English Words:
History and Structure

ROBERT STOCKWELL AND
DONKA MINKOVA

4   Smaller than words: morphemes and types of morphemes

1   The smallest meaningful units

We think of words as being the most basic, the most fundamental, units through which meaning is represented in language. There is a sense in which this is true. Words are the smallest free-standing forms that represent meaning. Any word can be cited as an isolated item. It can serve as the headword in a dictionary list. It can be quoted. It can be combined with other words to form phrases and sentences. In general the word is the smallest unit that one thinks of as being basic to saying anything. It is the smallest unit of sentence composition and the smallest unit that we are aware of when we consciously try to create sentences.

But actually there are even smaller units that carry the fundamental meanings of a language. Words are made up of these units. Consider just the unit *gen* in Figure 4.1. It is clearly not a free-standing word, but rather some kind of smaller unit which goes into the make-up, the composition, of words:
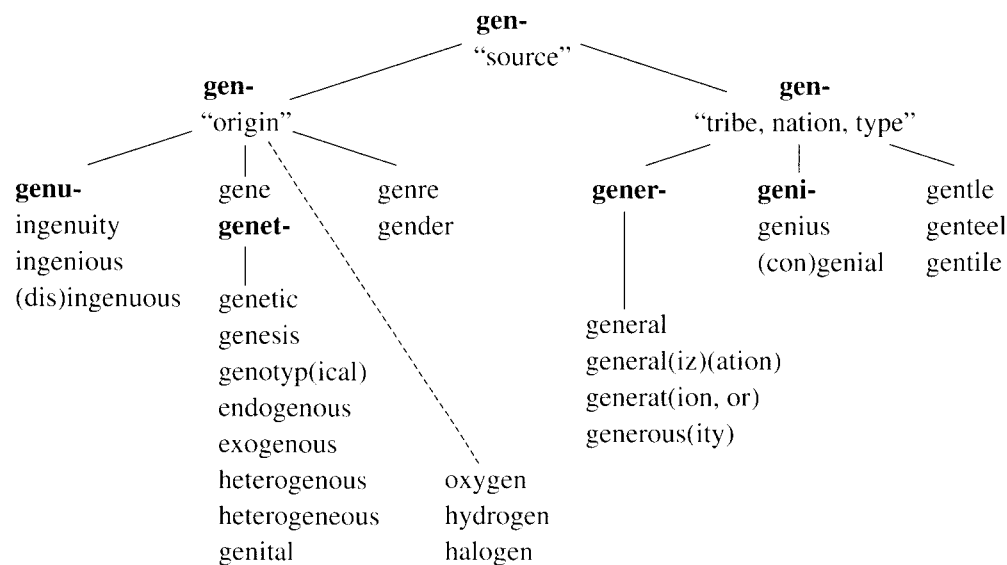

Figure 4.1

These smaller units are called **morphemes**. *Gen* is a morpheme. It has a basic single meaning "source" (at the top of the diagram) which has split into two now distinct meanings, "origin" and "tribe, nation, type." Looking at the words that appear under each of these meanings, one can readily see the difference. The meaning "origin" is most easily seen on the middle branch below it, in words like *genetic* or *genital*. The meaning "type" is most easily seen in words like "general." The meaning "gentle" is derived from the notion "belonging to a good tribe/family." There is some overlap: words like *generate* could just as reasonably be attached to the left branch ("origin"). *Gentile* originally, in the Vulgate translation of the Bible from Hebrew into Latin, simply meant "nations/tribes other than the Jews." A later, seventeenth-century re-borrowing of the same word, *gentil*, from French, originally also meaning "of good stock," resulted in our word *genteel*, which has changed its erstwhile positive meaning of "having proper breeding," "elegant, stylish," to its rather ironic and disparaging current meaning. *Genius* comes about through Greek mythology. It meant "The tutelary god or attendant spirit allotted to every person at his birth, to govern his fortunes and determine his character, and finally to conduct him out of the world" (*OED*). It is a short step from "attendant spirit" to "having a genius for music" to "being a musical genius." *Genre* is used by literary scholars to mean "a literary type." *Gender* refers to types, or categories, of nouns, in the usage of grammarians. In general usage it has recently become the accepted term, though it is an obvious euphemism, for two categories of humans, male and female, differentiated by virtue of their sex.

