equivalent to *eradicate*), proverbs such as *a stitch in time saves nine* and possibly quotations such as *to be or not to be, that is the question.*

2.2 ELEMENTS SMALLER THAN THE WORD

Now consider the following sentence:

(11) He was born stupid and greatly increased his birthright.

We could isolate all the word-forms in this sentence in the ways outlined above, but we can also look within the word-forms and isolate recurrent forms within the word-form. For instance, if we consider the word-form *birthright*, we can divide that into two parts. For the first part, *birth*, we could substitute within the word-form things like *copy* and *water*, for the second part, *right*, we could substitute things like *day*, *place* and *rate*. In a similar way, we could divide the word-form *greatly* up into two parts. For the first part, *great*, we could substitute other forms like *vast* and *incredible* and, within the word-form (though not within this particular sentence), we could substitute items like *-ness*, *-er*, *-est* for *-ly*. Again, in a similar way, we can divide *increased* in (11) into two portions: *increase* and *-d*. For the first, we could substitute items like *enlarge* and *minimise*, for the second, items like *-s* and *-ing*. In other words, the same techniques that allow us to segment sentences into word-forms also allow us to segment word-forms. The units which we arrive at within the word-form we will call **morphs.** A word-form may contain only one morph (*stupid*, *and*) or it may contain several (*great·ly*, *increase·d*, *birth·right*). In this book, the decimal point will be used to separate morphs. A

morph, then, is a unit which is a segment of a word-form. It has a constant form and realises or is related to a constant meaning.

Some morphs have the potential of being word-forms on their own. In

(12) Every·one live·s by sell·ing some·thing.

this applies to the morphs *every, one, live, by, sell, some* and *thing*. Such morphs are called **potentially free morphs**. Notice that the potentiality is not actually exploited for all of these morphs in (12). Indeed, only *by* is actually free in (12) but the others listed are potentially free. Morphs which cannot be word-forms by themselves but which need to be attached to other morphs are termed **obligatorily bound morphs**. In (12) only the morphs *-s* and *-ing* are obligatorily bound. Notice that there are words in English (and far more in many other languages) which are made up entirely of obligatorily bound morphs. Examples from English include: *Euro·crat, octo·pus, phil·anthrop·y, phonet·ic, quadra·phon·ic, wis·dom* and so on. In (12), it so happens that all of the potentially free morphs realise lexemes and none of the obligatorily bound morphs do. This is typically the case in English but not invariably the case, as the examples above show. In other languages, such as Latin, for example, it is typically the case that the morphs realising lexemes are also obligatorily bound.

In most cases in English (and, indeed, in all languages) and in all of the cases in (12) – though there are exceptions which will be dealt with below – the morph which realises the lexeme does not also realise anything else. Any morph which can realise a lexeme and which is not further analysable (except in terms of phonemes) is termed a **root**. Obligatorily bound morphs which do not realise lexemes and which are attached to roots to produce word-forms are called **affixes**. In a word like *dealings, deal* is the root and *-ing* and *-s* are affixes In *something* in sentence (12) there are two roots. Note that this implies that *some* and *thing* in *something* realise the lexemes SOME and THING, respectively, even though SOMETHING is also a lexeme in its own right. Affixes can be added directly to a root, as in *fool·ish,* or they can be added to a root and some already attached affix, as is the case with *-ness* in *fool·ish·ness*. We can call anything we attach affixes to, whether it is just a root or something bigger than a root, a **base**. So in the formation of *dealings* the root is *deal* but the base to which the *-s* is added is *dealing*. Note that in this case the final *-s* was not added to a root.

If an affix is attached before a base, it is called a **prefix**, if it is

attached after a base it is called a **suffix** and, if it is attached in the middle of a base, it is called an **infix**. In the word *prepacked*, there is a root *pack*, a prefix *pre-* and a suffix *-ed*. All of the affixes that have been illustrated in (11) and (12) have been suffixes, which are more common in English than prefixes are. There are no infixes in English: the closest we have is the use of expletives in the middle of words like *absobloominglutely* and *kangabloodyroo*.

