Basic concepts

2

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 words bear different meanings. Consider the examples in (2.1).

(2.1)

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*	

These words 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**.[1] 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 **monomorphemic**. Morphemes are the ultimate elements of morphological analysis; they are, so to speak, morphological atoms.

In this chapter we introduce some other fundamental concepts and their related terms, starting with *lexemes* and *word-forms*.

[1] Some approaches question the usefulness of the notion 'morpheme'. We will discuss these extensively in Chapters 3 and 4, but for the moment it is helpful to begin in this more conventional way.

2.1 Lexemes and word-forms

The most basic concept of morphology is of course the concept 'word'. For the sake of convenience, let us assume for the moment that a word is whatever corresponds to a contiguous sequence of letters.[2] Thus, in one sense the first sentence of this paragraph consists of twelve words, each separated by a blank space from the neighbouring word(s). And in another sense the sentence has nine words – there are nine *different* sequences of letters separated by spaces. 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 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 concrete instantiations of the 'same' word LIVE. Thus, there are three rather different notions of 'word'. When a word is used in some text or in speech, that occurrence of the word is sometimes referred to as a **word token**. In this sense the first sentence in the paragraph consists of twelve words. The other two senses of the term 'word' are not defined in reference to particular texts; they correspond to the 'dictionary word' and the 'concrete word'. Since this distinction is central to morphology, we need special technical terms for the two notions, *lexeme* and *word-form*, respectively.

A **lexeme** is a word in an abstract sense. LIVE is a verb lexeme. It represents the core meaning shared by forms such as *live, lives, lived* and *living*. In most languages, dictionaries are organized according to lexemes, so it is usually reasonable to think of a lexeme as a 'dictionary word'. Although we must assign names to lexemes to be able to talk about them, lexemes are abstract entities that have no phonological form of their own. LIVE is therefore just a convenient label to talk about a particular lexeme; the sequence of sounds [lɪv] is not the lexeme itself. Sometimes we will use the convention of writing lexemes in small capital letters.

By contrast, a **word-form** is a word in a concrete sense. It is a sequence of sounds that expresses the combination of a lexeme (e.g. LIVE) and a set

[2] Of course, we should really define words in terms of *sounds*, since language is primarily a spoken (not written) medium, and there are other problems with this definition as well. But it is sufficient for the present purposes. A more sophisticated approach is deferred to Chapter 9.

of **grammatical meanings** (or **grammatical functions**) appropriate to that lexeme (e.g. third person singular present tense). *Lives* is a word-form. Thus, word-forms are concrete in that they can be pronounced.

Lexemes can be thought of as sets of word-forms, and every word-form belongs to one lexeme. The word-forms *live, lives, lived*, and *living* all belong to the lexeme LIVE. Word-forms belonging to the same lexeme express different grammatical functions, but the same core concept. When a word-form is used in a particular text or in speech, this instance of use is a word token. The first sentence of this paragraph thus has sixteen word tokens, fifteen word-forms (*of* is repeated), and thirteen lexemes (e.g. *lexemes* and *lexeme* both belong to LEXEME).

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**. The paradigm of the Modern Greek noun lexeme FILOS 'friend' is given in (2.2). (Earlier we saw a partial paradigm of two Sumerian verb lexemes (Section 1.1).)

(2.2) The paradigm of FILOS

	singular	plural
nominative	*filos*	*fili*
accusative	*filo*	*filus*
genitive	*filu*	*filon*

This paradigm contains six different word-forms and expresses notions of number (singular, plural) and case (nominative, accusative, genitive).[3] By contrast, English nouns have no more than four word-forms (e.g. ISLAND: *island, islands* and perhaps *island's, islands'*), 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*).

It is not always immediately clear how many word-forms belong to a lexeme. This is shown by the paradigm of the Latin noun lexeme INSULA 'island' in (2.3). Are there ten word-forms in this lexeme's paradigm, or seven?

(2.3) The paradigm of INSULA

	singular	plural
nominative	*insula*	*insulae*
accusative	*insulam*	*insulās*
genitive	*insulae*	*insulārum*
dative	*insulae*	*insulīs*
ablative	*insulā*	*insulīs*

[3] The meanings of the cases are discussed in Chapter 5. They are also given in the Glossary.

