INTRODUCTION TO MORPHOLOGY

Rutgers and McGill

Draft – for class use

3. COMPOUNDING

3.1 The Definition of Compounds

So far, we have been looking at derivational morphology—the formation of new words by attaching affixes to roots. However, the grammars of many languages include a second way of forming new words, namely **compounding**. A compound can be defined as follows

DEFINITION: A *compound* is a word that contains more than one root.

Compounding has both similarities to derivation and differences. Compare, for example, the derived word *blackness* with the compound word *blackbird*. Both contain the adjective root *black*. Both words are nouns as a result of combining *black* with another morpheme with its own distinct meaning. The primary difference is that what is added in the second example—*bird*—is itself a root, capable of standing on its own as a complete word. It contrasts in this respect with *ness*, which is not a root, and cannot form the basis of a well-formed word. Thus, one can easily say *The bird flew over the fence*, but not **His ness impresses me*. For this reason, *blackbird* qualifies as a compound, and *blackness* does not.

Compounding is a very common and productive mode of word formation in English. Here are some representative examples that give a sense of the range of compounding found in English:

(1)

A	B	C	D
hubcap	greenhouse	overdose	swearword
sales manager	highschool	outhouse	rattlesnake
frogman	blackbird	underdog	pickpocket
bullfrog	blackboard	uprising	scarecrow

E	F	G	H
headstrong	icy-cold	overripe	intake
skin-deep	white-hot	underprivileged	overdo
jet-black	widespread	ingrown	uproot

Compounding is also found in many other languages. There are, however, differences in how important compounding is to the language as a whole. For example, compounding is the primary way of creating new words in Mandarin Chinese, a language that has very little true affixation. On the other hand, Greenlandic Inuktitut is very rich in derivational affixes of every kind, but it has essentially no compounding. We will also see that languages vary in the specific types of compounding that are found in them.

One might think that an equally good definition of compounding would be to say that a compound is a word that contains more than one word inside it. For most of the English words in (1), this alternative definition works just as well as our official one. For example, *bullfrog* is made up of *bull* and *frog*, each of which is a word in its own right (as well as a root). Similarly with the somewhat antiquated example *frogman* (i.e., a scuba diver). However, recall that there are two kinds of roots: bound and free. Free roots count as words by themselves, but bound roots do not. Nevertheless, compounds can be formed of bound roots also. For example, *scissor* is a bound root in English. It is not used as a word by itself, but it is an easily-identifiable part (together with the plural suffix *–s*) of the word *scissors*. It can also occur without -s, as the first constituent in a compound:

(2) scissor-kick (a component of the breast-stroke)

scissor-blade (half of a scissors)

scissor-hands (a character in a movie with scissors for hands)

Consider also the following examples:

(3) a. blúeberry góoseberry

b. cránberry húckleberry

c. erythrocyte chlorophyll

The words in (3b) look like normal English compounds, parallel in every respect to the examples in (3a). However, the first part of the examples in (3b) cannot be used as words in their own right, nor are they prefixes/affixes. On the contrary, they can occur only in these compounds. (A morpheme that is found in only one or two compounds is sometimes called a "cranberry morpheme" in honor of this example.) Words like those in (3c) (typically from specialized scientific language) also involve two elements that are unable to stand on their own. These are coined semi-productively from Latin or Greek morphemes which are not otherwise used in ordinary English.

The contrast between root compounding and word compounding is more striking in languages that are more richly inflected, where bound roots are the rule, rather than the exception. For example, verbs in German never occur without a suffix of some sort. Thus, the verb *treffen* 'to meet' has two morphemes; it consists of the root *treff-* and the "infinitive suffix" *-en*. As a finite verb, it may occur with other (inflectional) suffixes, *ich treff-e* 'I meet'; *ihr treff-t* 'You-pl meet'. But **treff* alone is not a word. Now consider the compounds in (4):

(4) *treffen* ‘to meet’ *Punkt* ‘point’ *Treffpunkt* ‘Meeting point’

essen ‘to eat’ *Lokal* ‘place’ *Esslokal* ‘Eating place’

brechen ‘to break’ *Hammer* ‘hammer’ *Brechhammer* ‘A hammer for breaking things’

but **treff* **ess* and **brech*

These are verb-plus-noun compounds, in which the bound root of the verb is combined with a noun to form a new word, a noun in each case. Similarly, all noun roots in Greek are bound; they must take a suffix that indicates the **number** of the noun (singular or plural) and its **case** (i.e., how it is used in a sentence; see section x.x).

(5)

lik-os	wolf	(subject form)
lik-oi	wolves	(subject form)
lik-o	wolf	(object form)
lik-ous	wolves	(object form)
lik-ou	wolf’s	(possessive form)
lik-on	wolves’	(possessive form)

Clearly the morpheme that means ‘wolf’ is *lik*, since that is the sequence of sounds that all of these words share. (The *o* is a theme vowel, part of the suffix.) However **lik* on its own is unacceptable to a Greek speaker. It can be used in a compound, however:

(6) likanthropos ‘werewolf’

from: *lik* ‘wolf’ + *anthrop* ‘person’ + *os* ‘singular subject form’

Finally, in Mohawk all well-formed nouns and verbs must contain both prefixes and suffixes of various kinds:

(7) a. ka-*natar*-u ‘bread’ (noun)

b. ye-*kwetar*-us ‘she cuts’ (verb)

Mohawk also has many Noun + Verb compounds. These are formed by combining the root of the noun and the root of the verb. Then the same prefixes and suffixes are attached to the resulting unit as are added onto a simple verb. This results in complex words like (8):

(8) ye-*natar*-a-*kwetar*-us ‘she cuts bread’, literally ‘she bread-cuts’

(The inflectional affixes associated with nouns are not used at all in this formation, although there is a special ‘linking vowel’ *a* between the two roots, that we will ignore here). In this case, there is no subpart of the word that counts as a valid word on its own, yet it is a well-formed

compound. These examples show that defining compounds as words that contain two roots is much more valid across languages than defining them as words that contain two words. There are languages in which compounds are constructed from two or more fully formed words,, but of course all such examples have two or more roots as well (since every word contains a root). Therefore the definition of compounding is general enough to include this case as well.

Practice Problem C-1. For each of the following English words, is it best analysed as a compound or an affixed word? Explain in each case by comparing to other common English words. (Do not feel constrained by what we have taken for granted elsewhere in this book for purposes of this problem. Having the definition of a morpheme clearly in mind will help.)

(i) breakable

(ii) hopeful

(iii) hopeless

(iv) counterattack

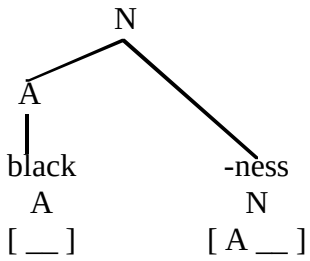
(v) outlive

3.1 Word Structure Trees for Compounds

Now that we know what a compound is, we can consider what it would take to extend the theory of word formation that we introduced in the last chapter so that it includes compounds as well. The first step in doing this is to construct a word structure tree for compounds. What should such a word structure tree look like?

To see how to do this, let us compare once again the derived word *blackness* with the compound *blackbird*. We know that the word structure tree for *blackness* looks like (9).