2.3 INFLECTION AND DERIVATION

Affixes can be of two kinds, inflectional or derivational. An **inflectional affix** is one which produces a new word-form of a lexeme from a base. A **derivational affix** is one which produces a new lexeme from a base. Take a word-form like *recreates*. This can be analysed into a prefix *re-*, a root *create*, and a suffix *-s*. The prefix makes a new lexeme RECREATE from the base *create*. But the suffix *-s* just provides another word-form of the lexeme RECREATE. The prefix *re-* is derivational but the suffix *-s* is inflectional. In English (though not in every language) prefixes are always derivational. Suffixes in English, though, may be either derivational or inflectional. In the word-form *formalises* the root is *form* and there are three suffixes: *-al*, *-ise* and *-s*. *Formal* belongs to a different lexeme from *form*, so *-al* is a derivational suffix; *formalise* belongs to a different lexeme from *formal*, so *-ise* is a derivational suffix; but *formalises* belongs to the same lexeme as *formalise*, so *-s* is an inflectional affix.

There are a number of ways of telling whether a suffix is inflectional or derivational if you are not sure whether or not it produces a new lexeme.

(a) If an affix changes the part of speech of the base, it is derivational. Affixes which do not change the part of speech of the base are usually (though not invariably) inflectional. So, because *form* is a noun and *formal* is an adjective, *-al* has changed the part of speech; it is, thus, a derivational affix. *Formal* is an adjective and *formalise* is a verb; *-ise* has changed the part of speech; it is a derivational suffix. *Formalise* is a verb *formalises* is still a verb; *-s* has not changed the part of speech so *-s* is likely to be an inflectional affix. Note, however, that, while all prefixes in English are derivational, very few of them change the part of speech of the base.

(b) Inflectional affixes always have a regular meaning. Derivational affixes may have an irregular meaning. If we consider an inflectional affix like the plural *-s* in word-forms like *bicycles*, *dogs*, *shoes*, *tins*, *trees*

and so on, the difference in meaning between the base and the affixed form is always the same: 'more than one'. If, however, we consider the change in meaning caused by a derivational affix like *-age* in words like *bandage, cleavage, coinage, dotage, drainage, haulage, herbage, mileage, orphanage, peerage, shortage, spillage* and so on, it is difficult to sort out any fixed change in meaning or even a small set of meaning changes.

(c) As a general rule, if you can add an inflectional affix to one member of a class, you can add it to all members of the class, while, with a derivational affix, it is not generally possible to add it to all members. That is, inflectional affixes are fully **productive** while derivational affixes are not. For example, you can add *-s* to any non-modal verb in English to make the 'third person singular of the present indicative' but you cannot add *-ation* to any non-modal verb to make a noun: *nationalis·ation* is a perfectly good word, so it works some of the time, but none of **com(e)·ation, *inflect·ation, *produc(e)·ation* or **walk·ation* are words of English. We can summarise this criterion in the following way: affixes which show limited productivity with large numbers of gaps are derivational; affixes which are fully productive (can be used with all members of a class) may be either inflectional or derivational.

In fact, the distinction between inflectional and derivational affixes is more complex than this suggests and the matter will be taken up again in Chapter 6. The criteria provided here, though, will cover most of the straightforward cases.

2.4 ALLOMORPHS AND MORPHEMES

Sometimes two or more morphs which have the same meaning are in complementary distribution. That is, the two can never occur in precisely the same environment or context, and between them they exhaust the possible contexts in which the morpheme can appear. For example, there are two morphs in English which can be glossed as 'indefinite article': *a* and *an*. Some examples of their distribution can be seen below.

(13)

(a)	(b)
a man	an oak
a horse	an elephant
a kettle	an uncle
a university	an apple
a green apple	an old man

From the examples in (13) and from your knowledge of the rest of the language, you can see that *a* occurs when the next word begins with a phonetic consonant and that *an* occurs when the next word begins with a phonetic vowel. The word 'phonetic' is important, since *university* begins orthographically with a *u* but phonetically with a consonant, /j/. (In some rather conservative varieties of English this rule is not quite true, since it is possible to say *an hotel* and *an historical novel*. Not all such speakers pronounce these words without an /h/, which would make them conform to the general rule. We shall, provisionally, ignore these varieties.)