Above we defined a word-form in terms of a lexeme and a set of grammatical functions. The importance of the latter part of the definition is seen in paradigms like INSULA. Although there are only seven *different* sequences of sounds in (2.3), we can still say that the paradigm of INSULA has ten word-forms, because ten different sets of grammatical functions are expressed (e.g. genitive singular and nominative plural are distinct, despite having the same form).

Not all morphological relationships are of the type illustrated in (2.2) and (2.3). 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*):

(2.4) 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, and thereby 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' analyzes. The nature of the difference between lexemes and word-forms will be the topic of Chapter 5, 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:

(2.5) 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**.

(Note that we are making a terminological simplification here: a lexeme is an abstract entity without phonological form so, strictly speaking, one lexeme cannot be derived from another. When morphologists talk about *derived lexemes*, they mean that form *a* (e.g. *reader*), corresponding to lexeme A (READER), is derived from form *b* (*read*), corresponding to lexeme B (READ). However, since this phrasing becomes quite clumsy, morphologists commonly simplify the terminology. We will do the same in this book.)

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 5. 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**, or just **compounds**, for short. 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.

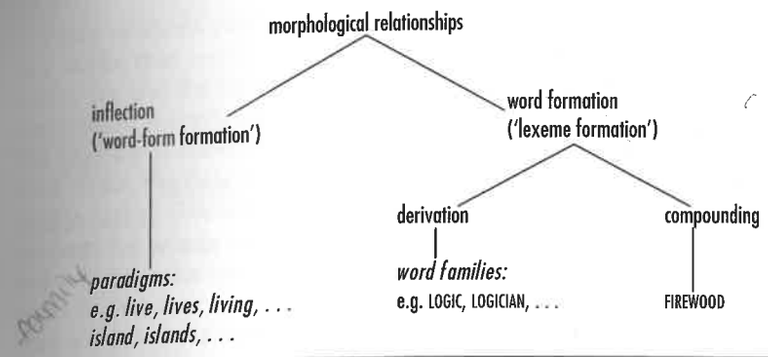

Figure 2.1 Subdivisions of morphology

2.2 Affixes, bases and roots

In both inflection and derivation, morphemes 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 (2.3)) serves to mark the direct object in a sentence, 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 noun phrase, 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.

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 additionally contain different shorter morphemes, called affixes. An **affix** attaches to a word or a main part of a word. It usually has an abstract meaning, and an affix cannot occur by itself. For instance, Russian nouns have different affixes in the paradigm in (2.6), which have case meaning (*-a* for nominative, *-u* for accusative, etc.), and Classical Nahuatl nouns have different affixes in the paradigm in (2.7) that indicate a possessor (*no-* for 'my', *mo-* for 'your', etc.).

(2.6) Russian case inflection (singular forms)

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

(2.7) 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 (2.6)), and affixes that precede it are called **prefixes** (e.g. the Classical Nahuatl possessor prefixes in (2.7)). 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. A base is also sometimes called a **stem**, especially if an inflectional (as opposed to derivational) affix attaches to it.

There are still other kinds of affixes, besides prefixes and suffixes, 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-* in *no-cal* 'my house' English *un-* in *unhappy*
infix:	occurs inside the base	Arabic *-t-* in *(i)š-t-aǧala* 'be occupied' (base: *šaǧala*) Tagalog *-um-* in *s-um-ulat* 'write' (base: *sulat*)
circumfix:	occurs on both sides of the base	German *ge-…-en*, e.g. *ge-fahr-en* 'driven' (base: *fahr*)

Table 2.1 Types of affixes

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 analyzed any further into constituent morphemes is called a **root**. 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'. (We will refine this definition of 'base' in the next chapter to account for words which are difficult to describe in terms of morphemes, but will keep the idea that bases are relative notions.) Affixes are similar to roots in that they cannot be further analyzed into component morphemes; they are primitive elements.

A base may or may not be able to function as a word-form. For instance, in English, *cat* is both the base of the inflected form *cats* and itself a word-form (*active* is a word-form and the base for the derived form *activity*, etc.). However, in Italian word-form *gatti* ('cats') can be broken up into the suffix *-i* ('plural') and the base *gatt-* ('cat'), but *gatt-* is not a word-form. Italian nouns must inflect for number, and even in the singular, an affix is required to express this information (e.g. *gatt-o* 'cat', *gatt-i* 'cats'). In this respect Italian differs from English. Bases that cannot also function as word-forms are called **bound stems**.

