Chapter 4

The internal structure of words and processes of word formation

Chapter preview

This chapter first introduces the criteria used for distinguishing a word from a phrase. It then considers the internal structure of words, making use of the abstract notion of a morpheme (meaningful unit) and the concrete notion of a morph. The different types of morphemes and morphs are described. It is shown that there is not always a correspondence between the morphemes and morphs of a word and that morphemes may be realized in different ways as morphs. The chapter then introduces allomorphs (predictable variants) of morphemes, and the writing of morphemic rules is explained. Both stem and root allomorphy is treated. The next section of the chapter explores the different processes of word formation in English, focusing on the complexities of derivation and compounding; minor processes of word formation – reduplication, conversion, blending, shortening, and root creations – are treated in less detail. The chapter ends with a brief discussion of idioms.

Commentary

1. Defining the word

We move now from an examination of the smallest segments of language (sounds) to the larger units: words. However, since speech is a phonetic continuum, without pauses between words (we generally pause between larger syntactic units such as phrases or

clauses), we need some means of determining the boundaries of words. We all have an intuitive feel for the words of the language and we think immediately of the written word, but even nonliterate speakers can divide the speech chain into words. Thus, there must be some formal criteria for wordhood which all speakers use. These might be of various kinds:

1. Orthographic: a word is what occurs between spaces in writing.
2. Semantic: a word has semantic coherence; it expresses a unified semantic concept.
3. Phonological:
   a. potential pause: a word occurs between potential pauses in speaking. Though in normal speech, we generally do not pause, we may potentially pause between words, but not in the middle of words.
   b. stress: a word spoken in isolation has one and only one primary stress.
4. Morphological: a word has an internal cohesion and is indivisible by other units; a word may be modified only externally by the addition of suffixes and prefixes.
5. Grammatical: words fall into particular classes.
6. Syntactic: a word has external distribution or mobility; it is moved as a unit, not in parts.

We can see the usefulness of these criteria if we look at some problematical examples of word delimitation:

supermarket	*jack-of-all-trades*
travel agency	*noteworthy*
runner-up	*try out*
forget-me-not	*pins and needles*

By the criterion of orthography, *supermarket* and *noteworthy* would be considered a single word, as would hyphenated forms such as *jack-of-all-trades, forget-me-not,* or *runner-up*, while phrases such as *travel agency*, *take out,* or *pins and needles* must be considered as multiple words, or phrases. Yet by the second criterion, semantic unity, the words and the phrases all appear to be equally unified conceptually. The discrepancy is especially apparent if you compare *supermarket* with related concepts such as *toy store* or *grocery store*. In fact, the conventions of spacing between words, as well as hyphenation practices, are often quite arbitrary in English. As well as being hyphenated, *forget-me-not, jack-of-all-trades,* and *runner-up* meet the syntactic criterion of wordhood: they are moved as a single unit. However, they differ in respect to the morphological criterion; while *forget-me-not* always behaves as a single word, with external modification (*two forget-me-nots*, *forget-me-not's*), *runner-up* is inconsistent, behaving as a single word when made possessive (*runner-up's*), but as a phrase, that is, with internal modification, when pluralized (*runners-up*); *jack-of-all-trades* is similarly inconsistent. The third criterion, a single primary stress, would seem to be the most reliable, but even here compound adjectives such as *noteworthy* pose a problem: they have two primary stresses and are phonologically phrases but are treated orthographically, morphologically, and syntactically as single words. "Phrasal verbs" such as

try out also present an interesting case.[1] Though having many of the qualities of a phrase – internal modification occurs (*tried out*), material may intercede between the parts (*try out the car*, but also *try the car out*), and both *try* and *out* receive primary stress – phrasal verbs seem to express a unified semantic notion, the same as expressed in this case by the single word *test*. As this chapter progresses, we examine these problems in more detail.

Another difficulty when treating words is the term *word* itself, which may be used in a number of different ways:

1. It may refer to the word form, the physical unit or concrete realization, either the orthographical word (the written form) or the phonological word (the uttered or transcribed form).
2. It may refer to the **lexeme**, which is rather like a dictionary entry. A lexeme includes all inflected forms of a word. It is thus a kind of abstraction or class of forms and is indicated by small capitals, as in the following examples:

   WALK – *walk*, *walks*, *walked*, *walking*
   RUN – *run, runs, ran, running*
   SING – *sing, sings, sang, sung, singing*

   Note that since the lexeme is an abstraction, it is conventional to choose one of the inflected forms to represent it, such as the infinitive of the verb or the singular of the noun. The same word form may in fact represent different lexemes:

   a. A homonym is a single orthographic and phonological word standing for two lexemes, as *bear* is either the verb or the noun.
   b. A homograph is a single orthographic word (but separate phonological words) standing for two lexemes, as *lead* is either the noun /lɛd/ or the verb /lid/.
   c. A homophone is a single phonological word (but separate orthographical words) standing for two lexemes, as /mit/ is either the noun *meat* or the verb *meet*.

   The same lexeme might also have quite distinct word forms, as in the case of the definite article *the*, represented by /ði/ or /ðə/, or the indefinite article *a/an*, represented by /eɪ/, /ə/, /ən/, or /æn/.
3. Finally, *word* may also refer to a morphosyntactic word (or grammatical word). A morphosyntactic word consists of a lexeme and associated grammatical meaning. For example, in:

   I take the garbage out every week. (TAKE + present)
   I took the garbage out yesterday. (TAKE + past)
   I have taken the garbage out already. (TAKE + past participle)

---

1. Phrasal verbs consist of a verb plus a following particle such as *up* or *out* (e.g. *fill up, fill out*); they will be discussed in more detail below and in Chapter 8.

the different morphosyntactic words are represented by different word forms (*take, took, taken*). But in

I put the garbage out every week. (PUT + present)
I put the garbage out yesterday. (PUT + past)
I have put the garbage out already. (PUT + past participle)

the different morphosyntactic words are represented by the same word form (*put*).

2. Morphemes

We begin the study of morphology by taking words as given and examining their internal structure.

2.1 Morpheme versus morph

We must start by identifying the **morpheme**, the smallest meaningful unit in a language; the morpheme is not necessarily equivalent to a word, but may be a smaller unit. For example, the word *headphones* consists of the three morphemes *head, phone,* and *-s*; the word *ringleader* consists of three morphemes, *ring, lead,* and *-er.* Some of these morphemes may stand alone as independent words (*head, phone, ring, lead*), others must always be attached to some other morpheme (*-er, -s*).

Like the phoneme, the morpheme refers to either a class of forms or an abstraction from the concrete forms of language. A morpheme has the following characteristics:

- it is internally indivisible; it cannot be further subdivided or analyzed into smaller meaningful units.
- it has internal stability since nothing can be interposed in a morpheme.
- it is externally transportable;
- it has positional mobility or free distribution, occurring in various contexts.

Morphemes are represented within curly braces { }.

HINT: Like the phoneme, the morpheme refers to either a class of forms or an abstraction from the concrete forms of language. For example, the feminine morpheme is an abstraction which can be realized in a number of different ways, by *-ess*, as in *actress*, or by a personal pronoun such as *she.* Because morphemes are abstractions we place them within curly braces { } using capital letters for lexemes and descriptive designations for other types of morphemes. For *actress*, the morphemes are {ACTOR} and {f} (for {feminine]).

Based on meaning, there are a number of types of morphemes, as shown in Figure 4.1.

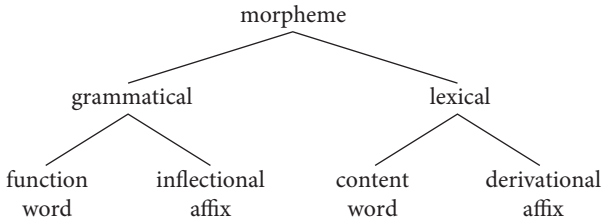


Figure 4.1. Types of Morphemes

Lexical morphemes express lexical, or dictionary, meaning. They can be categorized into the major lexical categories, or word classes: noun, verb, adjective, or adverb; these are frequently called "content words". They constitute open categories, to which new members can be added. Lexical morphemes are generally independent words (free roots) or parts of words (derivational affixes and bound roots). **Grammatical morphemes** express a limited number of very common meanings or express relations within the sentence. They do not constitute open categories; they can be exhaustively listed. Their occurrence is (entirely) predictable by the grammar of the sentence because certain grammatical meanings are associated with certain lexical categories, for example, tense and voice with the verb, and number and gender with the noun. Grammatical morphemes may be parts of words (inflectional affixes) or small but independent "function words" belonging to the minor word classes: preposition, article, demonstrative, conjunction, auxiliary, and so on, e.g. *of*, *the*, *that*, *and*, *may*.

In the case of the morpheme – which is an abstraction – we must also recognize the level of the **morph**, the concrete realization of a morpheme, or the actual segment of a word as it is spoken or pronounced. Morphs are represented by phonetic forms. We must introduce the concept of the morph distinct from the morpheme because sometimes although we know that a morpheme exists, it has no concrete realization (i.e, it is silent and has no spoken or written form). In such cases, we speak of a **zero morph**, one which has no phonetic or overt realization. There is no equivalent on the level of the phoneme. For example, plural *fish* consists of the morphemes {fish} + {pl}, but the plural morpheme has no concrete realization (i.e. the singular and plural forms of *fish* are both pronounced /fɪʃ/). Another example of a zero morph is the past tense of *let*; although the past consists of the morpheme {let} plus the morpheme {past}, the past tense morpheme has no concrete expression (i.e. the present and past forms of *let* are both pronounced /lɛt/).

