10 Building words

In the previous section, we have referred to both *derivational* and *inflectional* processes which enable us to form words from other words. The field of linguistics that examines the internal structure of words and processes of word formation is known as **morphology**, and in this section we shall introduce some of the important ideas in this domain by illustrating their application to English word structure.

Morphemes

Many words in English can easily be split into smaller components. Consider words like *reader*, *printer* and *illustrator*. These are all nouns related to the verbs *read*, *print* and *illustrate*, and they all mean roughly 'person or instrument that *Verb*-s'. Clearly, it is the ending *-er* (with its alternative spelling *-or* in certain words) which conveys this new aspect of meaning and we can say that *-er/-or* creates a new noun from a verb. We can also create new verbs from verbs, as illustrated by pairs such as *read ~ re-read, print ~ re-print* and *illustrate ~ re-illustrate*. Here, the new verb begins with *re-* and means 'to *Verb* something again'. In both these cases, the complex word consists of a number of components, each with its own meaning. We call such components **morphemes**, and to make them easier to identify we can separate them by means of a hyphen (e.g. *read-er*). You will often see the morpheme described as the **minimal linguistic sign**. What this means is that the morpheme is the smallest component of a word which contributes to its meaning. We will see that if we are to subscribe to this, we have to understand 'meaning' rather broadly.

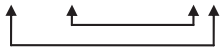
In *reader*, we have a morpheme *-er* attached to a word *read*. However, we cannot split *read* itself into smaller morphemes. This means that we can say that the word *read* is itself a single morpheme. A morpheme which can also stand as a word is called a **free morpheme**. By contrast, *-er/-or* and *re-* are unable to function as free-standing words and these are called **bound morphemes**. The verbs *read*, *print* and *illustrate* are the starting point for the derivation of *reader*, *printer* and *illustrator* in the sense that these verbs specify the activity undertaken by the person to whom *reader*, etc. refers. We therefore assume that *-er/or* and *re-* are attached to the morphemes *read*, *print* and *illustrate* to form the derived words. The ultimate starting point for deriving a word, that is, the most basic morpheme in a word, is its **root**. A morpheme such as *-er/or* added to the right of a root is a **suffix**. One added to the left of the root, such as *re-*, is a **prefix**. The general term covering suffixes and prefixes is **affix**.

We often find more than one affix added to a word. Consider *indecipherability*. The root is the noun *cipher*. From this, we form a verb *de-cipher* from which the adjective *de-cipher-able* is formed. This is then negated by the prefix *in-* to give *in-de-cipher-able*, and finally we create a noun from the adjective by adding *-ity* (and making a change to *-able-*, of which more later, pp. 151f.): *in-de-cipher-abil-ity*. The structures of the items in this sequence can be represented by **labelled bracketings** as in (104) (see section 9, exercise 5):

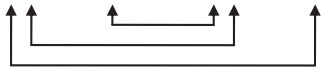
(104)

a. [$_N$ cipher]

b. [$_V$ de [$_N$ cipher]]

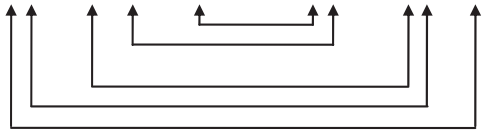


c. [$_A$ [$_V$ de [$_N$ cipher]] able]



d. [$_A$ in [$_A$ [λde [$_N$ cipher]] able]]

e. [$_N$[$_A$ in [$_A$ [$_V$ de [$_N$ cipher]] able]] ity]

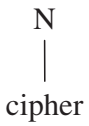


In (104), we have explicitly indicated paired brackets using double-headed arrows, although it should be noted that such arrows are *not* part of the conventional labelled bracketing notation. Taking (104c) for illustration, we have [$_A$ marking the beginning of the adjective *decipherable* and its paired unlabelled bracket marks the end of this word; [$_V$ marks the beginning of the verb *decipher* and the paired unlabelled bracket marks the end of this word; and [$_N$ marks the beginning of the noun *cipher*, the end of which is indicated by the paired unlabelled bracket.

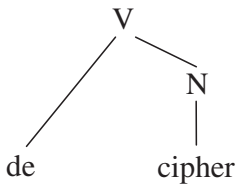
Alternatively, we can represent the same information using the **tree diagrams** in (105):

(105)

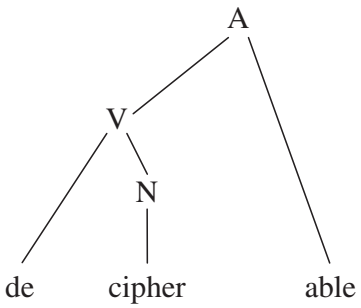
a.



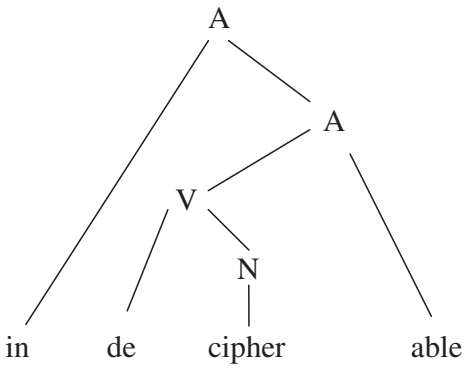
b.



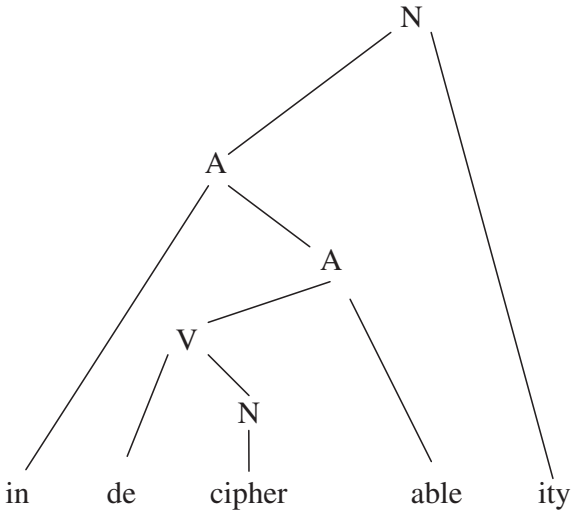
c.



d.



e.



To illustrate the interpretation of such trees, take (105c). This tells us that *cipher* is a noun (N), that *decipher* is a verb (V) formed by adding the prefix *de-* to the noun *cipher* and that *decipherable* is an adjective (A) formed by adding the suffix *-able* to the verb *decipher*.

Although English has a fair number of affixes, it also makes use of a morphological process whereby, *without any affixation*, a word of one syntactic category is used as though it belonged to a different category. This commonly happens when we treat nouns as verbs, as in the examples in (106):

(106) a. Smith *motored* along for three hours
b. Mary *codes* her messages skilfully
c. The tourists are *fishing* near the bridge

Furthermore, we are equally likely to find examples of verbs being used as nouns in such phrases as *a splendid **catch**, a dangerous **run**, a fitful **sleep***. This process is known as **conversion**, and in some cases it is difficult to tell which is the original category. For example, is *rain* basically a verb (107a) or a noun (107b), or is it more appropriate to regard it as having dual-category status, with neither the noun nor the verb being derived from the other?

(107) a. It *rained* every day on our holidays
b. This *rain* is good news for the farmers

Morphological processes – derivation and inflection

One of the key concepts in morphology is that of 'word'. Up to now, we have taken this concept for granted, but at this point we are going to have to be a little more careful. Note first that the term 'word', as it is used in ordinary language, hides an important ambiguity, which we must understand before we can proceed. Consider the following examples:

(108) a. cat
b. cats

(109) a. cat
b. dog

How many words are illustrated in (108) and in (109)? The answer seems clear: two in each example. However, while it is obvious that this is the only answer for (109), there is a sense in which only one word appears in (108). This is the word CAT, with (108a) being its singular form and (108b) the plural. This second sense of 'word' is the one we intend to convey when we say 'this dictionary contains 50,000 words' or 'I know 5,000 words of Greek.' The term we use for this more abstract notion of 'word' is **lexeme**, and when we wish to make it clear that we are discussing a lexeme, the convention is to write it with capital letters. Thus, (108) illustrates only the lexeme CAT, while (109) illustrates the two lexemes CAT and DOG. What, then, of *cat* and *cats* in (108)? These are the singular and plural forms of the lexeme CAT, and we say that (108) illustrates two **word forms** of one lexeme. The singular and plural forms of a lexeme are examples of **inflections**, and we say that CAT inflects for the plural by taking the suffix *-s*. In (109), we again have two word forms (*cat*, *dog*), but these are the singular forms of two lexemes, CAT and DOG. From the point of view of meaning, different lexemes refer to distinct concepts, whereas this is not so for word forms of the same lexeme. Up to this point, then, we have replaced the problematic 'word' with two distinct notions: lexeme and word form.

