Understanding Morphology

Martin Haspelmath

Max Planck Institute for Evolutionary Anthropology, Leipzig

Understanding Morphology

Martin Haspelmath

Max Planck Institute for Evolutionary Anthropology, Leipzig

What do you think about this book? Or any other Arnold title?
Please send your comments to feedback.arnold@hodder.co.uk

Contents

Preface

This book provides an introduction to the field of linguistic morphology. It gives an overview of the basic notions and the most important theoretical issues, emphasizing throughout the diversity of morphological patterns in human languages. Readers who are primarily interested in understanding English morphology should not be deterred by this, however, because an individual language can be understood in much greater depth when viewed against the cross-linguistic background.

The focus of this book is on morphological phenomena and on broad issues that have occupied morphologists of various persuasions for a long time. No attempt is made to trace the history of linguists' thinking about these issues, and references to the theoretical literature are mostly confined to the 'Further reading' sections. I have not adopted any particular theoretical framework, although I did have to opt for one particular descriptive format for morphological rules (see Section 3.2.2). Readers should be warned that this format is no more 'standard' than any other format, and not particularly widespread either. But I have found it useful, and the advanced student will soon realize how it can be translated into other formats.

Although it is often said that beginning students are likely to be confused by the presentation of alternative views in textbooks, this book does not pretend that there is one single coherent and authoritative view of morphology. Debates and opposing viewpoints are so much part of science that omitting them completely from a textbook would convey a wrong impression of what linguistic research is like. And I did not intend to remain neutral in these debates, not only because it would have been virtually impossible anyway, but also because a text that argues for a particular view is invariably more interesting than one that just presents alternative views.

A number of people have helped me in writing this book. My greatest thanks go to the series editors, Bernard Comrie and Greville Corbett, who provided countless suggestions for improving the book.

Abbreviations

ABL	ablative	EXCL	exclusive
ABS	absolutive	FOC	focus
ACC	accusative	F	feminine
AG	agent	FUT	future
ADJ	adjective	G	gender (e.g. G1 = gender 1)
ADV	adverb(ial)	GEN	genitive
AFF	affirmative	HYP	hypothetical
AGR	agreement	IMP	imperative
ALL	allative	IMPF	imperfect(ive)
ANTIC	anticausative	IMPV	imperative
ANTIP	antipassive	INCL	inclusive
AOR	aorist	INESS	inessive
ART	article	INF	infinitive
ASP	aspect	INSTR	instrumental
AUX	auxiliary	INTF	interfix
CAUS	causative	INTR/intr.	intransitive
CLF	classifier	LOC	locative
COMP	complementizer	M	masculine
COMPL	completive	N	noun
COND	conditional	N	neuter
CONT	continuative	NEG	negation, negative
CONV	converb	NP	noun phrase
DAT	dative	NOM	nominative
DECL	declarative	OBJ	object
DEF	definite	OBL	oblique
DEM	demonstrative	*OED*	*Oxford English Dictionary*
DET	determiner	PASS	passive
DO	direct object	PAT	patient
DU	dual	PERF	perfect
DUR	durative	PFV	perfective
ERG	ergative	PL	plural

POSS	possessive	REFL	reflexive
PP	prepositional phrase	REL	relative clause marker
PRED	predicate	SG	singular
PREF	prefix	SS	same-subject
PRES	present	SUBJ	subject
PRET	preterite	SUBORD	subordinator
PRIV	privative	SUF	suffix
PROG	progressive	TOP	topic
PROPR	proprietive	TR/tr.	transitive
PTCP	participle	V	verb
RECIP	reciprocal	VP	verb phrase

Introduction

1

1.1 What is morphology?

Morphology is the study of the internal structure of words. Somewhat paradoxically, morphology is both the oldest and one of the youngest sub-disciplines of grammar. It is the oldest because, as far as we know, the first linguists were primarily morphologists. The earliest extant grammatical texts are well-structured lists of morphological forms of Sumerian words, some of which are shown in (1.1). They are attested on clay tablets from Ancient Mesopotamia and date from around 1600 BCE.

(1.1) *badu*   'he goes away'   *iṅgen*   'he went'
       baduun   'I go away'   *iṅgenen*   'I went'
       bašidu   'he goes away to him'   *inšiṅen*   'he went to him'
       bašiduun 'I go away to him'   *inšiṅenen* 'I went to him'

(Jacobsen 1974: 53–4)

Sumerian was the traditional literary language of Mesopotamia, but by the second millennium BCE, it was no longer spoken as a medium of everyday communication (having been replaced by the Semitic language Akkadian), so it needed to be recorded in grammatical texts. Morphology was also prominent in the writings of the greatest grammarian of Antiquity, the Indian Pāṇini (fifth century BCE), and in the Greek and Roman grammatical tradition. Until the nineteenth century, Western linguists often thought of grammar as consisting primarily of word structure, perhaps because the classical languages Greek and Latin had fairly rich morphological patterns that were difficult for speakers of the modern European languages.

This is also the reason why it was only in the second half of the nineteenth century that the term *morphology* was invented and became current. Earlier there was no need for a special term, because the term *grammar* mostly evoked word structure, i.e. morphology. The terms *phonology* (for sound structure) and *syntax* (for sentence structure) had existed for

centuries when the term *morphology* was introduced. Thus, in this sense morphology is a young discipline.