We say that a morpheme is "realized" as a morph.

Based on form and distribution, there are different types of morphs (see Figure 4.2).

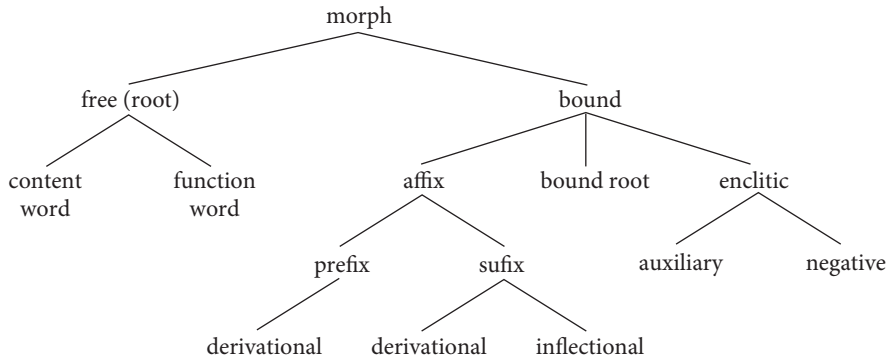


Figure 4.2. Types of Morphs

A **free morph** may stand alone as a word, while a **bound morph** may not; it must always be attached to another morph. A free morph is always a **root**.[2] That is, it carries the principal lexical or grammatical meaning. It occupies the position where there is greatest potential for substitution; it may attach to other free or bound morphemes. Examples of roots are underlined in the following words:

unavoidably	*overgrown*	*altruistic*	*decoration*	*provocative*
disheartened	*reclassify*	*heterosexuality*	*upbringing*	*realization*

Roots are also occasionally bound morphs. These are called **bound roots**. Bound roots are often foreign borrowings that were free in the source language, but not free in English. For example, in the following sets of words, we would all intuitively identify the root *-vert*, *-mit*, *-ceive*, or *-fer* (in part because it occurs in a number of words, as do the prefixes):

-vert	*convert, revert, subvert, introvert, pervert*
-mit	*transmit, commit, remit, admit, omit, submit*
-ceive	*conceive, perceive, receive, deceive*
-fer	*transfer, refer, prefer, defer, confer*

However *-vert*, *-mit*, *-ceive,* and *-fer* cannot stand alone as independent words, and we would also find it very difficult to state the meaning of any of these roots, unless we know

2. A root is often distinguished both from a "base" (a root plus associated derivational affixes, to which derivational affixes are added) and from a "stem" (a root plus associated derivational affixes, to which inflectional affixes are added). Thus, in the word *engagement, gage* is the root, *engage* is the base, and *engagement* is the stem.

Latin, from which these words derive: *-vert* is from Latin *vertere* meaning 'to turn', *-mit* is from Latin *mittere* meaning 'to send', *-ceive* is from Latin *capere* meaning 'to seize', and *-fer* is from Latin *ferre* meaning 'to bring'. Other examples of bound roots borrowed from the Romance languages include *disgruntled*, *nonchalant*, and *incognito*. Bound roots may also be native English, as with *-kempt* (< *unkempt*) and *-couth* (< *uncouth*), where the positive form no longer exists. You could say that the bound roots have a meaning only if you know their history, or etymology.[3]

Unlike a root, an **affix** does not carry the core meaning. It is always bound to a root. It occupies a position where there is limited potential for substitution; that is, a particular affix will attach to only certain roots. English has two kinds of affixes, **prefixes**, which attach to the beginnings of roots, and **suffixes**, which attach to the end of roots. Some languages regularly use "infixes", which are inserted in the middle of words. In Modern English, infixes are used only for humorous purposes, as in *im-bloody-possible* or *abso-blooming-lutely*. While it might initially be tempting to analyze the vowel alternation indicating plural (as in *man*, *men*) or past tense (as in *sing*, *sang*) in Modern English as a kind of infix, the vowels are not added or inserted but actually replace the existing vowels.

Self-Testing Exercise: To practice identifying roots, prefixes, and suffixes, do Exercise 4.1.

Affixes may be of two types, derivational or inflectional, which have very different characteristics. A **derivational affix** in English is either a prefix or a suffix. There may be more than one derivational affix per word. A particular derivational affix may attach to only a limited number of roots; which roots it attaches to is not predictable by rule, but highly idiosyncratic and must be learned. A derivational affix has one of two functions: to convert one part of speech to another (in which case, it is class changing) and/or to change the meaning of the root (in which case, it is class maintaining). Such affixes function, then, in word formation and are important in the creation of new lexemes in the language. They always precede an inflectional affix. An **inflectional affix** in English is always a suffix. A particular inflectional affix attaches to all (or most) members of a certain word class. The function of inflectional affixes is to indicate grammatical meaning, such as tense or number. Because grammatical meaning is relevant outside the word, to the grammar of the entire sentence, inflectional affixes always occur last, following the root and any derivational affixes, which are central to the meaning or class of the root. The differences between derivational and inflectional affixes are summarized in Table 4.1.

3. For this reason, they have been termed "etymemes".

Table 4.1. Derivational vs. Inflectional Affixes in English

Derivational affixes	Inflectional affixes
either prefixes or suffixes	only suffixes
optionally more than one per word	only one per word[4]
attach idiosyncratically to only a limited number of roots	attach to all (or most) members of a word class
have two functions 1. to convert one part of speech to another 2. to change the meaning of the root	have one function 1. to indicate grammatical meaning
precede the inflectional suffix	follow derivational suffix(es)

A distinction can be made between productive inflections, which would attach to any new word entering the language to express a particular grammatical category, and nonproductive, or remnant, inflections, which are found on select members of a class, but would never be added to a new word. There are only eight productive inflections in Modern English, as shown in Table 4.2. Some examples of nonproductive inflections are the plural vowel alternation *tooth-teeth*; the *-most* superlative of *foremost*; the *-en* past participle of *write-written*; or the past tense vowel alternation of *ring-rang*.

Table 4.2. The Productive Inflections of Modern English

plural number	*-s*	Noun
possessive case	*-s*	
present (nonpast) tense, 3rd p sg	*-s*	Verb
past tense	*-ed*	
past participle	*-ed*	
present participle	*-ing*	
comparative degree	*-er*	Adjective
superlative degree	*-est*	

HINT: As you can see in Table 4.2, the *-ed* ending has two distinct functions, as a past tense marker and as past participle marker. These two functions are distinguished as follow: past tense forms serve as the sole verb in a sentence (e.g. *He walked along the shore, The teacher helped the student with the problem*), while past participles must occur with a preceding auxiliary BE or HAVE (e.g. *He has walked along the shore for an hour; The student was helped with the problem (by the teacher)*). Depending on the verb, the past tense and past participle may be either the same (e.g. *walked~walked*) or different

4. A form such as *men's* has two inflectional morphemes (plural and possessive), though only one inflectional suffix (*-s*).

(e.g. *drove~driven*). Note also that the *-s* ending has different functions for nouns and verbs. This does not usually pose a problem unless the same word functions as either a noun or a verb (e.g. *His hopes to win the lottery were dashed* [*-s* plural on the noun] vs. *He hopes to win the lottery* [*-s* 3rd person present indicative on the verb]).

Self-Testing Exercises: To learn to identify inflectional suffixes, do Exercise 4.2. Then to better understand the difference between inflection and derivation, do Exercise 4.3.

An **enclitic** is a kind of contraction, a bound form which derives from an independent word and must be attached to the preceding word.[5] In English, we have two kinds of enclitics: contracted auxiliaries, which are attached to the preceding subject, and the negative contraction *-n't*, which is attached to the preceding auxiliary. Certain auxiliaries (e.g. *may, can, must, should, might, was*) cannot be contracted, while the contraction of *not* produces marginally acceptable forms in some cases (e.g. *?mayn't, ?mightn't*) or unacceptable forms in other cases (e.g. **am't*).

Contracted auxiliaries	**Contracted negatives -n't**
will, shall > 'll	*won't, ?shalln't*
would, had > 'd	*wouldn't, hadn't*
is, has > 's	*isn't, hasn't*
have > 've	*haven't*
am > 'm	**am't (ain't)*
are > 're	*aren't*
*was > *'s*	*wasn't*

Words are analyzed morphologically with the same terminology used to describe different sentence types:

- a "simple" word has one free root, e.g. *hand*;
- a "complex" word has a free root and one or more bound morphs e.g. *unhand*, *handy*, *handful*, or it has two or more bound morphs, e.g. *transference, reception, conversion*
- a "compound" word has two free roots, e.g. *handbook*, *handrail*, *handgun*; and
- a "compound-complex" word has two free roots and associated bound morphs, e.g. *handwriting*, *handicraft*.

2.2 The analysis of words into morphs and morphemes

The importance of the distinction between morph and morpheme is that there is not always a one-to-one correspondence between morph and morpheme, and morphemes can

5. Some languages have "proclitics", originally free words which must be attached to the word that follows; the articles in French are proclitics, e.g. *la auto > l'auto*. Also, the archaic forms in English *'twas* (< *it was*) or *'tis* (< *it is*) contain proclitics.

combine or be realized in a number of different ways. We can thus analyze words in two different ways:

1. into morphs following formal or structural divisions, or
2. into morphemes, recognizing the abstract units of meaning present.