Roots and affixes can generally be distinguished quite easily, but sometimes there are problems. For example, the Salishan language Bella Coola has a number of suffix-like elements that do not seem to have an abstract meaning (see 2.8). In (2.9), we see two examples of how these elements are used.

(2.8)

-us	'face'	*-lik*	'body'
-an	'ear'	*-altwa*	'sky, weather'
-uc	'mouth'	*-lt*	'child'
-aɬ	'foot'	*-lst*	'rock'
-ak	'hand'	*-lx̣s*	'nose'

(2.9) a. *quć-aɬ-ic*
wash-foot-I.him
'I am going to wash his foot' (lit.: 'foot-wash him')

b. *kma-lx̣s-c*
hurt-nose-I
'my nose hurts' (lit.: 'I nose-hurt')

(Mithun 1998: 300–5)

In these cases, it is not immediately clear whether we are dealing with suffix–root combinations or with root–root combinations, i.e. compounds. The elements in (2.8) do not occur as lexemes by themselves but must always be combined with other roots. In this respect they have a property that is typical of affixes, and scholars of Salishan languages have generally regarded them as such. However, if affixes are defined as 'short morphemes with an abstract meaning', then these elements are very atypical affixes, to say the least.

English has a number of morphemes that are similarly difficult to classify as roots or affixes. Some examples are given in (2.10).

(2.10)

biogeography	*aristocrat*
bioethics	*autocrat*
bioengineering	*democrat*
biorhythm	*Eurocrat*
bioterrorism	*plutocrat*
biomedicine	*technocrat*
biochip	*theocrat*

The elements *bio-* and *-crat* could be regarded as affixes because they do not occur as independent lexemes, but their very concrete meaning and also their (not particularly short) form suggest that they should be regarded as bound stems that have the special property of occurring only in compounds.

2.3 Morphemes and allomorphs

While the distinction between roots on the one hand and affixes on the other is by itself quite useful, these concepts turn out to be more complicated than the simple picture that we have seen so far. One of the most common complications is that morphemes may have different phonological shapes under different circumstances. For instance, the plural morpheme in English is sometimes pronounced [s] (as in *cats* [kæts]), sometimes [z] (as in *dogs* [dɒgz]), and sometimes [-əz] (as in *faces* [feisəz]). When a single affix has more than one shape, linguists use the term **allomorph**. Affixes very often have different allomorphs – two further cases from other languages are given in (2.11).

(2.11) a. Korean accusative suffix (marker of direct object): two allomorphs

-ul	*ton*	'money'	*ton-ul*	'money-ACC'
	chayk	'book'	*chayk-ul*	'book-ACC'
-lul	*tali*	'leg'	*tali-lul*	'leg-ACC'
	sakwa	'apple'	*sakwa-lul*	'apple-ACC'

b. Turkish first person possessive suffix: five allomorphs

-im	*ev*	'house'	*ev-im*	'my house'
	dil	'language'	*dil-im*	'my language'
-üm	*köy*	'village'	*köy-üm*	'my village'
	gün	'day'	*gün-üm*	'my day'
-um	*yol*	'way'	*yol-um*	'my way'
	tuz	'salt'	*tuz-um*	'my salt'
-ım[4]	*ad*	'name'	*ad-ım*	'my name'
	kız	'girl'	*kız-ım*	'my daughter'
-m	*baba*	'father'	*baba-m*	'my father'

[4] The Turkish letter *ı* corresponds to IPA [ɯ] (high back unrounded vowel).

Not only affixes, but also roots and stems may have different allomorphs (or, as linguists often say, 'exhibit allomorphy'). For instance, English verbs such as *sleep, keep, deal, feel, mean*, whose root has the long vowel [i:] in the present-tense forms, show a root allomorph with short [ɛ] in the past-tense forms (*slept, kept, dealt, felt, meant*). Cases of stem allomorphy from other languages are given in (2.12).