Returning now to the processes with which we introduced this section, we can ask about the status of *read* and *reader* with respect to the lexeme/word form distinction. Clearly, both *read* and *reader* are word forms, but in addition they refer to rather different (though related) concepts, one a process and the other a physical object taking part in that process. Thus, adding *-er* to a verb creates a new lexeme and READER and READ are distinct lexemes. Of course, each of them has a number of word forms: *reader* and *readers* in the case of READER, and *read* (/riːd/), *read* (/rɛd/), *reads*, *reading* in the case of READ. Moreover, the new lexeme is of a different syntactic category from that of the original lexeme (a verb has become a noun). The creation of new lexemes is the province of **derivational morphology** (or 'derivation'). Of the major lexical categories from section 9, prepositions (P) do not participate in derivation in English (or most other languages for that matter), while adverbs (ADV) are often derived, but only from adjectives, by the suffixation of *-ly* (*bad* ~ *bad-ly*, *noisy* ~ *noisi-ly*, etc.). The other three categories (N, V and A) can, however, readily be derived from each other.

We have already seen that verbs can give rise to nouns via *-er*/*-or* suffixation, and to other verbs via *re-* prefixation. The third possibility for verbs is illustrated by the suffix *-able*. Suffixed to verbs, this gives words such as *read-able*, *print-able*, *illustrat-able*, etc., which are adjectives with the meaning 'such that can be *Verb*-ed'. This suffix is also spelt *-ible* in cases such as *convert-ible*. Starting with adjectives, in *happi-ness*, *sad-ness*, *disinterested-ness*, etc., we create nouns by suffixation of *-ness*. We also find cases in which an adjective is turned into a verb, e.g. by suffixation of *-en* as in *short-en*, *weak-en*, *wid-en*, etc.; and the negative prefix *un-* creates a complex adjective from another adjective as in *un-happy*. Finally, if we take noun roots, we can create adjectives such as *boy-ish* and *child-ish* using the suffix *-ish*, verbs such as *motor-ise* and *demon-ise* with the suffix *-ise* and complex nouns such as *boy-hood*, *child-hood* and *nation-hood* by means of the suffix *-hood*. These options are summarised in table 15 (***exercise 1***).

To date, we have seen various examples of derivations enabling us to form new lexemes in English. Derivation is not the only function of morphology, however. In the previous section, we considered examples such as *Tom reads comics*, pointing out that the verb *reads* consists of the base form *read* and a suffix *-s*. However, this suffix doesn't create a new lexeme; rather it signals agreement with a third person singular subject of the sentence (as well as the fact that the verb is present rather than past tense). Realising agreement is an important function

Table 15 *Examples of derivational morphology in English*

	Derived form		
Basic form	Noun	Verb	Adjective
Noun	boy-*hood*	motor-*ise*	child-*ish*
Verb	print-*er*	*re*-write	read-*able*
Adjective	sad-*ness*	short-*en*	*un*-happy

of **inflectional morphology**, and it is much more widespread in some languages than in English.

The *-s* ending which signals agreement in English is often thought of as a morpheme. However, such a morpheme does not have a meaning in the way that *re-* or *-er*/*-or* have meanings. Rather, it is an inflection which expresses an inflectional category (of agreement) and the purpose of this category is to signal a syntactic relationship, that of the verb to its subject. It is in this sense that we have to interpret rather broadly the notion of a morpheme as a minimal sign *having a single meaning*. Indeed, it is often thought appropriate to resort to a more neutral terminology in such cases. Instead of regarding the English agreement suffix *-s* as a morpheme, we can refer to it as an **inflectional formative** (or simply an 'inflection'), and instead of saying that an inflection means, say, 'third person singular', we say that it is the **exponent** of the property 'third person singular'. As we will see in section 11, there is much more than just terminology at stake here.

A further important concept can now be introduced if we return to (108). We have already noted that (108a, b) illustrate two word forms of the lexeme CAT. However, both of these word forms 'contain' the word form *cat* – (108a) just is *cat*, whereas (108b) is *cat-s*. Thus, we need to observe that the word form *cat* is found in two distinct *functions* in (108). In (108a), it is simply the singular form of the noun, but in (108b), it is the form of the noun to which the plural suffix is added. The form obtained when we remove inflections is called the **stem**. In regular nouns in English the stem is always the same as the singular word form. However, in a plural form such as *knives* the stem is pronounced with a voiced final fricative [naɪv], while the singular ends in an unvoiced fricative [naɪf]. In other words, the plural form of the lexeme KNIFE has a special stem form. Note that the notion of stem is distinct from that of root. The root is the smallest morphological form associated with a lexeme, while a stem is that form to which inflections are added. Thus, the root of the word form *printers* is *print*, but the stem (of the plural form) is *printer-*, which itself consists of a root and a derivational suffix *print-er-*.

The important distinction between lexemes and word forms enables us to explain a widely observed phenomenon in morphology: inflectional affixes tend to appear outside derivational affixes. Thus, in English we have *painter* 'one who paints', a form of a derived noun lexeme (PAINTER), composed of a form of the verb lexeme PAINT and the suffix *-er*. The plural form of this new lexeme is *painter-s* and not **paint-s-er*. This makes sense if we regard plural formation as something which happens to the lexeme. The morphological rule of plural formation is to add *-s* to the end of the stem of the lexeme: *cat-s, painter-s*, and this rule doesn't need to worry about whether the lexeme itself is derived or not. Clearly, we can't form the plural of a derived lexeme such as *painter* until we have created that new lexeme, so we do not see forms such as **paint-s-er*.

A further complication concerning the notion of 'word' can be appreciated if we return to the inflectional categories of English verbs discussed in section 9. If we

take a regular verb lexeme such as CROSS, it has the word forms *cross*, *crosses*, *crossing* and *crossed*. Setting the base form *cross* and third person singular present form *crosses* aside, let's focus attention on *crossed*. As we have observed, one function of this form, illustrated in (110), is to express past tense:

(110) The dog crossed the road safely

Additionally, recall that a form such as *crossed* helps to form the perfect aspect construction with the auxiliary verb *have* as in (111) (see p. 136):

(111) I have crossed this road before

We have referred to the word form *crossed* in this construction as the *perfect participle*, and the same form is found with the passive voice combined with the auxiliary verb *be*:

(112) This river is crossed by three bridges

In (112), *crossed* is referred to as the *passive participle*. But now note that the terminology we have introduced to date for replacing the unclear concept 'word' does not enable us to come to terms with these distinctions. Focusing entirely on the lexical verb, there is only one lexeme in (110)–(112), namely CROSS. Furthermore, there is only one word form of this lexeme in these examples, namely *crossed*. It is necessary, then, to introduce a third sense of 'word' which is distinct from both lexeme and word form. We need to convey the fact that the single word form *crossed* corresponds to *two* distinct inflected forms, the past tense form of CROSS and the perfect/passive participle form of CROSS. We will call a description such as 'the past tense form of CROSS' a **grammatical word** or **morphosyntactic word**. This means that *crossed* corresponds to *two* grammatical words, though it is a *single* word form of a *single* lexeme. At this point, it is useful to recall that in the previous section, we insisted that perfect/passive participle forms should be referred to as the *-n* forms of verbs even when they were suffixed with *-ed*. It should now be clear that the distinction between the *-d* and *-n* forms of verbs which we introduced there is a distinction between two *grammatical words*. In many cases, this distinction corresponds to a distinction between two word forms (*ate* ~ *eaten*, *sang* ~ *sung*, *gave* ~ *given*); in the case of regular verbs, however, only one word form corresponds to these two grammatical words (*crossed* ~ *crossed*, *walked* ~ *walked*, *jumped* ~ *jumped*, etc.) (***exercise 2***).

The *-ing* suffix is also rather complex. Suffixed to a verb form which is combined with the auxiliary *be*, it forms the progressive participle in a progressive aspect construction, as in (113):

(113) Harriet is sending a text message

It is also used to create from a verb a form which has some of the characteristics of nouns, as (114) shows:

(114) Sending text messages is easy

In this example, the phrase *sending text messages* behaves rather like an ordinary noun such as *linguistics* in *linguistics is easy*. However, in a phrase such as *the person sending text messages*, the word seems to behave more like an adjective, in that it forms a phrase, *sending text messages*, which serves to describe *person*, rather like the adjective *responsible* in *the person responsible for this message*. The use of a participle form as an adjective-like modifier is even clearer in an expression such as *running water*.