(9)

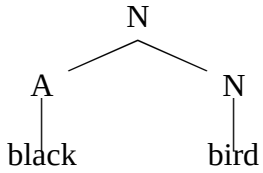


This tree is asymmetrical. The affix *–ness* is **immediately dominated** by the topmost node of the tree, whereas the root *black* is only indirectly dominated by that node. In between the root *black* and the top node there is an intermediate node, labelled A, that dominates only the root. The word structure tree is drawn in this way to respect the selectional restrictions of the morphemes included in the tree. Thus, *-ness* is a suffix that attaches to adjectives. This is respected in (9) because the left-hand sister of *–ness* under the dominating node N is a node

labelled “A”. On the other hand, *black* is not an affix but a root. Although morphemes may attach to it, it in itself does not need to attach to anything. Its selectional restrictions are vacuous. This is represented in (9) by the fact that the node immediately above *black* contains nothing else, neither to its right nor to its left—a sign of the relative morphological independence of this morpheme. In this way, the distinction between affixes and roots is built graphically into the word structure tree, as was discussed in Chapter D.

Now within this way of doing things, what should the word structure tree for a compound like *blackbird* look like? The A node above *black* should still be there, since *black* is an adjectival root with no selectional restrictions in this example too. The top N node should be there too, because the whole word is a noun. The only difference is that the second morpheme, *bird*, is a root rather than an affix. Like *black*, it has no specific selectional restrictions. Therefore, to be consistent, it too should have a node directly above it that dominates nothing else, either on the left or the right, to show that *bird* does not need to attach to anything. In other words, the word structure tree for a compound should look like (10):

(10)



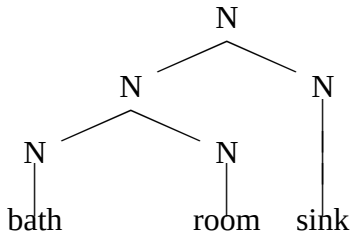
This is a symmetrical structure, with both morphemes treated the same way, which seems natural for a compound. Note again that each root in these trees is immediately dominated by a non-branching node, the convention we established in the previous chapter.

This basic tree structure will also be valid for all the other kinds of compounds show in (1) in English. More precisely, the shape of the word structure tree will always be the same, and the only thing that will change is the labels of the nodes in that tree to match the lexical properties of the particular roots involved. Here are two more compound word structure trees:

(11)

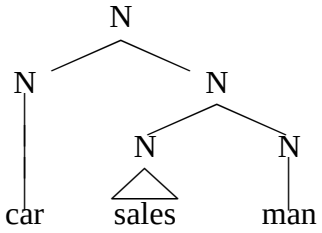
While the basic word structure tree for compounds is always the same, more complex structures can arise when a word contains more than two morphemes. For example, one can start with a compound noun like *bathroom* and compound it with another noun like *sink* to get the compound *bathroom sink*. This would have the following word structure tree:

(12)



Car salesman is also a compound noun made up of three noun roots, but the arrangement is a bit different. This time the simple noun the first part of the compound and the compound noun is the second part: a *car salesman* is most naturally *car* + *salesman*, not *car-sales* + *man*. Hence the word structure tree will look like (13):

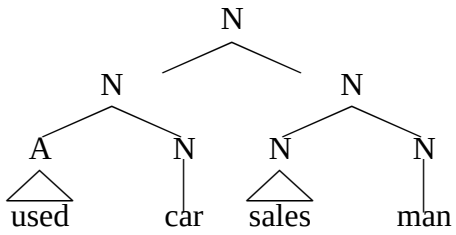
(13)



Note that we have abbreviated the tree somewhat, in particular, by ignoring the internal structure of the word *sales*. When we make an abbreviation, by convention, we draw a small triangle over the string whose internal structure we are not indicating. Since *sales* is not a root (it is a combination of the root *sale* plus the plural suffix *-s*), it would be technically incorrect to connect this to the node above it with a single straight line. A non-branching node is by definition a root, and a root is, by definition, not internally complex.

Both the first part of a compound and the second part of a compound can themselves be compounds. For example, a *used car salesman* is a man involved with the sales of cars that have been used – a four-part compound with the following structure:

(14)



And so on. Indeed, one of the striking properties of Germanic languages (as compared to Romance languages like French and Spanish, for example) is how productive noun-noun compounding is. This is especially striking in German, where the standard orthography consistently writes compounds as a single word, with no spaces or hyphens. English writing, on the other hand, is notoriously inconsistent on this point: some familiar compounds are written all together as one word, the parts are separated with a hyphen or a space in others. Thus, in

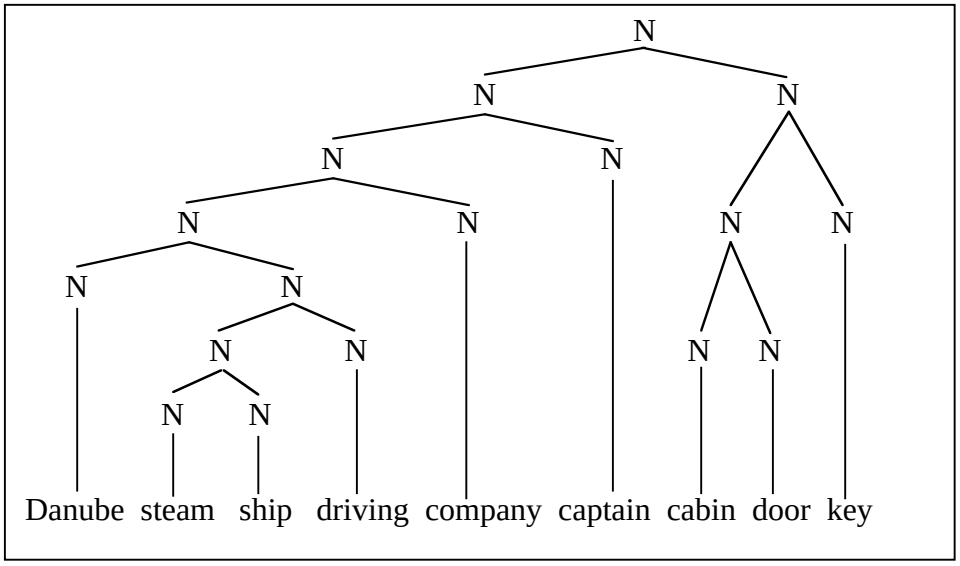
German a *typewriter salesman* is a *Schreibmaschinenvertreter*, written all as one word. (The major parts are *schreib* 'write', *machinen* 'machines' and *vertreter* 'salesman, vendor'—although these have internal complexity—thus it has essentially the same structure as the three-way compound in (12).) Or consider the still more impressive example in (15):

(14) Der Donaudampfschifffahrtsgesellschaftskapitänskajütentürschlüssel

Danube-steam-ship-driving-company-captain-cabin-door-key

This is a nine-root compound, containing the following nouns: *Donau* 'Danube', *Dampf* 'steam', *Schiff* 'ship', *Fahrt* 'driving', *Gesellschaft* 'company', *Kapitän* 'captain', *Kajüte* 'cabin', *Tür* 'door', and *Schlüssel* 'key'. It refers (of course) to the key that opens the door of the cabin belonging to the captain who works of a company that drives steam ships on the Danube. Simplifying the internal structure of some of these nouns, the best word structure tree for this example is:

(15)



Practice Problem C-2. We saw that there are two ways of making compounds out of three noun roots in English. How many different ways should there be of making a compound that contains four noun roots? Draw the word structure tree for each, and try to find or make an example of an English word that has that structure.

