WORD-FORMATION IN ENGLISH

INGO PLAG

University of Siegen

CAMBRIDGE
UNIVERSITY PRESS

Contents

1

Basic concepts

Outline

This chapter introduces basic concepts needed for the study and description of morphologically complex words. Since this is a book about the particular branch of morphology called word-formation, we will first take a look at the notion of 'word.' We will then turn to a first analysis of the kinds of phenomena that fall into the domain of word-formation, before we finally discuss how word-formation can be distinguished from the other sub-branch of morphology, inflection.

1.1   What is a word?

It has been estimated that average speakers of a language know from 45,000 to 60,000 words. This means that we as speakers must have stored these words somewhere in our heads, our so-called **mental lexicon**. But what exactly is it that we have stored? What do we mean when we speak of 'words'?

In non-technical everyday talk, we speak about 'words' without ever thinking that this could be a problematic notion. In this section we will see that, perhaps contra our first intuitive feeling, the 'word' as a linguistic unit deserves some attention, because it is not as straightforward as one might expect.

If you had to define what a word is, you might first think of the word as a unit in the writing system, the so-called **orthographic word**. You could say, for example, that a word is an uninterrupted string of letters which is preceded by a blank space and followed either by a blank space or a punctuation mark. At first sight, this looks like a good definition that can be easily applied, as we can see in the sentence in example (1):

(1)   Linguistics is a fascinating subject.

We count five orthographic words: there are five uninterrupted strings of letters, all of which are preceded by a blank space, four of which are also followed by a blank space, one of which is followed by a period. This count is also in accordance with

our intuitive feeling of what a word is. Even without this somewhat formal and technical definition, you might want to argue, you could have told that the sentence in (1) contains five words. However, things are not always that straightforward. Consider the following example, and try to determine how many words there are:

(2)   Benjamin's girlfriend lives in a high-rise apartment building

Your result depends on a number of assumptions. If you consider apostrophies to be punctuation marks, *Benjamin's* constitutes two (orthographic) words. If not, *Benjamin's* is one word. If you consider a hyphen a punctuation mark, *high-rise* is two (orthographic) words, otherwise it's one (orthographic) word. The last two strings, *apartment building*, are easy to classify, they are two (orthographic) words, whereas *girlfriend* must be considered one (orthographic) word. However, there are two basic problems with our orthographic analysis. The first one is that orthography is often variable. Thus, *girlfriend* is also attested with the spellings <girl-friend> and even <girl friend> (fish brackets are used to indicate spellings, i.e. letters). Such variable spellings are quite common (cf. *word-formation*, *word formation*, and *wordformation*, all of them attested), and even where the spelling is conventionalized, similar words are often spelled differently, as evidenced with *grapefruit* vs. *passion fruit*. For our problem of defining what a word is, such cases are rather annoying. The notion of what a word is, should, after all, not depend on the fancies of individual writers or the arbitrariness of the English spelling system. The second problem with the orthographically defined word is that it may not always coincide with our intuitions. Thus, most of us would probably agree that *girlfriend* is a word (i.e. one word) which consists of two words (*girl* and *friend*), a so-called **compound**. If compounds are one word, they should be spelled without a blank space separating the elements that together make up the compound. Unfortunately, this is not the case. The compound *apartment building*, for example, has a blank space between *apartment* and *building*.

To summarize our discussion of purely orthographic criteria of wordhood, we must say that these criteria are not entirely reliable. Furthermore, a purely orthographic notion of 'word' would have the disadvantage of implying that illiterate speakers would have no idea about what a word might be. This is plainly false.

What, might you ask, is responsible for our intuitions about what a word is, if not the orthography? It has been argued that the word could be defined in four other ways: in terms of sound structure (i.e. phonologically), in terms of its internal integrity, in terms of meaning (i.e. semantically), or in terms of sentence structure (i.e. syntactically). We will discuss each in turn.

You might have thought that the blank spaces in writing reflect pauses in the spoken language, and that perhaps one could define the word as a unit in speech surrounded by pauses. However, if you carefully listen to naturally occurring speech

you will realize that speakers do not make pauses before or after each word. Perhaps we could say that words can be surrounded by potential pauses in speech. This criterion works much better, but it runs into problems because speakers can and do make pauses not only between words but also between syllables, for example for emphasis.