At this point, it is appropriate to assess the implications of our discussion so far for the lexical entries which form a fundamental component of a grammar. We can now see that it is lexemes which appear in the mental lexicon. When we say that speakers of English know the word *walk*, we are saying that their lexicon contains a lexical entry WALK which provides several kinds of information. Firstly, there is information about the meaning of the lexeme (see section 12). Secondly, there is the syntactic information that it is a V and is intransitive. Thirdly, there is information about how to pronounce all the word forms associated with the lexeme. Now, the lexeme itself doesn't have a pronunciation; rather, it can be realised by one or more word forms and it is they that have a pronunciation. In regular cases the lexical entry just contains the pronunciation of the base form. For instance, the lexeme WALK has the base form *walk* which is pronounced /wɔːk/. Sometimes things are more complex and the lexical entry will contain the pronunciation of certain of the stem forms of a lexeme, as in the case of KNIFE, with its irregular plural stem. In other cases, it is necessary to include the pronunciation of a whole word form, as in the case of the irregular verb BRING with the past tense form, /brɔːt/.

In (115), we see highly simplified lexical entries for WALK, KNIFE and BRING:

(115) a. Lexical entry for WALK

Phonology:	/wɔːk/	base
Syntax:	V, intransitive	
Semantics:	'move on foot with alternate steps'	

b. Lexical entry for KNIFE

Phonology:	/naɪf/	base
	/naɪv/	plural stem
Syntax:	N	
Semantics:	'instrument for cutting'	

c. Lexical entry for BRING

Phonology:	/brɪŋ/	base
	/brɔːt/	[past tense]
Syntax:	V, transitive	
Semantics:	'carry something towards the speaker'	

Other types of information (e.g. the fact that the third person singular present forms of WALK and BRING end in *-s*) are predictable from the principles of

English morphology and therefore don't need to be included in the lexical entries. More subtly, we haven't mentioned the perfect/passive participle form (such as occurs in *has brought* and *was brought*) in (115c), even though this is also irregular. This is because, in the general case, the perfect/passive participle form is identical to the past tense form, and this generalisation of English morphology allows us to predict the perfect participle form of most verbs in the language. There are some exceptions. For example, *sang* is the past tense form of *sing* but *sung* is the perfect participle (*has sung*). In such cases, the lexical entry will have to contain the perfect participle form as well as the past tense form.

Having urged caution with respect to the concept of 'word' in the above discussion and introduced terminology which obviates confusion when precision is called for, we shall continue to use the word 'word' from here on, unless it is necessary to be circumspect.

Compounds

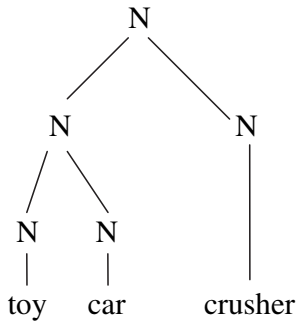
English shares with many languages the ability to create new words by combining old words. For instance, *blackbird* is clearly formed from the adjective *black* and the noun *bird*. However, a blackbird is a different thing from a black bird. Firstly, *blackbird* denotes a particular bird species, not just any old bird that happens to be black; and secondly, female blackbirds are brown, but a black bird has to be black. The expression *blackbird* is a type of word, just like *thrush* or *crow*, but it happens to consist of two words. It is therefore called a **compound word**.

A blackbird is a type of bird, a windmill a type of mill, a coffee table a type of table and so on. We say that *bird*, *mill* and *table* are **heads** of the compounds *blackbird*, *windmill* and *coffee table*. The other part of the compound is a **modifier**. It is possible to form compounds out of compounds. For instance, we can have *finance committee*, *finance committee secretary*, *finance committee secretary election*, *finance committee secretary election scandal* and so on. Now, the way these are written makes them look rather like phrases, but they behave in sentences just like single words. The above list consists of compound nouns and determiners such as *the*, and adjectives such as *efficient* have to precede these compounds just as they would a single non-compound noun: *the highly efficient finance committee secretary*. The fact that they are written with spaces between the elements of the compound is a fact about English orthography and an arbitrary one at that. There are no principled criteria that would tell us whether *windmill* has to be written as one word, as two words (*wind mill*) or as a hyphenated word (*wind-mill*).

There is no theoretical limit to the lengths of compounds because the process of forming compounds can feed itself ad infinitum: a compound noun is itself a noun and can be subject to further compounding. This property is called **recursion** and we say that compounding in English is **recursive**. This is an important property which makes compounding resemble the syntactic processes of phrase- and sentence-formation (see. pp. 3f. and section 19).

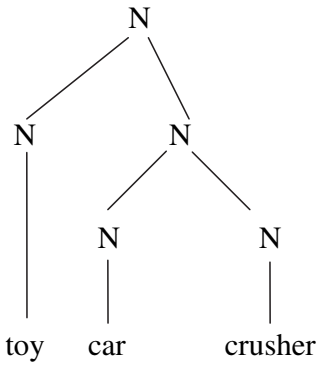
Another respect in which compounding is reminiscent of syntactic processes is in the types of ambiguities it permits. Consider a compound such as *toy car crusher*. This can refer to either a device for crushing toy cars (say, in a recycling factory) or a child's toy modelled after a car crusher. The ambiguity can be represented in terms of labelled brackets and tree diagrams as in (116):

(116) a. toy car crusher 'crusher for toy cars'



$[_N\ [_N\ [_N$ toy$]\ [_N$ car$]]\ [_N$ crusher$]]$

b. toy car crusher 'car crusher which is a toy'



$[_N\ [_N$ toy$]\ [_N\ [_N$ car$]\ [_N$ crusher$]]]$

An ambiguity of this sort, which results from the way the words are bracketed together, is called a **structural ambiguity** (see also section 23). It is an important type of phenomenon because it is very difficult to see how we could explain such ambiguities without resorting to something like the structures in (116). (***exercise 3***).

English permits a variety of compounds. We can combine adjectives with nouns (*sweetcorn*, *lowlife*), or nouns with nouns (*windmill*, *coffee table*). In these cases, it is the first element (*sweet-*, *low-*, *wind-*, *coffee*) which receives the most stress in the compound. We can also combine two adjectives (*dark blue*, *icy cold*) or nouns with adjectives (*canary yellow*, *iron hard*), but in these cases the stress usually falls on the last element. However, in English it is rare for a verb to participate in compounding. Examples such as *swearword* (verb + noun) and *babysit* (noun + verb) are exceptional.

We observed earlier that inflection generally appears outside derivation, a fact that we put down to derivation giving rise to new lexemes and regular inflectional processes such as pluralisation applying to lexemes. Now derivation can appear inside compounding in the sense that a derived word can be compounded with another word. Thus, in the compound *printer cable*, the first element, *printer*, consists of the verb *print* suffixed with *-er*, giving the overall structure [$_N$[$_N$[$_V$print]-er] [$_N$cable]]. We clearly don't first form a (non-existent) compound of the verb *print* and the noun *cable* (**print cable*) and then add *-er* to the *print* component. (Apart from other considerations, as we have just noted, it is virtually impossible in English to form a compound by adding a noun to a verb.)

The situation with regard to inflection is more revealing. Thus, with noun + noun compounds, we seldom find morphology on the first noun. A dog catcher is presumably someone who catches more than one dog, yet we don't say **dogs catcher*, and even if we had a cable for use with several printers we wouldn't call it a **printers cable*. The lack of plurals in this position even extends to words which only ever occur in the plural, so that although there is no noun **trouser*, we do have *trouser leg* and *trouser press*. There are a few cases of plurals inside compounds, e.g. *systems analyst*, *arms control*, but usually the plural form is more than just a simple plural and involves some change of meaning, suggesting that we have a different lexeme from that linked to the singular form. On the other hand, we do have *dog catchers* and *printer cables*. Here, the plural formation rule pluralises the whole compound (***exercise 4***).

Clitics

Another puzzle about words can be illustrated by the examples in (117). How many words are there in each of these examples?

(117) a. it's
b. they've
c. she'll
d. wasn't

It will come as no surprise that there are *two* correct answers. In one sense, *it's* is a single word (indeed, it's just a single syllable), homophonous with (that is, being pronounced identically to) *its*. However, while *its* means 'pertaining or belonging to it' (*its name*, *its function*), *it's* means the same as *it is* or *it has*. Thus, there is a sense in which it combines two distinct words. The *-s*, *-ve*, *-ll* and *-n't* components of the words in (117) correspond to the full words *is/has*, *have*, *will/shall* and *not* and can be thought of as words. However, they can't stand alone in a sentence and they can't be stressed – to be pronounced they have to be attached to some other word (much like an affix). For this reason, they are referred to as **bound words**.