If we start first with nouns, we would arrive at the two analyses of each of the following words:

	Morphs	**Morphemes**
writers	3 morphs *writ/er/s*	3 morphemes {WRITE} + {-er} + {pl}
authors	2 morphs *author/s*	2 morphemes {AUTHOR} + {pl}
mice	1 morph *mice*	2 morphemes {MOUSE} + {pl}
fish	1 morph *fish*	2 morphemes {FISH} + {pl}
children	2 morphs *child/ren*	2 morphemes {CHILD} + {pl}
teeth	1 morph *teeth*	2 morphemes {TOOTH} + {pl}
man's	2 morphs *man/s*	2 morphemes {MAN} + {poss}
men's	2 morphs *men/s*	3 morphemes {MAN} + {pl} + {poss}

HINT: Inflectional morphemes can often be realized by a number of different forms, or the same form may denote a number of different inflectional morphemes. Therefore, it is usual to use descriptive designations for inflectional morphemes, such as {pl} (rather than {-s} or {-es}) for the plural marker and {poss} (rather than {-s}) for the possessive marker. The descriptive designations that we will use should be self-evident in the following discussion (also see the list of abbreviations in Appendix I).

A noun such as *sheep* raises a difficulty for morphemic analysis, since it is either singular or plural. Should we postulate two morphemic analyses?

{SHEEP} + {pl}
{SHEEP} + {sg}

This seems a good idea. If we postulate a morpheme for singular, even though it is never realized, we can account for number systematically. Thus, we will analyze all singular nouns as containing an abstract {sg} morpheme, so that *man's* above would have the analysis {MAN} + {sg} + {poss}, *writer* the analysis {WRITE} + {-er} + {sg}, and *author* the analysis {AUTHOR} + {sg}.

Let us look at how morphological and morphemic analysis works in adjectives:

	Morphs	**Morphemes**
smaller	2 morphs *small/er*	2 morphemes {SMALL} + {compr}
smallest	2 morphs *small/est*	2 morphemes {SMALL} + {supl}
better	1 morph *better*	2 morphemes {GOOD} + {compr}
best	1 morph *best*	2 morphemes {GOOD} + {supl}

(Here, compr = comparative degree and supl = superlative degree, as will be discussed in the next chapter.) Again, we need to postulate a morpheme positive degree {pos}, even though it is never realized, to account systematically for the inflected forms of adjectives:

good 1 morph *good* 2 morphemes {GOOD} + {pos}

For verbs, the two analyses work as follows:

	Morphs	**Morphemes**
worked	2 morphs *work/ed*	2 morphemes {WORK} + {past}
		2 morphemes {work} + {pstprt}
wrote	1 morph *wrote*	2 morphemes {WRITE} + {past}
written	1 morph *written*	2 morphemes {WRITE} + {pstprt}
working	2 morphs *work/ing*	2 morphemes {WORK} + {prsprt}
		3 morphemes {WORK} + {gerund}+ {sg}
put	1 morph *put*	2 morphemes {PUT} + {past}
		2 morphemes {PUT} + {pstprt}

(Here, pstprt = past participle, prsprt = present participle; see further Chapter 9.) Note that we have to analyze *-ing* verbal forms not only as present participles, but also as "gerunds".[6] Since gerunds are functioning as nouns, they may sometimes be pluralized, e.g.:

readings 3 morphs *read/ing/s* 3 morphemes {READ} + {gerund} + {pl}

We need to postulate a morpheme {pres}, which is never realized, to account coherently for the distinction past versus present:[7]

work 1 morph *work* 2 morphemes {WORK} + {pres}
write 1 morph *write* 2 morphemes {WRITE} + {pres}

---

6. Gerunds are word forms that are derived from verbs – by the addition of the suffix *-ing* – but function grammatically as nouns; they are "verbal nouns". Thus, in *Swimming is good exercise, swimming* is derived from the verb *swim*, but since it serves as the subject of the sentence, it functions as a noun; in *I enjoy singing in the shower, singing* is derived from the verb *sing*, and the entire phrase *singing in the shower* functions as a noun phrase direct object. As shown above, (some) gerunds can be pluralized and function as sentential subjects or objects: e.g. *Readings by local authors are given at the local book store; I detest poetry readings.* (Gerunds will be discussed in more detail in Chapter 9.)

7. The 3rd person singular form *works* or *writes* causes some difficulty for our analysis. Would we need to propose the following analysis for *works*: {WORK} + {pres} + {sg}? If we do this, we would also have to postulate a {pl} morpheme, which is never realized. However, we won't do this, but will assume that *-s* is added by a rule of grammar, that of concord, which copies the feature of number from the noun subject to the verb.

The morphemic analysis of pronouns is somewhat more complicated:

	Morphs	**Morphemes**
we	1 morph *we*	3 morphemes {1st p} + {pl} + {nomn}
him	1 morph *him*	4 morphemes {3rd p} + {sg} + {m} + {obj}
its	2 morphs *it/s*	4 morphemes {3rd p} + {sg} + {n} + {poss}

(Here, nomn = nominative case and obj = objective case; see the following chapter.)

Morphemes combine and are realized by one of four **morphological realization rules**:

1. agglutinative rule: two morphemes are realized by morphs which remain distinct and are simply "glued" together, e.g. {WRITER} + {pl} > *writers*
2. fusional rule: two morphemes are realized by morphs which do not remain distinct but are fused together, e.g. {TOOTH} + {pl} > *teeth*
3. null realization rule: a morpheme is never realized as a morph in any word of the relevant class, e.g. {sg} on nouns, which never has concrete realization in English.
4. zero rule: a morpheme is realized as a zero morph in particular members of a word class, e.g. {SHEEP} + {pl} > *sheep*. Note that in most other members of the class noun, {pl} has concrete realization as *-s*.

Examples of the four different morphological realization rules, or combinations of these rules, are the following:

agglutinative	{WORK} + {past} > *worked*
fusional	{WRITE} + {past} > *wrote*
null	{WORK} + {pres} > *work*
zero	{PUT} + {past} > *put*, {PUT} + {pstprt} > *put*
fusional and agglutinative	{MAN} + {pl} + {poss} > *men's*

HINT: Distinguishing between the concept of a null rule and a zero rule can be difficult. Remember that in the case of a null rule, the morpheme is never concretely realized. For example, {pres} on verbs is always unmarked. No verb has an overt marker of the present. In contrast, when a morpheme is usually concretely realized, but is not realized on certain words, then we have a zero rule. For example {pl} on nouns is typically expressed by *-s*, but on a noun such as *deer*, it is not marked and hence a zero rule.

Self-Testing Exercise: Do Exercise 4.4.

2.3 Allomorphs and morphemic rules

Just as phonemes have predictable variants, called allophones, morphemes have predictable variants called **allomorphs**. Allomorphs are the members of the class, morpheme, or

the phonetic realizations of the abstraction, morpheme. Allomorphs are semantically similar and in complementary distribution. They needn't be phonologically similar, however. Allomorphs are predicted, or "conditioned", in one of three ways:

1. the appearance of a particular allomorph is predictable from the phonetic environment, hence phonologically conditioned;
2. the appearance is unpredictable phonologically but is determined by the grammar of the language, hence grammatically conditioned; or
3. the allomorphs are used interchangeably in all environments, hence in free variation.

Let's consider the following example involving regular plural formation in nouns in English, as shown in Table 4.3.

Table 4.3. Regular Plural Formation in Nouns

A	B	C	
bushes /ʃ/	maps /p/	knobs /b/	rays /eɪ/
buses /s/	cats /t/	rods /d/	sofas /ə/
mazes /z/	racks /k/	logs /g/	toys /ɔɪ/
judges /dʒ/	ropes /p/	seals /l/	keys /i/
matches /tʃ/	laughs /f/	mirrors /r/	news /ɪu/
boxes /s/	paths /θ/	pans /n/	lathes /ð/
garages /ʒ/		tombs /m/	coves /v/
rouges /ʒ/		rings /ŋ/	

Although the orthographic form of the plural is *s* or *es* in all cases, you will notice that the phonological form of the plural morpheme in column A is /əz/, in column B /s/, and in column C /z/. Thus, there are three allomorphs of the plural morpheme. These allomorphs are phonetically similar, as well as semantically similar, all expressing the concept 'more than one'. A speaker of English knows which of these three forms to choose in any particular case. For the made-up noun, *prat,* the speaker would know to add the /s/ plural, whereas with the made-up noun *stad*, the speaker would add /z/. Thus, the particular endings, or allomorphs, are predictable – but how? If they are phonologically conditioned, there must be something about the phonetic environment of the noun which determines the choice of allomorph. In fact, it is the final sound of the root of the noun which is the determining factor. Note that in column A, all of the nouns end with a fricative or an affricate, in column B, with a voiceless consonant, and in column C, with a voiced consonant or vowel. We can refine this information and state it in terms of a **morphemic rule** similar in form to a phonemic rule (as in Chapter 3) We must first recognize that the sounds found in column A /s, z, ʃ, ʒ, tʃ, dʒ/ constitute a natural class called sibilants. It would be inaccurate to say that /əz/ occurs after fricatives, since certain fricatives such as /f/ take the /s/ allomorph while

others such as /v/ take the /z/ allomorph. Once we recognize the class of sibilants, we can state the rule as follows: [8]

{pl} → [əz]/ sibilants —
[s]/ voiceless consonants —
[z]/ elsewhere

Remember that the rule is read downward, so that "voiceless consonants" in the second line would exclude any voiceless consonants already included in the first line among sibilants.

As with phonemic rules, we specify one allomorph as "elsewhere". This is the form with widest distribution or the one found in the most diverse phonetic environments, in this case, after voiced consonants and vowels. It should also be the form from which the other forms can be derived with the simplest set of phonological rules. For example, if we take /z/ as the underlying form,[9] then we can derive the other forms:

– by inserting schwa between two sibilants (giving the [əz] allomorph); and
– by devoicing [z] when it immediately follows a voiceless consonant (giving the [s] allomorph).