Our initial definition of morphology, as the study of the internal structure of words, needs some qualification, because words have internal structure in two very different senses. On the one hand, they are made up of sequences of sounds (or gestures in sign language), i.e. they have internal **phonological structure**. Thus, the English word *nuts* consists of the four sounds (or, as we will say, *phonological segments*) [nʌts]. In general, phonological segments such as [n] or [t] cannot be assigned a specific meaning – they have a purely contrastive value (so that, for instance, *nuts* can be distinguished from *cuts, guts, shuts*, from *nets, notes, nights*, and so on).

But often formal variations in the shapes of words correlate systematically with semantic changes. For instance, the words *nuts, nights, necks, backs, taps* (and so on) share not only a phonological segment (the final [s]), but also a semantic component: they all refer to a multiplicity of entities from the same class. And, if the final [s] is lacking (*nut, night, neck, back, tap*), reference is made consistently to only one such entity. By contrast, the words *blitz, box, lapse* do not refer to a multiplicity of entities, and there are no semantically related words *blit, *bok, *lap.[1] We will call words like *nuts* '(morphologically) **complex words**'.

In a morphological analysis, we would say that the final [s] of *nuts* expresses plural meaning when it occurs at the end of a noun. But the final [s] in *lapse* does not have any meaning, and *lapse* does not have morphological structure. Thus, morphological structure exists if there are groups of words that show identical partial resemblances in both form and meaning. Morphology can be defined as in Definition 1.

> **Definition 1**
> Morphology is the study of systematic covariation in the form and meaning of words.

It is important that this form–meaning covariation occurs systematically in *groups of words*. When there are just two words with partial form–meaning resemblances, these may be merely accidental. Thus, one would not say that the word *hear* is morphologically structured and related to *ear*. Conceivably, *h* could mean 'use', so *h-ear* would be 'use one's ear', i.e. 'hear'. But this is the only pair of words of this kind (there is no *heye* 'use one's eye', *helbow* 'use one's elbow', etc.), and everyone agrees that the resemblances are accidental in this case.

[1] The asterisk symbol (*) is used to mark nonexistent or impossible expressions.

Morphological analysis typically consists of the identification of parts of words, or, more technically, **constituents** of words. We can say that the word *nuts* consists of two constituents: the element *nut* and the element *s*. In accordance with a widespread typographical convention, we will often separate word constituents by a hyphen: *nut-s*. It is often suggested that morphological analysis primarily consists in breaking up words into their parts and establishing the rules that govern the co-occurrence of these parts. The smallest meaningful constituents of words that can be identified are called **morphemes**. In *nut-s*, both the suffix *-s* and the stem *nut* represent a morpheme. Other examples of words consisting of two morphemes would be *break-ing, hope-less, re-write, cheese-board*; words consisting of three morphemes are *re-writ-ing, hope-less-ness, ear-plug-s*; and so on. Thus, morphology could alternatively be defined as in Definition 2.

> **Definition 2**
> Morphology is the study of the combination of morphemes to yield words.

This definition looks simpler and more concrete than Definition 1. It would make morphology quite similar to syntax, which is usually defined as 'the study of the combination of words to yield sentences'. However, we will see later that Definition 2 does not work in all cases, so that we should stick to the somewhat more abstract Definition 1 (see especially Section 3.2.2 and Chapter 9).

In addition to its main sense, where morphology refers to a subdiscipline of linguistics, it is also often used in a closely related sense, to denote a part of the language system. Thus, we can speak of 'the morphology of Spanish' (meaning Spanish word structures) or of 'morphology in the 1980s' (meaning a subdiscipline of linguistics). The term *morphology* shares this ambiguity with other terms such as *syntax, phonology* and *grammar*, which may also refer either to a part of the language or to the study of that part of the language. This book is about morphology in both senses. It is hoped that it will help the reader to understand morphology both as a part of the language system and as a part of linguistics.

One important limitation of the present book should be mentioned right at the beginning: it deals only with spoken languages. Sign languages of course have morphology as well, and the only justification for leaving them out of consideration here is the author's limited competence. As more and more research is done on sign languages, it can be expected that these studies will have a major impact on our views of morphology and language structure in general.

1.2 Morphology in different languages

Morphology is not equally prominent in all (spoken) languages. What one language expresses morphologically may be expressed by a separate word or left implicit in another language. For example, English expresses the plural of nouns by means of morphology (*nut/nuts, night/nights,* and so on), but Yoruba (a language of south-western Nigeria) uses a separate word for expressing the same meaning. Thus, *ọkùnrin* means '(the) man', and the word *àwọn* can be used to express the plural: *àwọn ọkùnrin* 'the men'. But in many cases where several entities are referred to, this word is not used and plurality is simply left implicit.

Quite generally, we can say that English makes more use of morphology than Yoruba. But there are many languages that make more use of morphology than English. For instance, as we saw in (1.1), Sumerian uses morphology to distinguish between 'he went' and 'I went', and between 'he went' and 'he went to him', where English must use separate words. In Classical Greek, there is a dual form for referring to two items, e.g. *adelphō* 'two brothers'. In English it is possible to use the separate word 'two' to render this form, but most of the time one would simply use the plural form and leave the precise number of items implicit.

Linguists sometimes use the terms **analytic** and **synthetic** to describe the degree to which morphology is made use of in a language. Languages like Yoruba, Vietnamese or English, where morphology plays a relatively modest role, are called **analytic**. Consider the following example sentences.[2]

(1.2) Yoruba
 Nwọn ó   maa gbà   pọ́nùn mẹ́wǎ lọ́sọ̀ọ̀sẹ̀.
 they FUT PROG get   pound ten   weekly
 'They will be getting £10 a week.'