A similar phenomenon is represented by the possessive *-'s* of *Harriet's hat*. It is often thought that *Harriet's* is a suffixed form of *Harriet*, just as the plural form

hats is a suffixed form of *hat*. However, this is misleading, because we can have expressions such as *the man who Harriet met's hat* or *the girl I'm speaking to's hat*. Here, the *-'s* ends up attached to a verb form (*met*) or a preposition (*to*). This is not the normal behaviour of a suffix. What is happening here is that *-'s* is added to the last word of a whole phrase, *the man who Harriet met* or *the girl I'm speaking to*. Unlike the bound word, this type of element never corresponds to a full word and hence it is called a **phrasal affix**.

Bound words and phrasal affixes are examples of **clitics** (from a Greek word meaning 'to lean') and the word that a clitic 'leans' on is its **host**. Clitics such as *-'s* and *-'ve* appear to the right of their hosts, like suffixes. Such clitics are called **enclitics**. In other languages, we find clitics which attach to the left side of the host, as though they were prefixes, called **proclitics**. Pronouns in Romance languages behave like this. Thus, in (118), the Spanish unstressed pronouns *me* 'me' and *las* 'them' appear immediately before the verb:

(118) Me las enseña
me them (he) shows
'He shows them to me'

When the verb is in the imperative form, however, the clitics follow the verb (they are enclitics):

(119) ¡Enséñamelas!
show.me.them
'Show them to me!'

Notice that in Spanish orthography the proclitics are written separately, while the enclitics are written as one word with the verb. However, once more this is merely an orthographic convention, which does not bear at all on the status of these items as clitics.

Allomorphy

We noted earlier that when *-ity* is suffixed to *indecipherable*, a change occurs in the suffix *-able*. Specifically, there is a change in its pronunciation from [əbl̩] to [əbɪl], a change which is reflected by a change in spelling to *-abil-*. To look at what is going on here in a little more detail, we will consider a similar, but more regular, case involving the pronunciation of the suffix *-al*. This creates adjectives from nouns, and its pronunciation also changes when such an adjective is converted to another noun by the suffixing of *-ity*. So consider the sets of examples in (120):

(120) a. nation, nation-al, nation-al-ity
b. music, music-al, music-al-ity
c. tone, ton-al, ton-al-ity
d. origin, origin-al, origin-al-ity.

In each case, *-al* is pronounced as a syllabic /l̩/ at the end of the word and as /al/ before *-ity*. What is happening here is that *-ity* causes the word stress to move to the immediately preceding syllable. When *-al* is unstressed, it is pronounced as /l̩/ but when stressed, it is pronounced with a vowel /a/. This is a regular phonological alternation. Thus, we can say that the morpheme *-al* occurs in two shapes /l̩/ and /al/ depending on stress. The shapes of morphemes as they are actually pronounced in a word are referred to as **morphs**, and where two morphs are variants of one morpheme, we say they are **allomorphs** of that morpheme. The terminology here mirrors that of the phoneme, phone and allophone discussed in section 5.

We have said that the /al/ ~ /l̩/ alternation depends on stress. Since stress is an aspect of the phonology of a word, we can therefore say that the alternation is **phonologically conditioned**. This means that we can describe the difference between the two in purely phonological terms. However, this is not true of all allomorphy. In some cases, a word form will be idiosyncratic in that it contains unusual inflections. Thus, the plural form of the lexeme OX is *oxen*. This is simply a peculiar property of this particular lexeme, and so we say that the plural allomorph *-en* is **lexically conditioned** here (***exercise 5***).

A well-known irregular verb in English is GO. This has a base form /gou/ and a past tense form /wɛnt/, which is completely different. This change in form illustrates the phenomenon of **suppletion**. Since there is no overlap at all in form between *go* and *went*, this is a case of **total suppletion**. The example of *bring* ~ *brought* to which we have already referred (115c) is also a case of suppletion, but as the form /brɔːt/ bears a partial resemblance to the base form /brɪŋ/ (they have the same syllable onset), we say that it is **partial suppletion**. In these cases, we can't say that the allomorphy is triggered by some phonological factor such as stress. Again, we have idiosyncratic properties of the lexemes concerned and so further instances of lexically conditioned allomorphy. Of course, it is precisely such lexically conditioned allomorphs which must appear in lexical entries (***exercises 6*** and *7*).

The concept of allomorphy pertains to morphemes, and it encourages the view that complex word forms consist of strings of morphemes with the form of these morphemes (their allomorphs) being determined by either phonological or lexical factors. However, while this view is attractive in some cases, in others it proves difficult to sustain. We can illustrate the type of problem it confronts by considering again the exponents of the property 'perfect participle'. These include the endings *-ed* (*walked*) and *-en* (*taken*), and perhaps in these cases, it is appropriate to suppose that there is a morpheme PERF(ect) which enables us to analyse *walked* as *walk* + PERF and *taken* as *take* + PERF, with *-ed* and *-en* being treated as lexically conditioned allomorphs of this morpheme PERF. However, we also find forms such as *sing* ~ *sung*, where the perfect participle differs from the base form by virtue of a vowel change. Should we regard *sung* as analysable as *sing* + PERF, with something (what exactly?) being a distinct allomorph of PERF in *sung*? It doesn't make much sense to say this, but it's a question of a type that recurs continually with inflection. An alternative is to say that there is a **morphological process** of perfect participle formation and this can be **realised** in a variety of

ways, including affixation (*-ed* suffixation and *-en* suffixation) and a vowel change. We therefore speak of the affixes *-ed*/*-en* or the vowel change to /ʊ/ in *sung* as **realisations** of the morphological process. Morphologists sometimes also use **exponents**, a term we have already met, for referring to realisations. Adopting this perspective, it is common to represent morphological properties as *features*, similarly to the way we treated phonological properties in section 5, and so we can say that a perfect participle form of a verb has the feature [+perfect participle]. Thus, selecting a verb from the lexicon with this feature is a signal to trigger whatever phonological operation realises that function, whether regular affixation of *-ed*, the irregular *-en* suffixation, vowel change, or the choice of a suppletive form like *brought*.

One upshot of this reasoning is that we don't now have to say that complex words consist of morphemes, neatly strung out in a row, each with its own meaning. Instead, we regard the operations of affixation (if they are what the morphology requires) as separate from the morphological process which is realised by each affixal morpheme. The morphological function itself is then represented by the set of features the word bears. The idea that affixes don't necessarily have a fixed meaning in the way that words do is known as the **Separation Hypothesis**. For simple cases, of course, such as regular plurals or past tenses in English, it does no real harm to simplify the description and treat the affixes as things which have their own form and their own meaning. Thus, for many purposes in syntax it is sufficient to think of the past tense form *walked* as WALK + PAST TENSE, just as *coffee table* is COFFEE + TABLE. However, when we come to look at more complex inflectional systems in the next section, we will see that the notion of Separationism is an important idea.

Exercises

1. This is an exercise in English derivational morphology. Analyse the following words into root and derivational affix. Identify the function of each affix, the lexical category of the root (base category), V, N or A, and the lexical category of the derived word (output category):

   absorbent, defamation, freedom, ladylike, mishear, purify, unaware, accessible, motorise, Marxist, counter-example, encircle, expressive, greenish, broaden, unlock, Roman, obscurity, arrival

Model answer for *absorbent*

The word form *absorbent* comprises a root *absorb* and a suffix *-ent*. The root is a verb indicated by the fact that *absorb* has the forms *absorbs*, *absorbed*, *absorbing* that are characterstic of regular English verbs. Of these forms, *absorbs* signals third person singular agreement illustrated by the contrast between *he absorbs punishment* and

I absorbs punishment*/they absorbs punishment*; *absorbed* has a number of functions, serving as the past tense form of the verb (*he absorbed a lot of punishment yesterday*), the perfect participle form (*he has absorbed a lot of punishment*) and the passive participle form (*a great deal of information was absorbed in the session*); *absorbing* is the progressive participle form of the verb (*he is absorbing the lesson*). The derived form *absorbent* is an adjective, which can be used to modify nouns (*absorbent material*) and can follow forms of the verb *to be* (*this material is absorbent*). While it does not have *-er* and *-est* forms like some adjectives (**absorbenter*, **absorbentest*), the comparative and superlative senses can be expressed by using *more* and *most* (*more absorbent*, *most absorbent*). Thus, *-ent* is a suffix that converts verbs like *absorb* into adjectives like *absorbent*.

