four

Morphology: The Analysis of Word Structure

William O'Grady
Videa de Guzman

Carve every word before you let it fall.
–Oliver Wendell Holmes, Sr., *Urania: A Rhymed Lesson (1846)*

Objectives

In this chapter, you will learn:

- how we analyze the structure of words
- how we form words by adding prefixes, suffixes, and infixes
- how we form words by putting two or more existing words together
- how we mark words to show grammatical concepts such as number, case, agreement, and tense
- how we form words by other processes
- how the processes of word formation interact with phonology

Nothing is more important to language than words. Unlike phonemes and syllables, which are simply elements of sound, words carry meaning. And unlike sentences, which are created as needed and then discarded, words are permanently stored in a speaker's mental dictionary, or lexicon. They are arguably the fundamental building blocks of communication.

The average high school student knows about sixty thousand basic words—items such as *read*, *language*, *on*, *cold*, and *if*, whose meaning cannot be predicted from their component parts. Countless other words can be constructed and comprehended by

LANGUAGE MATTERS How Many Words Does English Have?

The *Oxford English Dictionary* (20 volumes), whose stated goal is to present all of English vocabulary "from the time of the earliest records to the present day," contains a total of 616,500 word forms. But no dictionary can ever be up to date, because new words and new uses of old words are being added to the language all the time. The *Oxford English Dictionary* offers quarterly updates (http://public.oed.com/the-oed-today/recent-updates-to-the-oed/), and the online *Urban Dictionary* (www.urbandictionary.com) adds hundreds of new definitions EVERY DAY!

Information from: The *Oxford English Dictionary*, 2nd ed. (1989).

the application of general rules to these and other elements. For example, any speaker of English who knows the verb *phish* ('fraudulently obtain sensitive information via e-mail') recognizes *phished* as its past tense form and can construct and interpret words such as *phisher, phishing*, and *unphishable*.

Linguists use the term **morphology** to refer to the part of the grammar that is concerned with words and word formation. As we will see, the study of morphology offers important insights into how language works, revealing the need for different categories of words, the presence of word-internal structure, and the existence of operations that create and modify words in various ways.

1 Words and Word Structure

As speakers of English, we rarely have difficulty segmenting a stream of speech sounds into words or deciding where to leave spaces when writing a sentence. What, though, is a word?

Linguists define the **word** as the smallest **free form** found in language. A free form is simply an element that does not have to occur in a fixed position with respect to neighboring elements; in many cases, it can even appear in isolation. Consider, for instance, the following sentence.

1) Dinosaurs are extinct.

We all share the intuition that *dinosaurs* is a word here and that the plural marker *-s* is not. But why? The key observation is that *-s* is not a free form: it never occurs in isolation and cannot be separated from the noun to which it belongs. (Elements that must be attached to something else are written here with a hyphen; an asterisk indicates unacceptability.)

2) *Dinosaur are -s extinct.

In contrast, *dinosaurs* is a word because it can occur both in isolation, as in example *3*, and in different positions within sentences, as in example *4*:

3) Speaker A: What creatures do children find most fascinating?
   Speaker B: Dinosaurs.

4) a. Paleontologists study *dinosaurs*.
   b. *Dinosaurs* are studied by paleontologists.
   c. It's *dinosaurs* that paleontologists study.

Some words—like *are*—normally do not occur in isolation. However, they are still free forms because their positioning with respect to neighboring words is not entirely fixed, as shown in *5*.

5) a. Dinosaurs *are* extinct.
   b. *Are* dinosaurs extinct?

1.1 Morphemes

Words have an internal structure consisting of smaller units organized with respect to each other in a particular way. The most important component of word structure is the **morpheme**, the smallest unit of language that carries information about meaning or function. The word *builder*, for example, consists of two morphemes: *build* (with the meaning 'construct') and *-er* (which indicates that the entire word functions as a noun with the meaning 'one who builds'). Similarly, the word *houses* is made up of the morphemes *house* (with the meaning 'dwelling') and *-s* (with the meaning 'more than one').

Some words consist of a single morpheme. For example, the word *tremble* cannot be divided into smaller parts (say, *tr* and *emble* or *t* and *remble*) that carry information about the word's meaning or function. Such words are said to be **simple words** and are distinguished from **complex words**, which contain two or more morphemes (see Table 4.1).