 (Rowlands 1969:93)

(1.3) Vietnamese
 Hai đú.a   bo? nhau   là tại   gia-đình thàng chông.
 two individual leave each.other be because.of family guy   husband
 'They divorced because of his family.'

 (Nguyen 1997:223)

When a language has almost no morphology and thus exhibits an extreme degree of analyticity, it is also called **isolating**. Yoruba and Vietnamese, but not English, are usually qualified as isolating. Languages like Sumerian, Swahili (a language of East Africa) or Lezgian (an eastern

---

[2] For each example sentence from an unfamiliar language, not only an idiomatic translation is provided, but also a literal ('morpheme-by-morpheme') translation. The abbreviations are found on pp. xii–xiii, and further notational conventions are explained in the Appendix to Chapter 2.

Caucasian language), where morphology plays a more important role, would be called **synthetic**. Let us again look at two example sentences.

(1.4) Swahili
 Ndovu   wa-wili wa-ki-song-ana   zi-umia-zo   ni   nyika.
 elephants PL-two 3PL-SUBORD-jostle-RECIP 3SG-hurt-REL is   grass
 'When two elephants jostle, what is hurt is the grass.'

 (Ashton 1947:114)

(1.5) Lezgian
 Marf-adi wiči-n   qalin st'al-ra-ldi   qaw   gata-zwa-j.
 rain-ERG self-GEN dense drop-PL-INSTR roof hit-IMPF-PAST
 'The rain was hitting the roof with its dense drops.'

 (Haspelmath 1993:140)

When a language has an extraordinary amount of morphology and perhaps many compound words, it is called **polysynthetic**. An example is Greenlandic Eskimo.

(1.6) Greenlandic Eskimo
 Paasi-nngil-luinnar-para   *ilaa-juma-sutit.*
 understand-not-completely-1SG.SUBJ.3SG.OBJ.INDIC come-want-2SG.PTCP
 'I didn't understand at all that you wanted to come along.'

 (Fortescue 1984:36)

The distinction between analytic and (poly)synthetic languages is not a bipartition or a tripartition, but a continuum, ranging from the most radically isolating to the most highly polysynthetic languages. We can determine the position of a language on this continuum by computing its degree of synthesis, i.e. the ratio of morphemes per word in a random text sample of the language. Table 1.1 gives the degree of synthesis for a small selection of languages.

Language	Ratio of morphemes per word
Greenlandic Eskimo	3.72
Sanskrit	2.59
Swahili	2.55
Old English	2.12
Lezgian	1.93
German	1.92
Modern English	1.68
Vietnamese	1.06

Table 1.1 The degree of synthesis of some languages

Source: based on Greenberg (1959), except for Lezgian.

Although English has much more morphology than isolating languages like Yoruba and Vietnamese, it still has a lot less than many other languages. For this reason, it will be necessary to refer extensively to languages other than English in this book.

1.3 The goals of morphological research

Morphological research aims to describe and explain the morphological patterns of human languages. It is useful to distinguish four more specific sub-goals of this endeavour: elegant description, cognitively realistic description, system-external explanation and a restrictive architecture for description.

(i) **Elegant description**. All linguists agree that morphological patterns (just like other linguistic patterns) should be described in an elegant and intuitively satisfactory way. Thus, morphological descriptions should contain a rule saying that English nouns form their plural by adding -s, rather than simply listing the plural forms for each noun in the dictionary (*abbot, abbots; ability, abilities; abyss, abysses; accent, accents; . . .*). In a computer program that simulates human language, it may in fact be more practical to adopt the listing solution, but linguists would find this inelegant. The main criterion for elegance is **generality**. Scientific descriptions should, of course, reflect generalizations in the data and should not merely list all known individual facts. But generalizations can be formulated in various ways, and linguists often disagree in their judgements of what is the most elegant description. It is therefore useful to have a further objective criterion that makes reference to the speakers' knowledge of their language.

(ii) **Cognitively realistic description**. Most linguists would say that their descriptions should not only be elegant and general, but they should also be cognitively realistic. In other words, they should express the same generalizations about grammatical systems that the speakers' cognitive apparatus has unconsciously arrived at. We know that the speakers' knowledge of English does not only consist of lists of singulars and plurals, but comprises a general rule of the type 'add -s to a singular form to get a plural noun'. Otherwise speakers would be unable to form the plural of nouns they have never encountered before. But they do have this ability: if you tell an English speaker that a certain musical instrument is called a *duduk*, they know that the plural is (or can be) *duduks*. The dumb computer program that contains only lists of singulars and plurals would fail miserably here. Of course, cognitively realistic description is a much more ambitious goal than merely elegant description, and we would really have to be able to look inside people's heads for a full understanding of the cognitive machinery. So this is mainly a programmatic goal at present, but it often affects the way linguists work. Sometimes they reject proposed descriptions because

they seem cognitively implausible, and sometimes they collaborate with psychologists and neurologists and take their research results into account.