2. For each of the following words, give a full grammatical description. Indicate those instances where you need more than one description of a single word form (for instance, *crossed*: 'past tense; perfect/passive participle of the lexeme CROSS').

walks
sheep
cut
left

(Hint: bear in mind what was said about conversion.)

3. Draw tree diagrams for the following compounds. Note that they all have more than one meaning and therefore require more than one tree. How does the tree structure relate to the difference in meaning?
(a) French history teacher
(b) criminal law firm
(c) senate inspection review committee

4. Analyse the following words into morphemes and explain their structure in terms of derivation, inflection, compounding, affixation and conversion. Give a brief explanation of the meaning or function of each bound morpheme.

incomprehensibility	disingenuosity	unhappier
reprivatised	counterintuitively	deforestation
babysitter	party hats	

5. (a) English regular plural allomorphy
Regular nouns in English form their plural by 'adding an *-s* (or sometimes *-es*)': *cats*, *dogs*, *cows*, *horses*, *ostriches*, *flamingos*, etc. However,

this *-(e)s* suffix undergoes phonologically conditioned allomorphy, appearing as [s], [z] or [əz]/[ɪz]. Use the following examples to identify the phonological conditions of this allomorphy (Hint: you will need to pay particular attention to the phonological nature of the final segment of the singular form.):

tops	pots	tabs	pads
packs	bags	cliffs	cloths
classes	clutches	crashes	cruises
cages	cows	quays	suckers
names	manes	rails	

(b) English third person singular and possessive *-'s* allomorphy
Collect together examples of uses of the third person singular ending and the possessive *-'s* phrasal affix, using 5a as a model. Like the English regular plural, this morpheme undergoes allomorphy. Describe this allomorphy and identify the conditioning factors. Compare your results with your answer to 5a. (Hint: don't forget the possessive forms of regular and irregular plural nouns.)
(c) English regular past tense allomorphy
Regular verbs in English form their past tense by 'adding a *-d* (or sometimes *-ed*)': *walked*, *played*, *waited*, etc. However, this *-(e)d* suffix undergoes phonologically conditioned allomorphy, appearing as [t], [d] or [əd]/[ɪd]. Use the following examples to identify the phonological conditions of this allomorphy. Comment on the relationship between this allomorphy and the allomorphy you have described in 5a and 5b.

caged	padded	rolled	crashed
classed	laughed	played	proved
tabbed	bagged	named	moaned
topped	potted	packed	clutched

6. The past tense and perfect/passive participle of *bring* is *brought*. However, children (and some adults) sometimes use the form *brung*. On the other hand, it is very rare for a child to coin a form such as **rought* for the past tense or perfect/passive participle of *ring* (although *ringed* is common in children's speech, see section 13). Why might this be so?

7. Take the verbs BE, HAVE, UNDERGO and SEND. Enumerate all their inflectional forms and transcribe them phonetically. Then segment each word form into morphemes. How many distinct stems do we need for each verb? How many forms show partial suppletion and how many show total suppletion? How many stems are used for more than one word form in each verb?

11 Morphology across languages

The previous section has concentrated almost entirely on English morphological phenomena. In fact, languages differ considerably in the extent and nature of the morphological processes employed in their grammars. Vietnamese, for example, has no bound morphemes, so that the only morphology in the language is compounding. By contrast, there are languages in which morphology is extremely intricate and accounts for much of the grammar's complexity. In this section, we will look at some examples of the types of morphological system that are found in the languages of the world, and the kinds of functions realised by that morphology. A range of the examples we consider will be seen to provide further support for the Separation Hypothesis introduced at the end of the previous section.

The agglutinative ideal

In the last century, linguists introduced a classification of morphological systems which is still often referred to today. This classification distinguished **isolating**, **agglutinating** and **inflectional** languages. We start with isolating languages. These, exemplified by Vietnamese, Chinese and a number of other Far Eastern languages, as well as a number of West African languages, have few, if any, bound morphemes. Thus, in Vietnamese, there is no morpheme corresponding to English *-er* in *driver*, this concept being conveyed by a compound with roughly the structure '*drive* + *person*'.

At the other extreme are languages such as Turkish, Finnish, Hungarian, the Bantu languages of Africa, many languages of the Americas and Australasia and most of the languages of Russia. Here, words of great complexity, consisting of many morphemes, are formed. A (fairly typical) word from the classic example of an agglutinating language, Turkish, appears in (121) (note that this example uses the orthographic system of Turkish):

(121) çalıştırılmamalıymış
'apparently, (they say) he ought not to be made to work'

The segmentation of this word into its component morphemes is indicated in (122):

(122)

çalış-	tır	–	ıl	–	ma	–	malıy	–	mış
work	cause		passive		negation		obligation		inference

The root, the verb *çalış* 'work', comes at the beginning and the suffixes each add their own component of meaning.

Languages such as Turkish give the impression that every morpheme has just one meaning and every meaning in the language is assigned its own unique morpheme. This is often thought of as a kind of morphological ideal, with the characterisation of such languages as agglutinating conveying the idea that morphemes are glued together one by one.

It is indeed the case that a 'perfect' isolating or agglutinating language would have the property that every morpheme would have just one meaning and every individual component of meaning expressible in the language would correspond to just one morpheme. The difference between the two types would be that in an agglutinating language some of the morphemes would be bound, giving the possibility of the construction of complex words like that in (121), whereas in an isolating language they would all be free. In practice, however, there are innumerable deviations from such ideals, and it's unlikely that any language has ever met the ideal. Moreover, there are many languages which show, say, agglutinating tendencies in some areas of grammar and isolating tendencies in others. For this reason, it is much more interesting to ask whether *specific morphological processes* are isolating, agglutinating or something else. Whether a language can be so categorised is something of a non-question. With this background, we can now ask more detailed questions about the kinds of inflectional systems we find in the world's languages.

We begin by contrasting two languages, Latin and Turkish. In tables 16 and 17, we see sets of forms of the Turkish noun EV 'house' and the Latin noun VILLA 'villa, country house'.

Table 16 *Forms of the Turkish noun EV 'house'*

	singular	plural
nominative	ev	evler
accusative	evi	evleri
genitive	evin	evlerin
dative	eve	evlere
ablative	evden	evlerden

Table 17 *Forms of the Latin noun VILLA 'country house'*

	singular	plural
nominative	viːlla	viːllae
accusative	viːllam	viːllaːs
genitive	viːllae	viːllaːrum
dative	viːllae	viːliːs
ablative	viːllaː	viːliːs

These nouns each have sets of singular and plural forms, but in addition they have **case** forms. A case form of a noun is a special form used to indicate various types of grammatical relationship. Roughly speaking, the functions of the cases are as follows: nominative – the basic form of the word; accusative – the form used when the noun is the object of the verb undergoing the action denoted by the verb (e.g. *They painted **the house***); genitive – possession, *of **the house***, dative – *to/for **the house***, ablative – *away from **the house***. However, the meanings are not so important here; our focus is on the way the words are constructed.

The first thing we notice about the Turkish forms is that there is a single set of case endings which are used for both singular and plural: *-i*, *-in*, *-e*, *-den*. Moreover, the exponent of the plural for all case forms is the suffix *-ler*. When we turn to the Latin forms, however, the picture is much less clear. Firstly, there's no single suffix which expresses the property 'plural'. Moreover, the case endings for the singular and plural don't correspond to each other at all. In fact, it's worse than this because the nominative plural form is identical to the genitive and dative singular forms. And yet, when we investigate the Latin noun system, it's clear that we need to distinguish the five cases and the two numbers, because all nouns have sets of distinct forms for the various case/number combinations. The problem is that each Latin noun is only able to take a single suffix. Therefore, each suffix has to be simultaneously the exponent of *two* properties, number and case. When a single affix expresses more than one property within a word form in this way, we say that the affix **cumulates** those properties, and the phenomenon in general is called **cumulation**.