(2.12) a. German: a voiced obstruent becomes voiceless in syllable-final position

Tag	[ta:k]	'day'	*Tage*	[ta:gə]	'days'
Hund	[hʊnt]	'dog'	*Hunde*	[hʊndə]	'dogs'
Los	[lo:s]	'lot'	*Lose*	[lo:zə]	'lots'

b. Russian: when the stem is followed by a vowel-initial suffix, the vowel *o/e* is often dropped if it is the last vowel in the stem

zamok	'castle'	*zamk-i*	'castles'
kamen'	'stone'	*kamn-i*	'stones'
nemec	'German'	*nemc-y*	'Germans'
nogot'	'nail'	*nogt-i*	'nails'

The crucial properties which define the German stems [ta:k] and [ta:g] or the Korean suffixes [-ul] and [-lul] as being allomorphs are that they have the same meaning and occur in different environments in **complementary distribution**. Additionally, all our examples so far have shown only fairly small differences in the shapes of morphemes, which can by and large be regarded as mere differences in pronunciation. Being phonologically similar is a common property of allomorphs, but is not a necessary one. Allomorphs that have this property are **phonological allomorphs**. The formal relation between two (or more) phonological allomorphs is called an **alternation**.

Linguists often describe alternations with a special set of **morphophonological rules**, which were historically phonetically motivated, but affect morphology. Morphophonological rules and the difference between them will be discussed more extensively in Chapter 10, and we will consider them only briefly here.

Metaphorically, it is often convenient to think about phonological allomorphy in terms of a single **underlying representation** that is manipulated by rules under certain conditions. The end result, i.e. what is actually pronounced, is the **surface representation**. For instance, the alternations in (2.12a, b) can be described by the underlying representations in the (a) examples below, and by the respective rules in the (b) examples. The surface representations (resulting word-forms) are given in (c).

(2.13) a. underlying: [ta:g] 'day.SG'
b. rule: a voiced obstruent becomes voiceless in syllable-final position (application: [ta:g] → [ta:k])[5]
c. surface: [ta:k] 'day.SG'

[5] In this (morpho)phonological context, the arrow ('X→Y') means that X turns into Y.

(2.14) a. underlying: [ta:g-ə] 'day-PL'
b. rule: a voiced obstruent becomes voiceless in syllable-final position (doesn't apply)
c. surface: [ta:g-ə] 'day-PL'

(2.15) a. underlying: [zamok] 'castle.SG'
b. rule: *o/e* in the final stem syllable disappears when the stem is followed by a vowel-initial suffix (doesn't apply)
c. surface: [zamok] 'castle.SG'

(2.16) a. underlying: [zamok-i] 'castle-PL'
b. rule: *o/e* in the final stem syllable disappears when the stem is followed by a vowel-initial suffix (application: [zamok-i] → [zamk-i])
c. surface: [zamk-i] 'castle-PL'

Notice that for (2.13) and (2.14), the underlying representation (morpheme) meaning 'day' is the same, and the rule applies only when its conditions are met. The same is true for (2.15) and (2.16). That the alternation is produced by the morphophonological rule is made particularly clear in this way: the underlying representation shows no allomorphy at all.

In many cases of phonological allomorphy, it is evident that the historical reason for the existence of the morphophonological rule and thus for the allomorphy is to facilitate pronunciation. For instance, if the English plural were uniformly [-z], words such as *cats* and *faces* would be almost unpronounceable (try to pronounce [kætz] and [feisz]!). Since this is a textbook on morphology, we cannot go into greater phonological detail here, but phonological allomorphs will be taken up again in Chapter 10.

Overall, the main point here is that at some level, phonological allomorphs represent a single morpheme whose form varies slightly depending upon the phonological context created by combining morphemes. For this reason, it is common to think of the morpheme as the more abstract underlying representation, rather than the more concrete surface word-form. The underlying and surface representations may be the same, or they may differ as a result of the application of morphophonological rules. However, it is important to remember that the underlying representation is a tool used by linguists. It may or may not reflect the kinds of generalizations that language users make. There are examples where it seems unlikely that there is a single underlying representation in the minds of speakers; we see this in another type of allomorphy: **suppletion**.