Note that we must make these changes in this order. Although we could assume [s] or [əz] as the underlying form, the phonological changes that would need to occur are less natural than the ones suggested above.

Table 4.3 above gives the forms of noun plural in English that are phonologically conditioned, but certain noun plurals are grammatically conditioned:

Ø	*fish, sheep, deer*
vowel alternation	*mice, lice, geese*
-en	*children, brethren, oxen*
foreign plurals	
-a	*phenomena, data, criteria*
-i	*stimuli, foci*
-ae	*alumnae, formulae*
-ices	*indices, appendices*
-es	*bases, axes*
-im	*kibbutzim, cherubim*

These endings are not productive: they are either linguistic fossils (remnant forms from an earlier stage of English) or foreign borrowings. Note that if a noun such as *mouse* took a productive ending, it would be the [əz] allomorph, *child* would take /z/, and *tooth* would take /s/.

---

8. This rule will also account for the allomorphs of the possessive morpheme (as well as of the 3rd p sg pres morpheme on verbs and contractions of 3rd p sg pres of HAVE and BE).

9. The underlying form need not correspond to the actual historical form.

Let's look at one set of forms that does not seem to follow the morphemic rule for plural allomorphs given above. We would expect the plural allomorph of words ending in /f/ (a voiceless non-sibilant consonant) to be /s/, as in the following words:

belief – beliefs *chief – chiefs*
proof – proofs *safe – safes*

However, what we find in the following set of forms is not /s/, but instead the plural allomorph /z/, with a simultaneous voicing of the final root consonant:

wolf – wolves *leaf – leaves*
knife – knives *loaf – loaves*
sheaf – sheaves *wife – wives*
elf – elves *life – lives*
shelf – shelves *calf – calves*
thief – thieves *self – selves*

In some cases, we also find variation between the phonologically expected and unexpected forms:

wharf – wharfs/wharves *dwarf – dwarfs/dwarves*
hoof – hoofs/hooves *scarf – scarfs/scarves*

A similar irregularity appears in the following words ending in /s/; the expected /əz/ allomorph is found, but there is also voicing of the final root /s/:[10]

house – houses *blouse – blouses*

How do we account for these irregularities in the plural forms? We could have a morphological realization rule which changes final voiceless fricatives to voiced fricatives when {pl} is added. However, such a rule would have to apply generally to all roots ending in voiceless fricatives, and it does not. Instead, we say that there are two predictable variants of the root, what is called **root allomorphy**. The two allomorphs of the root are grammatically conditioned, by the presence of either a following {sg} and {pl} morpheme. The rule for *leaf/leaves* is as follows:

{lif} → [liv]/ -{pl}
[lif]/ elsewhere

Note that "elsewhere" would include the environment before both {sg} and {poss}. Hence, this form has the widest distribution. Actually, the -{pl} environment is too restricted since

10. It is interesting to note that in these cases, the possessive morpheme *-s* is altogether regular: *wolf's, knife's, life's, thief's, elf's*, and so on.

we also have voicing when a verb is formed from the noun (for example, *to shelve*, *to calve*, *to halve*).[11]

A similar kind of root allomorphy is thus seen in cases of shifts from noun to verb where (a) the nominal forms have /s/ and the verbal forms have /z/ (Table 4.4a), or (b) the nominal forms have /θ/ while the verbal forms have /ð/ (Table 4.4b).

Table 4.4. Root Allomorphy

a. N: /s/	V: /z/	b. N: /θ/	V: /ð/
house	to house	bath	to bathe
blouse	to blouse	cloth	to clothe
use	to use	breath	to breathe
excuse	to excuse	mouth	to mouthe
advice	to advise	teeth	to teethe
abuse	to abuse	wreath	to wreathe

Finally, it is interesting to note that bound roots may show root allomorphy; for example, *-cept* is a predictable variant of *-ceive* before *-ion*, as in *conception*, *perception*, *reception*, and *deception*.

Generally, English is not rich in allomorphy, though we have inherited quite a lot of it with the Latinate vocabulary that we borrowed, as you will see in Exercise 4.5. However, two other examples of native allomorphy are the [ðə]/[ði] variants of the definite article {THE} – can you determine how these are conditioned? A further example of root allomorphy is *staves/staffs* (< *staff*), where the root-allomorphic plural and the regular plural have become semantically distinguished, the former being restricted to music.

Self-Testing Exercise: To practice writing morphemic rules, do Exercise 4.5.

3. Processes of word formation

English has a number of means by which morphs combine or are altered to form new words. However, only two of these processes of word creation, derivation and compounding, are responsible for significant numbers of new words.

3.1 Derivation

The addition of a word-forming affix is called **derivation**. We have already looked at the features of derivational affixes (in contrast to inflectional suffixes) (see Table 4.1). The addition

11. Voicing even occurs in some cases where there is no voicing in the noun plural (as in *to believe, to prove, to grieve*).

of a derivational affix to a root produces a new word with one or more of the following changes:

- a phonological change (including stress change): *reduce* > *reduction*, *clear* > *clarity*, *fuse* > *fusion*, *include* > *inclusive*, *drama* > *dramatize*, *relate* > *relation*, *permit* > *permissive*, *impress* > *impression*, *eléctric* > *electrícity*, *phótograph* > *photógraphy*;
- an orthographic change to the root: *pity* > *pitiful*, *deny* > *denial*, *happy* > *happiness*;
- a semantic change, which may be fairly complex: *husband* > *husbandry*, *event* > *eventual*, *post* > *postage*, *recite* > *recital, emerge* >*emergency*; and/or
- a change in word class: *eat* (V) > *eatable* (A), *impress* (V) > *impression* (N) (see further below).

In English, derivational affixes are either prefixes or suffixes. They may be native (deriving from Old English) or foreign (borrowed along with a word from a foreign language, especially French). Their productivity may range from very limited to quite extensive, depending upon whether they are preserved in just a few words and no longer used to create new words or whether they are found in many words and still used to create new words. An example of an unproductive suffix is the *-th* in *warmth*, *width*, *depth*, or *wealth*, whereas an example of a productive suffix is the *-able* in *available*, *unthinkable*, *admirable*, or *honorable*. But which affix attaches to which root is always quite arbitrary and unpredictable; it is not a matter of rule but must be stated separately for each root (as, for example, in a dictionary). That is, derivation is part of the lexicon, not part of the grammar of a language.

Only three prefixes, which are no longer productive in English, systematically change the part of speech of the root:

a-	N/V > A	*ablaze*, *asleep*, *astir, astride, abed, abroad*
be-	N > V	*betoken*, *befriend*, *bedeck, becalm, besmirch*
en-	A/N > V	*enlarge*, *ensure*, *encircle*, *encase*, *entrap*

Other prefixes change only the meaning of the root, not its class. Prefixes fall into a number of semantic classes in English, depending upon the meaning that they contribute to the root, as shown in Table 4.5. Note the difference between privation and negation: a privative prefix expresses the reverse of an action (as in *undo*) or the absence of a quality (as in *amoral* 'without morals'), whereas the negative prefix expresses 'not' (as in *immoral* 'not moral'). The list given in Table 4.5 is not an exhaustive one; other semantic categories would be needed to classify all the prefixes of English, such as "completeness" (e.g. *fulfill*), "reversal" (e.g. *counterattack*), or "subordination" (e.g. *vicechair*). Furthermore, some prefixes may fit into more than one category; e.g. *under-*, expresses both degree (in *underpayment*) and place (in *underwater*). Prefixes may often attach to more than one part of speech, e.g. *mislead* (V) and *misfortune* (N).

Table 4.5. Semantic Classes of Prefixes in English

a.	Time	
	pre-	prearrange, presuppose, preheat
	after-	aftershock, afterthought, afterglow
b.	Number	
	tri-	tricycle, triannual, triconsonantal
	multi-	multinational, multilingual, multimillionaire
c.	Place	
	in-	infield, in-patient, ingrown
	inter-	interconnect, interbreed, interlace
d.	Degree	
	super-	supersensitive, supersaturated, superheat
	over-	overanxious, overconfident, overdue
e.	Privation	
	a-	amoral, apolitical, asymmetric
	un-	unlock, untie, unfold
f.	Negation	
	un-	unafraid, unsafe, unwise
	anti-	antisocial, antitrust, antiwar
g.	Size	
	micro-	microcosm, microchip, microfilm
	mini-	miniskirt, minivan, minimall

Of the prefixes given in Table 4.5, *after-*, *in-*, *over-*, and *un-* are native English, while *pre-*, *inter-*, *super-*, and *mini-* are Latin and *tri-*, *a-*, *micro-*, and *anti-* are Greek. Note that many of the native prefixes, such as *over-*, *under-*, *out-*, or *in-* (as in *over-skilled*, *underpayment*, *outcast*, *infield*), also exist as independent prepositions and adverbs.

Suffixes have two functions: to change the meaning of the root and to change the part of speech of the root. Many suffixes attached to nouns change their meaning but not their class:

- the diminutive suffixes *-ling*, *-let*, *-y, -ie* (as in *princeling*, *piglet*, *daddy, hoodie*),[12]
- the feminine suffixes *-ess*, *-ette*, *-rix*, *-ine* (as in *actress*, *usherette*, *aviatrix*, *heroine*) – which, for social and cultural reasons, are now falling out of use,
- the abstract suffixes *-ship*, *-hood*, *-ism*, making abstract nouns out of concrete nouns (as in *friendship*, *neighborhood*, *hoodlumism*), or
- suffixes denoting people such as *-(i)an, -ist, -er* (in *librarian, Texan, Canadian, Marxist, Londoner*).