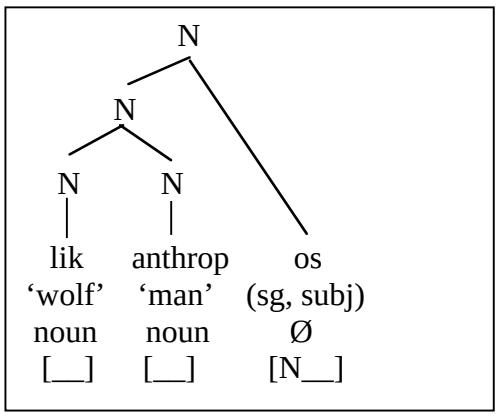
Practice Problem C-3. What is the longest compound in English that you can find in a printed source or that you can make up and have someone else understand? What do you think its word structure tree would be and why?

end of problem

It is also possible to have complex words that containing both affixation and compounding. Indeed, we have already seen a few examples of this in section 1, from

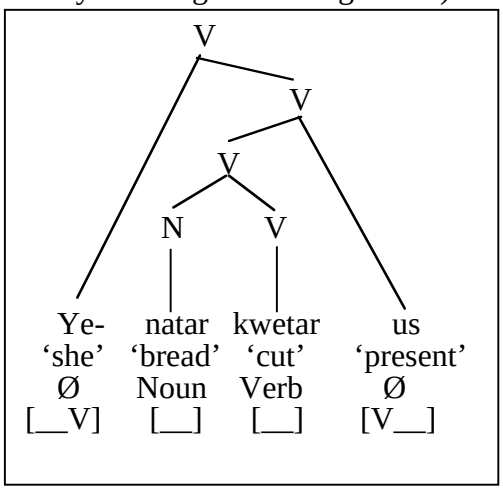
languages like Greek and Mohawk, in which all roots are bound roots. This means that one or more affixes must attach to them before they can be used in sentences. The same affixes have to attach to compounds, in general. Thus, *likanthropos* 'werewolf' in Greek is a compound formed by two noun roots, with a singular nominative case affix attached at the end. Its word structure tree would look like this:

(xx)



Presumably English world like *bathrooms* and *icy-coldness* have the same structure, with the plural suffix *–s* attaching to the compound *bathroom* and the noun forming affix *-ness* attaching to the adjectival compound *icy-cold*. The Mohawk example *yenatarakwetarus* 'she bread-cuts' is similar, except that it has a prefix that attaches to the compound as well as a suffix (we have simplified by omitting the linking vowel).

(xx)

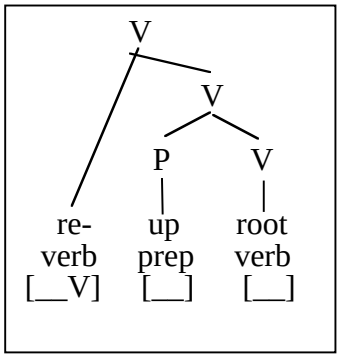


There is no way of telling whether the tense suffix or the feminine subject prefix attaches first without more information, because both affixes attach to verbs. However, it is very clear that the prefix *ye-* (at least) attaches after the noun root *natar* 'bread' and the verb root *kwetar* 'cut' have been compounded. *Ye-* never attaches to nouns in Mohawk; its selectional restrictions stipulate that it must prefix to a verb stem. Therefore, it cannot attach directly to *natar* 'bread', but only to whole compound *natarakwetar* 'cut bread', which has the feature V. (In contrast, it could be debated whether the compounding or the affixation happens first in (xx); can you see why?) A similar example in English is the coined but easily understandable word *reuproot*. This word is perfectly understandable in a sentence like the following:

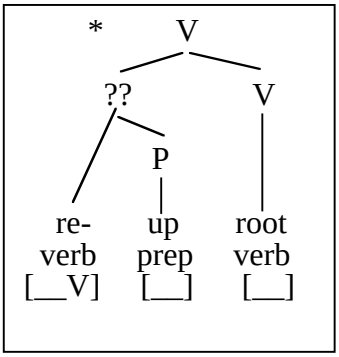
(xx) A new outbreak of fighting **reuprooted** the refugees just as soon as they had been resettled on their land.

Re- is a prefix that attaches to verbs, never to prepositions. Therefore the word structure for *re-uproot* must be (xxa), not (xxb).

(xx) a.

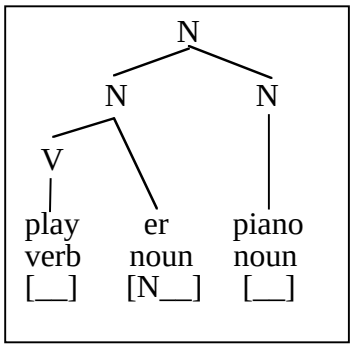


b.

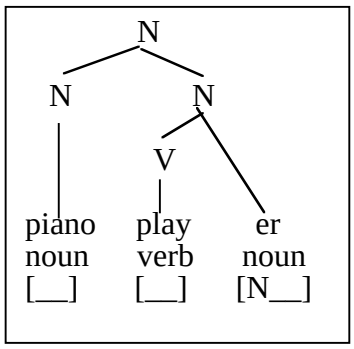


It is perfectly possible for affixation to take place prior to compounding, however. For example, a piano that was mechanically rigged to play songs without anyone touching the keys was a *player piano*. Here the first part of the compound contains a derivational affix *–er*. Conversely, a person who plays pianos is a *piano player*; in this case it is the second part of the compound that contains the affix. The word structure trees for these examples would be as follows (compare also the example *used car salesman* in which both *used* and *sales* have affixation feeding compounding):

(xx) a.



b.



Thus, we have seen how to draw the word structure tree for basic compounding examples, and how to extend this notion of word structure to include complex iterated compounds and combinations of compounding and affixation. Compounding and affixation can generally apply in any order, so long as the selectional restrictions of the morphemes involved are respected. The resulting word structure trees can be quite complex. However, for any node in such a tree, it is easy to tell whether it represents a case of compounding or a case of affixation. A branching node with a morpheme directly under it always stands for an affixing step (the affix is the item directly under it), whereas a node with two other nodes directly under it stands for an instance of compounding.

Practice Exercise C-4. For one of the words in (xx), there is a second word structure tree that is consistent with the selectional restrictions of the morphemes. Which word is it, and what would the alternative word structure tree look like? Even though two word structure trees are possible

in this case, can you see a reason for preferring the version given in (xx)? Explain. (A careful look at (1) might help. We will return to this topic below.)

3.1 Labelling word structure trees with compounding.

One striking fact about compounding is that native speakers often know how to use and understand a newly formed compound immediately, without any learning process or experimentation. For example, even if you have never heard the word *turkey-stranglers* before, you understand it and recognize it as being a noun. Thus, (xxa) will strike you as a well-formed sentence of English, because *turkey-stranglers* is used in a position typical of nouns, whereas (xxb) and (xxc) sound unacceptable (they attempt to use *turkey-stranglers* as a verb or an adjective, respectively).