1.1   Morphemes and syllables

Since morphemes are the smallest carriers of meaning, each word must contain at least one morpheme. Some words are made up of more than one morpheme. The word *morpheme* itself is made up of two morphemes: *morph* "form, shape" and *-eme* "meaningful." So a *morpheme* is a meaningful unit of form. The essential point about morphemes is that they cannot be dissected further into smaller meaningful units: they are themselves the smallest ones. But one might challenge this claim by pointing out that *morph* itself consists of the sound <m> plus the sound <o> plus the sound <r> plus the sound <ph> (= <f>). Why are these not smaller units? The answer is that they are not units of **meaning**. They are units of **sound** (or spelling) which serve together to **represent** the morpheme. The relationships between sounds, morphemes, and meanings is like this, taking *gen* as our example.
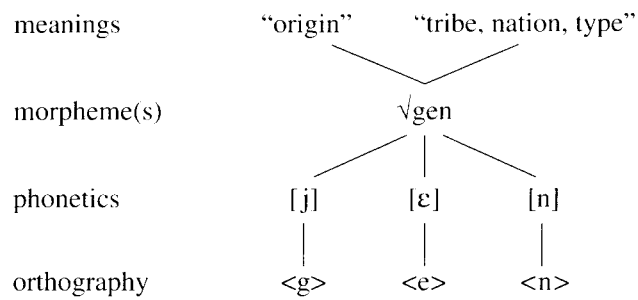
meanings   "origin"   "tribe, nation, type"

morpheme(s)   √gen

phonetics   [j]   [ε]   [n]

orthography   <g>   <e>   <n>

Figure 4.2

In Figure 4.2 we find some technical symbols. The checkmark in front of the morpheme *gen* simply means, "This is a morpheme." Although the common mathematical meaning of the symbol is "root," we use "√" for all morphemes, roots and affixes both. We could not write, for example, √genetic, because *genetic* is more than one morpheme. We would have to write something like √gen √et √ic, indicating that it is made up of three morphemes. The square brackets on the level below √gen indicate pronunciation. They are used for phonetic writing, also known as phonetic transcription. Don't worry if you don't recognize all the phonetic symbols: we will explain them later. Finally, in the orthography row, the angle brackets mean "This is a letter of the alphabet traditionally used to spell this sound in this word."

1.2   The properties of morphemes

We summarize below the properties of morphemes in an effort to show how they differ from other linguistic units like syllables, words, and individual sounds. The properties which uniquely differentiate morphemes from other linguistic units are these:

(1) A morpheme is the smallest unit associated with a meaning. As an example, consider the following words:

car	**care**	**car**pet	**car**digan
carrot	**care**ss	**car**go	**car**amel
vi**car**	s**care**	dis**card**	pla**card**

Each of these words contains the spelling <c a r>. How can we determine whether this fact is significant or not? Answer: Ask whether there is some constant meaning in each word that can be attributed to a morpheme having the form <car>. It is obvious that *care* has nothing to do with *car* – the meaning of *car* is completely independent of the meaning of *care*. Take *caress*: although superficially *caress* resembles *car* the way *princess* resembles *prince*, there is clearly no shared meaning in the first pair. *Carpet*, on the other hand, looks as though

somehow (imaginably) there might be a connection: perhaps *carpet* could be a little *carp* (as *cigarette* is a little *cigar*), but of course it is not. So *carpet* is a single morpheme, and can be written √carpet. We can also write √care, √cardigan, √carrot, √caress, √cargo, √caramel, √scare, and √√vicar, following the same logic in each instance. They merely "accidentally" contain <car>, but they do not contain the morpheme √car. (Why can we not do this also with *discard* and *placard*?)

(2) Morphemes are recyclable units. One of the most important properties of the morpheme is that it can be used again and again to form many words. The morpheme √care can be used to form *uncaring*, *careful*, *careless*, *caregiver*, and we saw in Figure 4.1 that √gen is used in forming dozens of words. If you did not know the meaning of the words *cardigan* and *caramel*, and if you thought that they might contain the morpheme √car, one way to test your conjecture would be to see whether the remaining material can be used in other words, – i.e., whether it is another morpheme. Obviously, *-digan*, and *-amel* do not meet our first definition of a morpheme, they are not contributors of independent meanings, nor are they recyclable in the way in which the morphemes √care + √ful, or √un + √care + √ing, or √care + √give + √er are. One should be careful, however: recyclability can be deceptive, as it was in the case of *carrot, carpet, caress, cargo*. Though all morphemes can be used over and over in different combinations, non-morphemic parts of words may accidentally look like familiar morphemes.

The test defined in (1) above, namely that what makes a sequence of sounds a morpheme is its ability to convey independent meaning, or add to the meaning of the word, should always be applied first. However, there are some interesting cases for which the decision on whether some part of a word is a morpheme or not requires a combination of tests (1) and (2). If we try to parse the word *happy*, we can easily isolate √-y as a morpheme: it adds to the grammatical meaning of the word by turning it into an adjective. But what about *happ-*? Taken in isolation, it does not mean anything to a speaker whose knowledge of etymology does not extend to Old Norse. In Old Norse there was a noun *happ*, meaning "luck, chance." The word was borrowed into English in the twelfth century. That morpheme is no longer likely to appear by itself, but it has kept its ability to turn up in various words and to form the core of their meaning: *mishap, happen, happenstance, hapless, unhappiness*. In other words, the recyclability of √hap(p)- in the language today confirms its status as a morpheme, even without the etymological information. As you will see, many of the classical morphemes we will be dealing with in this book are of the √hap(p)- type.