Table 4.1 Words consisting of one or more morphemes

One	*Two*	*Three*	*More than three*
and			
couple	couple-s		
hunt	hunt-er	hunt-er-s	
act	act-ive	act-iv-ate	re-act-iv-ate

Free and Bound Morphemes

A morpheme that can be a word by itself is called a **free morpheme**, whereas a morpheme that must be attached to another element is a **bound morpheme**. The morpheme *boy*, for example, is free because it can be used as a word on its own; plural *-s*, however, is bound.

Concepts that are expressed by free morphemes in English do not necessarily have the same status in other languages. For example, in Hare (an Athabaskan language spoken in Canada's Northwest Territories), morphemes that indicate body parts must always be attached to a morpheme designating a possessor, as shown in Table 4.2. (The diacritic [´] marks a high tone.)

Table 4.2 Some body part names in Hare

Without a possessor		*With a possessor*	
*fí	'head'	sefí	'my head'
*bé	'belly'	nebé	'your belly'
*dzé	'heart'	ʔedzé	'someone's heart/a heart'

In English, of course, body part names are free morphemes and do not have to be attached to another element.

Conversely, there are also some bound forms in English whose counterparts in other languages are free. The notion 'past' or 'completed' is expressed by the bound morpheme *-ed* in English (as in *I washed the car*, or *a washed car*), but by the free morpheme *lɛɛw* in Thai. As the following sentence shows, *lɛɛw* can even be separated from the verb by an intervening word. (Tone is not marked here.)

6) Boon thaan khaaw lɛɛw.
Boon eat rice PAST
'Boon ate rice.'

Allomorphs

The variant pronunciations of a morpheme are called its **allomorphs**. The morpheme used to express indefiniteness in English has two allomorphs—*an* before a word that begins with a vowel sound and *a* before a word that begins with a consonant sound.

7) an orange a building
an accent a car
an eel a girl

Note that the choice of *an* or *a* is determined on the basis of pronunciation, not spelling, which is why we say *an M.A. degree* and *a U.S. dollar*.

Another example of allomorphic variation is found in the pronunciation of the plural morpheme *-s* in the following words.

8) cats
dogs
judges

Whereas the plural is /s/ in *cats*, it is /z/ in *dogs*, and /əz/ in *judges*. Here again, selection of the proper allomorph is dependent on phonological facts. (For more on this, see Section 6.)

Yet another case of allomorphic variation is found in the pronunciation of the prefix *in-*, with the meaning 'not'. The final consonant is pronounced as /n/ in most cases—*indirect, inactive*, and so on. But it is pronounced as /m/ in front of another labial consonant (*impossible, immodest*), as /l/ in front of another /l/ (*illegal*), and as /ɹ/ in front of another /ɹ/ (*irregular*). These changes are easy to spot because of the spelling, but remember that allomorphic variation involves pronunciation. In some cases, this is reflected in the spelling, but in other cases (such as plural *-s*), it is not.

1.2 Analyzing Word Structure

In order to represent the internal structure of words, it is necessary not only to identify each of the component morphemes but also to classify them in terms of their contribution to the meaning and function of the larger word.

Roots and Affixes

Complex words typically consist of a **root** morpheme and one or more **affixes**. The root constitutes the core of the word and carries the major component of its meaning. Roots typically belong to a **lexical category**, such as noun (N), verb (V), adjective (A), or preposition (P).

Unlike roots, affixes do not belong to a lexical category and are always bound morphemes. For example, the affix *-er* is a bound morpheme that combines with a verb such as *teach*, giving a noun with the meaning 'one who teaches'. The internal structure of this word can be represented in Figure 4.1. (*Af* stands for affix.)

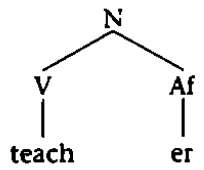

Figure 4.1 The internal structure of the word *teacher*

LANGUAGE MATTERS Having Trouble Figuring Out a Word's Category?

Here are some traditional rules of thumb that you may have learned in school:

- Nouns typically refer to people and things (*citizen, tree, intelligence*, etc.).
- Verbs tend to denote actions, sensations, and states (*depart, teach, melt, remain*, etc.).
- Adjectives usually name properties (*nice, red, tall*, etc.).
- Prepositions generally encode spatial relations (*in, near, under*, etc.).

These traditional rules of thumb can suffice for now, but in Chapter 5 we will look at how to categorize words in a more sophisticated way.

Figure 4.2 provides some additional examples of word structure.