But there is another way in which the sound structure can tell us something about the nature of the word as a linguistic unit. Think of stress. In many languages (including English) the word is the unit that is crucial for the occurrence and distribution of stress. Spoken in isolation, every word can have only one **main stress**, as indicated by the acute accents (´) in the data presented in (3) below (note that we speak of linguistic 'data' when we refer to language examples to be analyzed).

(3)   cárpenter   téxtbook
         wáter   análysis
         féderal   sýllable
         móther   understánd

The main stressed syllable is the syllable which is the most prominent one in a word. Prominence of a syllable is a function of loudness, pitch and duration, with stressed syllables being pronounced louder, with higher pitch, or with longer duration than the neighboring syllable(s). Longer words often have additional, weaker stresses, so-called **secondary stresses**, which we ignore here for simplicity's sake. The words in (4) now show that the phonologically defined word is not always identical with the orthographically defined word.

(4)   Bénjamin's
         gírlfriend
         apártment building

While *apártment building* is two orthographic words, it is only one word in terms of stress behavior. The same holds for other compounds like *trável agency*, *wéather forecast*, *spáce shuttle*, etc. We see that in these examples the phonological definition of 'word' comes closer to our intuition of what a word should be.

We have to take into consideration, however, that not all words carry stress. For example, function words like articles or auxiliaries are usually unstressed (*a cár*, *the dóg*, *Máry has a dóg*) or even severely reduced (*Jane's in the garden*, *I'll be there*). Hence, the stress criterion is not readily applicable to function words and to words that hang on to other words, so-called **clitics** (e.g. *'ve*, *'s*, *'ll*).

Let us now consider the integrity criterion, which says that the word is an indivisible unit into which no intervening material may be inserted. If some modificational element is added to a word, it must be done at the edges, but never inside the word. For example, plural endings such as *-s* in *girls*, negative elements

such as *un-* in *uncommon* or endings that create verbs out of adjectives (such as *-ize* in *colonialize*) never occur inside the word they modify, but are added either before or after the word. Hence, the impossibility of formations such as **gi-s-rl*, **com-un-mon*, **col-ize-onial* (note that the asterisk indicates impossible words, i.e. words that are not formed in accordance with the morphological rules of the language in question).

However, there are some cases in which word integrity is violated. For example, the plural of *son-in-law* is not **son-in-laws* but *sons-in-law*. Under the assumption that *son-in-law* is one word (i.e. some kind of compound), the plural ending is inserted inside the word and not at the end. Apart from certain compounds, we can find other words that violate the integrity criterion for words. For example, in creations like *abso-bloody-lutely*, the element *bloody* is inserted inside the word, and not, as we would expect, at one of the edges. In fact, it is impossible to add *bloody* before or after *absolutely* in order to achieve the same effect. *Absolutely bloody* would mean something completely different, and **bloody absolutely* seems utterly strange and, above all, uninterpretable.

We can conclude that there are certain, though marginal counterexamples to the integrity criterion, but surely these cases should be regarded as the proverbial exceptions that prove the rule.

The semantic definition of 'word' states that a word expresses a unified semantic concept. Although this may be true for most words (even for *son-in-law*, which is ill-behaved with regard to the integrity criterion), it is not sufficient in order to differentiate between words and non-words. The simple reason is that not every unified semantic concept corresponds to one word in a given language. Consider, for example, the smell of fresh rain in a forest in the fall. Certainly a unified concept, but we would not consider *the smell of fresh rain in a forest in the fall* a word. In fact, English simply has no single word for this concept. A similar problem arises with phrases like *the woman who lives next door*. This phrase refers to a particular person and should therefore be considered as something expressing a unified concept. This concept is however expressed by more than one word. We learn from this example that although a word may always express a unified concept, not every unified concept is expressed by one word. Hence the criterion is not very helpful in distinguishing between words and larger units that are not words. An additional problem arises from the notion of 'unified semantic concept' itself, which seems to be rather vague. For example, does the complicated word *conventionalization* really express a unified concept? If we paraphrase it as 'the act or result of making something conventional,' it is not entirely clear whether this should still be regarded as a 'unified concept.' Before taking the semantic definition of 'word' seriously, it would be necessary to define exactly what 'unified concept' means.