In this case, the choice between the two morphs *a* and *an* is determined or conditioned by the following phonetic sound. We can say that their distribution is **phonetically** (or **phonologically**) **conditioned.** In other cases, the distribution of morphs may be determined by other factors. For example, consider the way we make adjectives from nouns describing people which end in *-or*. While there are words with no corresponding adjective (like *juror, vendor*), most such words make an adjective by adding *-ial*, as in *editor·ial* and *professor·ial*. In a very few cases, though, we simply add *-al*: *doctor·al* is the clearest example. From the examples cited, it might seem that this is a phonologically conditioned difference but a pair such as *doctoral* versus *censorial* suggests that the difference has nothing to do with the phonology of the bases concerned. It is simply a fact about DOCTOR that its corresponding adjective is DOCTORAL; the choice between *-ial* and *-al* in such cases is **lexically conditioned**. Morphs can also be **grammatically conditioned**. In a language like German, adjectives change their form depending on the gender of the noun they modify. Thus, in the nominative singular, we find the following pattern:

(14)	ein gross·er Wagen	'a big car (masculine)'
	ein gross·er Fisch	'a big fish (masculine)'
	ein gross·es Haus	'a big house (neuter)'
	ein gross·es Tier	'a big animal (neuter)'
	eine gross·e Feder	'a big feather (feminine)'
	eine gross·e Schlange	'a big snake (feminine)'

The suffix on the adjective is not determined by the phonetic shape of the base or of the next word, nor is it determined by the particular lexeme following, but by the grammatical gender class that lexeme belongs to. The conditioning is thus neither phonetic nor lexical but grammatical.

But, if English *a* and *an*, or *-al* and *-ial*, or German *-er*, *-es*, and *-e* are clearly separate morphs because of their different shapes, they, nevertheless, have things in common. They have their meaning in common: 'indefinite article', 'adjective' or 'nominative singular'. Between them, they divide up a single distribution: always before a singular countable noun, always on the end of a noun ending in *-or*, or always on the end of an attributive adjective following an indefinite determiner. They even have a similarity in form. There is a sense, therefore, in which *a* and *an* (and the other sets) are 'the same thing'. We will say that these various sets of morphs realise the same **morpheme**.

As with lexemes, the name we use to refer to the morphemes is arbitrary. We could call the morpheme realised by *a* and *an* '9,673' or '99/0/7245-T2' or 'Zoë' – whatever made sense in our classification. But, to simplify matters, we will agree to call a morpheme either by a label describing its meaning (in this case 'indefinite article') or by one of its morphs, say 'a'. If we make the latter choice, we will usually use the morph that occurs most widely in the language. To show we are talking about a morpheme we will enclose this name in braces when we write it: {indefinite article} or {a}.

Notice that the morpheme, like the lexeme and the phoneme, is realised by something else. You cannot hear a morpheme or say a morpheme (just as you cannot say or hear a lexeme or a phoneme): you can only say or hear something which realises a morpheme (or a lexeme or a phoneme). You can hear or say a morph (or a word-form or a phone) but not what it realises. Morphemes (like lexemes and phonemes) are abstract units. The morpheme {a} is whatever all the morphs which can realise {a} have in common.

Morphs which realise a particular morpheme and which are conditioned (whether phonetically or lexically or grammatically) are called the **allomorphs** of that morpheme. (If we consider the written form of the language, it is also possible to talk about orthographically conditioned allomorphs of a morpheme, as in *come* and *com-*, the latter of which occurs in *coming*.) *A* and *an* are the two phonetically conditioned allomorphs of the morpheme {a}. Notice that every allomorph is a morph. The term allomorph is simply more informative than is morph on its own because it says that the morph is one of several realisations of the same morpheme.

It should be noted that the terms 'morpheme' and 'allomorph' are frequently used by other linguists in rather different senses from the ones they have been given here. We will consider this in more detail in Chapter 7. The implication of this is that you will have to take care,

when you meet these terms in other works, that you know precisely what is intended.

This abstractness of morphemes frequently causes problems of understanding for students who are new to morphology.

A parallel might be helpful. Consider the various symbols presented below:

(15) R r ℛ

What do these symbols have in common? The answer, fairly simply, is that they are all kinds of 'R'. What is it about them that shows, then, that they are all kinds of 'R'? How can you tell that they are 'R's? The answer lies not in their particular shape, which varies depending on the kind of script they happen to occur in. The answer lies in their function; they all function as 'R', they all have the same value in the system, they all, for instance, can be used at the beginning of the word *rat*. But this is an abstract quality. The 'R-ness' of the symbols presented in (15) is something abstract. The 'R-ness' corresponds, in the analogy, to the morpheme. The different forms for these 'R's which have the same function in the system correspond to morphs. The symbol 'r' is just one of the morphs that can realise the morpheme, the abstract notion of what 'R' is.