(3) Morphemes must not be confused with syllables. A morpheme may be represented by any number of syllables, though typically only one or two, sometimes three or four. Syllables have nothing to do with

meaning. Syllables are units of pronunciation. In most dictionaries, hyphens are used to indicate where one may split the word at the end of a line. Hyphens are also used to separate the word into syllables. A syllable is the smallest independently **pronounceable** unit into which a word can be divided. The number of morphemes in a word is very likely to differ from the number of syllables. √Car and √care are one syllable each; √carpet, √caress, √carrot, and √cargo are two syllables each; and √cardigan and √caramel have three syllables. But each of these words is a single morpheme.

Morphemes may be **less** than a syllable in length, too. Consider *car* vs. *cars*. *Cars* is one syllable, but two morphemes, namely √car + √-s, where √-s is the morpheme that means "plural," i.e., more than one. Other examples of morphemes which are not syllables include the √-ed of *cared*, *caressed*, and the √-th of *growth*, *warmth*. Generally, however, morphemes are independently pronounceable and are at least one syllable in length, like √gen, √morph, √hap(p), and √-y. In some (few) instances, a morpheme may be present only by inference. If we say *The sheep are grazing*, we have to infer that *sheep* is plural, even though its form is the same as the form of the singular. If we say *I cut some flowers yesterday*, we understand *cut* to be in the past tense because of *yesterday*, not because of any morpheme attached to *cut*. The longest morphemes tend to be names of places or rivers or Indian nations, like *Mississippi*, or *Potawatomi*, or *Cincinnati*. In the indigenous languages of America from which these names were borrowed, the words were polymorphemic, but that information is completely lost to us, as English speakers.

(4) One and the same morpheme may take phonetically different shapes. Different forms of the same morpheme are called "allomorphs" (which means "other forms" – √allo "other" + √morph "form"). This general property of allomorphic variation is called **allomorphy**. Recognizing different allomorphs of the same morpheme is one of the surest ways to extend one's vocabulary and to identify relationships between words. Any speaker of English will identify the nouns *cares*, *caps*, and *classes*, as sharing the plural morpheme √-s, though both the spelling and the pronunciation of the morpheme vary in the three words. That is, the morpheme has three allomorphs. But although the allomorphy of the plural √-s is part of everyone's core knowledge of English, there are many morphemes where this knowledge is not at all automatic. Consider the morpheme meaning "take" or "contain" whose most familiar allomorph is √cap, as in words like:

    capable   **cap**sule   **cap**tive   **cap**acity

It also has the allomorph √cep in words like

    ac**cep**t   de**cep**tion   inter**cep**t   per**cep**tible   re**cep**tacle

It has a third allomorph √cip in words like

    anti**cip**ate   eman**cip**ate   in**cip**ient   parti**cip**ate   prin**cip**al
    re**cip**e   re**cip**ient

The fourth allomorph is √cup in a few words like *occupy* and *recuperate*. In Chapter 9 we will discuss the fact that the meaning of the morpheme *cap* is **transparent** in some of these words – e.g., *captive* is "one who has been taken," *capable* is "able to take," *participate* is "to take part." But it is **opaque** in others. For example, it is not obvious what *perceptible* has to do with the basic meanings "take" or "contain." Perhaps the association is something rather vague like "able to take in [through one of the senses]." It is even more opaque in *recuperate* – perhaps something like "to take back [one's health]," and *anticipate* "take beforehand." Finally, the connection between *recipe* and √cap is recoverable only with the help of an etymological dictionary: it is the imperative form of the verb "take" in Latin. That's how the word *recipe* acquired its first meaning of a "formula for a medical prescription." Later, it was extended further into cooking and, of course, today, into general transferred use as a "list of ingredients and a set of actions" in phrases such as "a recipe for success, a recipe for disaster" – a far cry from the basic meaning of √cap. The interplay of formal variation and semantic variation is very complex; we will be returning to these issues in the following chapters.

These, then, are the four essential properties of all morphemes: (1) they are packaged with a meaning; (2) they can be recycled; (3) they may be represented by any number of syllables; and (4) morphemes "morph," i.e., they may have phonetically different shapes.

2   Types of morphemes

2.1   Roots

Not all morphemes are equally central to the formation of a word. Morphemes are of two main types: **roots** and **affixes**. We turn our attention first to roots. Every word has at least one root. Roots are at the center of word-derivational processes. They carry the basic meaning from which the rest of the sense of the word can be derived. Morphemes such as *chair*, *green*, *ballet*, *father*, *cardigan*, *America*, *Mississippi*, are roots; these roots also happen to be free forms, i.e., independent words. But more often, roots are like *seg* in *segment*, *gen* in *genetics*, *card* in *cardiac*, *sequ* in *sequence*, *brev* in *brevity*, *pter* in *pterodactyl*. These cannot stand alone as words. They are called **bound root morphemes**, as distinct from **free root morphemes** (the ones that are also independent words).