Latin nouns illustrate a further important feature of inflecting languages. The endings of the Turkish word KEDI 'cat' are essentially the same as those in table 16: *kediden* 'from a cat', *kedilerin* 'of the cats' and so on. In table 18, we see the forms of the Latin noun FELES 'cat':

There are only vague similarities between the endings for VILLA and those for FELES. Now, it turns out that there is a very large number of words which take the same endings as VILLA and quite a few which take the same endings as FELES, so this is not just a case of isolated irregularity. Moreover, there are other patterns of endings for other groups of nouns (traditionally, five such classes are recognised). Distinct groups of words with different inflections to express the same sets of properties are called **inflectional classes**. The traditional term for inflectional

Table 18 *Forms of the Latin noun FELES 'cat'*

	singular	plural
nominative	feːleːs	feːleːs
accusative	feːlem	feːliːs
genitive	feːlis	feːlium
dative	feːliː	feːlibus
ablative	feːle	feːlibus

classes of nouns (and adjectives) is **declension**, and the facts of Turkish, briefly referred to above, indicate that it lacks declensions. For verbs, if we find that inflections expressing agreement, tense, etc. fall into distinct classes, as they do in Latin, we speak of **conjugations** (***exercise 1***).

Although it's not immediately apparent, the two Latin nouns we have cited illustrate a further characteristic feature of Latin declensions. If we look at the dative and ablative plural forms of VILLA and FELES, we find that they are identical: *viːlliːs*, *feːlibus*. This identity obtains for all nouns in Latin, and therefore it is a fact about the grammar of Latin. Here we have to say, then, that we have a single word form but that form corresponds to two grammatical words, much like the past tense and perfect/passive participles of English regular verbs (see p. 146). This is a widespread phenomenon in languages such as Latin, and morphologists refer to it as syncretism. We say that the forms *viːlliːs*, *feːlibus* are **syncretic**, and that they **syncretise** the dative/ablative plural.

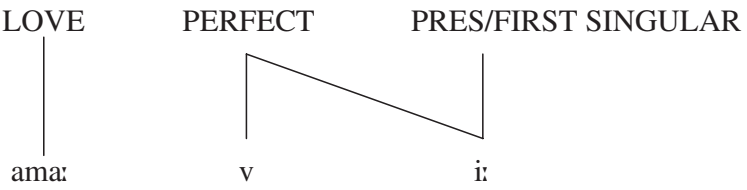
A rather different morphological phenomenon can be observed in Latin verbs. In (123), we see various forms of the verb AMO 'I love':

(123)

am-oː	'I love'
amaːb-oː	'I will love'
amaːb-am	'I was loving'
amaːv-iː	'I have loved'
amaːver-am	'I had loved'

These forms are based on a stem form *amaː*- (or *am*- in the present tense). The final suffix is the exponent of the first person singular form, but notice that it's a different suffix depending on the tense/aspect of the form. In the present and future forms, we have *-oː* but in the two past tense forms, the ending is *-m*, while in the present perfect form, it *-iː*. This kind of variation is different from that illustrated by the different noun suffixes in tables 17 and 18, because here we are dealing with forms of a single lexeme (and, moreover, an example of a completely regular verb in Latin). When we come to analyse a form such as, say, *amaːviː* 'I have loved', what we find is that the *-iː* suffix is not just an exponent of the property first singular – it is also telling us that the verb is present tense and perfect aspect. This is diagrammed in (124):

(124)



In (124), we can see that the property PERFECT is extended over two distinct suffixes. This situation is referred to as **extended exponence** (***exercises 2*** and ***3***).

An interesting fact about English is that a single base form such as *walk* or *book* is in most cases a perfectly good word. Therefore, we are tempted to think that we take the base form of a word and then add inflections to it (e.g. *walk-ing*, *book-s*),

or conversely that we can get to the base form by stripping off the inflections. This makes sense for English and a number of other languages including German, Hungarian and Turkish, but for many inflecting languages, stripping off all the inflections often produces something which cannot function as a word. Thus, the Latin word forms we've seen all need some sort of inflectional ending to form a proper word. The bare root can't be used on its own: **viːll*, **feːl* or **am(a)* are not words in Latin. In other words, the root of a word in such languages is a bound form, not a free form. The same is true, broadly speaking, of Russian, Spanish, Greek, Latin, Japanese, Swahili, Chukchee, Navajo (for verbs at least) and many other languages. Moreover, we sometimes get a different form depending on which set of inflections we strip off. For instance, in Latin the noun meaning 'body' has a basic (nominative) form *corpus*, but its other forms are based on the stem *corpor-* (e.g. *corporis* 'of a body', *corporibus* 'to/from bodies').

While English has small numbers of examples justifying more than one stem appearing in the representation of a lexeme (see the discussion of KNIFE in section 10), we can generally think of its inflection (or that of German, Hungarian, etc.) as being **word-based**, while Latin (or Spanish or Russian, etc.) inflection is **stem-based**. The distinction has implications for psycholinguistic theories of the way that words are processed by the mind/brain and the way that language processing develops in children or is disturbed by brain damage (see sections 15 and 26).

The properties of Latin that we have briefly sampled here are what lead to its exemplifying the class of inflectional languages, and a fourth class of language, often added to the traditional typology, is the class of **polysynthetic** languages. This class is illustrated by Chukchee (also spelled *Chukchi*), a language spoken in NE Siberia. In (125b), we see a *word* which corresponds to the *phrase* in (125a):

(125) a. nəteŋqin ŋelgən
good hide
'a good skin, hide'
b. teŋŋelgən
'a good skin, hide'

In (125a) *teŋ* is the adjective root and *nə-...-qin* combines with this to form an adjective *nəteŋqin* 'good'. In (125b), the adjective root has formed a compound with the noun *ŋelgən* 'hide' to make a single word. There are various ways in which we can show that this is a single word and not just a closely knit phrase, one of which is the fact that adjective roots like *teŋ* never appear without their prefix *nə-* and suffix *-qin* except in compounds.

In (126), we see a similar phenomenon:

(126) a. tə-lʔu-gʔen ŋelgən
I-saw-it hide
'I saw a/the hide'
b. tə-ŋelgə-lʔu-k
I-hide-saw-I
'I saw a/the hide'

In (126a), the verb form *təlʔugʔen* has a prefix *tə-* marking a first person singular subject ('I') and a suffix *-gʔen* marking a third person singular object, agreeing with the direct object *ŋelgən* 'hide'. In (126b), three things have happened. Firstly, the object has now joined the verb and formed a compound verb stem *ŋelgə-lʔu* 'hide-saw'. Secondly, in so doing it has lost the *-n*, which in fact is a case suffix. Thirdly, the verb now ends in a suffix referring again to the first person singular subject. This suffix occurs with *intransitive* verbs in Chukchee, but this is explicable as the verb in (126b) is intransitive. This is because its original object has actually formed a compound with it (to have this compound appear with an object would produce a structure equivalent to the English **I saw the hide the tent* with too many complements).

Compounding of this kind, functioning as an alternative to a syntactically formed phrase, is known as **incorporation**. The noun incorporates its adjective in (125b) and the verb incorporates its object in (126b). Adjective incorporation is not very widespread (though in Chukchee itself it is extremely common), but object incorporation or noun incorporation is very frequently found in the world's languages. In fact, in Chukchee, object incorporation can apply to the result of adjective incorporation:

(127) a. tə-lʔu-gʔen nəteŋqin ŋelgən
I-saw-it good hide
'I saw a/the good hide'

b. tə-lʔu-gʔen teŋ-ŋelgən
I-saw-it good-hide
'I saw a/the good hide'

c. tə-teŋ-ŋelgə-lʔu-k
I-good-hide-saw-I
'I saw a/the good hide'

Here, we first incorporate the adjective into the noun in (127b). Then, this compound noun, which functions as an object in (127b), is incorporated into the verb in (127c). Words like *təteŋŋelgəlʔuk* are not especially uncommon or exotic in Chukchee.

Incorporation is found in a large number of language groups; many languages of the Americas, such as the Iroquoian languages, the Mayan languages, Nahuatl (the language of the Aztecs), large numbers of languages of the Pacific including Maori, Samoan and Tongan, a number of Australian languages, certain of the languages of India and a host of others exhibit incorporation.

What are referred to as polysynthetic languages are those that make use of incorporation in their morphology, though they may also have agglutinating or inflectional processes, or even show isolating tendencies. Chukchee, for instance, is typical in having a large number of very regular derivational processes, which are relatively agglutinating, just like Turkish. However, it also has a rich inflectional system showing cumulation, extended exponence, syncretism and so on (***exercise 4***).