Besides phonological allomorphs, morphemes may also have allomorphs that are not at all similar in pronunciation. These are called **suppletive allomorphs**. For instance, the English verb *go* has the suppletive stem *wen* in the past tense (*wen-t*), and the English adjective *good* has the suppletive stem *bett* in the comparative degree (*better*). The Russian noun *čelovek* 'human being' has the suppletive stem *ljud'* in the plural (*ljud-i* 'people').

The Spanish verb *ir* 'go' has the suppletive stem *va-* in the present tense (*vas* 'you go', *va* 's/he goes', *vamos* 'we go', etc.). The term *suppletion* is most often used to refer to stem shape (*ir* and *va-* are both verbal stems), and some linguists reserve the term for this use, but others also talk about affixes as being potentially suppletive (see (2.17) later in this chapter for an example from Persian).

It is not always easy to decide whether an alternation is phonological or suppletive, because the categories are end points on a continuum of traits, rather than a clear-cut binary distinction. Some examples are therefore intermediary. For instance, what about English *buy/bought*, *catch/caught*, *teach/taught*? The root allomorphs of these verbs ([bai]/[bɔ:], [kætʃ]/[kɔ:], [ti:tʃ]/[tɔ:]) are not as radically different as *go/wen-t*, but they are not similar enough to be described by phonological rules either. In such cases, linguists often speak of **weak suppletion**, as opposed to **strong suppletion** in cases like *go/went*, *good/better*.

For both weak and strong suppletion, it is theoretically possible to posit an underlying representation from which suppletive allomorphs are derived by rule. However, considering that suppletive allomorphs share little or no form, the underlying representation would need to be very abstract, and the rules converting the underlying representation to surface representations could not exist to make pronunciation easier. There is no evidence that language users make such abstractions, so underlying representations are perhaps best treated as useful metaphors.

Type of allomorphy	Description	Example
Phonological allomorphy	Alternation could be described by a rule of pronunciation	English plural [-z], [-s],[-əz]; Russian *zamok/zamk-*
Weak suppletive allomorphy	Allomorphs exhibit some similarity, but this cannot be described by phonological rules	English *buy/bough-*, *catch/caugh-*, etc.
Strong suppletive allomorphy	Allomorphs exhibit no similarity at all	English *good/bett-*

Table 2.2 Types of allomorphy: summary

When describing the allomorphy patterns of a language, another important dimension is the **conditioning** of the allomorphy, i.e. the conditions under which different allomorphs are selected. Phonological allomorphs typically have **phonological conditioning**. This means that the phonological context

determines the choice of allomorph. For instance, the English plural allomorphs [-z], [-s] and [-əz] are strictly phonologically conditioned: [-əz] appears after a sibilant (i.e. [s], [z], [ʃ], [ʒ], [tʃ] or [dʒ], e.g. *face-s, maze-s, bush-es, garage-s, church-es, badge-s*), [-s] appears after a voiceless non-sibilant obstruent (e.g. *cat-s, book-s, lip-s, cliff-s*) and [-z] appears elsewhere (e.g. *bag-s, bell-s, key-s*). The Korean accusative allomorphs *-ul/-lul* (see (2.11a)) are also phonologically conditioned: *-ul* appears after a consonant, *-lul* after a vowel.

By contrast, stem suppletion usually has **morphological conditioning**, meaning that the morphological context (usually, grammatical function) determines the choice of allomorph (e.g. Spanish *ir* 'go' in the infinitive and future tense, *va-* in the present and imperfective past tense and *fu-* in the perfective past tense).[6]

And, finally, we find **lexical conditioning**, where the choice of a suppletive affix allomorph is dependent on other properties of the base, for instance semantic properties as in (2.17).

(2.17) Persian plural marking: human nouns *-an*, non-human nouns *-ha*

-an	*mœrd*	'man'	*mœrd-an*	'men'
	geda	'beggar'	*geday-an*	'beggars'
-ha	*gorbe*	'cat'	*gorbe-ha*	'cats'
	ettefaq	'incident'	*ettefaq-ha*	'incidents'

(Mahootian 1997: 190)

Lexical conditioning is also involved where the choice of allomorph cannot be derived from any general rule and must be learned individually for each word. This is the case for the English past participle suffix *-en*: speakers must simply learn which verbs take this suffix and not the more common suffix *-ed*.