2.5 RECAPITULATION

Armed with all this terminology, consider the word-form *was* in the sentence

(16) While he was not dumber than an ox, he was not any smarter.

Was is a word-form which is probably not further analysable into morphs. This word-form realises, among other things, the lexeme BE, as we can see if we change the syntactic environment of (16), but not its meaning:

(17) Being not dumber than an ox, but being no smarter, he was not a brilliant conversationalist.

(18) I hope, at least, to be not dumber than an ox.

But there are also other morphemes realised by the word-form *was*. Firstly, *was* is singular, because if it had a plural subject it would be replaced by *were*:

(19) While they were not dumber than oxen, they were not any smarter.

You cannot say in standard English

(20) While they was not dumber than an ox, they was not any smarter.

Secondly, *was* also marks past tense. You cannot say:

(21) While he was not any dumber than an ox, he was not any smarter at the moment.

You have to say:

(22) While he is not any dumber than an ox, he is not any smarter at the moment.

In other verbs, pastness is shown by a separate morph (as, for example, in the difference between *compliment* and *compliment·ed*) but, here, this difference is included in the single word-form. So *was* is a single morph that realises not only the lexeme BE (which is made up of a single morpheme {be}) but also the morphemes {singular} and {past tense}. A morph which realises more than one morpheme in this way is called a **portmanteau morph**.

To summarise, we can say that an actual speech event can be analysed into a series of word-forms, some of which will be made up of a single morph, others of which will be made up of more than one morph. The morphs realise (though not necessarily in a one-to-one manner) morphemes. Morphemes are abstract units of grammatical and semantic analysis. One or more morphemes may form a lexeme, an abstract vocabulary item. This is presented in tabular form (with a certain amount of simplification) in Figure 2.1.

LEXEMES	made up of one or more	MORPHEMES
realised by		realised by
WORD-FORMS	made up of one or more	MORPHS

Figure 2.1: Summary of basic terms

REFERENCES AND FURTHER READING

Most introductory linguistics textbooks cover the kind of material that has been covered in this chapter. It should be noted, however, that the terminology used here derives from a British tradition, which is, in this area, distinct from the American tradition. In particular, many

American sources do not distinguish between 'morph' and 'morpheme' in the way that has been done here. For a discussion of the various ways in which the term 'morpheme' has been used in Linguistics, see Mugdan (1986). It is clear from that article just how careful you have to be when reading works which discuss morphemes.

Among the best books to consult for consolidation of the discussion here are Brown & Miller (1980), Lyons (1968) and Matthews (1991). Of these, I would particularly recommend Brown & Miller (1980) for beginners. Note that Matthews's use of the term 'word' is different from the one in this book.

The discussion of abstractness in terms of letter-shapes is loosely based on Lass (1984).

Not all authorities agree that allomorphs can be lexically conditioned. Lyons (1968), for example, is in a tradition of excluding the use of the term 'allomorph' in such circumstances. It is true that it is far less clear that there is 'conditioning' involved in lexical conditioning, which is far more random than other kinds of conditioning. There are, therefore, advantages to this kind of restriction. But there are also advantages (albeit less obvious ones) to the wider usage of the term adopted in this book.

The sentences used as examples are, of course, not my own. Many of them can be found in Bentley & Esar (1951). The data from Yingkarta in the exercises comes from Dench (1998).

EXERCISES

1. Which of the following are lexemes of English: *superstition*, CAT, BEFORE, {neat}.

2. For each of (a), (b) and (c):
   i. How many orthographic words does the sentence contain?
   ii. How many different word-forms does the sentence contain?
   iii. How many different lexemes does the sentence contain?
   iv. How many different grammatical words does the sentence contain?
   a. I walked to town yesterday, she has walked there this morning.
   b. Not many banks have branches on the banks of the Avon.
   c. She stoops to refill the stoops of wine and, having refilled them, wishes the wine-waiter refilled them himself.

3. In the example sentence (b) in (2) above, is *banks* the same word-form occurring twice or two different word-forms?
4. Consider the following data from English, presented in transcription.