(iii) **System-external explanation**. Once a satisfactory description of morphological patterns has been obtained, many linguists ask an even more ambitious question: why are the patterns the way they are? In other words, they ask for explanations. But we have to be careful: most facts about linguistic patterns are historical accidents and as such cannot be explained. The fact that the English plural is formed by adding -s is a good example of such a historical accident. There is nothing necessary about plural -s: Hungarian plurals are formed by adding -k, Swedish plurals add -r, Hebrew plurals add -im or -ot, and so on. Only non-accidental facts, i.e. universals of human language, can be explained, so, before asking why-questions, we must find out which morphological patterns are universal. Clearly, the s-plural is not universal, and, as we saw in the preceding section, not even the morphological expression of the plural is universal – Yoruba is an example of a language that lacks morphological plurals. So even the fact that English nouns have plurals is not more than a historical accident. But there is something about plurals that is not accidental: nouns denoting people are quite generally more likely to have plurals than nouns denoting things. For instance, in Tzutujil (a Mayan language of Guatemala), only human nouns have regular morphological plural forms (Dayley 1985: 139). We can formulate the universal statement in (1.7).

(1.7) *A universal statement:*
    If a language has morphological plural forms of nouns at all, it will
    have plurals of nouns denoting people.

(Corbett 2000: ch. 3)

Because of its 'if . . . then' form, this statement is true also of languages like English (where most nouns have plurals) and Yoruba (where nouns do not have a morphological plural). Since it is (apparently) true of all languages, it is in all likelihood not a historical accident, but reflects something deeper, a general property of human language that can perhaps be explained with reference to system-external considerations. For instance, one might propose that (1.7) is the case because, when the referents of nouns are people, it makes a greater difference how many they are than when the referents are things. Thus, plurals of people-denoting nouns are more useful, and languages across the world are thus more likely to have them. This explanation (whatever its merits) is an example of a system-external explanation in the sense that it refers to facts outside the language system: the usefulness of number distinctions in speech is such a system-external fact, because it concerns exclusively the sphere of language use.

(iv) **A restrictive architecture for description**. Many linguists see an important goal of grammatical research in formulating some general design principles of grammatical systems that all languages seem to adhere to. In other words, linguists try to construct an architecture for description (also

called **grammatical theory**) that all language-particular descriptions must conform to. For instance, it has been observed that rules by which constituents are fronted to the beginning of a sentence can affect syntactic constituents (such as whole words or phrases), but not morphological constituents (i.e. morphemes that are parts of larger words). Thus, (1.8b) is a possible sentence (it can be derived from a structure like 1.8a), but (1.9b) is impossible (it cannot be derived from 1.9a). (The subscript line __ stands for the position that the question word *what* would occupy if it had not been moved to the front.)

(1.8) a. *We can buy cheese.*
     b. *What can we buy __ ?*

(1.9) a. *We can buy a cheeseboard.*
     b. **What can we buy a __-board?*

This restriction on fronting (which seems to hold for all languages that have such a fronting rule) follows automatically if fronting rules (such as *what*-fronting) and morpheme-combination rules (such as compounding, which yields *cheeseboard* from *cheese* and *board*) are separated from each other in the descriptive architecture. A possible architecture for grammar is shown in Figure 1.1, where the boxes around the grammatical components 'syntax', 'morphology' and 'phonology' symbolize the separateness of each of the components.
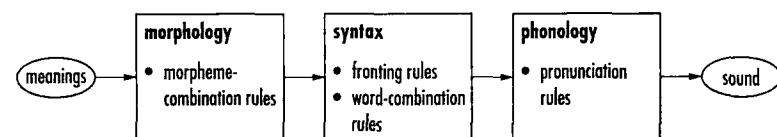
Figure 1.1 A possible descriptive architecture for grammar

This architecture is restrictive because it automatically disallows certain logically possible interactions of rules (see Section 8.5 for more discussion). Many linguists assume that the architecture of grammar is innate – it is the same for all languages because it is genetically fixed for the human species. The innate part of speakers' grammatical knowledge is also called **Universal Grammar**. To sum up, one goal of morphological research is to discover those principles of the innate Universal Grammar that are relevant for word structure.

The goals (iii) and (iv) are similar in that both ask deeper, **theoretical** questions, and both exclusively concern universal aspects of morphology. And both are more ambitious than (i) and (ii) in that they involve

explanation in some sense. Thus, one might say that Universal Grammar-oriented research asks questions such as 'Why cannot constituents of words be fronted to the beginning of the sentence?' and answers them with reference to a hypothesis about the innate architecture of grammar ('Because fronting rules are part of the syntactic component, and morpheme-combinations are part of morphology, and syntax and morphology are separate'). However, explanations of this kind are strictly system-internal, whereas explanations of the kind we saw earlier are even more general in that they link universal properties of grammars to general facts about human beings that are external to the grammatical system.

It is a curious observation on the sociology of science that currently most linguists seem to be concerned either with system-external explanation or with formulating an architecture for grammatical description, but not with both goals simultaneously. There are thus two primary orientations in contemporary theoretical morphological research: the **functionalist** orientation, which aims at system-external explanation, and the **generative** (or **formalist**) orientation, which seeks to discover the principles of the innate grammatical architecture. However, it does not seem wise to divide the labour of morphological research in this way, because neither system-external factors nor innate principles can explain the whole range of morphological patterns. Accordingly, both goals will be simultaneously pursued in the more theoretically oriented parts of this book.

1.4 A brief user's guide to this book

Sources of data

In this book I give examples from many different languages. When they are from well-known and widely studied languages such as Modern English, Russian, Standard Arabic or Old English, I do not give a reference because the data can easily be obtained from any standard reference book. But for examples from less widely known languages, the reference is given after the example.