(xx) a. The turkey-stranglers were arrested last night at 10:00pm.

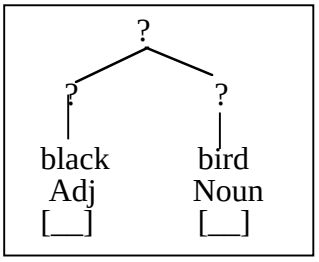
b. *Chris turkey-stranglers for us on every major holiday.

c. *You seem very turkey-stranglers today, by the look in your eyes.

This is a sign that there must be general principles by which the properties of a compound can be calculated from the properties of its parts.

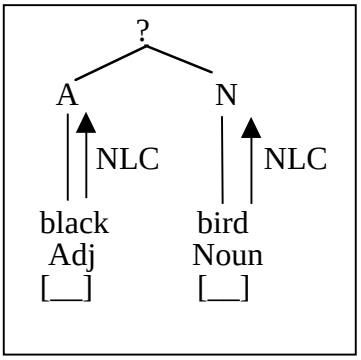
In this respect also, compounding is similar to derivational morphology but not identical to it. For derived words, too, native speakers often have intuitions about the syntactic category even if it is one that they have not heard before, as long as they know the morphemes it is made of. Our account of this depended on two conventions—the Node Labelling Convention and the Feature Percolation Convention—which regulate how the properties of the individual morphemes are combined to give the properties of the word as a whole. Something similar must happen in compounds. However, a closer look at the word structure trees of a compound show that it cannot be exactly the same. Consider again the word structure tree for a simple compound like *blackbird*, but now without the node labels that we have taken for granted so far:

(xx)



The first step of labelling this tree is easy: the features of each root are passed up to the nonbranching nodes immediately above them—an ordinary use of the Node Labelling Convention.

(xx)



The crucial question that is distinctive to compounding is how to label the topmost node. The Node Labelling Convention does not apply here. This convention gives priority to the lexical node immediately dominated by the node under consideration. But for the node marked ? in (xx), there is no such lexical node. While this is never the case in affixation, this will always be the case in a compound structure. Therefore, node ? gets no features from the Node Labelling Convention. Given this, we might expect the Percolation Convention to step in, as it does for words that contain category-less affixes. It can apply in principle, but it does not resolve the questions about how the top node will be labelled. The PC could place features from either or *both*r of the nonbranching nodes on the topmost node. But this is wrong. *Blackbird* is not some new kind of blended word that is simultaneously a noun and adjective. Nor is it an indeterminate construction, which can be used as a noun or an adjective, according to the whims of the speaker. Rather, it is a perfectly ordinary noun, as shown by the examples in (xx).

(xx) a. A *blackbird* is singing in the reeds.

b. *That creature seems very *blackbird* from here.

Apparently, we need a third convention in order to determine a well-defined label for the top node of the symmetrical structures found in compounding.

The gist of the new convention that breaks this deadlock is fairly obvious from the examples that we have seen so far. Apparently, it is the node on the right hand side that determines the category of the compound in English:

OBSERVATION: The category of the whole is (almost) always that of the right hand member of the compound.

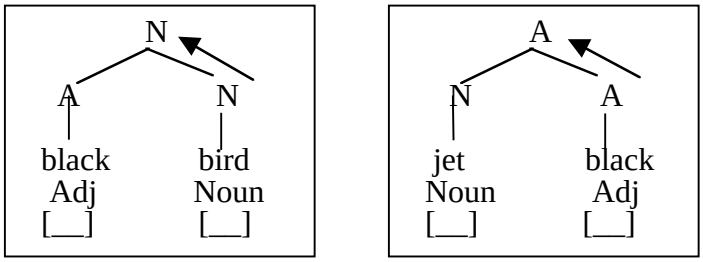
A nice pair that illustrates this is to compare *blackbird* with *jet-black*. Both are compounds containing a noun and the adjective *black*. Yet, *jet-black* is an adjective, whereas *blackbird* is a noun:

(xx) a. A jet-black cat crossed in front of me.

b. *A blackbird cat crossed in front of me.

The word structures trees look like this:

(xx)



Clearly, it is the position of the adjective *black* inside the compound that makes all the difference as to whether its features determine the features of the larger word or not.

This generalization is valid over an impressive range in English. In section C.1 we presented a longish list of the kinds of compounds found in English, to give a sense of the range of the phenomenon. We repeat that list here:

(xx)

A	B	C	D
hubcap	greenhouse	overdose	swearword
sales manager	highschool	outhouse	rattlesnake
frogman	blackbird	underdog	pickpocket
bullfrog	blackboard	uprising	scarecrow

E	F	G	H
headstrong	icy-cold	overripe	intake
skin-deep	white-hot	underprivileged	overdo
jet-black	widespread	ingrown	uproot

The list contains words that are nouns (A-D), adjectives (E-G), and verbs (H). Inside the words are noun roots (A-E), adjective roots (B, E-G), verb roots (D, H), and prepositions (C, G, H). However, in every case it is the second part of the compound that determines the category of the compound as a whole. This is shown for some illustrative examples in (xx):

(xx)

N + N = N	hub + cap = hubcap	N + A = A	skin + deep = skin-deep
A + N = N	green + house = greenhouse	P + A = A	over + ripe = over-ripe
P + N = N	over + dose = overdose	A + A = A	white + hot = white-hot
V + N = N	swear + word = swearword	P + V = V	over + do = overdo

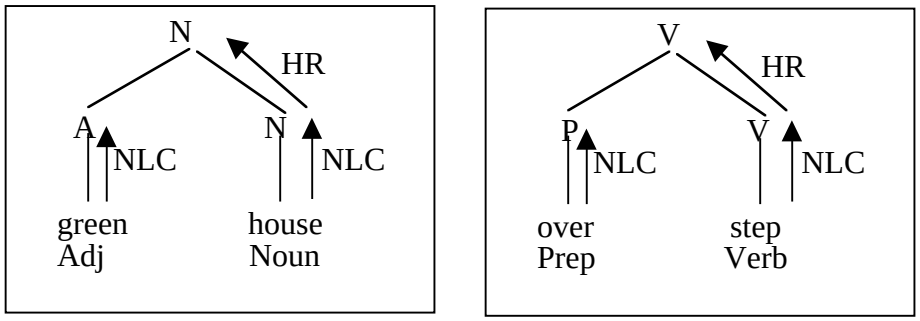
Thus, there is a systematic asymmetry in compounds, but it is an asymmetry of linear order, not of tree shape. As far as syntactic category is concerned, the righthand member of the compound is *stronger* or *privileged*. Thus, we can refer to it as the *head* of the compound, and state a third labelling convention as follows.

(1) The (Right-Hand) Head Rule: [HR]

The features of the **right**-hand member of a compound determine the features of the whole. [English, German, …]

Here are two more examples of the Head Rule at word in fully annotated word structure trees:

(xx)



It is easy to see that all of the fancier word structure trees that we saw in the previous section also obey this rule.

The Right-Hand Head Rule is valid not only for English, but also for many other languages, including German, Greek, and Mohawk. For example, we saw the following examples of verb-plus-noun compounding in German; note that the result is consistently a noun.