Type of conditioning	Description	Example
Phonological conditioning	Choice of allomorphs depends on phonological context	English plural depends on final sound in stem
Morphological conditioning	Choice of allomorphs depends on the morphological context	Spanish *ir*, *va-* or *fu-*, depending on tense
Lexical conditioning	Choice of allomorphs depends on the individual lexical item	English past participle *-en*/*-ed* is unpredictable and depends on individual verbs

Table 2.3 Types of conditioning: summary

[6] It is clear that phonological allomorphs can also have morphological conditioning. However, whether suppletive allomorphs can have phonological conditioning is subject to ongoing debate.

Summary of Chapter 2

This chapter introduced several concepts that are basic to morphology. Three different notions of word have to be distinguished: the word token ('use of a word in a text or in speech'), the lexeme ('abstract, dictionary word') and the word-form ('concrete word'). Inflectional morphology describes the relationship between the word-forms in a lexeme's paradigm, and derivational morphology describes the relation between lexemes. Complex words can often be segmented into morphemes, which are called affixes when they are short, have an abstract meaning, and cannot stand alone, and roots when they are longer and have a more concrete meaning. When two or more morphemes express the same meaning and occur in complementary distribution, they are often considered allomorphs. Allomorphs come in two types, phonological and suppletive, depending on the degree to which they are similar in form. Suppletive allomorphs are further subdivided into examples of strong suppletion and weak suppletion. The distinction between strong suppletion, weak suppletion and non-suppletion is a continuum. Allomorphs may have phonological, morphological, or lexical conditioning.

Appendix. Morpheme-by-morpheme glosses

When presenting longer examples (such as sentences or entire texts) from a language that the reader is unlikely to know, linguists usually add **interlinear morpheme-by-morpheme glosses** to help the reader understand the structure of the examples. We saw instances of such glosses in (1.2)–(1.6), and we will see more examples later in this book. Interlinear morpheme-by-morpheme glosses are an important aspect of 'applied morphology', and they are needed in other areas of linguistics as well (e.g. by syntacticians and fieldworkers). We will therefore explain the most important principles involved.

The conventions used in this book are based on the Leipzig Glossing Rules (www.eva.mpg.de/lingua/resources/glossing-rules.php; accessed July 2010). The Leipzig Glossing Rules are more detailed than the principles presented here, but include the following:

(i) One-to-one correspondence. Each element of the object language is translated by one element of the metalanguage (in the present context, this is English). Hyphens separate both the word-internal morphemes in the object language and the gloss, e.g.

Japanese

Taroo	*ga*	*hana*	*o*	*migotoni*	*sak-ase-ta.*
Taro	NOM	flower	ACC	beautifully	bloom-CAUS-PST

'Taro made the flowers bloom beautifully.'

(Shibatani 1990: 309)

Object-language words and their glosses are aligned at their left edges. The interlinear gloss is usually followed by an idiomatic translation in quotation marks.

(ii) Grammatical-value abbreviations. Grammatical elements (both function words and inflectional affixes) are not translated directly, but are rendered by grammatical-value labels, generally in abbreviated form (see the list of abbreviations on pp. xv–xvi). To highlight the difference between the value labels and the ordinary English words, the value labels are usually printed in small capitals, as seen in the above example.

(iii) Hyphens and periods. Hyphens are used to separate word-internal morphemes in object-language examples, and each hyphen in an example corresponds to a hyphen in the gloss. Periods are used in the gloss when two gloss elements correspond to one element in the example. This may be when a single example element corresponds to a multi-word expression in the gloss, e.g.

Turkish
çık-mak
come.out-INF
'to come out'

or it may be when a single example element corresponds to several inflectional meanings:

Latin
insul-ārum
island-GEN.PL
'of the islands'

or it may be when an inflectional meaning is expressed in a way that cannot be segmented, e.g.

Albanian

fik	*fiq*
fig.SG	fig.PL
'fig'	'figs'

(The Albanian letter *q* corresponds here to IPA [c] (voiceless palatal stop), and *k* corresponds to [k] (voiceless velar stop).)

The period is omitted when the two meanings are person and number, e.g.