Sources of ideas

In this book, I focus on morphological data and problems of analysis, not on the history of thinking about these issues in linguistics. Thus, I rarely mention names of particular authors in the text, and references to sources of ideas are given only in a few very specific cases (as in Table 1.1 and ex. (1.7)). In general, the reader is referred to the section 'Further reading', where all the most important works on theoretical morphology are mentioned.

Glossary

The glossary contains the technical terms relating to morphology that are used in this book. In addition to giving a brief definition, the glossary also refers the reader to the most important places where the term is discussed in the text.

Language index

Many languages mentioned in this book will be unfamiliar to the reader. The language index simultaneously serves to give information on each language, in particular about its genealogical affiliation and the place where it is spoken.

Spelling and transcription

Morphology of spoken languages deals with spoken words, so ideally all the examples should be in phonetic transcription in this book. But since many languages have a conventional spelling that renders the pronunciation more or less faithfully, it was more practical and less confusing to adopt that spelling for the examples here. (Although English spelling is not particularly close to the pronunciation, English examples will usually be given in the spelling, because it is assumed that the readers know their pronunciation.) Examples cited in the spelling (or conventional transcription) are always printed in italics, whereas examples cited in phonetic transcription are printed in ordinary typeface and are usually included in square brackets.

Abbreviations

A list of abbreviations (especially abbreviations of grammatical terms) is found on pp. xii–xiii.

Summary of Chapter 1

Morphology is most simply defined as the study of the combination of morphemes to yield words, but a somewhat more abstract definition (as the study of systematic covariation in the form and meaning of words) will turn out to be more satisfactory. Different languages vary strikingly in the extent to which they make use of morphology. The goals of morphological research are (on the descriptive level) elegant and cognitively realistic description of morphological structures, plus (on the theoretical level) system-external explanation and the discovery of a restrictive architecture for description (perhaps based on innate knowledge).

Further reading

For an elementary introduction to morphology, see Coates (1999).

Other morphology textbooks that are somewhat similar in scope to the present book are Bauer (1988) and Bubenik (1999) (as well as Scalise (1994), in Italian, and Plungian (2000), in Russian). Spencer (1991) is a very thorough introduction that concentrates on the generative orientation in morphology. Matthews (1991) puts particular emphasis on the definition of morphological concepts. Carstairs-McCarthy (1991) gives an excellent overview of the theoretical debates in the 1970s and 1980s.

The most comprehensive work on morphology that has ever been written by a single author is Mel'čuk (1993–2000) (five volumes, in French). Although its style is somewhat unusual, it is very readable.

Reference works that are devoted exclusively to morphology are Spencer and Zwicky (1998) and Booij, Lehmann and Mugdan (2000–2). A bibliography is Beard and Szymanek (1988).

The complementarity of the functionalist and the generative approaches to morphology is explained and emphasized in the introductory chapter of Hall (1992).

An introduction to a sign language that also discusses morphology is Sutton-Spence and Woll (1999).

A note on the history of the term *morphology*: in the biological sense ('the study of the form of animals and plants'), the term was coined by Johann Wolfgang von Goethe (1749–1832), and, in the linguistic sense, it was first used by August Schleicher (1859).

Exercises

1. Which of the following English words are morphologically complex? For each complex word, list at least two other words that provide evidence for your decision (i.e. words that are both semantically and formally related to it).

   nights, owl, playing, affordable, indecent, reprimand, indolent, bubble, during, searched, hopeful, redo

2. Identify the morphological constituents and describe their meanings in the following Standard Chinese nouns.

chàngcí	'libretto'	*dǐngdēng*	'top light'
chàngjī	'gramophone'	*diànchē*	'streetcar, tram'
chuánwěi	'stern'	*diàndēng*	'electric lamp'
cíwěi	'suffix'	*diànjī*	'electrical machine'

diànlì	'electric power'	*qìchē*	'car'
diànshì	'television'	*qìchuán*	'steamship'
dòngwùxué	'zoology'	*shāndǐng*	'summit'
dòngwùyóu	'animal oil'	*shìchàng*	'sightsinging'
dòngwùyuán	'zoo'	*shìlì*	'eyesight'
fángdǐng	'roof'	*shùcí*	'number word'
fángkè	'tenant'	*shuǐchē*	'watercart'
fēichuán	'airship'	*shuǐlì*	'waterpower'
fēijī	'aeroplane'	*shùxué*	'mathematics'
fēiyú	'flying fish'	*wěidēng*	'tail light'
huāchē	'festooned vehicle'	*wěishuǐ*	'tail water'
huāyuán	'flower garden'	*yóudēng*	'oil lamp'
jīchē	'locomotive'	*yóuzhǐ*	'oil paper'
jiǎolì	'strength of one's legs'	*yúyóu*	'fishoil'
kèfáng	'guest house'	*zhǐhuā*	'paper flower'

3. Identify the morphological constituents and their meanings in the following Tzutujil verbs (Dayley 1985:87) (A note on Tzutujil spelling: *x* is pronounced [ʃ], and 7 is pronounced [ʔ].)

xinwari	'I slept'	*xoqeeli*	'we left'
neeli	'he or she leaves'	*ninwari*	'I sleep'
ne7eeli	'they leave'	*xixwari*	'you(PL) slept'
nixwari	'you(PL) sleep'	*xe7eeli*	'they left'
xateeli	'you(SG) left'	*xwari*	'he or she slept'
natwari	'you(SG) sleep'		

How would you say 'I left', 'he or she sleeps', 'we sleep'?