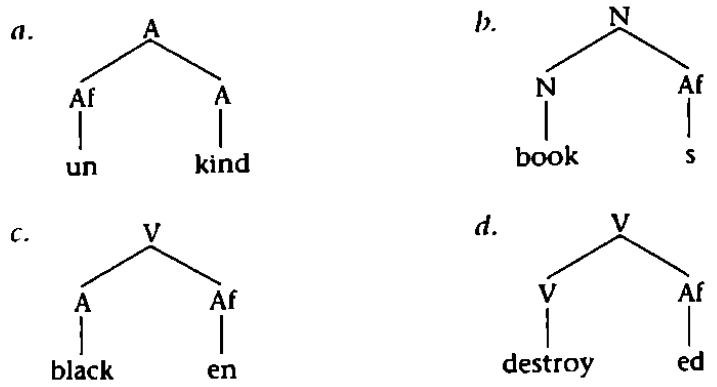

Figure 4.2 Some other words with an internal structure consisting of a root and an affix

The structural diagrams in Figures 4.1 and 4.2 are often called **trees**. The information they depict can also be represented by using labeled bracketing—[$_{A}$ [$_{Af}$ un] [$_{A}$ kind]] for *unkind* and [$_{N}$ [$_{N}$ book] [$_{Af}$ s]] for *books*. (This is somewhat harder to read, though, and we will generally use tree structures in this chapter.) Where the details of a word's structure are irrelevant to the point being considered, it is traditional to use a much simpler system of representation that indicates only the location of the morpheme boundaries: *un-kind*, *book-s*, and so on.

Bases

A **base** is the form to which an affix is added. In many cases, the base is also the root. In *books*, for example, the element to which the affix *-s* is added corresponds to the word's root. In other cases, however, the base can be larger than a root, which is always just a single morpheme. This happens in words such as *blackened*, in which the past tense affix *-ed* is added to the verbal base *blacken*—a unit consisting of the root morpheme *black* and the suffix *-en*.

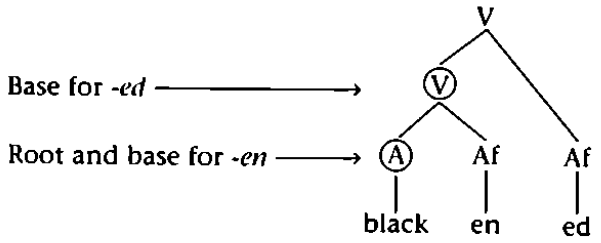

Figure 4.3 A tree diagram illustrating the difference between a root and a base

In this case, *black* is not only the root for the entire word but also the base for *-en*. The unit *blacken*, however, is simply the base for *-ed*.

Types of Affixes

An affix that is attached to the front of its base is called a **prefix**, whereas an affix that is attached to the end of its base is termed a **suffix**. Both types of affix occur in English, as shown in Table 4.3.

Table 4.3 Some English prefixes and suffixes

Prefixes	*Suffixes*
de-activate	faith-*ful*
re-play	govern-*ment*
il-legal	hunt-*er*
in-accurate	kind-*ness*

We will consider the nature and properties of English affixes in more detail in Sections 2.1 and 4.1.

Far less common than prefixes and suffixes are **infixes**, a type of affix that occurs within another morpheme. The data in Table 4.4 from the Philippine language Tagalog contains examples of the infix *-in-*, which is inserted after the first consonant of the root to mark a completed event.

Table 4.4 Examples of the Tagalog infix *-in-*

Base		*Infixed form*	
bili	'buy'	b-*in*-ili	'bought'
basa	'read'	b-*in*-asa	'read' (PAST)
sulat	'write'	s-*in*-ulat	'wrote'

Beginning students sometimes think that a morpheme such as *-en* in *black-en-ed* is an infix since it occurs between two other morphemes (*black* and *-ed*), but this is not right: *-en* is a suffix that combines with the adjective *black* to give the verb *blacken*, to which the suffix *-ed* is then added (see Figure 4.3). To be an infix, an affix must occur *inside* another morpheme (as when Tagalog *-in-* appears inside *sulat* 'write'). Nothing of this sort happens in the case of *-en*.

A very special type of infixing system is found in Arabic and other Semitic languages, in which a typical root consists simply of three consonants. Various combinations of vowels are then inserted among the consonants to express a range of grammatical contrasts. (In the examples that follow, the segments of the root are written in boldface.)

9) **kataba** **kutib** **aktub**
'wrote' 'has been written' 'am writing'

One way to represent the structure of such words is as follows, with the root and affixal vowels assigned to different **tiers**, or levels of structure, that combine with each other to give the word's pronunciation (see Figure 4.4).