Infinitive	*Past tense*	*Infinitive*	*Past tense*
kɪl	kɪld	kɒf	kɒft
fɪz	fɪzd	wɒʃ	wɒʃt
əlaʊ	əlaʊd	pju:k	pju:kt
græb	græbd	pɑ:s	pɑ:st
hʌm	hʌmd	slaʊʧ	slaʊʧt

What determines whether the past tense is marked by a /d/ or a /t/? What kind of conditioning is that? /d/ and /t/ both mean 'past tense' and are both alveolar plosives (and so share form). What does that imply about these two forms? Consider the past tense of the verbs WANT and PRETEND. How do they fit into the pattern? Using appropriate notation and terminology, write a statement of what is going on here. Confirm the predictions made in your statement by considering what happens with the verbs PRODUCE, INSULT and CONTAIN.

5. In Yingkarta, an aboriginal language of Western Australia, nouns and adjectives are marked for a case which has a number of functions and which is, therefore, called the genitive/dative case. On pronouns, the marker for this case is *-ngu*. On nouns and adjectives, the markers are as shown in the table below (which is presented in orthography).

Base	*Genitive/Dative*	*Gloss*
mantu	mantuwu	'meat'
papa	papawu	'water'
pika	pikawu	'sick'
pipi	pipiwu	'mother'
thuthu	thuthuwu	'dog'
kurtan	kurtanku	'bag'
majun	majunku	'turtle'
mangkurr	mangkurrku	'three'
nginthirn	nginthirnku	'cold'
nyanyjil	nyanyjilku	'woman'

What are the allomorphs of the morpheme {genitive/dative}? How is each allomorph conditioned? Given the base forms *purlijiman* 'policeman' and *kartu* 'man', what would you expect the genitive/dative forms to be?

6. Affixes are usually defined as obligatorily bound morphs. Choose one of the sets of data provided below (preferably from a language with which you have some familiarity) and decide whether this must always be the case.

(a) English
absobloodylutely
confronbloodytation
guaranbloodytee
imbloodypossible
inbloodyfallible
incanbloodydescent
kangabloodyroo
unbebloodylievable
unibloodyversity

(b) French

sous-entendre	surchauffer	suspendre
sous-exposer	surclasser	
sous-louer	suréquiper	
sous-tendre	surestimer	
soustraire	surfaire	
sous-traiter	surmonter	

(c) German

auf·drucken	aus·drucken	be·drucken
auf·fallen	aus·fallen	be·fallen
auf·geben	aus·geben	be·geben
auf·halten	aus·halten	be·halten
auf·nehmen	aus·nehmen	be·nehmen
auf·schreiben	aus·schreiben	be·schreiben
auf·stehen	aus·stehen	be·stehen
auf·steigen	aus·steigen	be·steigen
auf·stellen	aus·stellen	be·stellen
auf·tragen	aus·tragen	be·tragen

(d) Latin

ac·curro:	re·curro:
ac·ce:do:	re·ce:do:
ad·du:co:	re·du:co:
ad·fero:	re·fero:
ad·ficio:	re·ficio:
ad·lego:	re·lego:
ad·mitto:	re·mitto:

ap·po:no:	re·po:no:
ap·porto:	re·porto:

7. Are derivational affixes such as *-al* and *-ise* lexemes?

CHAPTER 3

The Morphological Structure of Words

In this chapter, we shall consider the various processes by which words can be built. I shall illustrate these processes from a number of languages, some of which will be familiar to you and others of which will not be familiar to you. It is the wide range of ways in which it is possible to build words which is the central focus of this chapter. In passing, attention will also be drawn to some of the difficulties that arise in morphological description, to show why linguists find morphology interesting. One reviewer said about morphology recently that 'we do not understand all that we know'. This is part of the interest and the challenge provided by morphology.

3.1 WORD-BUILDING PROCESSES USING AFFIXES

By far the most common way of building new words in the languages of the world is by using affixes. The commonest type of affix by far is the suffix. There are several languages in the world which use suffixes to the exclusion of any other type of affix (Basque, Finnish and Quechua are examples) but only very few which use prefixes to the exclusion of other types of affix (Thai is frequently cited as an example) and none which use any other type of affix exclusively. Thus, the obligatorily bound morph *par excellence* in the languages of the world is the suffix.