Incorporation processes like those described above strike us as 'exotic'. It is noteworthy, then, that a similar phenomenon is found with a very common type of compound in English. This is illustrated in (128):

(128) a. Tom drives taxis
b. Tom is a taxi-driver

The compound in (128b) includes the object of the verb *drive* from which the deverbal noun *driver* is derived. Similar examples are *taxi-driving*, *insect repellent* and *motorcycle maintenance*. In these compounds, the head is derived from a verb (*drive*, *repel*, *maintain*). The non-head of the compound functions effectively as the object of the verb (see *drive taxis*, *repel insects*, *maintain motorcycles*). This is referred to as **synthetic compounding**. If it were possible to form a verb from these, as in (129), we would have proper noun incorporation in English:

(129) a. *Tom taxidrove yesterday
b. *Agent Orange insectrepels very effectively
c. *Bikers should motorcyclemaintain regularly

Even where it looks as though we have such a case, as in *Dick babysat for Tom and Harriet*, we generally find that there is no syntactic (analytic) equivalent in which the object and the verb are separated: **Dick sat the baby for Tom and Harriet*. The verb *babysit* is just an idiosyncratic form, not a regular compound, and we are justified in concluding that English does not exhibit proper incorporation.

Types of morphological operations

We have already seen numerous examples of prefixation and suffixation, and the examples of vowel changes and suppletions, as in English past tense forms *sang* and *brought*, have indicated that there are additional ways in which the morphological structure of a word can be modified. The Chukchee example in (125) provides another case, where the root *teŋ* in the word *nə-teŋ-qin* is *simultaneously* prefixed and suffixed to form the adjective. A similar phenomenon is seen in German. In regular verbs, the perfect/passive participle is formed by simultaneously adding a prefix *ge-* and a suffix *-t* to the verb stem. Thus, from the stem *hab* 'have' we get *ge-hab-t* 'had'. Since the prefix and suffix are added together, we can think of *nə-...-qin*, or *ge-...-t* as a composite, discontinuous morpheme. Such a morpheme is called a **confix** or **circumfix**.

The languages of the Philippines illustrate another type of affixation. Here are some verb forms in the major language of those islands, Tagalog:

(130)

	verb stem	infinitive	meaning
a.	aral	umaral	'teach'
b.	sulat	sumulat	'write'
c.	basa	bumasa	'read'
d.	gradwet	grumadwet	'graduate'

The crucial thing about these examples is that *aral*, *sulat*, *basa* and *gradwet* are single, undecomposable morphemes. In (130a), we see the prefix *um-* added to a vowel-initial stem. However, (130b, c, d) do not have the infinitive forms **umsulat*, **umbasa*, **umgradwet*. Rather, when the stem begins with a consonant, the affix goes *inside* the stem morpheme, after the onset of the first syllable. This is a regular and pervasive process in Tagalog and several hundred related languages, as can be seen from the fact that it applies to the recent English loan word from *graduate* (130d). An affix which is inserted strictly inside another affix or stem like this is known as an **infix**.

Prefixes and suffixes (and circumfixes) behave like things which are added to stems. This is like compounding in that we simply concatenate two entities, and, indeed, such affixation often develops historically from compounding. Morphology of this type is called **concatenative**, and it encourages the view, briefly discussed in section 10, that complex word forms consist simply of strings of morphemes. However, very often a morphological process seems to be realised by a *phonological operation* performed on the stem itself, as in the case of the vowel changes in *sing ~ sung ~ sang*. Indeed, infixation can be construed in this way as involving first affixation, then a phonological operation which moves the affix to a position inside the stem. It should also be clear that infixation represents another type of deviation from strict agglutination.

Tagalog illustrates a further way in which affixation looks more like a process than a straightforward concatenation of morphemes. Here are some more verb forms in this language:

(131)

	verb stem	future	meaning
a.	sulat	susulat	'write'
b.	basa	babasa	'read'
c.	trabaho	tatrabaho	'work'

From (131) we can see that the future tense form of the verb involves taking the first syllable and copying the first consonant from its onset and its vowel to create a new syllable which appears as a prefix. This type of process is known as **reduplication**, and it provides a rather vivid demonstration of the inappropriateness of suggesting that Tagalog has a morpheme FUTURE with various lexically conditioned allomorphs. Obviously, the list of such allomorphs would be rather long and such a list would fail to make explicit the fundamental fact about Tagalog future formation. This fact is acknowledged by suggesting that there is a morphological feature, say [+future], which can attach to verb lexemes. When this happens, a phonological process is triggered which produces the correct future form of the verb by consulting the syllable structure of the stem form and performing the appropriate operations (***exercise 5***).

On several occasions, in this and the previous section, we have invoked examples of vowel changes in English verb forms as another type of phonological operation which subserves a morphological purpose. Alongside *sing ~ sang ~ sung*, we find

ring ~ rang ~ rung, *hang ~ hung*, *fling ~ flung*, etc., and it is now time to introduce the technical term for this sort of process. It is known as **ablaut** (sometimes called **apophony**). A larger number of English verbs combine a vowel change with suffixation, especially in the participle, so we find such sets of forms as the following: *write ~ wrote ~ written*, *give ~ gave ~ given*, *take ~ took ~ taken*, *do ~ did ~ done*. Each of these simply involves a vowel change in forming the past tense form (the second member of each set); for the participles (the third member of each set), however, there is suffixation of *-en* with or without a vowel change. A specific kind of ablaut, which is particularly common in Germanic languages (and a number of other language groups), occurs when a back vowel is replaced by a front vowel. A number of German plurals are formed this way: /apfl ~ epfl/ 'apple', /fogl ~ føgl/ 'bird', /brudr ~ brydr/ 'brother'. This type of vowel fronting is known as **umlaut**, and there are vestiges of this in English irregular plurals such as *men, teeth* and *geese*.

The last morphological process we shall consider here is represented marginally by some English verbs which are derived from nouns. The difference between *a mouth* and *to mouth* or *a house* and *to house* is that the final consonant is voiced in the verb: /maʊθ ~ maʊð/, /haʊs ~ haʊz/. In the Nilotic language DhoLuo, spoken in Western Kenya, much more systematic use is made of this process in the formation of plurals. Here are some singular and plural forms of nouns in this language:

(132) DhoLuo plurals

	singular		plural	
a.	kede	'twig'	kete	'twigs'
b.	got	'hill'	gode	'hills'
c.	luθ	'stick'	luðe	'sticks'
d.	puoðo	'garden'	puoθe	'gardens'
e.	buk	'book'	buge	'books'
f.	ʧogo	'bone'	ʧoke	'bones'
g.	apwojo	'rabbit'	apwoʧe	'rabbits'
h.	kwaʧ	'leopard'	kwaje	'leopards'

One way of forming a plural involves adding a suffix *-e* as in these examples. In general, when this occurs, the voicing of the final consonant of the stem changes from voiced to voiceless or vice versa (with the palatal glide /j/ being treated as the voiced correlate of the voiceless palato-alveolar affricate /ʧ/).

The above phenomenon exemplifies what is often called **consonant mutation**, and this is even more obvious and varied in its effects in Celtic languages. Look at the way adjectives behave in Literary Welsh when modifying masculine nouns and feminine nouns (adjectives come after nouns in Welsh):

(133) Welsh consonant mutation

	masculine nouns		feminine nouns	
a.	dur klir	'clear water'	nos glir	'clear night'
b	gwint poeθ	'hot wind'	teisen boeθ	'hot cake'

c.	hogin tal	'tall lad'	geneθ dal	'tall girl'
d	ti glan	'clean house'	calon lan	'clean heart'
e	ɬivr bax	'little book'	ferm vax	'little farm'
	[/ɬ/ is a voiceless /l/]			

Operations such as reduplication, ablaut and consonant mutation are rather different from the concatenative types of morphological operation discussed earlier because they do not involve adding anything (such as an affix) to a stem or base in any obvious sense. This type of morphology is often referred to as **non-concatenative morphology**, and, as we have observed, it is very difficult to interpret in terms of the morpheme concept. For instance, in the past tense form *sang*, what is the past tense morpheme? Or in the plural form *men*, what is the plural morpheme? We don't want to say that it is the /a/ or the /ɛ/, because this would imply that the non-past form of SING was */sng/* and the singular form of MAN */mn/*, which is clearly not the case. Earlier, we pointed out that a single morph may realise several different functions at once. Thus, the *-iː* ending of the Latin verb form *amaːviː* 'I have loved' in (124) realises present tense, perfect aspect and first person singular, while the inflectional suffix of a Latin noun realises simultaneously noun declension, number and case. Equally, we have found that a single function may be realised by several different morphs. In the Latin *amaːviː* 'I have loved', both the suffix *-v-* and the suffix *-iː* help to realise the property PERFECT. Similarly, in the English perfect participle form *driven* (/drɪvn̩/), PERFECT is realised by the *-n* suffix and by the process of ablaut applied to the verb root: /aɪ/ ⇒ /ɪ/ (cf. *drive* (/draɪv/). These phenomena are more intelligible if we appeal to Separationism and distinguish the abstract morphological processes of tense formation, agreement, perfect participle formation, plural formation and so on, from the concrete operations of suffixation, ablaut and so on (***exercises 6*** and *7*).