4. In the following list of Hebrew words, find at least three sets of word pairs whose two members covary formally and semantically, so that a morphological relationship can be assumed. For each set of word pairs, describe the formal and semantic differences.

kimut	'wrinkling'	*maḥšev*	'computer'
diber	'he spoke'	*masger*	'lock'
ḥašav	'he thought'	*dibra*	'she spoke'
sagra	'she shut'	*milmel*	'he muttered'
ḥašav	'she thought'	*kimta*	'she wrinkled'
kalat	'he received'	*milmla*	'she muttered'
maklet	'radio receiver'	*sagar*	'he shut'
kalta	'she received'	*dibur*	'speech'
kimet	'he wrinkled'		

Basic concepts

2.1 Lexemes and word-forms

The most basic concept of morphology is of course the concept 'word'. The possibility of singling out words from the stream of speech is basic to our writing system, and for the moment let us assume that a word is whatever corresponds to a contiguous sequence of letters (a more sophisticated approach to this problem will be deferred to Chapter 8). Thus, the first sentence of this chapter consists of twelve words, each separated by a blank space from the neighbouring word(s). But when a dictionary is made, not every sequence of letters is given its own entry. For instance, the words *live*, *lives*, *lived* and *living* are pronounced and written differently and are different words in that sense. But a dictionary would contain only a single entry LIVE. The dictionary user is expected to know that *live*, *lives*, *lived* and *living* are different instantiations of the 'same' word LIVE.

Thus, there are two rather different notions of 'word': the 'dictionary word' and the 'text word'. Since this distinction is central to morphology, we need special technical terms for the two notions, **lexeme** and **word-form**.

(2.1) Definitions of *lexeme* and *word-form*

Lexeme: A 'dictionary word' is called a *lexeme* (this is because the mental dictionary in our heads is called the *lexicon* by linguists). Lexemes are abstract entities and can be thought of as sets of word-forms. Sometimes we will use the convention of writing lexemes in small capitals (e.g. LIVE is a lexeme).

Word-form: A 'text word' (i.e. whatever is separated by spaces in writing) is called a *word-form*. Word-forms are concrete in that they can be pronounced and used in texts. Every word-form belongs to one lexeme, e.g. the word-form *lived* belongs to the lexeme LIVE.

In the most interesting case, lexemes consist of a fair number of word-forms. The set of word-forms that belongs to a lexeme is often called a **paradigm**. As an example, the paradigm of the Latin noun lexeme INSULA 'island' is given in (2.2). (Earlier we saw a partial paradigm of two Sumerian verb lexemes (see Section 1.1).)

(2.2) The paradigm of INSULA

	singular	plural
nominative	*insula*	*insulae*
accusative	*insulam*	*insulas*
genitive	*insulae*	*insularum*
dative	*insulae*	*insulis*
ablative	*insula*	*insulis*

Latin nouns have at least ten different word-forms and express notions of number (singular, plural) and case (nominative, accusative, etc.). By contrast, English nouns generally have only two or three word-forms (e.g. ISLAND: *island*, *islands* and perhaps *island's*), but the notional distinction between lexemes and word-forms is no less important when the paradigm is small. In fact, for the sake of consistency we have to make the distinction even when a lexeme has just a single word-form, as in the case of many English adjectives (e.g. the adjective SOLID, which has only the word-form *solid*). Since the lexeme is an abstract entity, its name is quite arbitrary. Usually a particularly frequent word-form is selected from the paradigm to represent the lexeme. Thus, in Latin dictionaries, verbs are listed in the first person singular present form, so SCRIBO stands for the lexeme that means 'write' (*scribo* 'I write', *scribis* 'you write', etc.). In Arabic, by contrast, the third person singular perfect is used in dictionaries, so KATABA stands for the lexeme 'write' (*kataba* 'he wrote', *katabtu* 'I wrote', etc.). This form is called the **citation form**, and it is a purely practical convention with no theoretical significance.

Not all morphological relationships are of the type illustrated in (2.2). Different lexemes may also be related to each other, and a set of related lexemes is sometimes called a **word family** (though it should more properly be called a *lexeme family*):

(2.3) Two English word families
   a. READ, READABLE, UNREADABLE, READER, READABILITY, REREAD
   b. LOGIC, LOGICIAN, LOGICAL, ILLOGICAL, ILLOGICALITY

Although everyone recognizes that these words are related, they are given their own dictionary entries. Thus, the difference between word-forms and lexemes, and between paradigms and word families, is well established in the practice of dictionary-makers, which is known to all educated language users.

At this point we have to ask: why is it that dictionaries treat different morphological relationships in different ways? And why should linguists recognize the distinction between paradigms and word families? After all,

linguists cannot base their theoretical decisions on the practice of dictionary-makers – it ought to be the other way round: lexicographers ought to be informed by linguists' analyses. The nature of the difference between lexemes and word-forms will be the topic of Chapter 4, but the most important points will be anticipated here.

(i) Complex lexemes (such as READER or LOGICIAN) generally denote new concepts that are different from the concepts of the corresponding simple lexemes, whereas word-forms often exist primarily to satisfy a formal requirement of the syntactic machinery of the language. Thus, word-forms like *reads* or *reading* do not stand for concepts different from *read*, but they are needed in certain syntactic contexts (e.g. *the girl reads a magazine; reading magazines is fun*).