Most bound roots found in the language today are of classical origin – i.e., they were borrowed into English from Latin or Greek during the Renaissance, or through French. Moreover, we usually borrowed words from these languages wholesale, i.e., the classical roots came into the language nested inside derived forms such as *segment, genetics, cardiac* etc. Sometimes, though not very frequently, borrowed roots do make their way into the inventory of free forms too: *contra* in the meaning of "counterrevolutionary" (1981), *graph* (1878), *phone* (1866) are some examples. As you can see from the dates of their first recorded appearance in English, such roots which are also independent words are fairly recent and formed by shortening of the classical words or phrases that contain them, i.e. their transition from bound to free roots has occurred on English soil. So, it would be fair to say that roots borrowed from classical sources are nearly always bound roots.

On the other hand, the number of bound roots of Germanic origin like *happ* "luck, fortune," as in *hapless, happy*, is comparatively small. Of Germanic origin are the bound roots of *feckless, reckless, ruthless, listless, uncouth, unkempt*. What has happened in all these cases is a straightforward historical change: a root, which used to be also a word at earlier times, became obsolete or disappeared completely, leaving behind only a derivative. Thus *feckless* is derived from a sixteenth-century Scots word *feck*, a shortened form of *effect*. Later *effect* was reintroduced into Scots, replacing the form *feck*, yet its derivative *feckless* "ineffective" is still around. We consider *ruth* in *ruthless* as a bound morpheme today, but it used to be a common word in English meaning "pity, sorrow" well into the eighteenth century. Though it appears as an entry in many dictionaries, the word is obsolete in present-day English, but note its connection with the verb "to rue," from which the noun was obviously derived originally. In any case, the historical processes we are illustrating here are not recoverable without the aid of specialized dictionaries. For the ordinary speaker of English *feck-, hap(p)-, ruth-* etc. are bound roots.

To be completed, bound root morphemes require that another morpheme be attached to them. This additional morpheme may be either another root or an affix. If it is another root, the result is a **compound**. Some issues related to compounds and compounding were discussed in Chapter 1: you will remember that words like *airship, birdcage, bookmark, flagship, hemisphere, hydrogen, phonograph, polymath, telephone* etc. are compounds. They all contain two roots. If a bound root is not attached to another root, as in **brevity, capable, cardiac, gentile**, etc., it must be accompanied by an **affix**.

Affixes carry very little of the core meaning of a word. Mainly affixes have the effect of slightly modifying the meaning of the stem – a stem is either a root or a root plus an affix, or more than one root with or without affixes – to which more affixes can be attached. The most common such modification is to change the word-class, the part of speech, to which the word belongs. Thus *child* (a noun) becomes an adjective in *childish*. That adjective can in turn be changed to an adverb: *childishly*, or to a different kind of noun – an "abstract" noun – by adding another affix, as in *childishness*. This process, known as *affixation*, is one of the two most fundamental processes in word formation (the other is compounding, discussed below). Let us therefore examine more closely the properties of affixes.

2.2 Affixes

All morphemes which are not roots are affixes. Affixes differ from roots in three ways.

(1) They do not form words by themselves – they have to be added on to a **stem**.

(2) Their meaning, in many instances, is not as clear and specific as is the meaning of roots, and many of them are almost completely meaningless.

(3) Compared with the total number of roots, which is very large (thousands or tens of thousands in any language), the number of affixes is relatively small (a few hundred at most).

In English, all the productive affixes ("productive" in the sense that they do a lot of work) are either attached at the end of the stem – **suffixes** – or at the front of the stem – **prefixes**. Here are examples of common prefixes where the meaning is clear:

co + occur "occur together"	**peri** + meter "measure around"
mid + night "middle of the night"	**re** + turn "turn back"
mis + treat "treat badly"	**un** + filled "not filled"

And here are examples of common suffixes where the meaning is also clear:

act + **ion** "state of acting"	child + **ish** "like a child"
act + **or** "person who acts"	child + **hood** "state of being a child"
act + **ive** "pertaining to being in action"	child + **less** "without a child"

The majority of affixes are, unfortunately, less clear than these. We will provide more detailed information on them later, matching them to some of the possible meanings they may have.

All affixes, by definition, are bound morphemes. Historically it is quite normal for free morphemes to lose their independence and become "bound." One transparent example is the suffix *-less*: its origin in the adjective *less* "devoid of" and its connection with the word *less* do not require specialized knowledge. The suffixes *-dom*, *-hood*, and *-ship* once had independent meaning as nouns. *Dom* meant "doom, judgment, statute" and is the ancestor of the modern word *doom* as well as the suffix *-dom*. The suffix *-hood* meant "condition" or "state of affairs"; it has no modern independent counterpart, however, and is unrelated to our word *hood* "covering for the head." The basic numerals in the classical languages, which were free forms in Greek and Latin just as the corresponding numerals are in English, have provided the bound roots out of which many English compounds are formed: e.g., *penta-* in *pentagon* "having five angles," *sept-* in *septet* "a group of seven," *oct-* in *octagonal* "having eight angles"; *uni-* in *unilateral* "one-sided."