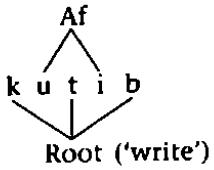

Figure 4.4 Two tiers used to represent the structure of the infixed word meaning 'has been written' in Arabic

Tagalog and Arabic offer examples of **nonconcatenative morphology**, in that words are not always built in the linear, additive fashion illustrated by English words such as *travel-er-s* and *creat-iv-ity*.

Problematic Cases

The majority of complex words in English are built from roots that are free morphemes. In the words *re-do* and *treat-ment*, for example, the root (*do* and *treat*, respectively) can itself be used as a word. Because most complex words work this way in English, English morphology is said to be **word-based morphology**.

There is no such requirement in morpheme-based systems. In Japanese and Spanish, for instance, verbal roots are always bound and can therefore not stand alone: *camin* is not a word in Spanish, *arui* is not a word in Japanese, and so on.

10) a. Spanish

camin-ó	**escuch**-ó	**limpi**-ó
walk-PAST	listen-PAST	wipe-PAST

b. Japanese

arui-ta	**kii**-ta	**hui**-ta
walk-PAST	listen-PAST	wipe-PAST

English too has a sizeable number of bound roots. For example, the word *unkempt* seems to consist of the prefix *un-* (with the meaning 'not') and the root *kempt* (meaning 'groomed' or 'combed'), even though *kempt* cannot be used by itself. *Kempt* was once a word in English (with the meaning 'combed'), and it was to this base that the affix *un-* was originally attached. However, *kempt* later disappeared from the

LANGUAGE MATTERS Word Play

The following excerpt from the humorous essay "How I Met My Wife" by Jack Winter plays on the fact that certain English roots are bound and cannot be used as words:

> I was **furling** my **wieldy** umbrella for the coat check when I saw her standing alone in the corner. She was a **descript** person, a woman in a state of total **array**. Her hair was **kempt**, her clothing **shevelled**, and she moved in a **gainly** way. (From *The New Yorker*, July 25, 1994).

language, leaving behind the word *unkempt*, in which an affix appears with a bound root.

Still other words with bound roots were borrowed into English as whole words. *Inept*, for instance, comes from Latin *ineptus*, 'unsuited'. Its relationship to the word *apt* may have been evident at one time, but it now seems to consist of a prefix meaning 'not' and a bound root.

Another class of words that are problematic for morphological analysis includes items such as *receive, deceive, conceive,* and *perceive,* or *permit, submit,* and *commit.* These items were borrowed into English from Latin (usually via French) as whole words, and their component syllables have no identifiable meaning of their own. Unlike the *in-* of *inept*, which retains the meaning of negation, the *re-* of *receive* does not have the sense of 'again' that it does in *redo*, and no specific meaning can be assigned to *-ceive* or *-mit*. For this reason, we will not treat these word parts as morphemes.

2 Derivation

Derivation uses an affix to build a word with a meaning and/or category distinct from that of its base. One of the most common derivational affixes in English is the suffix *-er*, which combines with a verb to form a noun with the meaning 'one who *Vs*' as shown in Table 4.5. (Do not confuse this suffix with the *-er* that applies to a noun in cases such as *New Yorker* and *islander* or the *-er* that combines with an adjective in cases such as *taller* and *smarter*.)

Table 4.5 The derivational affix *-er*

Verb base	*Derived noun*
sell	sell-er (one who sells)
write	writ-er (one who writes)
teach	teach-er (one who teaches)
sing	sing-er (one who sings)
think	think-er (one who thinks)

Other examples of derivation include *treatment*, in which the suffix *-ment* combines with the verb *treat* to give the noun *treatment*; *unkind*, in which the prefix *un-* combines with the adjective *kind* to give a new adjective with a different meaning; and the other derived words illustrated in Figure 4.5.

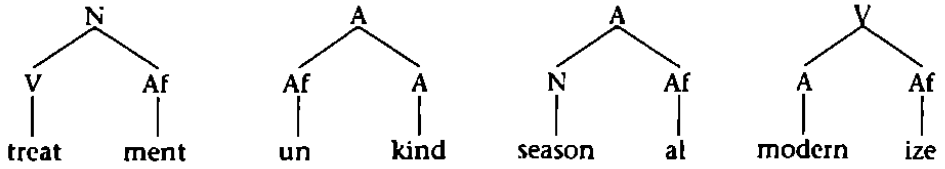

Figure 4.5 Some words formed by derivation