3.1.1 Suffixes

Suffixes are used for all purposes in morphology. They are used derivationally as in:

(1)

English:	constitut·ion·al·ity
Finnish:	asu·nno·ttom·uus
	live·noun·without·abstract-noun
	'houselessness'
Mam:	txik·eenj
	cook·patient
	'something cooked'

and inflectionally as in:

(2)

Finnish:	talo·i·ssa·an
	house·plural·in·3rd-person-possessive
	'in their houses'
Turkish:	gel·é·miy·eceğ·im
	come·be-able·negative·future·1st-person
	'I will not be able to come.'

Notice that all of the suffixes in (2) are inflectional, even though some of them are translated into English by separate lexemes. The meaning of an affix is not sufficient to tell you whether that affix is inflectional or derivational. This will be taken up again in Chapter 6. Neither is it the case that a given type of meaning is always realised in the same kind of way across languages. Even plurality may not always be an inflectional category. In Diyari, an aboriginal language of South Australia, plurality is marked optionally by a derivational suffix.

As is clear from the examples given above, suffixes can occur in sequences although there is no expectation that they will. When both inflectional and derivational suffixes co-occur in the same word-form, the general rule (although it is by no means exceptionless, see below section 6.5) is that the derivational suffixes precede the inflectional ones, so that the cases in (3) are typical.

(3)

Diyari:	jiŋki·mali·ji
	give·reciprocal (deriv)·present (infl)
	'give one another'
Finnish:	kirja·sto·sta·mme
	book·collective (deriv)·out-of (infl)·our (infl)
	'out of our library'
French:	égal·is·a
	equal·verb (deriv)·3rd-person-singular-past

(infl)
'[he/she/it] equalised'

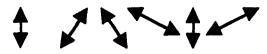
Portmanteau morphs are very common as suffixes in highly inflecting languages. This is illustrated by the case and number marking on the nouns in many Indo-European languages. The paradigm for the Latin noun ANNUS 'year', given below, will provide an example.

(4)

	Singular	*Plural*
Nominative	ann·us	ann·i:
Vocative	ann·e	ann·i:
Accusative	ann·um	ann·o:s
Genitive	ann·i:	ann·o:rum
Dative	ann·o:	ann·i:s
Ablative	ann·o:	ann·i:s

In this paradigm, it can be seen that there is neither a consistent realisation of singularity, nor one of plurality. Neither is there a single realisation of any one of the cases (and if the other genders were taken into consideration, this would be even more striking). Rather, the final morph analysed in the word-forms in (4) has to be seen as a portmanteau morph, realising simultaneously the category of number and that of case. The alternative position, where a single morpheme is realised in more than one morph, can also be illustrated from Latin. Consider the realisation relations shown by the arrows in the Latin word-form *re:ksisti:* 'you (sing) ruled'.

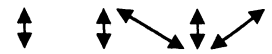
(5) RULE perfective 2nd singular



re:k s is ti:

Each of the morphs analysed in (5) can be motivated by comparison with other forms of Latin and the realisation relations can be justified since, if any of the morphemes were changed, the morphs realising those morphemes would also change. Compare, for example, the form for 'I ruled':

(6) RULE perfective 1st singular



re:k s i:

3.1.2 Prefixes

Although they are rarer than suffixes, prefixes work in very much the same way. They can be derivational, as in:

(7) *English*: dis·en·tangle
Mam: aj·b'iitz
agent·song
'singer'
Tagalog: pan·ulat
instrument·write
'pen'

or inflectional as in:

(8) *Mam*: t·kamb'
3rd-singular-possessive·prize
'his prize'
Swahili: a·si·nga·li·jua
he·negative·concessive·past·know
'if he had not known'
Tagalog: i·sulat
modal·write
'writing (participle)'

These examples show that, like suffixes, prefixes can occur in sequences. The norm is for derivational prefixes to occur to the right of inflectional prefixes within the same word-form, as is shown below in the data from Achenese, a language of Sumatra.

(9) (a) jih ji·langũ
he 3rd-person-(younger) (infl)·swim
'he swims (transitive, e.g. swims the river)'
(b) jih ji·mɨ·langũ
he 3rd-person-(younger) (infl)·intransitive (deriv)·swim
'he swims (intransitive)'
(c) jih ji·pɨ·langũ
he 3rd-person-(younger) (infl)·causative (deriv)·swim
'he makes [someone] swim'

Prefixes can, of course, co-occur in the same word with suffixes and all possible combinations of derivational and inflectional are found in such cases.

(10) *English*: un·thank·ful
(deriv) (deriv)