This leaves us with the syntactically oriented criterion of wordhood. Words are usually considered to be syntactic atoms, i.e. the smallest elements in a sentence. Words belong to certain syntactic classes (nouns, verbs, adjectives, prepositions, etc.), which are called **parts of speech, word classes,** or **syntactic categories**. The position in which a given word may occur in a sentence is determined by the syntactic rules of a language. These rules make reference to words and the class they belong to. For example, *the* is said to belong to the class called articles, and there are rules which determine where in a sentence such words, i.e. articles, may occur (usually before nouns and their modifiers, as in *the big house*). We can therefore test whether something is a word by checking whether it belongs to such a word class. If the item in question, for example, follows the rules for nouns, it should be a noun, hence a word. Or consider the fact that only words (and groups of words), but no smaller units, can be moved to a different position in the sentence. For example, in 'yes/no' questions, the auxiliary verb does not occur in its usual position but is moved to the beginning of the sentence (*You **can** read my textbook* vs. ***Can** you read my textbook?*). Thus syntactic criteria can help to determine the wordhood of a given entity.

To summarize our discussion of the possible definition of 'word' we can say that, in spite of the intuitive appeal of the notion of 'word,' it is sometimes not easy to decide whether a given string of sounds (or letters) should be regarded as a word or not. In the treatment above, we have concentrated on the discussion of such problematic cases. In most cases, however, the stress criterion, the integrity criterion and the syntactic criteria lead to sufficiently clear results. The properties of words are summarized in (5):

(5)   Properties of words
        – words are entities having a part of speech specification
        – words are syntactic atoms
        – words (usually) have one main stress
        – words (usually) are indivisible units (no intervening material possible)

Unfortunately, there is yet another problem with the word *word* itself, namely its ambiguity. Thus, even if we have unequivocally decided that a given string is a word, some insecurity remains about what exactly we refer to when we say things like

(6)   a. The word *be* occurs twice in the sentence.
        b. [ðəwɚdbiəkɚztwaɪsɪnðəsentəns]

The utterance in (6), given in both its orthographic and its phonetic representation, can be understood in different ways, it is ambiguous in a number of ways. First,

<be> or the sounds [bi] may refer to the letters or the sounds which they stand for. Then sentence (6) would, for example, be true for every written sentence in which the string <BLANK SPACE be BLANK SPACE> occurs twice. Referring to the spoken equivalent of (6a), represented by the phonetic transcription in (6b), (6) would be true for any sentence in which the string of sounds [bi] occurs twice. In this case, [bi] could refer to two different 'words,' e.g. *bee* and *be*. The next possible interpretation is that in (6) we refer to the grammatically specified form *be*, i.e. the infinitive, imperative or subjunctive form of the linking verb BE. Such a grammatically specified form is called the **grammatical word** (or **morphosyntactic word**). Under this reading, (6) would be true of any sentence containing two infinitive, two imperative or two subjunctive forms of *be*, but would not be true of a sentence which contains any of the forms *am*, *is*, *are*, *was*, *were*.

To complicate matters further, even the same form can stand for more than one different grammatical word. Thus, the **word-form** *be* is used for three different grammatical words, expressing subjunctive, infinitive or imperative, respectively. This brings us to the last possible interpretation, namely that (6) may refer to the linking verb BE in general, as we would find it in a dictionary entry, abstracting away from the different word-forms in which the word BE occurs (*am*, *is*, *are*, *was*, *were*, *be*, *been*). Under this reading, (6) would be true for any sentence containing any two word-forms of the linking verb, i.e. *am*, *is*, *are*, *was*, *were*, *be*, and *been*. Under this interpretation, *am*, *is*, *are*, *was*, *were*, *be*, and *been* are regarded as realizations of an abstract morphological entity. Such abstract entities are called **lexemes**. Coming back to our previous example of *be* and *bee*, we could now say that BE and BEE are two different lexemes that simply sound the same (usually small capitals are used when writing about lexemes). In technical terms, they are **homophonous** words, or simply **homophones**.

In everyday speech, these rather subtle ambiguities in our use of the term 'word' are easily tolerated and are often not even noticed, but when discussing linguistics, it is sometimes necessary to be more explicit about what exactly one talks about. Having discussed what we can mean when we speak of 'words,' we may now turn to the question of what exactly we are dealing with in the study of word-formation.