The opposite development, whereby a bound morpheme escapes into the list of free morphemes, is unusual. This is even more true of affixes than it is of roots. There has been a recent trend in the language to detach affixes and elevate them to the status of roots. A typical example is *anti*. We can say things like, "It doesn't matter what the principle is, he is so stubborn that he's bound to be anti." There are even a few forms, originally affixes, that have been detached to become independent words themselves: e.g., the form *pro* from the word *professional*, which originally meant "one who declares (*fess*) forth (*pro-*)." We no longer think of *pro* in a phrase like *pro golfer* as having anything to do with the prefix *pro-* that occurs in *process, provide, profess*. Since the mid-eighties, the negative prefix *dis-* has been used as a verb meaning "insult, show disrespect, criticize," as in *dissed, dissing*. "The Lady is a Trans," meaning "a transgendered person," was a 1996 musical hit.[1] The latter two examples are still considered "non-standard." Other examples of this kind, more or less acceptable, include *hyper, mini, maxi, stereo*. In any case, the status of these items is still in flux and their occasional encroachment into the realm of free roots does not change the basic norm that affixes are bound forms which must be attached to stems.

2.3   Functions of affixes

Affixes have two quite different functions. The first is to participate in the formation of new words. The affixes which do this are called

derivational affixes. We can think of the root as the **nucleus** of the derivation. The affixes are like satellites; furthermore, they have to circle the nucleus at different distances, vaguely like the solar system. As an example, consider the word *uninhabitableness*. The stem is *habit*. Now, can we add *un-* to *habit*? Of course not: *unhabit* is not a word. So in this derivation, *in-* must be added first. Again we ask: can *un-* be added to *habit*? Same answer.[2] And so on: we keep adding morphemes on the right-hand side until we get to *inhabitable*. Now, finally, we can put *un-* in front of the formation. Why is this? Because after we added *-able*, we had finally created an adjective. It is one of the properties of *un-* that it normally attaches only to adjectives. There are a few funny counterexamples like *uncola*, but precisely the reason why that formation is so effective is that it violates the normal rules of word formation in English. And of course we can't attach *un-* to just any adjective whatever: with few exceptions such as *unable, unkind, unwise*, simple adjectives can't take *un-*: *unbad, *unglad, *ungood, *unstrong are not possible. Derived adjectives, on the other hand, take the negative prefix freely: *uninteresting, unreliable, unimportant, unsympathetic*, etc. We can represent this hierarchical property of word formation by affixation in the tree-diagram in Figure 4.3.
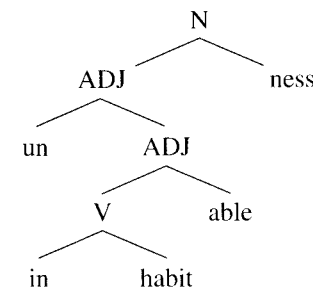

Figure 4.3

Note that the root *habit* can have the prefix *in-* attached first, followed by the suffix *-able*, because both *inhabit* and *inhabitable* are words. By the same logic, *-ness* can only be added after the word has become an adjective, or an adjectival participle. Thus the affix *-ness* is constrained in its use: it cannot be attached to roots with a clear verbal meaning, therefore *habitness*, or *inhabitness* are as impossible as *eatness, *jumpness, *sleepness. The isolated example of *forgiveness* is not really an

exception because it is probably a simplification of the adjectival past participle *forgiven* + *-ness*, which would be a regular formation. Thus, with *-ness* too, in order to get to the proper form of the noun we have to attach the adjective suffix first.

The affixes which have the function of deriving new words, then, are called derivational affixes. The other type of affix, which does not participate in word formation at all, is called **inflectional**. Inflectional affixes, of which English has only a very small number compared with Latin or Greek or Old English are really part of syntax, though some inflectional affixes are the indicators of very broad semantic categories like tense (*plays, played*) or number (singular–plural, as in *girl, girls*). The most typical inflectional affixes, in most languages, serve to indicate which word is the subject of the sentence or which word is the object of the verb. Thus Latin:

Ama-t puer puella-m
"love-s boy–NOM girl–ACCUS"

NOM means "nominative," the case for marking the subject of the sentence. ACCUS means "accusative," the case for marking the object of the verb. The three words can be arranged in any order without changing the meaning – they all just mean "The boy loves the girl":

Puer puellam amat
Puellam puer amat
Puellam amat puer
Puer amat puellam

Since inflectional affixes are nothing more than markers of sentence structure and organization, they are not involved in the derivation of new words and hence of no further interest in the present context. It should be acknowledged, however, that this is a simplified picture: there are cases where the addition of an inflectional affix can result in the development of a new meaning, detached from the original semantic category represented by the affix, as in *customs, news, spectacles*. Similarly, the present and past participle affixes: *-ing* and *-ed*, lead a dual life: They can be purely inflectional as in "They were *building* the new dorm," "They *painted* the wall," or they can behave more like derivational affixes and produce new parts of speech: "The *building* on the corner, the *painted* walls."