Some suffixes attached to adjectives likewise change only their meaning

- *-ish* means 'nearly, not exactly' in *greenish, fortyish, coldish*
- *-ly* express 'resemblance' in *goodly, sickly, lonely*

12. Diminution (e.g. *doggy*) is not the only use for the diminutive suffix; it may also express degradation (e.g. *dummy*), amelioration (e.g. *hubby*), and intimacy (e.g. *Jenny* < *Jennifer*).

More often, however, suffixes change the word class of the root. As shown in Table 4.6a and b, the suffix may produce a noun from a verb or an adjective. Any suffix which produces a noun is called a **nominalizer**. This constitutes the largest set of class-changing suffixes. A highly productive nominalizer is the agentive suffix *-er*, which may be added to many verbs to produce agent nouns. A suffix which produces a verb from a noun or an adjective is called a **verbalizer**, as exemplified in Table 4.6c, while one which produces an adjective from a noun, a verb, or another adjective is called an **adjectivalizer** and is exemplified in Table 4.6d and e. The smallest set of class-changing suffixes is the **adverbializer**, as shown in Table 4.6f.

Table 4.6. Derivational Suffixes in English

Nominalizer	a. V > N	-ment	arrangement, judgment, advancement
		-er	worker, helper, leader
		-(c)ation	legalization, simplification, taxation
		-al	disposal, refusal, arrival, trial
		-ance/-ence	ignorance, performance, reference
	b. A > N	-dom	freedom, officialdom, Christendom
		-ness	happiness, cleverness, bitterness
		-ity	legality, purity, equality
Verbalizer	c. A/N > V	-ify	pacify, simplify, purify
		-ize	prioritize, publicize, centralize
		-ate	hyphenate, orchestrate, chlorinate
		-en	lighten, soften, tighten, moisten
Adjectivalizer	d. N > A	-y	flowery, thirsty, bloody
		-ous	poisonous, famous, glamorous
		-ful	delightful, sinful, pitiful
	e. V > A	-ive	supportive, generative, assertive
		-able	acceptable, livable, changeable
		-ful	hopeful, thankful, useful
		-ent/-ant	absorbent, flippant, repellent
Adverbializer	f. A/N > Adv	-ward	homeward, eastward, downward
		-ly	quickly, terribly, gradually
		-way(s)	sideway(s), anyway(s), someway

The formation of complex words is not always entirely predictable or regular. For example, *-ist* is typically added to common nouns (e.g. *cyclist*) and occasionally to proper nouns (e.g. *Platonist*). The addition of *-ist* results in a phonetic change /-ɪst/ and a semantic change 'one connected with X'. However, some words have an additional phonetic change, as in *Platonist* above or in *publicist*, *historicist* (< *public*, *historic*), where the final /k/ consonant changes to /s/. The semantics of *-ist* words is also more complex than first suggested since such words may denote persons adhering to a theory (e.g. *anarchist*, *realist*,

hedonist), persons exercising a scientific profession (e.g. *linguist*, *dentist*, *psychiatrist*, *botanist*), or persons addicted to an ideology (e.g. *perfectionist*, *extremist*, *nationalist*, *fascist*). In the last category would fall *racist*, *sexist*, *lookist*, or *ageist*, which denote not just people addicted to race, sex, looks, or age, but those who make discriminations or hold prejudices based on these qualities. Note too that many such words have acquired negative connotations (as has, in fact, the notion of 'addiction'). The morphology of *-ist* words is also not entirely regular; some *-ist* words are related to abstract nouns ending in *-y* (e.g. *botany*, *psychiatry*) and some to ones in *-ism* (e.g. *realism*, *fascism*), while some *-ist* words take an *-ic* adjectivalizer and others do not (e.g. *hedonistic*, **dentistic*). And the combination of *-ist* with a particular root is not predictable: while *balloonist* and *cyclist* are possible, for example, **boatist* and**skatist* are not. The suffix *-ist* is often in competition with either *(i)an* (e.g. *pedestrian*, *grammarian*, *barbarian*) and *-ite* (e.g. *suburbanite*, *socialite*, *Troskyite*). In the case of a follower of Darwin, all three forms – *Darwinian*, *Darwinist*, and *Darwinite* – exist.

Finally, the false morphological division of words may result in more or less productive suffixes, which one scholar calls "splinters", as in the following:

ham/burger	>	*cheeseburger*
		fishburger
		mushroomburger
		vegieburger
alc/oholic	>	*workaholic*
		chocaholic
		rageaholic
mar/athon	>	*workathon*
		telethon
		swimathon
		walkathon
pano/rama	>	*autorama*
		motorama
caval/cade	>	*aquacade*
		motorcade
heli/copter	>	*heliport*
		helidrome
		helistop

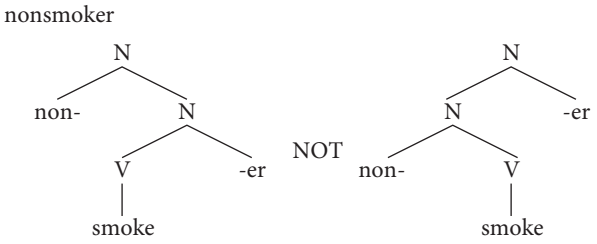
For instance, in *alcohol*, a word derived from Arabic, *al* is a definite article and the root is *-cohol-* (related to the word *kohl* 'powder used to darken eyelids'); therefore, dividing the word between *alco-* and *-holic* (or *alc-* and *-oholic*) is a mistake. *Helicopter* is derived from Greek roots meaning 'spiral' (cf. *helix*) + 'wing' and should correctly be divided between *helic-* and *-opter* (or *helico-* and *-pter*); that is, the *-c-* belongs with the root.[13]

13. These forms might also be analyzed as blends; see below.

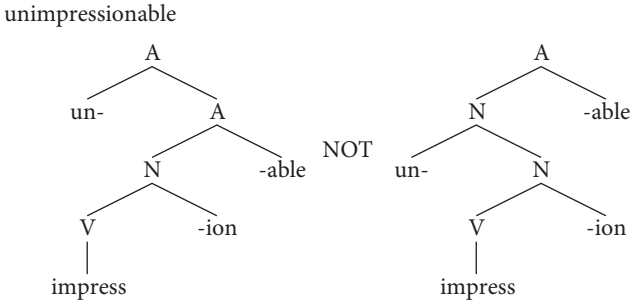
Derivation can be stated in terms of lexical rules:

mis- + align (V) + -ment	>	*misalignment* (N)
image (N) + -ine + -ary	>	*imaginary* (A)
false (A) + -ify	>	*falsify* (V)

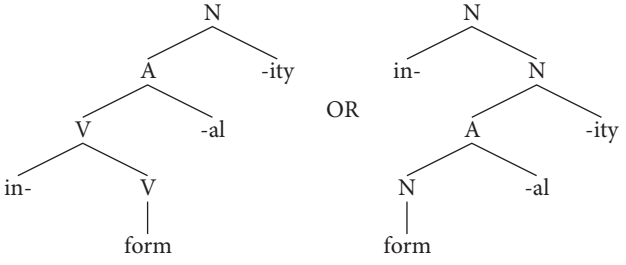
Or they can be expressed by tree diagrams, which have the advantage that they indicate the hierarchical arrangement and order of derivation of complex words. Possible representations of the derived form *nonsmoker* are the following:



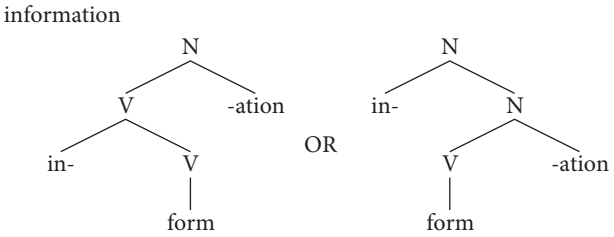
The reason that the second derivation is impossible is that one must be able to stop at any point in the derivation and still have a word of English. The second derivation produces the nonword **nonsmoke*. The form *unimpressionable* has two possible derivations:



(Here we are not analyzing *impress* into its bound root *-press* and prefix *im-*.) Again, the second derivation produces the nonword **unimpression*. Look at the two possible trees for *informality*:



While neither derivation produces a nonword, the reason for preferring the second derivation in this case is semantic. *Informality* means 'not having the quality of form' and is related to the noun *form*. The prefix is negative. It is not related to the verb *inform*, which does not contain a negative prefix. Compare the following derivations of *information*:



Again, there are semantic reasons for preferring the first derivation here, since *information* is related to the action of informing, not to a formation.

HINT: In writing or reading these tree diagrams, you should work from the bottom up; that is, you should begin with the root and then add the prefixes and/or suffixes. They should yield the word when read from left to right.

Self-Testing Exercises: Do Exercises 4.6 and 4.7 on derivation.

3.2 Reduplication

Reduplication is a process similar to derivation, in which the initial syllable or the entire word is doubled, exactly or with a slight phonological change. Reduplication is not a common or regular process of word formation in English, though it may be in other languages. In English it is often used in children's language (e.g. *boo-boo*, *putt-putt*, *choo-choo*) or for humorous or ironic effect (e.g. *goody-goody*, *rah-rah*, *pooh-pooh*). Three different kinds of reduplication can be identified:

1. exact reduplication: *papa*, *mama*, *goody-goody*, *so-so*, *hush-hush*, *never-never, tutu, fifty-fifty, hush-hush*;
2. ablaut reduplication in which the vowel alternates while the consonants are identical: *criss-cross*, *zig-zag*, *flip-flop*, *mish-mash*, *wishy-washy*, *clip-clop*, *riff-raff*; *achy-breaky,* and
3. rhyme reduplication in which the consonants change while the vowel remains the same: *hodge-podge*, *fuddy-duddy*, *razzle-dazzle*, *boogie-woogie*, *nitty-gritty*, *roly-poly, hob-nob, hocus-pocus*.