1.2   Studying word-formation

As the term 'word-formation' suggests, we are dealing with the formation of words, but what does that mean? Let us look at a number of words that fall into the domain of word-formation and a number of words that do not:

(7)	a.	employee	b.	apartment building	c.	chair
		inventor		greenhouse		neighbor
		inability		team manager		matter
		meaningless		truck driver		brow
		suddenness		blackboard		great
		unhappy		son-in-law		promise
		decolonialization		pickpocket		discuss

In columns (7a) and (7b) we find words that are obviously composed by putting together smaller elements to form larger words with more complex meanings. We can say that we are dealing with morphologically **complex words**. For example, *employee* can be analyzed as being composed of the verb *employ* and the ending *-ee*, the adjective *unhappy* can be analyzed as being derived from the adjective *happy* by the attachment of the element *un-*, and *decolonialization* can be segmented into the smallest parts *de-, colony, -al, -ize*, and *-ation*. We can thus decompose complex words into their smallest meaningful units. These units are called **morphemes**.

In contrast to those in (7a) and (7b), the words in (7c) cannot be decomposed into smaller meaningful units, they consist of only one morpheme, they are monomorphemic. *Neighbor*, for example, is not composed of *neighb-* and *-or*, although the word looks rather similar to a word such as *inventor*. *Inventor* ('someone who invents (something)') is decomposable into two morphemes, because both *invent-* and *-or* are meaningful elements, whereas neither *neighb-* nor *-or* carry any meaning in *neighbor* (a neighbor is not someone who neighbs, whatever that may be . . .).

As we can see from the complex words in (7a), some morphemes can occur only if attached to some other morpheme(s). Such morphemes are called **bound morphemes**, in contrast to **free morphemes**, which do occur on their own. Some bound morphemes, for example *un-*, must always be attached before the central meaningful element of the word, the so-called **root**, **stem**, or **base**, whereas other bound morphemes, such as *-ity*, *-ness*, or *-less*, must follow the root. Using Latin-influenced terminology, *un-* is called a **prefix**, *-ity* a **suffix**, with **affix** being the cover term for all bound morphemes that attach to roots. Note that there are also **bound roots**, i.e. roots that only occur in combination with some other bound morpheme. Examples of bound roots are often of Latin origin, e.g. *later-* (as in combination with the adjectival suffix *-al*), *circul-* (as in *circulate, circulation, circulatory, circular*), *approb-* (as in *approbate, approbation, approbatory, approbator*), *simul-* (as in *simulant, simulate, simulation*), but occasional native bound roots can also be found (e.g. *hap-*, as in *hapless*).

Before we turn to the application of the terms introduced in this section, we should perhaps clarify the distinction between 'root,' 'stem,' and 'base,' because

these terms are not always clearly defined in the morphological literature and are therefore a potential source of confusion. One reason for this lamentable lack of clarity is that languages differ remarkably in their morphological make-up, so that different terminologies reflect different organizational principles in the different languages. The part of a word which an affix is attached to is called **base**. We will use the term **root** to refer to bases that cannot be analyzed further into morphemes. The term 'stem' is usually used for bases of inflections, and occasionally also for bases of derivational affixes. To avoid terminological confusion, we will avoid the use of the term 'stem' altogether and speak of 'roots' and 'bases' only.

The term 'root' is used when we want to explicitly refer to the indivisible central part of a complex word. In all other cases, where the status of a form as indivisible or not is not at issue, we can just speak of **bases** (or, if the base is a word, of **base words**). The derived word is often referred to as a **derivative**. The base of the suffix *-al* in the derivative *colonial* is *colony*, the base of the suffix *-ize* in the derivative *colonialize* is *colonial*, the base of *-ation* in the derivative *colonialization* is *colonialize*. In the case of *colonial* the base is a root, in the other cases it is not. The terminological distinctions are again illustrated in (8), using *colonialization* as an example:

(8)