3 Compounds

A compound is a word which contains two or more roots. It may also contain affixes, because a compound is a stem, just like a simple non-compound root. The roots in a compound may be either free or bound, but there must be at least two of them. Thus *orthodontist* is a compound consisting of two bound roots and a suffix:

√orth "right" + √odont "tooth" + √ist "one who"

Pterodactyl is a compound consisting of one bound root and one free root:

√ptero "wing" + √dactyl "finger"

Dactyl is a free root only in one special sense, having to do with poetic meter. The *OED* gives no current citations in any other sense, and the only sense listed in the *Merriam-Webster Tenth Collegiate Dictionary* is this one. So in the sense "finger," *dactyl* should be viewed as a bound root. Less controversial examples would be *stratosphere, hemisphere,* and *biosphere*, in all of which the second root is *sphere*, which is clearly a free root:

√strato "spread out"
√hemi "half" —————— + √sphere "round, ball"
√bio "life, living"

Airport, backpack, getaway, leftmost, and *killjoy* are familiar and transparent compounds in which both members are free morphemes. This is an enormously productive pattern: *downsize, laptop, shareware, trackball,* are all late twentieth-century compounds – speakers continue to create them on a daily basis.

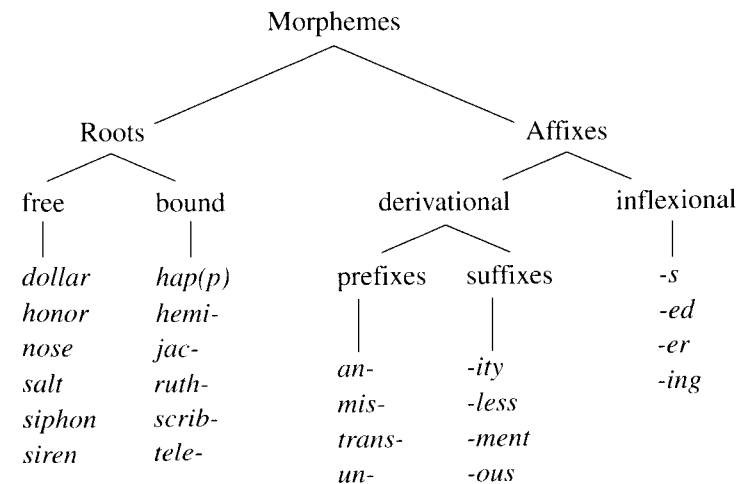Figure 4.4 summarizes all the types of morphemes discussed above.

Figure 4.4

4 Hyphens

Throughout this book we have indicated the position at which an affix attaches to a stem by placing a hyphen either before or after it, depending on which side the affix is added to: thus *trans-*, *un-*, but *-ment*, *-ous*, *-ly*, *-ness*. Affixes have a fixed position: they are either prefixes or suffixes, and speakers are not at liberty to switch around their place in order to create a different meaning. There is very strict ordering of the morphemes in terms of types, i.e. what affix can attach to what kind of stem, and the strictness is applicable to their linear ordering too. A simple example: *un + happ + y + ness* is the only option we have for combining the morphemes in this word; any other combination is linguistic garbage: **y + happ*, **un + ness + y + happ*, **happ + ness + un + y*, **un + ness + happ + y*. The principle of fixed linear ordering is valid for all affixes, so note the position of the hyphen when you first encounter an unfamiliar affix.

If a morpheme is not marked with a hyphen, it may still be a bound morpheme that can occur only in compounds or with affixes. Even when root morphemes are bound forms, however, it is not possible to predict whether other forms will occur necessarily before, or necessarily after, them. This is true of all roots, native and borrowed, free and bound, compare *baby*-*sit* to *crybaby*, *crowbar* to *scarecrow*, *horsepower* to *race-horse*. Similarly, with borrowed and bound roots, *tele* generally occurs as the first root in a compound, but *phon* may occur in either position: compare for example *telephone* with *phonology*, *dictaphone* with *verdict*.