(4) *treffen* 'to meet' + *Punkt* 'point' = *Treffpunkt* 'Meeting point'

essen 'to eat' + *Lokal* 'place' = *Esslokal* 'Eating place'

brechen 'to break' + *Hammer* 'hammer' = *Brechhammer* 'A hammer for breaking things'

On the other hand, *ye-na'tar-a-kwetar-us* is a compound of *na'tar* 'bread' (a noun) and *kwetar* 'cut' (a verb) in Mohawk; the result of this combination is a verb. The second member of the compound determines the category of the whole in these examples as well.

The Right-Hand Head Rule is not universal, however. In a sense, the rule is arbitrary. Even if it is convenient for a language to have a consistent rule for compounding, there is nothing sacred about the right hand side. One can easily imagine languages working just as well by consistently picking the first member of a compound to be the head that determines the features of the word as a whole. And indeed there are languages that work this way. One is Vietnamese, as shown in the following examples (tones are not indicated).

(xx) nha = ‘establishment [N] + thuong ‘be wounded’ [V]= nha-thuong ‘hospital’ [N]

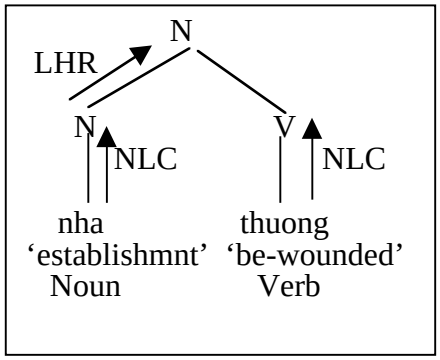
nguoti = ‘person’ [N] + o' ‘be located’ = nguoti-o’ ‘servant’ (N)

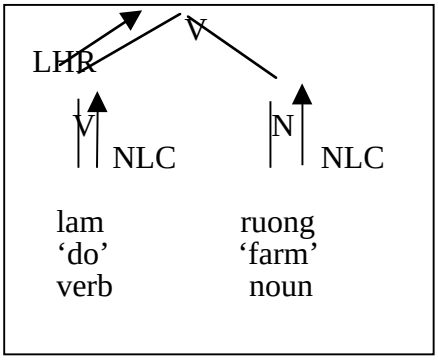
lam = ‘do’ [V] + viec ‘matter, thing’ [N] = lam-viec ‘work’ [V]

lam = ‘do’ [V] + ruong ‘rice-field’ [N] = lam-ruong ‘(to) farm’ [V]

Here if the first part of the compound is a noun, the compound is a noun; if the first part is a verb, the whole is a verb. The word structure trees would look like this:

(xx)





Hebrew also has left headed compounding involving nouns and adjectives as shown in (xx).

(xx) yom = ‘day’ [N] + tov ‘good’ [Adj] = yom-tov ‘holiday’ [N]

kesef = ‘money’ [N] + katan ‘small’ [Adj] = kesef-katan ‘change’ [N]

tov = ‘good’ [Adj] + lev ‘heart’ [N] = tov-lev ‘kind’ [Adj]

raze = ‘thin’ [Adj] + panim ‘face’ [N] = raze-panim ‘thin-faced’ [Adj]

kaxol = ‘blue’ [Adj] + kehe ‘dark’ [Adj] = kaxol-kehe ‘dark-blue’ [Adj]

These Hebrew words are the exact reverse of what one has in English: Hebrew has ‘day-good’, and English has ‘*holiday*’ (=holy+day); Hebrew has ‘blue-dark’ and English has ‘*dark-blue*’. (Notice that we have written the Hebrew words in the Roman alphabet. In traditional Hebrew orthography, one writes from right to left. This fact confuses what we mean by “right side” and “left side” in principles like the Head Rule. By convention, linguists always write transcriptions from left to right, so “left” is equivalent to “first” and “right” is equivalent to “last” in linguistic principles.) This alternative form of the Head Rule is stated in the obvious way, as follows.

(2) The Left-Hand Head Rule: [LHR]

The features of the **left**-hand member of a compound determine the features of the whole. [Vietnamese, Hebrew,..]

OPTIONAL SECTION: It is natural to wonder whether the difference between languages with right-headed compounds and languages with left-headed compounds is a basic choice that languages make, or whether it is related to other known differences between the languages in question. The most likely thing it could be related to is word order in the syntax.

The idea of a head is important in syntax as well. The verb, for example is considered to be the head of the verb phrase. In English, the verb comes to the left inside the verb phrase, whereas in languages like German the verb comes to the right inside the verb phrase.

(xx) a. Fido will [$_{VP}$ bite the Mailman]. [$_{VP}$ V Object]

b. Bello wird [$_{VP}$ den Briefträger beißen]. [$_{VP}$ Object V]
<dog> will the mailman bite

This syntactic difference is reminiscent of the difference we have been studying with respect to compounding. The question then is whether the two are related.

There are some cases that show immediate promise for the idea that they are. Two of the most common types of compounds in English are N+N compounds and A+N compounds, where the first noun or adjective acts as a kind of modifier of the head noun:

(xx) a. the doghouse

b. a blackbird

Adjectives can also function as modifiers of nouns in the syntax, and when they do they come immediately before the noun. Similarly, a possessive noun phrase can come just before the head noun in English. Thus, there is an obvious parallelism between the syntactic constructions in (xx) and the compounds in (xx).

(xx) a. the dog's house.

b. a black bird

Edo is an African language, spoken in Nigeria. Unlike English, both adjectives and possessor nouns follow the head noun in syntactic constructions.

(xx) a. ágá pẹ̀rhẹ̀
chair flat
'a flat chair'

b. ágá Ózó
chair Ozo
'Ozo's chair'

Edo also has productive N+N compounding (adjectives are more rare), and the head of the compound that defines its essential meaning is always on the left hand side.

(xx) a. èkón!khókhò from ekèn + òkhókhò
'chicken egg' egg chicken

b. ìkpákpé!ní from ìkpákpà + èní
'elephant hide' 'hide' 'elephant'

c. ékpétió!tá from ékpétìn + òtá
'radio' 'box' 'speech'

The order in the compounds in (xx) is the reverse of the normal English order, just as the order in the phrases in (xx) is the opposite of the English order. Examples like these are reasonably common, and lend credence to the idea that there is a single head rule that is valid for both the syntax of a language and the morphology.

However, there are problems for this unified approach as well. Indeed, the most notorious problem is staring us in the face. Recall that verbs consistently come before objects in the syntax of English. The same is true for Edo.

(xx) a. John [drives trucks (for a living)].

b. Òzó lé èvbàré
Ozo cooked food

So we would expect verbs to come before nouns in compounds in these languages. Now compounds made from nouns and verbs are not terribly productive in either language. However, one can make nouns from verbs in both languages: in English, this is done by adding the suffix –ing or -er; the Edo equivalent is the prefix u- and the suffix –mwen.

(xx) a. driving

b. driver

c. ù-lé-mwèn. 'cooking'

The object of the verb can still appear with the noun form, although now it needs to be marked with the preposition of. (This will be an important topic in the next chapter.) The word order is the same as in verb phrases, with the derived noun coming before its "object".