Exercises

1. For the regular Spanish verb forms below, which have been segmented into their constituents, indicate the functions of the suffixes and comment on any difficulties there are in finding a single meaning or function for each suffix. (The accent over a vowel marks exceptional stress, which would otherwise fall on the previous syllable.)

habl-a-r	'to speak'	com-e-r	'to eat'
habl-o	'I speak'	com-o	'I eat'
habl-a-s	'you speak'	com-e-s	'you eat'
habl-a-n	'they speak'	com-e-n	'they eat'
habl-é	'I spoke'	com-í	'I ate'
habl-a-ste	'you spoke	com-i-ste	'you ate'
habl-a-ron	'they spoke'	com-ie-ron	'they ate'

habl-aba	'I was speaking'	com-ía	'I was eating'
habl-aba-s	'you were speaking'	com-ía-s	'you were eating'
habl-aba-n	'they were speaking'	com-ía-n	'they were eating'
habl-a-r-é	'I shall speak'	com-e-r-é	'I shall eat'
habl-a-r-ás	'you shall speak'	com-e-r-ás	'you shall eat'
habl-a-r-án	'they will speak'	com-e-r-án	'they shall eat'

Model answer

The verb forms illustrate three tenses (traditionally called present, preterite and imperfect) and two inflectional classes, the first and second conjugations (there's also a third conjugation not illustrated here). The outermost endings indicate the person/number of the subject of the verb (the person speaking/eating). These are the same for both conjugations. However, they differ depending on the tense as indicated below:

	present	preterite	future
1sg.	-o	-é/í	-é
2sg.	-s	-ste	-ás
3pl.	-n	-ron	-án

In addition, the first singular preterite form depends on the conjugation class. The imperfect tense is indicated by the suffix *-aba* (first conjugation) or *-ía* (second conjugation). In the present and the preterite, there is no special tense marker. However, the two tenses are kept apart by their distinct person/number suffixes. In the future, yet another set of person/number endings is added to a form which is identical to the infinitive form. In the infinitive and present tense forms, the root of the verb is followed immediately by a vowel, *-a* or *-e*, which distinguishes the two conjugations. In the preterite, the second conjugation has instead the vowel *-i* (*-ie* in the third plural form). These conjugation class vowels are traditionally called 'theme vowels'. In the imperfect, there is no separate ending for first singular. In the present and preterite, there is no theme vowel in the first singular forms.

These paradigms illustrate a large number of dependencies. Firstly, all the person/number endings also serve to help indicate the tense, so these can be said to cumulate tense properties as well as expressing their own basic person/number properties. The first singular preterite endings additionally cumulate conjugation class information. The imperfect tense suffixes also indicate conjugation class information, so these cumulate inflectional class with tense properties. The lack of first singular ending (zero morph) in the imperfect serves as an indirect signal of tense.

The data also illustrate extended exponence. The *-ie* theme vowel for the second conjugation preterite form is unique to the third plural form, so this person/number property is signalled twice in the form *comieron* (as is the preterite tense information). The unique future tense endings are added to a special form which is almost always identical to the infinitive. Thus, the property 'future tense' is spread over the *-ar/er* form and the endings themselves.

2. Analyse the following English verb forms to show how they illustrate cumulation, syncretism, inflectional allomorphy and extended exponence. (Hint: you may find it useful to transcribe the verb forms into IPA.)

(she) walk**s**
(they have) dr**i**v**en**
(we) walk
(he) walk**ed**
(you have) sp**o**k**en**

3. Here are some verb forms in Italian (a language closely related to Spanish). Segment the words into their components. In some cases, this will not be straightforward, so comment on any difficulties you have in deciding where the boundaries fall between suffixes. Then indicate any instances of cumulation and extended exponence in the data. Finally, identify any syncretisms you find in these paradigms. (The present subjunctive is a form used in contexts where the speaker isn't entirely certain of the truth of the statement.)

number	person	present indicative	present subjunctive	past indicative
parlare 'to speak'				
	1	parlo	parli	parlai
sing.	2	parli	parli	parlasti
	3	parla	parli	parlò
	1	parliamo	parliamo	parlammo
plural	2	parlate	parliate	parlaste
	3	parlano	parlino	parlarono
credere 'to believe'				
	1	credo	creda	credei
sing.	2	credi	creda	credesti
	3	crede	creda	credè
	1	crediamo	crediamo	credemmo
plural	2	credete	crediate	credeste
	3	credono	credano	crederono

finire 'to finish'

	1	finisco	finisca	finii
sing.	2	finisci	finisca	finisti
	3	finisce	finisca	finì
	1	finiamo	finiamo	finimmo
plural	2	finite	finiate	finiste
	3	finiscono	finiscano	finirono

4. Below are some Chukchee words, slightly simplified. Segment them into their component morphemes and provide a rough meaning for each morpheme. Comment on the types of affixation found and on any allomorphy you observe.

ekwetək	to set off
eretək	to fall
nəwilək	to come to a halt
rəgelək	to go in
rənwiletək	to stop someone
rərgeletək	to introduce
rərgelewək	to lure in
rərultetək	to move something away
rətejŋetək	to feed (something to someone)
rətenmawək	to prepare (something)
rekwetewək	to send someone off (on a journey)
reretək	to drop
rultək	to step aside
runtəmewetək	to calm someone
tejŋetək	to eat (something)
tenmawək	to get oneself ready
untəmewək	to calm oneself down

5. In the data below we see examples of reduplication in the Palan dialect of Koryak (a language closely related to Chukchee). What is the rule for forming a noun of this kind in Koryak?

ʧajʧaj	'tea'	həlwehəl	'wild reindeer'
jiŋejiŋ	'mist'	jilhejil	'gopher'
kalikal	'book'	liŋliŋ	'heart'
mətqmət	'fat'	milgmil	'fire'
nutenut	'tundra'	tərgtər	'meat'
wətwət	'leaf'	wiruwir	'seal'
ʔawtaʔaw	'flint'		

6. Some plural forms in Arabic are very difficult to predict from the singular form. However, there are patterns. What is the common, invariant component of the following Arabic nouns (the forms are slightly

simplified in some cases)? How can the plural be constructed from the singular form in each case? (A doubled vowel, e.g. *aa*, represents a long vowel, e.g. [aː]; representing long vowels in this way may make it easier to see the principles that underlie this system. Note that the nouns come in two groups depending on the form of the singular.)

singular	plural	meaning
qidħ	qidaaħ	arrow
ʤamal	ʤimaal	camel
ħukm	ħakaam	judgement
ʔasad	ʔusuud	lion
jundub	janaadib	locust
raʤul	riʤaal	man
ʕinab	ʕanaab	grape
nafs	nufuus	soul
saħaabat	saħaaʔib	cloud
ʔumθulat	ʔamaaθil	example
ʤaziirat	ʤazaarʔir	island
ħaluubat	ħalaaʔib	milch-camel
kariimat	karaaʔim	noble
marħalat	maraaħil	stage

7. What deviations from agglutination are exhibited by the Swahili verb forms shown below? (The data are slightly simplified.)

(a)	i.	nilitaka	I wanted	tulitaka	we wanted
		ulitaka	you (sg.) wanted	mlitaka	you (pl.) wanted
		alitaka	he/she wanted	walitaka	they wanted
	ii.	nitataka	I shall want	tutataka	we shall want
		utataka	you (sg.) shall want	mtataka	you (pl.) shall want
		atataka	he/she shall want	watataka	they shall want
	iii.	ninataka	I want	tunataka	we want
		unataka	you (sg.) want	mnataka	you (pl.) want
		anataka	he/she wants	wanataka	they want
(b)	i.	sikutaka	I did not want	hatukutaka	we did not want
		haukutaka	you (sg.) did not want	hamkutaka	you (pl.) did not want
		haakutaka	he/she did not want	hawakutaka	they did not want
	ii.	sitataka	I shall not want	hatutataka	we shall not want
		hautataka	you (sg.) shall not want	hamtataka	you (pl.) shall not want
		haatataka	he/she/it shall not want	hawatataka	they shall not want
	iii.	sitaka	I do not want	hatutaka	we do not want
		hautaka	you (sg.) do not want	hamtaka	you (pl.) do not want
		haataka	he/she/it does not want	hawataka	they do not want