(ii) Complex lexemes must be listed separately in dictionaries because they are less predictable than word-forms. For instance, one cannot predict that the lexeme *illogicality* exists, because by no means all adjectives have a corresponding *-ity* lexeme (cf. nonexistent words like **naturality, *logicality*). It is impossible to predict that a specialist in logic should be called a *logician* (rather than, say, a **logicist*), and the meaning of complex lexemes is often unpredictable, too: a *reader* can denote not just any person who reads, but also a specific academic position (in the British system) or even a kind of book. By contrast, the properties of word-forms are mostly predictable and hence do not need to be listed separately for each lexeme.

Thus, there are two rather different kinds of morphological relationship among words, for which two technical terms are commonly used:

(2.4) Kinds of morphological relationship
   inflection   (= inflectional morphology)
              the relationship between word-forms of a lexeme
   derivation   (= derivational morphology)
              the relationship between lexemes of a word family

Morphologists also use the corresponding verbs *inflect* and *derive*. For instance, one would say that the Latin lexeme INSULA is inflected (or inflects) for case and number, and that the lexeme READER is derived from the lexeme READ. A derived lexeme is also called a **derivative**.

It is not always easy to tell how word-forms are grouped into lexemes. For instance, does the word-form *nicely* belong to the lexeme NICE, or does it represent a lexeme of its own (NICELY), which is in the same word family as NICE? Issues of this sort will be discussed in some detail in Chapter 4. Whenever it is unclear or irrelevant whether two words are inflectionally or derivationally related, the term *word* will be used in this book instead of *lexeme* or *word-form*. And for the same reason even the most technical writings on morphology often continue to use the term *word*.

Some morphologically complex words belong to two (or more) word families simultaneously. For instance, the lexeme FIREWOOD belongs both in the family of FIRE and in the family of WOOD. Such relationships are called

compounding, and lexemes like FIREWOOD are called **compound lexemes.** Compounding is often grouped together with derivation under the category of **word formation** (i.e. lexeme formation). The various conceptual distinctions that we have seen so far are summarized in Figure 2.1.
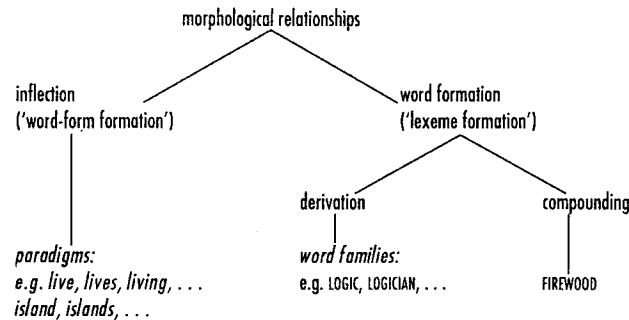
Figure 2.1 Subdivisions of morphology

2.2 Morphemes

We have seen that morphological structure exists if a group of words shows partial form–meaning resemblances. In most cases, the relation between form and meaning is quite straightforward: parts of word-forms bear different meanings. Consider the examples in (2.5).

(2.5) read   read-s   read-er   read-able
       wash   wash-es   wash-er   wash-able
       write   write-s   writ-er   writ-able

       kind   kind-ness   un-kind
       happy   happi-ness   un-happy
       friendly   friendli-ness un-friendly

The words in (2.5) are easily **segmented,** i.e. broken up into individually meaningful parts: read + s, read + er, kind + ness, un + happy, and so on. These parts are called **morphemes.** Words may of course consist of more than two morphemes, e.g. un-happi-ness, read-abil-ity, un-friend-ly, un-friend-li-ness.

Morphemes can be defined as the **smallest meaningful constituents** of a linguistic expression. When we have a sentence such as Camilla met an unfriendly chameleon, we can divide it into meaningful parts in various ways, e.g. Camilla/met an unfriendly chameleon, or Camilla/met/an/unfriendly/chameleon, or Camilla/met/an/un/friend/ly/chameleon. But further division is not possible. When we try to divide chameleon further (e.g. cha/meleon), we do not obtain parts that can be said to be meaningful, either because they

are not found in any other words (as seems to be the case with meleon), or because the other words in which they occur do not share any aspect of meaning with chameleon (cf. charisma, Canadian, caboodle, capacity, in which it would be theoretically possible to identify a word part cha/ca-). Thus, chameleon cannot be segmented into several morphemes, it is **mono-morphemic.** Morphemes are the ultimate elements of morphological analysis; they are, so to speak, **morphological atoms.**

Morphemes can have various kinds of meanings. Some meanings are very concrete and can be described easily (e.g. the meanings of the morphemes wash, logic, chameleon, un-), but other meanings are abstract and more difficult to describe. For instance, the morpheme -al in logic-al can perhaps be said to mean 'relating to' (cf. logic-al, mathematic-al, physic-al, natur-al), -able in read-able can be said to mean 'capable of undergoing a process', and the meaning of -ity is 'quality' (e.g. readability is 'the quality of being readable'). Some meanings are so abstract that they can hardly be called meanings. For example, the Latin morpheme -m in insula-m (see (2.2)) serves to mark the direct object, but it is difficult to say what its meaning is. And English -s in read-s is required when the subject is a third person singular NP, but again it is unclear whether it can be said to have meaning. In such cases, linguists are more comfortable saying that these morphemes have certain **grammatical functions.** But, since the ultimate purpose of grammatical constructions is to express meaning, we will continue to say that morphemes bear meaning, even when that meaning is very abstract and can be identified only in the larger grammatical context.