(xx) a. [The driving of trucks] becomes tedious after a while.

b. [A driver of trucks] must get a special license.

b. ùlémwèn óghé èvbàré ….
cooking of food

Since noun-noun compounding is very free in both languages, we are not surprised that these –ing/u- -mwen nouns can easily form compounds in both languages. Moreover, we would

expect that the deverbal noun should come before its "object" noun, mirroring the order in the syntactic construction. For Edo, this expectation is entirely correct:

(xx) a. ùdémwí!yán from ù-dé-mwèn + ìyán
'yam-buying' 'buying' 'yam'

b. òsúkpòn from ò + sé + ùkpòn
'tailor, clothes-sewer' er 'sew' 'clothes'

However, for English the expectation is entirely wrong. On the contrary, the "object" noun aways comes before the "head" noun in this kind of compound:

(xx) a. Truck-driving becomes tedious after a while.
b. Dish-washing is fun.
c. Cigar-smoking leads to throat cancer.
d. Truck-drivers/dish-washers/cigar-smokers often get throat cancer.

To the extent that nouns and verbs can compound to form verbs in English, we find the same reversal:

(xx)	a.	John tends (the) bar on Saturday nights for some extra cash.	Syntax
	b.	John bartends on Saturday nights …	Compound
	c.	*John tendbars on Saturday nights …	

So this is a sign that compounding order cannot always be identified with syntactic order in the simplest imaginable way.

However, the success stories for this account are enough that one should probably not give up on it too quickly. First of all, languages like Edo that fit the generalization are probably more common than languages like English that do not. Second, there is evidence from children learning language that there is something strange about English. Thus, small children who are first learning to construct compounds for themselves spontaneously make up compounds like 'driver-truck' or 'drive-truck-er' when they mean 'truck-driver'. (Occasionally examples like this are found even in adult language; thus, someone who does about doing good is a do-gooder (not a good-doer).) Only later do they consistently use the correct English pattern. This suggests that they believe that English should be like Edo in this respect, until they learn some further complication of the language.

One suggestion that has been proposed to explain why English seems to be exceptional in this respect points to a quirk in the history of English. It so happens that in Old English verbs did systematically come after their noun phrase objects, as is still the case in English's cousin languages Dutch and German. Compound orders like truck-driver and dish-washing were thus perfectly regular in Old English, as are similar compounds in Modern Dutch and Modern German (cf. Dutch *schoenmaker* 'shoemaker' from *schoen* 'shoes' and *maker* 'maker'). In the last thousand years the word order in English syntax has changed, but the older order is still preserved in compounds, according to this proposal.

While there may be an element of truth in this historical theory, it seems odd that morphology should lag so far behind syntax in this way. Most of us are quite oblivious to the fact that English used to have object-verb order, and this is certainly not a fact which children have access to in determining their grammars, yet compounds like truck-driver do not strike us as at all odd. Moreover, it is clear that *truck-driver* is not an archaic word left in the language since 800, the way *werewolf* is (*were* is an old root meaning 'man', which has long been lost). Rather, truck-driver has been coined in the last 100 years, in accordance with a still-robust pattern. A more promising way to explain this property of English might be to take a closer look at noun phrase syntax in English. Consider the following range of examples:

(xx) a. John [drives trucks]

b. [The driver of the truck] was pulled over for speeding.

c. [The truck's driver] was pulled over for speeding.

d. [The truck-driver] was pulled over for speeding.

While it is true that the object of the noun *driver* can follow it in English ((xxb)), it is also possible for the object to come first, as in (xxc). Whereas, (xxb) is parallel to the verbal construction in (xxa), (xxc) is parallel to normal possessor-plus-noun phrases in English. The compound in (xxd) could then be seen as analogous to the syntactic construction in (xxc), whose order it preserves, rather than to the construction in (xxb). Perhaps, then, the order found in these compounds in English is not so strange after all. Edo has no equivalent of the structure in (xxc), but only (xxb). Therefore, it uses that order also in compounding. In this way, we might be able to retain the idea that the order of words in compounds is predictable from the order of words in syntax after all. To pursue this idea in detail would require more knowledge of syntax than we assume here, but it looks like a promising avenue of research.

3.4 The Head Rule with Other Features

Much of the power and beauty of the Node Labelling Convention and the Backup Percolation convention came from the fact that they worked not only for category features like Noun, Verb, and Adjective, but also for many other kinds of features. Now we will see that the same is true for the Head Rule.

Consider first the number feature of nouns. In English, this feature has two values, singular or plural. This feature is important not only for aspects of meaning, but for matters of syntactic agreement. Thus, both demonstratives and verbs necessarily change their form depending on whether a noun used as the subject is singular or plural.

(xx) This dog is fierce

These dogs are fierce.

*These dog are fierce.

*This dogs is fierce.

In English compounds where the first member is a noun, that noun typically shows up in its bare (i.e., singular) form. Thus, we say *truck-driver* and *dish-washer*, not *trucks-driver* or *dishes-washer*, although the person referred quite probably washes dishes, not just a single dish.

There are, however, some nouns which appear in their plural form inside compounds. These come in two types. First, there are roots that are canonically used only in the plural (such as *clothes*) or which take on a specialized meaning in the plural, for example *arm*s with the meaning of *weapons* (if you sell one bazooka to a terrorist, you did not sell him *an arm*). Second, there are nouns that have a plural form which is not related to thhe singular by affixation. For example, *people* can be seen as a plural root, corresponding to the singular root *person*. (One can ask whether the same is true of pairs such as *goose-geese*; we return to these in later chapters). .

The following examples have a plural noun as the first member of the compound and a singular noun as the second member.

(xx) clothes store vs. shoe store *shoes store

Arms race vs. *arm race

people-watcher vs. bird-watcher, not *birds-watcher

Parks commissioner

Sales person

people person as in: “I’m a people person”

These compounds are grammatically singular:

(xx) This arms race has gone on too long.

Not: *These arms race have gone on too long.

On the other hand, the following compounds have a singular first part and a plural second part.

(xx) Canada geese

fieldmice

hill people

side arms

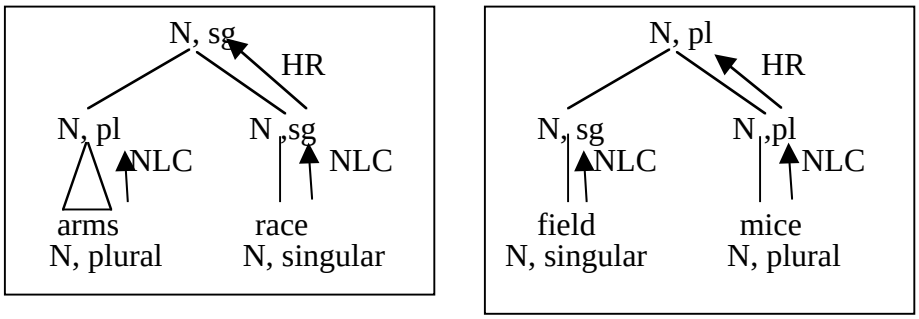
These are grammatically plural:

(xx) These fieldmice are a nuisance.

NOT: This fieldmice is a nuisance.