Reduplications can be formed with two meaningful parts, for example, *flower-power*, *brain drain*, *culture vulture*, *boy toy,* or *heart smart*.

Reduplication has many different functions: it can express disparagement (*namby-pamby*), intensification (*super-duper*), diminution (*teeny-weeny*), onomatopoeia (*tick-tock*), or alternation (*ping-pong*), among other uses.

3.3 Conversion or functional shift

A **functional shift** involves the **conversion** of one part of speech to another without the addition of a suffix, as in *a phone* (N) > *to phone* (V). It is sometimes said that a zero (Ø) derivational suffix is added (since it is usual for derivational suffixes to change the part of speech, as discussed above). The only concrete change that may occur in a functional shift is a change in stress.

The following kinds of functional shifts are most common in English:

V > N	*(a) run, drive, walk, bruise, cut, look, call, dump, spy, bite, sneeze*
N > V	*(to) man, head, shoulder, telephone, lust, contact, ship, sign, skin, mail*
A > V	*(to) weary, better, empty, idle, dirty, bare, quiet, tame, lower*
A > N	*(a) daily, double, private, commercial, formal, red, elder, roast*
Prt > V	*(to) down, up, off, thwart, out*

("Prt" denotes "particle", a super-class of words including prepositions, adverbs, and some conjunctions, as will be discussed in Chapter 5.) Less common kinds of conversions are a preposition to a noun (*ins and outs, ups and downs*), an adverb to a noun (*whys and wherefores*, *the hereafter*), or even a prefix to a noun (*pros and cons*). Once a word has been converted, it can normally take the inflections of the new class, for example, *two runs*, *telephoned/telephoning*, *dirtied*.

What happens semantically when a word is converted may be quite varied. For example, in the V > N shift, an action is treated as:

- an object or thing, though the emphasis may be on the action (e.g. *an attack*, *a fight*, *a kiss*, *a kick*, *a groan*),
- the result of the action (e.g. *an award*, *a find*, *a bruise*, *a crease*),
- the person performing the action (e.g. *a spy, a cook*),
- the time of the action (e.g. *the spring*, *the fall*),
- the place of the action (e.g. *a sink*, *a drain*, *a speak-easy*), or
- the range of the action (e.g. *an overlap*).

Note that with such shifts, actions become easily countable (e.g. *two kisses*, *several fights*).

With the N > V shift, the noun used as a verb denotes:

- the thing moved to a location (e.g. *to paint*, *to water*) or from a location (e.g. *to milk*, *to skin*),
- the location to which something is moved (e.g. *to bottle*, *to box*) or from which it is moved (e.g. *to mine*), or
- the instrument used to perform an action (e.g. *to lock*, *to mail*, *to whistle*, *to rattle*).

A > V gives the inchoative ('to become X') or the causative ('to cause to become X') meaning (see Chapter 10). The shift A > N treats a quality as an entity (and hence quantifiable).

It is often difficult to know in conversions which is the original (or basic) form and which the converted form. Sometimes semantics or morphological modification will offer a clue. When the noun is primary, the verb necessarily includes the meaning of the noun. Thus, *to butter toast with margarine* sounds odd because the converted verb *butter* includes the meaning of the noun. Similarly, *to garage the car in the shed* is not entirely natural. However, *to anchor the ship with a rock* or *to comb one's hair with one's fingers* is acceptable because the verb is original and the noun derived by conversion. What do you think in the case of *to hammer the stake with a rock*? Is the noun or the verb original? Another distinguishing feature is the regularity of inflection. Converted forms will always take the regular, productive inflection, never a remnant or irregular inflection. For example, *grandstand*, *highlight*, or *highstick* may originally be either nouns or verbs. Since the past tense forms are *grandstanded*, *highlighted*, and *highsticked*, we conclude that the verb must be derived from the noun; otherwise, the past tense would be *grandstood*, *highlit*, or *highstuck*. Similarly, the past tense of *ring (the city)* is *ringed*, not *rang (the city)*, showing that *ring* in this case is a verb converted from a noun rather than an original verb. In contrast, the past tense of *deepfreeze* is *deepfroze*, not **deepfreezed*, and thus the noun (*a deepfreeze*) must derive from the verb.

Stress changes accompany the conversion of phrasal verbs to nouns. The primary stress on the particle is lost in each case, as was discussed in Chapter 3:

V + Prt > N *cómeback, rúnoff, tákeover, mákeup, rúndown, stándby, shówoff, wríteoff, lóckout, pútdown, chéckup*

Another fairly large set of converted forms shows a difference in stress, with stress on the first syllable (prefix) of the noun and the second syllable (root) of the verb (see Chapter 3):

Verb	**Noun**
condúct	*cónduct*
rebél	*rébel*
permít	*pérmit*
recórd	*récord*
objéct	*óbject*

Note that unstressing of the syllable may also lead to reduction of the vowel, e.g. *rebél* (V) /rəbɛl/ (stress on second syllable, hence full vowel) vs. *rébel* (V) /rɛbəl/ (stress on first syllable, hence reduced vowel in second syllable).

A special kind of functional shift is what we may call **commonization**, in which a proper noun is converted into a common word. A proper noun, naming a real or fictional person or place, tribe, or group, may undergo commonization to a noun, verb, or adjective, often with no phonological change:

N: *cashmere, china, sandwich, odyssey, valentine, bourbon, Braille, madras, spa, Chablis, dunce, canary*
V: *lynch, pander, canter, welsh, boycott, meander, hector*
A: *maudlin, zany, frank, bantam*

In other cases, however, a derivational suffix is added to convert the noun into the appropriate part of speech:

N: *sadism, chauvinism, marionette, tangerine, bayonet, bobby* ('policeman'), *panic*
V: *tantalize, pasteurize, mesmerize*
A: *quixotic, platonic, spartan, machiavellian, jovial*

3.4 Compounds

A **compound** is the combination of two or more free roots (plus associated affixes). It can sometimes be difficult to distinguish a compound – which is considered a single word – from a syntactic phrase consisting of a number of distinct words. As we saw at the beginning of the chapter, English orthography is often unhelpful: compounds may be written as a single word or as two words, hyphenated or not, e.g. *icecream*, *ice cream*, *ice-cream*. Phrases may likewise be written as separate words or hyphenated. Both compounds and phrases may express semantically cohesive notions. Compare *shipyard* and *automobile assembly plant*; the meanings of the compound and the phrase might be considered equally unified. It is sometimes pointed out that the order of elements in a compound tends to be nonliteral, while in a phrase it is literal, as in the difference between *forthcoming* and *come forth*, or *offputting* and *put off*, but this rule cannot be extended very far.

A better means of differentiation is internal coherence, since compounds are externally modified (at the single word boundary), whereas phrases may be internally modified (at any of the word boundaries). For example, the plural of the compound *manhole* is *manholes* not **menhole*, with the plural marker at the end, whereas the plural of the phrase *man-of-war* is *men-of-war* not **man-of-wars*, with the plural marker internal to the phrase. Another good means of distinguishing compounds is their external mobility; that is, they move in a sentence as a whole, not in parts. For example, the compound *cross-examination* moves as a unit (*The lawyer conducted the cross-examination*, *The cross-examination was conducted by the lawyer*), while part of the phrase *check out* may be moved (*He checked out the witness*, *He checked the witness out*). However, stress seems to offer the most reliable means of distinguishing a compound from a phrase. As a single word, a compound will carry only one primary stress, whereas a phrase, as a group of words, will carry more than one primary stress. The second half of the compound carries secondary stress and the vowel may be reduced (see Chapter 3). Compare the stress patterns in the following sets:

Compound	**Phrase**
stónewall	*stóne wáll*
sáfeguard	*sáfe guárd*
bréakdown	*bréak dówn*

This principle holds for compound nouns and some compound verbs. Compound adjectives, however, may carry more than one primary stress, as *dúty-frée* or *chíld-próof*.

Both the semantics and the syntax of compound are complex. Often the semantics of compounds are not simply a sum of the meaning of the parts; that is, if we know the meaning of the two roots, we cannot necessarily predict the meaning of the compound, as in *firearm*, *highball*, *makeup*, or *handout*. Note the various ways in which the meanings of the roots of these compounds interact with *home*:

homeland	'land which is one's home'
homemade	'something which is made at home'
homebody	'someone who stays at home'
homestead	'a place which is a home'
homework	'work which is done at home'
homerun	'a run to home'
homemaker	'a person who makes (cares for) the home'

The syntax of compounds is even more complex. Any combination of parts of speech seems possible, with almost any part of speech resulting. One principle which holds is that the word class of the compound is determined by the head of the compound, or its rightmost member, whereas the leftmost member carries the primary stress. The only exception to this rule is a converted compound or one containing a class changing suffix. Look at the syntactic patterns of compounding shown in Table 4.7.

Note that in addition to combining two roots, compounds may contain derivational or inflectional affixes; when the present or past participle inflectional suffix (represented by *-ing* and *-en* in Table 4.7) is added to a verb, the resulting unit functions as an adjective. Compounds may also involve conversions and back formations (discussed later in this chapter).

Self-Testing Exercise: Do Exercise 4.8 on compounding.