<div align="center">

derivative of *-ize*/base of *-ation*

colony   *-al*   *-ize*   *-ation*

root/base of *-al*

derivative of *-al* / base of *-ize*

derivative of *-ation*

</div>
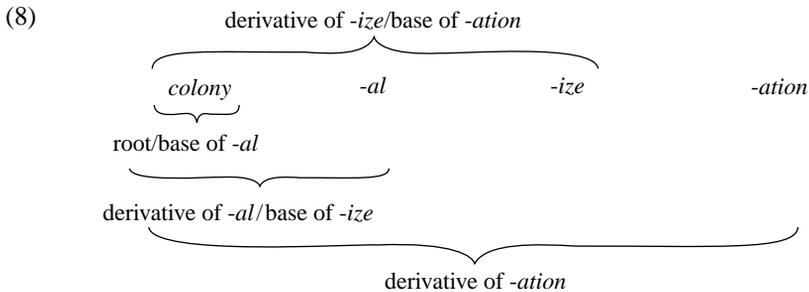
While suffixes and prefixes are very common in English, there are also rare cases of affixes that cannot be considered prefixes or suffixes, because they are inserted not at the boundary of another morpheme but right into another morpheme. Compare again our formation *abso-bloody-lutely* from above, where *-bloody-* interrupts the morpheme *absolute* (the base *absolutely* consists of course of the two morphemes *absolute* and *-ly*). Such intervening affixes are called **infixes**. Now, shouldn't we analyze *-al* in *decolonialization* also as an infix (after all, it occurs inside a word)? The answer is 'no.' True, *-al* occurs inside a complex word, but crucially it does not occur inside another morpheme. It follows one morpheme (*colony*), and precedes another one (*-ize*). Since it follows a base, it must be a suffix, which, in this particular case, is followed by another suffix.

One of the most interesting questions that arise from the study of affixed words is which mechanisms regulate the distribution of affixes and bases. That is, what exactly is responsible for the fact that some morphemes easily combine with each other, whereas others do not? For example, why can't we combine *de-* with *colony* to form **de-colony* or attach *-al* to *-ize* as in **summarize-al*? We will frequently return to this fundamental question throughout this book and learn that – perhaps unexpectedly – the combinatorial properties of morphemes are not as arbitrary as they may first appear.

Returning to the data in (7), we see that complex words need not be made up of roots and affixes. It is also possible to combine two bases, a process we already know as **compounding**. The words in (7b) (*apartment building*, *greenhouse*, *team manager*, *truck driver*) are cases in point.

So far, we have only encountered complex words that are created by **concatenation**, i.e. by linking together bases and affixes as in a chain. There are, however, also other, i.e. **non-concatenative**, ways to form morphologically complex words. For instance, we can turn nouns into verbs by adding nothing at all to the base. To give only one example, consider the noun *water*, which can also be used as a verb, meaning 'provide water,' as in *John waters his flowers every day*. This process is referred to as **conversion**, **zero-suffixation**, or **transposition**. Conversion is a rather wide-spread process, as is further illustrated in (9), which shows examples of verb-to-noun conversion:

(9)   *to walk*   take *a walk*
        to go   have *a go*
        to bite   have *a bite*
        to hug   give *a hug*

The term 'zero-suffixation' implies that there is a suffix present in such forms, only that this suffix cannot be heard or seen, hence *zero*-suffix. The postulation of zero elements in language may seem strange, but only at first sight. Speakers frequently leave out entities that are nevertheless integral, though invisible or inaudible, parts of their utterances. Consider the following sentences:

(10)   a. Jill has a car. Bob too.
         b. Jill promised Bob to buy him the book.

In (10a), *Bob too* is not a complete sentence, something is missing. What is missing is something like *has a car*, which can however be easily recovered by competent speakers on the basis of the rules of English grammar and the context. Similarly, in (10b) the verb *buy* does not have an overtly expressed subject. The logical subject (i.e. the buyer) can however be easily inferred: it must be the same person that is the logical subject of the superordinate verb *promise*. What these examples show us is that under certain conditions meaningful elements can indeed be left unexpressed

on the surface, although they must still be somehow present at a certain level of analysis. Hence, it is not entirely strange to posit morphemes which have no overt expression. We will discuss this issue in more detail in section 2.1.2 and section 5.1.2 when we deal with non-affixational word-formation.

Apart from processes that attach something to a base (affixation) and processes that do not alter the base (conversion), there are processes involving the deletion of material, yet another case of non-concatenative morphology. English christian names, for example, can be shortened by deleting parts of the base word (see (11a)), a process also occasionally encountered with words that are not personal names (see (11b)). This type of word-formation is called **truncation**, with the term **clipping** also being used.