The fact that all these compounds are nouns is not remarkable: since both roots that make them up are nouns, what else could they be? However, the Right hand Head Rule applies in a nontrivial way to the number feature, ensuring that the number of the right hand root is the number of the whole word. Here are the Word Structure Trees:

(3)



Similarly, if we are going to make the plural of a compound made up of two singular noun roots, the Righthand Head Rule implies that we must attach the affix –s to the second noun (or to the compound as a whole). Attaching *–s* to the left-hand member of the compound will not do the trick:

(4) a. We have a lot of [beehive]-s. c. Look at all the bullfrog-s.

b. *We keep a lot of beeshive. d. *Look at all the bullsfrog.

Affixing *–s* to the noun root *bee* will make it plural, but the plural feature will not be passed up to the whole word *beeshive*. Rather, the singular feature of hive will win out, thanks to the head rule.

Another important feature associated with nouns in many languages is **gender**. This feature is marginal at best in English, but is prevalent in languages like Russian and German, as we saw in the previous chapter. Indeed, gender is essential to picking the proper determiner forms and adjective forms to go with a given noun in these languages, much as number is in English. How then do gender features behave in compounds? The answer is the Head Rule applies. (xx) shows two compound nouns in German, both of which consist of a feminine root and a masculine root. When the masculine root is the righthand member of the compound, the

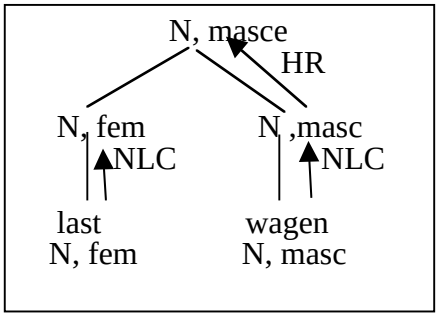
whole compound is masculine, and when the feminine root is the righthad member, the compound is feminine:

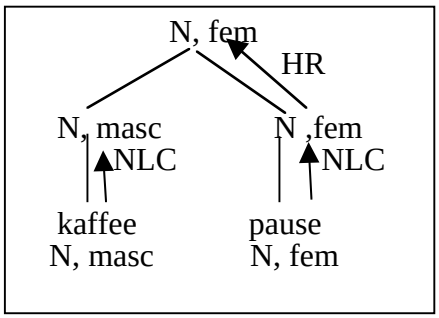
(5) die Last [fem] ‘cargo’ die Pause [fem] ‘break, pause’

der Wagen [masc] ‘vehicle’ der Kaffee [masc] ‘coffee’

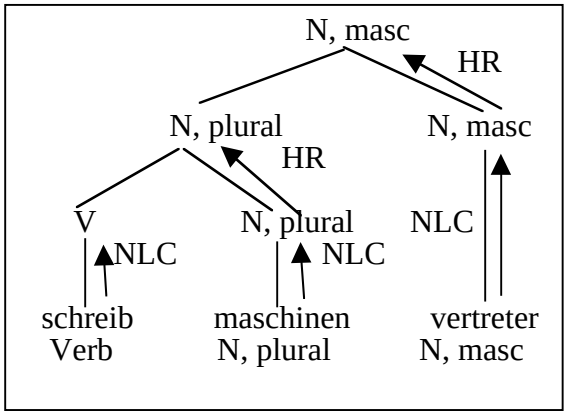
der Lastwagen [masc] ‘truck’ die Kaffeepause [fem] ‘coffee break’

The WSTs look like this:



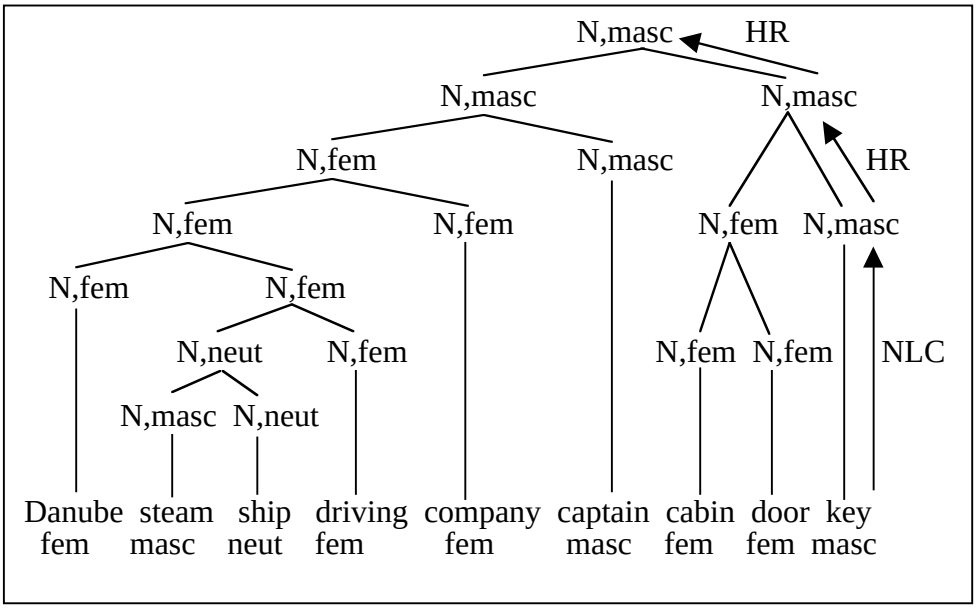


These principles generalize smoothly to German’s more impressive compounds. For example, der Schreibmaschinenvertreter means ‘The typewriter salesman’. It contains three roots: *schreib-* ‘write’, a verb; *maschin-* ‘machine’, occurring in *maschinen* a plural noun; and *tret-* occurring in *vertreter* ‘agent or salesman’, a singular masculine noun. The compound as a whole is masculine and singular, because that is what the right most element of the compound is:



Or take the fanciest compound we have considered: Der *Donaudampfschifffahrtsgesellschaftskapitänskajütentürschlüssel* (the Danube-steam-ship-driving-company-captain-cabin-door-key). This includes one neuter noun, three masculine nouns, and five feminine nouns. However, compounding is not a democracy: the noun as a whole is masculine, because the rightmost part of the rightmost part of the noun is. Here is the whole WST, now annotated for gender features.

(xx)



Note that the Head Rule applies eight times in this word, not just the two times that we have shown explicitly.

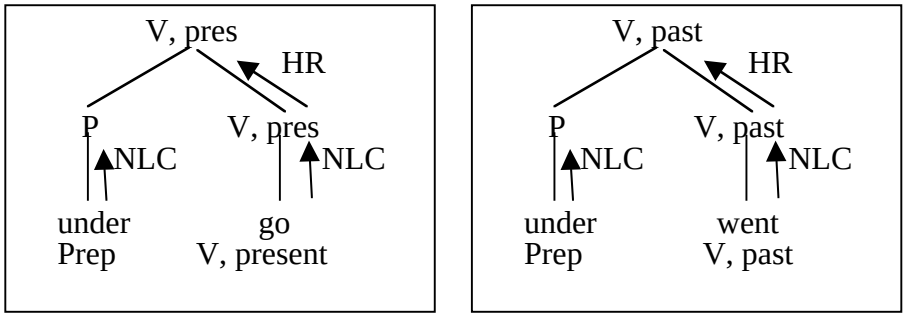
Consider now the features of verbs in compounds. Much as nouns are marked for number in English, verbs can be marked for tense. Most verb roots are not marked for tense, and are made past by adding the suffix *–ed*. However, some verbs have special past tense forms (we will return to this in chapter INFL-xx in more detail). Some examples are:

(xx)

	Present stem	Past stem
a.	go	went
b.	do	did
c.	stand	stood
d.	buy	bought
e.	freeze	froze

Now when an past tense form is used as the right hand member of a compound in English, the result is a past tense verb. (Note: these trees are simplified somewhat—in chapter Infl we will see that the treatment of past tense must be slightly more complex).