A problem for the differentiation of compounds and phrases is the **phrasal verb**. Older English preferred prefixed verbs, such as *forget*, *understand*, *withdraw*, *befriend*, *overrun*, *outdo*, *offset*, and *uproot* (note the position of stress on the root morpheme rather than on the prefix), but prefixing of verbs is not productive in Modern English, except for those with *out-* and *over-*. Modern English favors verbs followed by postverbal particles, such as *run over*, *lead on*, *use up*, *stretch out*, and *put down*. Like compounds, phrasal verbs have semantic coherence, evidenced by the fact that they are sometimes replaceable by single Latinate verbs, as in the following:

break out – erupt, escape	*think up – imagine*
count out – exclude	*put off – delay*
take off – depart, remove	*egg on – incite*
work out – solve	*put out – extinguish*
bring up – raise	*put away – store*
go on – continue	*take up – adopt*

Table 4.7. Syntactic Patterns in English Compounds

Compound Nouns	
N + N > N	airplane, lipstick, gold-mine, deathblow, figurehead, peppercorn
V + N > N	cut-throat, pickpocket, spoil-sport, leapfrog, drawbridge, crybaby
A + N > N	madman, blackbird, fast-food, software, hotbed, mainland, busybody
Prt + N > N	background, in-crowd, off-Broadway, afternoon
Prt + V > N	outcast, downpour, outbreak, offspring (converted prefixed or compound V)
V + Prt > N	put-down, drop-out, lockout, sit-in, fallout, runaway, drawback (converted phrasal V)
N + V > N	bloodshed, fleabite, bus-stand, sunrise, handshake, nosebleed, earthquake (converted V)
N + -'s + N > N	bachelor's degree, bull's eye, cow's milk, housemaid's knee
V + -ing + N > N	mocking bird, spending money, closing time, freezing point
N + V + ing > N	handwriting, housekeeping, foxhunting (gerund)
N + V + -er > N	hairdresser, nutcracker, landowner, peacemaker
Compound Verbs	
N + V > V	babysit, carbon-date, head-hunt, skydive, housekeep, proofread (backformations)
A + V > V	free-associate, double-book, fine-tune, whitewash, ill-treat (back-formations)
Prt + V > V	outdo, overcook, underrate, overeducate
V + V > V	blow-dry, play-act, sleep-walk, tap-dance, force-feed
A + N > V	strong-arm, blacklist, brownbag, mainstream (converted N)
Compound Adjectives	
N + A > A	headstrong, colorblind, childproof, duty-free, lifelong, carsick
A + A > A	bittersweet, icy-cold, red-hot, blue-green
N + N > A	seaside, coffee-table, back-street (converted N)[14]
A + N > A	redneck, blue-collar, solid-state (converted N)
V + Prt > A	tow-away, see-through, wrap-around (converted phrasal V)
N + V + -ing > A	man-eating, seed-bearing, heart-breaking, card-carrying, life-giving
A + V + -ing > A	easygoing, hard-hitting, good-looking, quick-cooking, high-flying
N + V + -en > A	manmade, hand-woven, housebroken, crest-fallen
A + V + -en > A	high-born, widespread, far-fetched, new-found
A + N + -ed > A	cold-blooded, thick-skinned, double-barreled, old-fashioned, public-spirited, heavy-handed

Furthermore, the meaning of the combination of verb and particle in the phrasal verb may be opaque, that is, not predictable from the meaning of the parts. Often, the difference in

14. Bauer (1983, p. 210) observes of forms like these that it might be "misleading to term them adjectives at all". They function in only very limited ways as adjectives. At the end of Chapter 5 we will look at a different way to analyze these forms.

meaning between the simple and the phrasal verb is 'completive'; the phrasal verb expresses termination or completion of the action:

burn vs. *burn down, up, on, out*	*work* vs. *work out, up*
eat vs. *eat up, through*	*wash* vs. *wash up, down, out*
pay vs. *pay up, off*	*read* vs. *read through*

Unlike compounds, however, phrasal verbs exhibit internal modification (*burn down/ burned down*, *burning down*), carry two primary stresses (*wórk óut*), and behave syntactically like phrases since the particle may move after the object, or an adverb may intercede between the verb and the particle:

He burned down the house.
He burned the house down.
He burned the house right down.
cf. **He burned right down the house. *He burned right the house down.*

For these reasons, we must conclude that phrasal verbs are phrases, not compounds.

A further problem in the analysis of compounds is **phrase compounds**, formed from entire phrases, such as *lady-in-waiting*, *dog-in-the-manger*, *forget-me-not*, *has-been*, *run-of-the-mill, break-and-enter, nuts-and-bolts, whiskey-and-soda*, *bubble-and-squeak*, or *son-in-law*, which are generally written as compounds (hyphenated) and have semantic unity. Many of these behave normally as compounds by being externally modified, such as *all has-beens, five whiskey-and-sodas* (rather than *whiskies-and-soda*). But some are internally modified like a phrase, as in the *all her ladies-in-waiting* or *our two sons-in-law*.[15] When they are inflected for the possessive, however, they show external modification like a compound, as in *son-in-law's (new car)*.[16] What precedes the possessive ending need not be a single-word compound but can be a phrase, as in *my neighbor next door's dog*, or even a clause, as in *a woman I know's niece*. By no criteria would *my neighbor next door* be considered a compound. Thus, phrase compounds seem to be phrasal in nature.

Another problem for analysis is **amalgamated compounds**. These are words which in origin are compounds, but which in the course of time have become fused and no longer separable into two distinct parts. Some examples are the following:

barn < bere 'barley' + ærn 'place'
halibut < hālig 'holy' + butte 'flatfish'
garlic < gar 'spear' + lēac 'leek'
neighbor < neah 'near' + gebur 'dweller'
cobweb < coppe 'kind of spider' + web

---

15. These forms are increasingly taking external modification, e.g. *our two son-in-laws*.

16. Historically, this has not always been so: prior to the sixteenth century, such phrases had internal modification in the possessive, as in *kings crown of England* (= 'king of England's crown'), which has the possessive ending *-s* on *king*. Then it became possible to add the possessive ending to an entire phrase, a construction called the "group genitive".

midrif < mid + hrif 'belly'
earwig < ear + wicga 'one that moves'
mildew < mele 'honey' + dew

Since these words are no longer recognizable as compounds, all are considered single, unanalyzable morphemes.[17]

3.5 Blends

A **blend** involves two processes of word formation, compounding and "clipping" (see below). Two free words are combined and blended, usually by clipping off the end of the first word and the beginning of the second word, although sometimes one or the other morpheme is left intact. Blends are sometimes called "portmanteau" words. Examples of blends are the following:

sm(oke) + (f)og	>	*smog*
mo(tor) + (ho)tel	>	*motel*
info(rmation) + (com)mercial	>	*infomercial*
simul(taneous) + (broad)cast	>	*simulcast*
trans(fer) + (re)sistor	>	*transistor*
sky + (hi)jacker	>	*skyjacker*
motor + (caval)cade	>	*motorcade*
perma(nent) + frost	>	*permafrost*
docu(mentary) + drama	>	*docudrama*
para(chutist) + trooper	>	*paratrooper*
film + (bi)ography	>	*filmography*

In the last six examples, where one half remains intact, it might also be possible to analyze *-jacker*, *-cade*, *perma-*, *docu-*, *para-*, and *-ography* as new (and perhaps productive) derivational affixes attached to free roots (think of *paramilitary/paralegal/parademic, discography*).

A rather interesting blend is *blog* < *web* + *log*, where only the final sound of the first word is part of the blended word.

3.6 Back formations

In **back formation**, speakers derive a morphologically simple word from a form which they analyze, on the basis of derivational and inflectional patterns existing in English, as a morphologically complex word. For example, by analogy with the very common derivational pattern in English in which the agentive suffix *-er* is added to a verb to produce a noun (*sing* + *-er* > *singer*, *work* + *-er* > *worker*, *buy* + *-er* > *buyer*), verbs have been formed from the following nouns by the removal of an agentive suffix, as in *sightseer* – *-er* > *sightsee*,

17. In the last four examples only half of the compound is opaque (*cob-, -rif, -wig, mil-*); the other half is identifiable.

babysitter – -er > baby-sit, or *typewriter – -er > typewrite*. Since the nouns predate the verbs in these cases, we say that the verbs are "back-formed". Back formation is thus the opposite of derivation: C – B > A as opposed to A + B > C. Without knowledge of the history of an individual word, it is usually impossible to know whether related forms result from derivation or back formation. In many cases of back formation a presumed affix is removed which is in fact not truly an affix, as in the following words where the *-or*, *-ar*, and *-er* are not the agentive suffix, but part of the root:

orator – -er > *orate*	*lecher* – -er > *lech*	*peddler* – -er > *peddle*
escalator – -er > *escalate*	*editor* – -er > *edit*	*swindler* – -er > *swindle*
sculptor – -er > *sculpt*	*hawker* – -er > *hawk*	*burglar* – -er > *burgle*
housekeeper – -er > *housekeep*	*babysitter* – -er > *babysit*	*panhandler* – -er > *panhandle*

These new words are called back formations. Note that some of them are colloquial or marginal, while others are fully accepted.

Other examples of back formations are the following, where a presumed derivational suffix has been removed:

-ion	*intuition > intuit*	-ive	*sedative > sedate*
	resurrection > resurrect	-al	*paramedical > paramedic*
	emotion > emote	-asm	*enthusiasm > enthuse*
	transcription > transcript	-y	*sleazy > sleaze*
	orientation > orientate		*lazy > laze*

In the case of *joyride < joyriding* or *henpeck < henpecked*, inflectional affixes (*-ing* and *-ed*) have been removed.