Equipped with the notion of morpheme, we can now say that morphologically complex words consist of a string of morphemes, in much the same way as sentences consist of a string of words, and morphemes themselves consist of a string of phonemes. This apparent parallelism between sentences, morphemes and phonemes is shown in Figure 2.2.
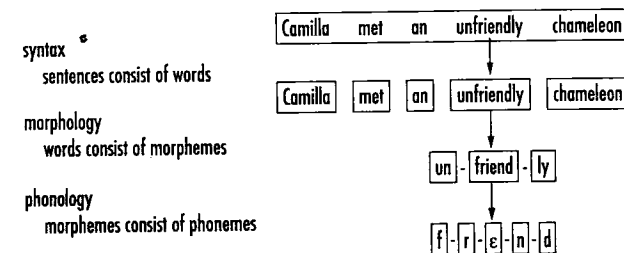
Figure 2.2 A simple picture

However, seductive as the neat picture in Figure 2.2 may be, reality turns out to be more complicated. These complications will occupy us on several occasions later in this book (see especially Section 3.2 and Chapter 9).

For the moment, consider just one example of such a complication. In German, one way of forming the plural of a noun is by replacing a back vowel of the singular form (e.g. [ʊ], [aː], [ɔ]) by a front vowel (e.g. [ʏ], [eː], [œ], spelled *ü, ä, ö*). Some examples are given in (2.6).

(2.6)

singular	plural	
Mutter	*Mütter*	'mother(s)'
Vater	*Väter*	'father(s)'
Tochter	*Töchter*	'daughter(s)'
Garten	*Gärten*	'garden(s)'
Nagel	*Nägel*	'nail(s)'

Here we have a clear-cut example of morphological structure in that a recurrent meaning ('plural') corresponds to a recurrent aspect of form (the front vowel), but the plural word-forms cannot be segmented. There is no segmental part of *Mütter* that could be assigned the meaning 'plural'. Thus, morphology is more than the concatenation of morphemes to form words.

Still, most kinds of morphological structuring can be described in terms of morphemes, and in practical terms the notion of morpheme is very important in morphology.

2.3 Affixes, bases and roots

Word-forms in an inflectional paradigm generally share (at least) one longer morpheme with a concrete meaning and are distinguished from each other in that they in addition contain different shorter morphemes with an abstract meaning. Such short morphemes with an abstract meaning are called **affixes**. For instance, Russian nouns have different case affixes in the paradigm in (2.7) (*-a* for nominative, *-u* for accusative, etc.), and Classical Nahuatl nouns have different possessor prefixes in the paradigm in (2.8) (*no-* for 'my', *mo-* for 'your', etc.).

(2.7) Russian case inflection

nominative	*ruk-a*	'hand'
accusative	*ruk-u*	
genitive	*ruk-i*	
dative	*ruk-e*	
locative	*ruk-e*	
instrumental	*ruk-oj*	

(2.8) Nahuatl possessor inflection

1SG	*no-cal*	'my house'
2SG	*mo-cal*	'your (SG) house'
3SG	*i-cal*	'his/her house'
1PL	*to-cal*	'our house'
2PL	*amo-cal*	'your (PL) house'
3PL	*in-cal*	'their house'

(Sullivan 1988: 26)

Morphologists often use special terms for different kinds of affixes, depending on their position within the word. Affixes that follow the main part of the word are called **suffixes** (e.g. the Russian case suffixes in (2.7)), and affixes that precede it are called **prefixes** (e.g. the Classical Nahuatl possessor prefixes in (2.8)). There are still other kinds of affixes, which are briefly described and illustrated in Table 2.1.

Types of affixes		Examples
suffix:	follows the base	Russian *-a* in *ruk-a* 'hand'
		English *-ful* in *event-ful*
prefix:	precedes the base	Classical Nahuatl *no-cal* 'my house'
		English *un-* in *unhappy*
infix:	occurs inside the base	Arabic *-t-* in *iš-t-aġala* 'be occupied'
		Tagalog *-um-* in *s-um-ulat* 'write'
circumfix:	occurs on both sides of the base	German *ge-...-en*, e.g. *ge-geb-en* 'given'

Table 2.1 Types of affixes

The part of the word that an affix is attached to is called the **base**, e.g. *ruk-* in Russian, or *-cal* in Classical Nahuatl. Affixes and bases can, of course, be identified both in inflected word-forms and in derived lexemes. For instance, in *read-er*, *read-able* and *re-read*, *read* is the base, *-er* and *-able* are suffixes, and *re-* is a prefix. In inflected word-forms, a base is also called a **stem**, and occasionally this term is also used for bases of derived lexemes.

Bases or stems can be complex themselves. For instance, in *activity*, *-ity* is a suffix that combines with the base *active*, which itself consists of the suffix *-ive* and the base *act*. A base that cannot be analysed any further into constituent morphemes is called a **root**. Thus, in *readability*, *read* is the root (and the base for *readable*), and *readable* is the base for *readability*, but it is not a root. Thus, the base is a relative notion that is defined with respect to the notion 'affix' (but we will propose a revised definition of 'base' in the next section). Affixes are similar to roots in that they are primitive elements.

It should be noted that, here and in the following, we are making a terminological simplification: we are talking about lexemes as if they could be broken up into morphemes just like word-forms. But in fact, a lexeme is an abstract entity (see (2.1)), and the lexeme itself cannot be segmented. What we mean when we say that a derived lexeme consists of an affix and a base is that the **stem** of the derived lexeme consists of an affix and a base. In