5 Cognates

In Chapter 3 we defined cognates as words which start out from the same ancestral root, but which develop into separate dictionary entries. Both roots and affixes can have cognate relations: historically *hyper* is cognate with *super*, *hemi-* is cognate with *semi-* in the same way that *bha-* is cognate with *fa-* and *phe-*. The word *cognate* actually contains a form of the root with which this chapter started, namely √gen. The word breaks down like this:

co- "together" + √gn (= √gen) "origin" + *-ate* "having"

or, roughly, "having a common origin." All root morphemes that can be traced back to a common origin are said to be cognate. Being cognate does not at all entail that such roots would today be viewed as examples of allomorphy. Being cognate is a historical relationship. Allomorphy is a synchronic relationship usually recognized by the

speakers of the language as it is today. The two sometimes merge rather closely into each other, as we will see. First, let us remind ourselves of a fairly extreme example of obscured common origin where the divergence is so extreme that the relationship of allomorphy cannot be invoked without reference to sophisticated historical knowledge. This is the story of the IE **bha-* root which surfaces in a number of common words: *bandit*, *fame*, *infant*, *phonetic*, and *symphony* all contain this root (in the forms √ban, √fa, and √phon) and these roots are cognates. But would we invoke the relationship of allomorphy between them? Not automatically, and probably not at all for the ordinary speaker of the language, though it is always interesting to know what the relationship is. Let us reproduce here the family tree of the root **bha* which will be familiar to you from Chapter 3.
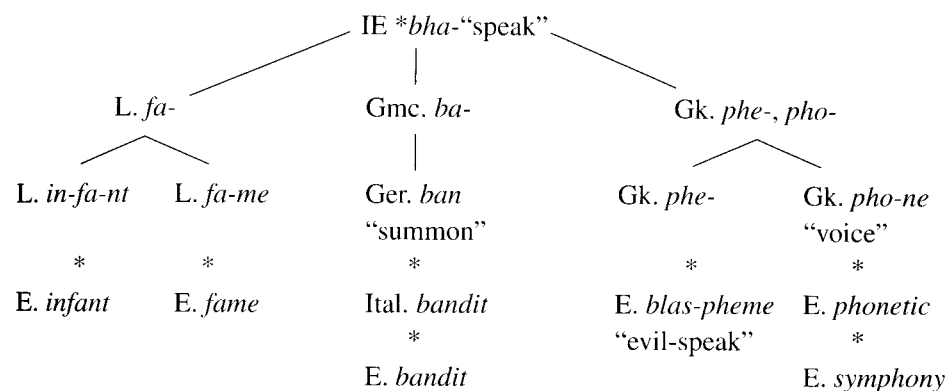
IE **bha-*"speak"

L. *fa-*   Gmc. *ba-*   Gk. *phe-*, *pho-*

L. *in-fa-nt*   L. *fa-me*   Ger. *ban* "summon"   Gk. *phe-*   Gk. *pho-ne* "voice"

*   *   *   *   *

E. *infant*   E. *fame*   Ital. *bandit*   E. *blas-pheme* "evil-speak"   E. *phonetic*

*   *

E. *bandit*   E. *symphony*

Figure 4.5

Looking at the "family tree" of the morpheme **bha-* in Figure 4.5 may give you the impression that there is a continuous line of inheritance from the older to the modern languages. But this is not always true, especially for words which have come into English more recently, say within the last four or five centuries. The Germanic form of **bha-*, *ban* "to summon, to proclaim, to band together" did not produce the word *bandit* directly. What happened was that first (Late) Latin borrowed the Germanic allomorph *ban*, in addition to its own allomorph *fa-*, then Italian used the new allomorph to create the word *bandit*, and finally English imported the word from Italian at the end of the sixteenth century. Thus, at a certain point, the line of transmission can break, a fact which we represent with an asterisk: *. What this means is that the link between the IE form and the form we use in English today is not a natural inheritance, but rather a borrowing. When borrowing takes place, it may easily happen, as here, that the borrowed form no longer resembles the cognate forms in the borrowing language. It is therefore no longer recognizable as related through allomorphy. It is just a different

root altogether, from the point of view of the borrowing language. The "natural" root here is √ban. It does not look at all like √fam or √phon, and when the latter two are borrowed they are not associated with √ban in the minds of the speakers of English who borrowed them in the first place or learned them later.

For English speakers the most accessible and fairly rich account of cognate relationships among English roots is to be found in the Appendix of the *American Heritage Dictionary* (first edition 1973, and third edition 1992) entitled "Indo-European Roots." It is also available as a separate publication, but that version is extremely difficult to use because one needs the cross-references to the Appendix that are found in the regular dictionary entries.