Tzutujil
x-in-wari
COMPL-1SG-sleep
'I slept'

(Dayley 1985: 87)

Here '1SG' is used instead of '1.SG'. (The period is felt to be redundant because person and number combine so frequently.)

(iv) Possible simplifications. Sometimes the precise morpheme division is irrelevant or perhaps unknown. Authors may still want to give information on the inflectional meanings, and again periods are used to separate these elements, e.g.

Japanese	Latin
sakaseta	*insulārum*
bloom.CAUS.PST	island.GEN.PL
'made to bloom'	'of the islands'

Sometimes morpheme-by-morpheme glosses are used also when the example is not set off from the running text. In such cases the gloss is enclosed in square brackets, e.g. 'the Japanese verb *saka-se-ta* [bloom-CAUS-PST] "made to bloom" ....'.

Comprehension exercises

1. Somali exhibits a great amount of allomorphy in the plural formation of its nouns. Four different allomorphs are represented in the following examples. Based on these examples, formulate a hypothesis about the phonological conditions for each of the plural allomorphs. (In actual fact, the conditions are more complex, but for this exercise, we have to limit ourselves to a subset of the data and generalizations.)

SINGULAR	PLURAL	
awowe	*awowayaal*	'grandfather'
baabaco	*baabacooyin*	'palm'
beed	*beedad*	'egg'
buug	*buugag*	'book'
cashar	*casharro*	'lesson'
fure	*furayaal*	'key'
ilmo	*ilmooyin*	'tear'
miis	*miisas*	'table'
qado	*qadooyin*	'lunch'
shabeel	*shabeello*	'leopard'
waraabe	*waraabayaal*	'hyena'
xidid	*xididdo*	'eagle'

Based on the generalizations found, form the plural of the following nouns:

tuulo	'village'
tog	'river'

albaab	'door'
buste	'blanket'

(Berchem 1991: 98–117)

2. The English past participle suffix spelled *-ed* has three different alternants: [d], [t], and [əd]. Are these phonologically or morphologically conditioned? Try to describe the conditioning factors in an approximate way.

3. Italian inhabitant nouns (e.g. *Anconetano* 'person from Ancona') exhibit different degrees of similarity to the corresponding city names. Order the following pairs of city names and inhabitant names on a scale from clear suppletion in the base form to clear non-suppletion, depending on the number of segments in which the base for the inhabitant noun differs from the base for the city name (see Crocco-Galèas 1991). Assume that word-final vowels are suffixes in Italian; the base for *Ancona* would thus be *Ancon-*. Additionally, inhabitant nouns contain the suffixes *-an*, *-in*, or *-es*, so the base for *Anconetano* is *Anconet-*.

CITY NAME	INHABITANT NOUN	
Ancona	*Anconetano*	
Bologna	*Petroniano*	
Bressanone	*Brissinese*	
Domodossola	*Domese*	
Gubbio	*Eugubino*	
Ivrea	*Eporediese*	
Milano	*Milanese*	'Milan'
Napoli	*Partenopeo*	'Naples'
Palermo	*Palermitano*	
Palestrina	*Prenestino*	
Piacenza	*Piacentino*	
Savona	*Savonese*	
Trento	*Trentino*	
Treviso	*Trevigiano*	
Venezia	*Veneziano*	'Venice'
Volterra	*Volaterrano*	

Exploratory exercise

This chapter introduced the idea that the set of word-forms belonging to the same lexeme is known as a *paradigm*. Readers may have noticed that a table-like format was used to list members of paradigms. The Modern Greek noun paradigm that we encountered in (2.2) is repeated below as (2.18). Here, the rows list cases and the columns list numbers. This format is sometimes called a **grid**. The grid format will be used elsewhere in the book, especially in Chapters 5 and 8, where inflectional morphology is discussed.

(2.18) The paradigm of FILOS 'friend'

	singular	plural
nominative	*fílos*	*fíli*
accusative	*fílo*	*fílus*
genitive	*fílu*	*fílon*

The grid format subtly implies that for a given lexeme there should be a word-form corresponding to each combination of case and number. The format makes sense because the expectation is usually fulfilled; in Greek, noun lexemes almost always have six word-forms corresponding to the six **cells** in the grid. There is thus some sense in which paradigms can be 'complete' or 'incomplete'.