(11)   a. Ron (← Aaron)   b. condo (← condominium)
            Liz (← Elizabeth)   demo (← demonstration)
            Mike (← Michael)   disco (← discotheque)
            Trish (← Patricia)   lab (← laboratory)

Sometimes truncation and affixation can occur together, as with formations expressing intimacy or smallness, so-called **diminutives**:

(12)   Mandy (←Amanda)
            Andy (← Andrew)
            Charlie (← Charles)
            Patty (← Patricia)
            Robbie (← Roberta)

We also find so-called **blends**, which are amalgamations of parts of different words, such as *smog* (← **sm**oke/**f**og*) or *modem* (← **mo**dulator/**dem**odulator*). Blends based on orthography are called **acronyms**, which are coined by combining the initial letters of compounds or phrases into a pronounceable new word (*NATO*, *UNESCO*, etc.). Simple **abbreviations** like *UK* or *USA* are also quite common. The classification of blending either as a special case of compounding or as a case of non-affixational derivation is not so clear. In section 5.2.2 we will argue that it is best described as derivation.

In sum, there is a host of possibilities speakers of a language have at their disposal (or had so in the past, when the words were first coined) to create new words on the basis of existing ones, including the addition and subtraction of phonetic (or orthographic) material. The study of word-formation can thus be defined as the study of the ways in which new complex words are built on the basis of other words or morphemes. Some consequences of such a definition will be discussed in the next section.

1.3   Inflection and derivation

The definition of 'word-formation' in the previous paragraph raises an important problem. Consider the italicized words in (13) and think about the question whether *kicks* in (13a), *drinking* in (13b), or *students* in (13c) should be regarded as 'new words' in the sense of our definition.

(13)   a. She *kicks* the ball.
         b. The baby is not *drinking* her milk.
         c. The *students* are not interested in physics.

The italicized words in (13) are certainly complex words, all of them are made up of two morphemes. *Kicks* consists of the verb *kick* and the third-person-singular suffix -*s*, *drinking* consists of the verb *drink* and the participial suffix -*ing*, and *students* consists of the noun *student* and the plural suffix -*s*. However, we would not want to consider these complex words 'new' in the same sense as we would consider *kicker* a new word derived from the verb *kick*. Here the distinction between word-form and lexeme is again useful. We would want to say that suffixes like participial -*ing*, plural -*s*, or third-person-singular -*s* create new word-forms, i.e. grammatical words, but they do not create new lexemes. In contrast, suffixes like -*er* and -*ee* (both attached to verbs, as in *kicker* and *employee*), or prefixes like *re-* or *un-* (as in *rephrase* or *unconvincing*) do form new lexemes. On the basis of this criterion (i.e. lexeme formation), a distinction has traditionally been made between **inflection** (i.e. conjugation and declension in traditional grammar) as part of the grammar on the one hand, and **derivation** and compounding as part of word-formation (or rather: lexeme formation).

Let us have a look at the following data which show further characteristics by which the two classes of morphological processes, inflection vs. word-formation, can be distinguished. The derivational processes are on the left, the inflectional ones on the right.

(14)   a. **derivation**   b. **inflection**
            work*er*   (she) work*s*
            use*less*   (the) worker*s*
            *un*truth*fulness*   (is) colonializ*ing*
            *inter*view   (we) pick*ed*
            curios*ity*   (the) child*ren*
            passiv*ize*   John*'s* (house)
            terror*ism*   Emily*'s* (job)

As already indicated above, the most crucial difference is that inflectional morphemes encode grammatical categories such as plural (*workers*), person (*works*), tense (*picked*), or case (*John's*). These categories are relevant for the building of

sentences and are referred to by the grammar. For example, there is a grammatical rule in English that demands that a third-person-singular subject is followed by a verb that is also marked as third-person-singular. This is called subject-verb agreement, which is also relevant for plural marking in sentences (*The flowers are/*is wonderful*). The plural and person suffixes are therefore syntactically relevant, hence inflectional.

One might argue that the suffix *-er* in *worker* is also syntactically relevant, in the sense that it is important for the syntax whether a word is a noun or a verb. That is of course true, but only in a very limited way. Thus, it is not relevant for the syntax whether the noun ends in *-er*, *-ee*, *-ion*, or whether the noun is morphologically complex at all. In that sense, derivational suffixes are not relevant for the syntax.