3.7 Shortening

The three types of shortening – acronyms, initialisms, and clipped forms – have in common the deletion of sound segments without respect to morphological boundaries. That is, parts of words, but not usually entire morphemes, are deleted.

Clipping. A **clipping** is the result of deliberately dropping part of a word, usually either the end or the beginning, or less often both, while retaining the same meaning and same word class, as in the following examples:

end

ad/advert < advertisement	*hack < hackney*
mike < microphone	*porn < pornography*
tarp < tarpolin	*condo < condominium*
rehab < rehabilitation	*fax < facsimile*
fan < fanatic	*mitt < mitten*

beginning

burger < *hamburger*	*venture* < *adventure*
spite < *despite*	*gin (cotton gin)* < *engine*
cello < *violoncello*	*phone* < *telephone*

beginning and end

fridge < *refrigerator*
flu < *influenza*
shrink < *head-shrinker*

The word *taxicab* has provided two clipped forms, *taxi* and *cab*, depending on whether the beginning or the end was clipped. Sometimes a word or part of a word in a phrase is clipped:

women's lib < *women's liberation*	*soap* < *soap opera*
high tech < *high technology*	*movie* < *moving picture*
narc < *narcotics agent*	*chauvinist* < *male chauvinist*

A diminutive affix may be attached to the clipped form, as in *movie*, *jammies*, *hankie*, and *nightie*. A clipping may leave behind a prefix or suffix rather than (part of) the root:

ex < *ex-husband*
bi < *bi-sexual*
bus < *omnibus*[18]

Clipping is generally not sensitive to morphological boundaries, though it does usually reflect phonological processes, selecting the longest possible syllable, what is called a maximal syllable, such as *narc* rather than *nar*.

Clippings often begin life as colloquial forms, such as the clipped forms *prof* (< *professor*), *gym* (< *gymnasium*), *chem* (< *chemistry*), *psych* (< *psychology*), or *lab* (< *laboratory*) one hears at university, but many have become fully accepted in the standard language and are no longer recognized as clipped forms.

Acronyms and initialisms. An extreme form of clipping results in acronyms and initialisms. In an **acronym**, the initial letters of words in a phrase are pronounced as a word, as in the following examples:

WASP	<	W(hite) A(nglo)-S(axon) P(rotestant)
SALT	<	S(trategic) A(rms) L(imitation) T(alks)
NATO	<	N(orth) A(tlantic) T(reaty) O(rganization)
AIDS	<	a(cquired) i(mmune) d(eficiency) s(yndrome)
radar	<	ra(dio) d(etecting) a(nd) r(anging)
laser	<	l(ight) a(mplification) (by) s(timulated) e(mission) (of) r(adiation)
sonar	<	so(und) na(vigation) r(anging)

18. In the last example, *bus* is actually part of the dative plural inflectional ending *-ibus* of the Latin word *omnis*, meaning 'all'.

Note that acronyms are not formed in an entirely systematic way; a word or words may be skipped, or the first two letters of a word may be chosen, always in order to produce a word which conforms to English phonotactics. Acronyms are written with capital letters when formed from a proper noun. In an **initialism**, the initial letters of words in a phrase are pronounced as letters, as in *r.s.v.p.*, *a.m.*, *p.m.*, *B.C.*, *A.D.*, *v.d.*, *b.m.* (What are the sources of these initialisms? Check a dictionary if you are uncertain.) Sometimes an initialism may involve only a single word, as in *i.d.* or *t.v.* Spelling and punctuation conventions are somewhat inconsistent: periods are normally used between the letters with proper nouns indicated with capitals. In a few cases such as *okay*/*o.k.*, *emcee*/*m.c.*, or *jeep* (for *G. P. = General Purpose*), the form is treated variously as an acronym or an initialism.

3.8 Root creations

The rarest form of word formation is **root creation**, the invention of an entirely new root morpheme. Brand names are the most likely examples of root creations (e.g. *Xerox(graphy), Oreo*), but when examined closely, they often prove to be based on existing words or names (e.g. *Levis,* named for the inventor Levi Strauss; *McDonald's*, named for the original owners Maurice and Richard McDonald; *Perrier*, named for the spring in France, itself named for Dr. Perrier; *Spandex* based on *expand*; *Thermos*, from the Greek word for 'warm'), or to follow patterns of word formation such as shortening or blending (e.g. *IBM*, an initialism for *International Business Machines*; *Q-tip*, an initialism for *quality tip*; *coke*, a clipped form of *Coca-cola*; *Jello*, a modified form of *jelly*; *Alka-Seltzer*, a blend of *alkali* + *seltzer*; *Amtrak*, a blend of *American* + *track*; *Botox*, a blend of *botulinum toxin*; *Cineplex*, a blend of *cinema* + *complex*; *Spam*, a blend of *spiced* + *ham*). A few recent root creations are *granola*, *quark*, and *googol* 'ten raised to the hundredth power'.[19] Onomatopoeic words, which in their pronunciation are imitative of animal sounds (e.g. *bow-wow*, *baa*, *cuckoo*, *moo*, *meow*) or other natural sounds (e.g. *twitter*, *gulp*, *hiss*, *sizzle*, *squeak*, *boom*, *blab*), can presumably be created at will as the need arises, though they are highly conventionalized and language-specific. Some new words are considered **literary coinages**, such as Shakespeare's *dwindle*, Milton's *sensuous*, and Spenser's *blatant* or *askance*. However, it is often difficult to know whether an author actually invented the word or whether he or she was simply the first to record it in writing.

Self-Testing Exercise: Using a dictionary, when necessary, identify the processes of word formation in Exercise 4.9.

19. This word is thought to be the basis for the internet search engine, *Google*, because of the large amount of information available on the internet.

4. Idioms

A final consideration in regard to words is the existence of special kinds of phrases called idioms. An **idiom** is a sequence of words which functions as a single unit; it is syntactically fixed and semantically conventionalized. Examples include the following:

spill the beans	*saw logs*	*shoot the breeze*
keep tabs on	*add fuel to the fire*	*lose one's cool*
steal the show	*bite the dust*	*rock the boat*
take stock of	*flog a dead horse*	*hold your horses*
search high and low for	*find fault with*	*take heart*
take fright	*hit the road*	*run the gamut*
be under the weather	*let the cat out of the bag*	*be dead to the world*

Idioms have the following characteristics:

- No, or little, variation is allowed in the words that constitute the phrase, so that you can't say, for example, **hold your stallions*, **bite the dirt*, **shoot the wind*, or **spill the rice*. (Note that when the wording of an idiom is changed, as in *spill the rice* above, the phrase can be interpreted only literally.)
- The semantics of the idiom are usually not predictable from the meaning of the individual words; this is what linguists call "noncompositionality". For example, you can't calculate the meaning of 'being sick' or 'feeling ill' from the meanings of *under* and *weather*.
- The meaning of idioms is often thought to be metaphorical or proverbial; they are emotionally-charged rather than neutral in meaning.
- Idioms are frequently quite colloquial.

Note that some idioms, such as *sit tight* or *easy does it*, are syntactically ill-formed.

Since idioms are not like free syntactic phrases – which can be accounted for the syntactic and semantic rules of the grammar – but are rather more like single words, the question arises as to whether they should be treated in the morphological component of the grammar, that is, whether they should be treated as unanalyzable wholes. The difficulty with doing so is that there appear to be degrees of idiomaticity, with some idioms permitting syntactic changes and some being more literal in meaning than others. *Pull some strings*, for example, seems to be much less idiomatic than *shoot the breeze* in respect to its flexibility:

internal modification:
She pulled some important strings for him.
?*They shot a little breeze today.*
fronting on object:
Those strings, he won't pull for you.
**The breeze, we shot yesterday.*

passive:

Some strings were pulled for him.
**The breeze was shot yesterday by us.*

Like *pull some strings* are *keep tabs on, lay eyes on, break the ice*, whereas like *shoot the breeze* are *saw logs, be under the weather, kick the bucket*. We don't intend to propose an answer to the question of the analysis of idioms here, but to leave it for you to ponder. Some scholars distinguish between "collocations", fixed groups of words, and true idioms. How do you think we should account for them in our grammar?

Chapter summary

Now that you have completed this chapter, you should be able to:

1. give the criteria used to distinguish a word from a phrase;
2. define the different types of morphemes and morphs;
3. analyze a word into its constituent morphs and morphemes and specify how the morphemes are concretely realized as morphs;
4. differentiate a morpheme from an allomorph;
5. write a morphemic rule for the allomorphs of a morpheme in English;
6. explain and identify the most common processes of word formation in English; and
7. analyze derived and compound words into their constituent morphs.

Recommended additional reading

Still the most complete treatment of English morphology is Marchand (1969). Bauer (1983, see especially Chapter 7) remains a useful treatment of the topic. More up-to-date treatments include Minkova and Stockwell (2009), Carstairs-McCarthy (2002), and Harley (2006). A classic discussion of morphology in the structuralist tradition may be found in Bloomfield (1933, Chapters 13 and 14) with many interesting examples. General linguistic accounts of morphology are Matthews (1991) and Bauer (2003), and an advanced treatment in a generative framework is Spencer (1991).

Textbooks which you might want to consult for a somewhat different perspective are Plag, Braun, Lappe, and Schraum (2009, Chapter 3), Jeffries (2006, Chapter 3), Fromkin, Rodman, and Hyams (2007, Chapter 3), Klammer, Schulz, and Della Volpe (2010, Chapter 3), Finegan (2008, Chapter 2), Curzan and Adams (2009, Chapter 4), or O'Grady and Archibald (2009, Chapter 4). A workbook, with exercises and answers, is Coates (1999).