In this exercise, you will explore whether the same notion applies to word families. Do word families usually have an equal number of members and parallel content? For instance, if the verb lexeme READ has READABLE, UNREADABLE, READER and REREAD in its word family, does every other verb lexeme X also have XABLE, UNXABLE, XER and REX? Or do word families often have some lexemes, but not others that seem equally possible? Finally, does the notion of 'completeness' apply to word families? Is it reasonable to talk about a word family as being incomplete? We will address some of these questions in Chapters 5 and 6, but in this exercise you will anticipate that discussion with some exploratory analysis.

English is used here for demonstration purposes because it is familiar to all readers, but you are encouraged to investigate your native language, whatever that might be.

Instructions

Step 1: Create a list of at least 20 adjectival (or nominal, verbal…) lexemes, e.g. CLEAR, FALSE, HAPPY. For each one, list all of the lexemes belonging to its word family. Use a dictionary to prod your memory if needed, but do not rely on dictionary entries when they contradict your own judgements. For instance, the *Oxford English Dictionary* lists the following entries (among others) as being related to the adjective *happy*: *happify*, *happiless*, *happily*, *happiness*, *happious*, *happy-go-lucky*, *happy-slappy*, *enhappy*, *mishappy*, and *trigger-happy*. Some of these, like *happiness*, are quite normal, but others, like *happify*, *happious*, and *enhappy*, seem odd at best. For the authors, the word family of HAPPY does not contain these three lexemes. For each word family in your data set, decide its content for yourself, according to your own usage and judgements about whether a given lexeme is possible.

Step 2: Compare the sets. The lexeme CLEAR is in the same set as CLARIFY, and FALSE is related in a parallel fashion to FALSIFY, but the word family for HAPPY does not contain HAPPIFY (despite being in the dictionary). Do the word families in your data set mostly have parallel content, or mostly not?

Step 3: Discuss the content of these word families in terms of the following questions:

1) In your data, was it ever hard to decide whether two lexemes belong to the same word family? If so, why? Discuss the issues surrounding any choices you had to make.

2) What kinds of meanings are expressed by the derivationally related lexemes? It is not important at this stage to be precise about terminology – describe them as best you can. How do these compare to the inflectional meanings that you have seen in this chapter? Can derivationally related lexemes be organized into grids in the way inflectionally related word-forms are? Why or why not?

3) Does it make sense to talk about word families as complete, or at least potentially complete? Are paradigms and word families similar or different in this respect? Explain your reasoning. (Both a 'yes' and a 'no' answer to the question is probably possible. The important part is that you explain and justify your answer.)

3 Rules

So far we have talked about morphological structure in mostly static terms: words 'have' affixes or 'share' parts, they 'exhibit' resemblances and they 'consist of' a base and an affix. However, it is often convenient to describe complex words as if they were the result of a process or event. Thus, we said that affixes 'are attached' to the base or that they 'combine' with it. Linguists use such process terms very frequently. They talk about elements 'being affixed' to bases, or about a complex word 'being **derived** from' (i.e. built on the basis of) a simpler one.[1]

Most linguists agree that complex words need not be derived from simpler ones each time they are used. Instead, frequently used words are probably listed in the **lexicon**. The lexicon is the linguist's term for the mental dictionary that language users must be equipped with, in addition to the grammatical rules of their language. When a linguist says that something is listed in the lexicon, this means that it must be stored in speakers' memories. If a complex word has its own **lexical entry** (i.e. listing in the lexicon), it does not need to be actively derived from a simpler form; it can simply be retrieved from memory when needed. (The content of the lexicon will be discussed in Chapter 4.)

Still, speakers have the capacity to create, and hearers can understand, an almost unlimited number of new words. The set of words in a language is never quite fixed. There must therefore be some processes by which new complex words are created. And even when a complex word is likely to be listed in the lexicon, it is useful to think of the relationship between it and its base in terms of these same processes. These processes, and how they can be formally described using **morphological rules**, are the topic of this chapter.

[1] The term *derive* is somewhat confusing because it is also commonly applied to inflectional morphology, not just to derivational morphology. Thus, one would say that the comparative form *warmer* is derived from the positive form *warm*, or that the past-tense form *played* is derived from the present-tense form *play*.