Let us turn to the next set of properties that unites the words on the left and differentiates them from the words on the right. These properties concern the position of the morphemes: in English derivational morphemes can occur at either end of the base words whereas regular inflection is always expressed by suffixes. Only irregular inflection makes use of non-affixational means, as for example in *mouse–mice* or *sing–sang*. There is no inflectional prefix in English. Furthermore, forms like *workers* or *colonializing* indicate that inflectional morphemes always occur outside derivational morphemes, they close the word for further (derivational) affixation (**workers-hood*, **colonializing-er*). As evidenced by derivatives like *un-truth-ful-ness* or the famous textbook example *dis-establish-ment-arian-ism*, derivational suffixes can and do occur inside other derivational suffixes.

Another interesting difference between the words in (14a) and (14b) concerns the part of speech. The suffixes in (14a) change the part of speech of the base word. For instance, the suffixation of *-less* makes an adjective out of a noun, the suffix *-ity* makes a noun out of an adjective, and the suffix *-ize* turns an adjective into a verb. The inflectional suffixes don't change the category of the base word. A plural marker on a noun does not change the category, nor does the past-tense marker on the verb. However, not all derivational affixes are category-changing, as is evidenced, for example, by most prefixes (as e.g. in *post-war*, *decolonialize*, *non-issue*), or by the nominal suffix *-ism*, which can attach to nouns to form nouns (e.g. *terrorism*).

The final property of derivation to be discussed here is exemplified by the two derivatives *interview* and *curiosity* in (14a), as against all inflectional forms. Both forms in (14a) show a property which is often found in derivation, but hardly ever in inflection, and which is called **semantic opacity**. If you consider the meaning of *interview* and the meaning of the ingredient morphemes *inter-* and *view*, you can observe that the meaning of *interview* is not the sum of the meaning of its parts. The meaning of *inter-* can be paraphrased as 'between,' that of (the verb) *view* as 'look at something' (definitions according to the *Longman Dictionary of Contemporary*

English), whereas the meaning of (the verb) *interview* is 'to ask someone questions, especially in a formal meeting.' Thus the meaning of the derived word cannot be inferred on the basis of its constituent morphemes; it is to some extent **opaque**, or **non-transparent**. The same holds for *curiosity*, a noun that has two related meanings: it can refer to a personal attribute 'the desire to know or learn about anything,' which is transparent, but it can also mean 'object of interest' (cf., for example, the definitions given in the *OED*), which is certainly less transparent. Non-transparent formations are quite common in derivational morphology, but rare in inflection.

Closely related to this generalization is the fact that inflectional categories tend to be fully **productive**, whereas derivational categories often show strong restrictions as to the kinds of possible combinations. What does 'fully productive' mean? A productive morpheme is one that can be attached regularly to any word of the appropriate class. For example, the morpheme expressing past tense can occur on all regular main verbs. And the morpheme expressing plural on nouns can be said to be fully productive, too, because all count nouns can take plural endings in English (some of these endings are irregular, as in *ox-en*, but the fact remains that plural morphology as such is fully productive). Note that the 'appropriate class' here is the class of count nouns; non-count nouns (such as *rice* and *milk*) regularly do not take plural. In contrast to the inflectional verbal and nominal endings just mentioned, not all verbs take the adjectival suffix *-ive*, nor do all count nouns take, say, the adjectival suffix *-al*:

(15)   a. *walk-ive   exhaust → exhaustive
            *read-ive   operate → operative
            *surprise-ive   assault → assaultive
         b. *computer-al   colony → colonial
            *desk-al   department → departmental
            *child-al   phrase → phrasal

The nature of the restrictions that are responsible for the impossibility of the asterisked examples in (15) (and in derivational morphology in general) are not always clear, but are often a complex mixture of phonological, morphological, and semantic mechanisms. The point is that, no matter what these restrictions in derivational morphology turn out to be, inflectional domains usually lack such complex restrictions.
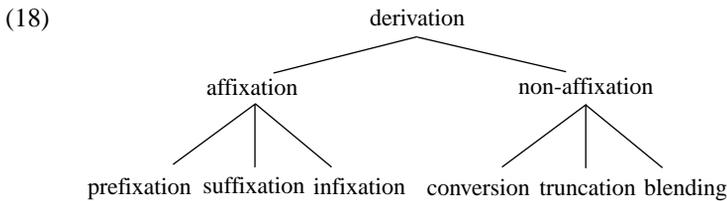
As a conclusion to our discussion of derivation and inflection, I have summarized the differences between inflection and derivation in (16). Exercise 1.6 below focuses on the problems in the application of these criteria and on the general nature of the dichotomy of inflection and derivation:

(16) **derivation**   **inflection**
　　　– encodes lexical meaning   – encodes grammatical categories
　　　– is not syntactically relevant   – is syntactically relevant
　　　– can occur inside derivation   – occurs outside all derivation
　　　– often changes the part of speech   – does not change part of speech
　　　– is often semantically opaque   – is rarely semantically opaque
　　　– is often restricted in its productivity – is fully productive
　　　– is not restricted to suffixation   – always suffixational (in English)

Based on these considerations we can conclude this sub-section by schematically conceptualizing the realm of morphology, as described so far:

(17)　　　morphology
　　　　　╱　　＼
　　inflection　word-formation
　　　　　　　　╱　＼
　　　　derivation　compounding
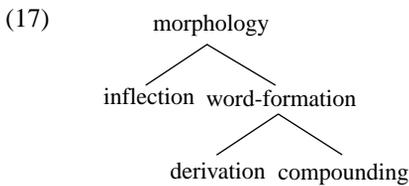
The formal means employed in derivational morphology and discussed so far can be classified in the following way:

(18)　　　　　　　　　　derivation
　　　　　　　　╱　　　　　　＼
　　　　affixation　　　　　non-affixation
　　　╱　｜　＼　　　　　　╱　｜　＼
prefixation suffixation infixation conversion truncation blending

1.4   Summary

In this chapter we have looked at some fundamental properties of words and the notion of 'word' itself. We have seen that words can be composed of smaller units, called morphemes, and that there are many different ways to create new words from existing ones by affixational, non-affixational and compounding processes. Furthermore, it became clear that there are remarkable differences between different types of morphological processes, which has led us to postulate the distinction between inflection and word-formation.

We are now equipped with the most basic notions necessary for the study of complex words, and can turn to the investigation of more (and more complicated) data in order to gain a deeper understanding of these notions. This will be done in the next chapter.

Further reading

Introductions to the basics of morphological analysis can also be found in other textbooks, such as the more elementary Bauer (1983, 1988), Katamba (1993), and Haspelmath (2002), and the more advanced Matthews (1991), Spencer (1991), and Carstairs-McCarthy (1992). All of these contain useful discussions of the notion of 'word' and introduce basic terminology needed for the study of word-formation. There are also two handbooks of morphology available, which contain useful state-of-the-art articles on all aspects of word-formation: Spencer and Zwicky (1998) and Booij et al. (2000).

Those interested in a more detailed treatment of the distinction between inflection and derivation can consult the following primary sources: Bybee (1985, ch. 4), Booij (1993), Haspelmath (1996). Note that these are not specifically written for beginners and as a novice you may find them harder to understand (this also holds for some of the articles in the above-mentioned handbooks).

Exercises

Basic level

Exercise 1.1

Explain the notions of grammatical word, orthographic word, word-form and lexeme. Use the italicized words in the following examples to show the differences between these notions:

a.   Franky *walked* to Hollywood every morning.
b.   You'll never *walk* alone.
c.   Patricia had a new *walking stick*.

Exercise 1.2

Define the following terms and give three examples illustrating each term:

morpheme   prefix   suffix   affix
compound   root   truncation

Exercise 1.3

Identify the individual morphemes in the words given below and determine whether they are free or bound morphemes, suffixes, prefixes or roots:

computerize   bathroom   numerous
unthinkable   intersperse   actors

Exercise 1.4

Consider the following sentence:

Textbook writers are sometimes grateful for comments and scholarly advice.

a. List all morphemes in this sentence. How many morphemes can you detect?
b. List all complex words and state for each of them which type of morphological process (inflection, derivation, or compounding) it is an example of.

Advanced level

Exercise 1.5

Consider again the notions of orthographic word, grammatical word, and lexeme as possible definitions of 'word.' Apply each of these notions to the words occurring in the following example and show how many words can be discerned on the basis of a given definition of 'word':

My birthday party's cancelled because of my brother's illness.

How and why does your count vary according to which definition you apply? Discuss the problems involved.

Exercise 1.6

Consider the status of the adverbial suffix *-ly* in English. Systematically apply the criteria summarized in (16) in chapter 1 and discuss whether *-ly* should be considered an inflectional suffix or a derivational one. You may want to take the following data into account:

slowly   aggressively   smoothly   hardly
rarely   intelligently   purposefully