5.1   Shared derivation

If several words share the same root, they are cognates, but what if words share a common derivation: that is, they are based on different roots but they share a common derivational affix or set of affixes? It is rather like saying, "All the members of this group have red hair, though they may differ in every other way." Consider the sets *childless, humorless, painless, timeless,* or *apologize, dramatize, digitize, terrorize.* These words are not cognates. They resemble each other because they contain a recognizable recurring part of the word, the affixes *-less,* and *-ize.* When a set of words shares an affix or even several affixes, they are said to have a shared derivation. Suppose we make a list of some of the words, or parts of words, that end in *-ology.* (There are hundreds of them including many "sportive nonce-words," as the *OED* describes them, such *as nothingology, commonsensology, keyboardology.*) It might look like this:

bio-	-ology	phil-	-ology
cosm-	-ology	phon-	-ology
herpet-	-ology	physi-	-ology
immun-	-ology	rheumat-	-ology
music-	-ology	robot-	-ology
neonat-	-ology	ur-	-ology
ornith-	-ology	vir-	-ology

It is clearly not the case that all of these words are cognates: they simply share a form, namely the form *-ology,* which is actually *-o-log-y,* that is, three affixes together. All of the suffixes are the same and therefore obviously cognate with each other (being identical is sort of an empty sense of being cognate). But the roots are not cognate. The part of the word which is important to establish a cognate relationship is the root – the part on the left, in this list. *Infant* and *fame* are cognate in this sense:

they share a root. But *biology* and *virology* are not cognates, even though they share a form *-ology,* because they are not descended from a common ancestral root.

5.2   Shared form and meaning

If two words contain a shared form with a constant meaning, there will necessarily be a semantic relation between them – the meanings will be similar in some clear way. In the list above, *-ology* means "the study of." These words mean, therefore, "the study of life," ". . . the cosmos," ". . . snakes," ". . . immune systems," etc. It is one of the curiosities of natural languages that a pair of words may actually be true cognates, and even have completely shared forms, yet still be, at least on the surface, quite different in meaning. Such a pair is *graceful* vs. *disgraceful.* Knowing pairs like *organized* vs. *disorganized,* where it is obvious that the entire difference in meaning depends on the negative polarizing sense of *dis-,* we should be able to infer that *disgraceful* is somehow the opposite of *graceful.* Even more dramatic is the difference between the cognates *ease* and *disease*; indeed, for most people the connection between these two words comes as a surprise. But such mismatched pairs are not typical: usually we can infer the meaning of the whole from the meanings of the parts.

6   Finding roots in a dictionary

Before we end the discussion of morphemes and the types of morphemes that make up the words of English, we should find out how roots and affixes are identified in dictionaries. All dictionaries mark words into syllables. The division into syllables is done mainly to enable the reader to decide where to hyphenate at the end of a line. Most dictionaries also indicate pronunciation, including especially marking where the accent falls in polysyllabic words. And of course they provide meanings and usage notes and all sorts of encyclopedic information. But curiously, most dictionaries do not divide up words into morphemes, and even if they do, they do not label roots and affixes as such. Rather, they give some etymological information from which the user is supposed to figure out what the root morphemes and other morphemes are. There is no dictionary, anywhere, in which the editors have neatly marked up the words of English into their constituent morphemes, though there are dictionaries of roots and dictionaries of affixes. Here is the relevant part of a dictionary entry (from the *Concise Oxford*) for *municipal*:

municipal †pert. to the internal affairs of a state; pert. to local self-government, esp. of a town. xvi. -L *municipalis*, f. *municipium* Roman city of which the inhabitants had Roman citizenship, f. *municeps*, *-cip*, f. *munia* civic offices + *capere* take.

From this information how do we find the root(s)? We have to know what the abbreviations mean, first. They're all listed in the front of the dictionary. In this entry, the symbols and abbreviations we need to know are the dagger, "†" for "obsolete" (i.e. the first meaning is obsolete, not the word), the Roman numeral xvi, which tells us that the word appears in English for the first time in the sixteenth century, L for "Latin," f. for "formed on," and most important the hyphen which precedes the "L." The hyphen marks the beginning of the etymological information. This means "adoption of" – that is, the word was borrowed from the Latin form *municipalis*, with just a small change, namely we dropped the final syllable *-is*. Now we look at the etymology itself – the material that follows the hyphen. This is where we find the roots, but only indirectly. We must look at the ultimate source of the word – the very last entries in the etymology, where in this case it says "f. *muni-a* civic offices + *capere* take." This tells us, first, that this word is formed on two roots: the word is a compound, combining the root that means "civic offices" and the root that means "take." Now comes the hard part: we have to be able to subtract from the forms that are given here any part of these forms that are not themselves roots. In the case of *munia*, for example, we have to know that *-a* is a suffix. We don't have to know very much Latin to know this, but we do have to know this much. It doesn't matter what suffix it is, or what it means: it only matters that it is a suffix, and can be subtracted to find the root. In this case, then, the first root is *muni*. The second root is given as *capere* "to take." Once again we have to know a little Latin: we have to know that *-ere* is some kind of suffix. It happens to be the suffix that marked the infinitive in a certain class of verbs. One thing we can count on, in looking at etymologies that contain Latin sources, is that whenever we find *-are*, *-ere*, or *-ire* at the end of a word, and the word is translated as a verb, then it is an infinitive, and the *-are*, *-ere*, *-ire* can be subtracted to get the stem. That stem will also be the root if the whole word has been maximally broken down, as in this case.