(xx)



Similarly, *overdo* is present tense, and *overdid* is past; *understand* is present tense, and *understood* is past. On the other hand, using a past tense root as the first member of a compound does not make the compound as a whole past. English has a few compound verbs that are made up of two verb roots: stir-fry, drop-kick, slam-dunk, and so on. One of these is freeze-dry:

(xx) I always freeze-dry the vegetables from my garden.

Now *freeze* happens to correspond to the past tense stem *froze*. However, the past tense of (xx) is (xxa), and (xxb) is unacceptable as an attempt to make the past tense of (xx).

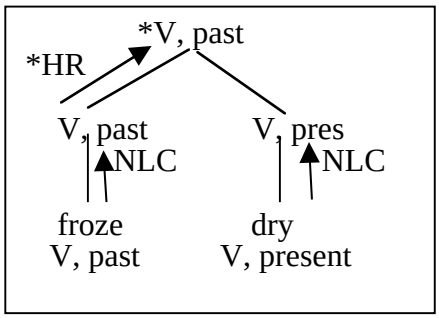
(xx) a. Last year I freeze-dried the vegetables from my garden.

b. *Last year I froze-dry the vegetables from my garden.

<JDB: (b) is grammatical as resultative + HNPS>

A past tense feature on the left hand member of a compound simply is not inheritable by the compound as a whole, according to the Right hand head rule.

(6)



More or less the same thing is seen by where the regular past tense affix –ed must attach in verb-verb compounds made up of two regular verbs. The affix must attach to the second verb root to make the whole verb past, never to the first.

(7) a. *The player on the field **drop** the ball. must be: dropped (or drops)

b. The player on the field [drop kick]-ed the ball.

c. *The player on the field dropped kick the ball.

(8) a. *I stirred-fri(ed) the veggies.

b. *I stir-fried the veggies.

Overall, the distribution of past tense marking in compounds with verbs is identical to the distribution of plural marking in compounds with nouns, showing that both are subject to the same head rule.

Practice Problem C-X. Suppose that transitivity is considered a feature of verbs similar in status to the tense feature, or to the number feature in nouns. Verbs that take a direct object in the syntax would be +transitive, and verbs that do not take a direct object in the syntax would be –transitive. Are the following examples consistent with the claim that this transitivity feature obeys the head rule in English? Explain.

(xx) a. I fried the steak

I pan-fried the steak.

b. I washed the sweater.

I hand-washed the sweater.

c. I fed the corn to the baby.

I spoon-fed the corn to the baby.

d. I napped (all afternoon).

I cat-napped (all afternoon).

e. I walk on Thursdays

I power-walk on Thursdays.

Can you find any counter examples to the general pattern among the compounds already mentioned in this chapter? If so, why do you think those examples work differently?

C.5 Distinguishing Compounds from Separate Words

In many languages, it is not difficult to recognize compounds. For example, in languages where most roots are bound, and require one or more inflectional affixes, compounding of two roots is generally obvious. This is what we saw in section C.1 above in languages like Greek and Mohawk. However, difficulties can arise in languages with many free roots, where two free roots combine to make a compound. Such a compound can look very much like a two word phrase—especially since the order of roots in a compound is often similar to the order of words in a phrase (see the end of section C.3). Spelling conventions can even add to the potential confusion in some languages. Thus, we saw in section C.2 that German very cooperatively always writes compounds together as one word, with no dividing spaces or hyphens but English does not. In English, we sometimes write compounds together *blackboard*, sometimes with a hyphen *pan-fry*, and sometimes as two units *drop kick, sales manager*. Thus, it is important to have some tests to show whether a given combination is a true compound or a syntactically formed phrase. To some extent these tests will vary from language to language, depending on the particular resources of the language in question. We will use English as a case study.

In fact, this is nothing more than a special case of the general issue discussed back in section xx. There we saw that words are distinguished units with respect to syntax, semantics, and phonology. This is true of compounding as well. Thus, we might expect to find syntactic, semantic, and phonological tests for distinguishing compounds from phrases.

Compare, for example, adjective-noun compounds like blackboard or greenhouse with noun phrases that include an adjective and a noun, such as black board and green house. These are clearly very similar, but a syntactic test shows the difference. Compounds are words made up of two roots, by definition. Phrases are syntactic units made up of smaller phrases. The potential for confusion comes because the smaller phrases that make up a phrase can be simple words. But there is a difference: the two parts of a compound must be single words, whereas the two parts of a phrase in general need not be. For example, in the syntax a noun can be modified by a simple adjective, but it can also be modified by an adjective phrase made up of an adjective plus and intensifying word such as very.

(9) This is a very black board. [The board is very black]

This is a very green house The house is very green.

However, the first part of a compound cannot be very plus an adjective; in this case, only a bare adjective will do:

(xx) *This is a very bláckboard. Cf. This is a very black blackboard.

*This is a very greenhouse.

Similarly the noun in a possessive phrase that modifies another noun can itself be modified by an adjective:

(xx) The nice dog's house (I.e. a house that belongs to a nice dog.)

However the noun that is the first part of a compound cannot be modified in this way:

(xx) #The nice doghouse (NOT a house intended for a nice dog.)

(Can only mean a doghouse that is nice, where nice modifies the whole compound noun.)

This is one rather reliable way to tell compounds from phrases.

Extra Note, though, that there is a specific class of exceptions, illustrated by the examples in (10).

(10) a. a "keep-off-the-grass" sign
b. a "keep-your-nose-out-of-my-business" attitude
c. The "leggo-my-Eggo" commercial
d. The "ka-ching-ka-ching" machine, the "bbbbllllllpppp" sound [i.e., "raspberry"]

It is readily possible in English to embed an entire quoted expression inside the left-hand member of a compound, when the compound as a whole can be interpreted as meaning something that is associated with the quote, e.g. a sign which says “keep off the grass”, or an attitude which says (figuratively speaking) “keep your nose out of my business” or a machine which makes a particular sound.

These look like syntactic phrases embedded inside compounds, and of course in some sense they are. Nevertheless, there are reasons to suspect that these are not simply syntactic constituents as such, but rather direct-quotations, treated by the morphology of compounding as single chunks, with no internal analysis.

One reason is phonological: there is a distinctive intonation in such compounds, where the quoted expression does not have normal sentence intonation (stress) but rather is pronounced in somewhat of a monotone, with the words rushed together. This is often indicated in writing with a string of hyphens. This special intonation suggests that there is something peculiar about these particular constructions.

The syntax and semantics also provide an argument that these compounds embed a direct quotation and not a simple syntactic constituent. In syntactic expressions, including embeddings of various types, a particular category of expressions known as *indexicals* have a fixed discourse function. Thus, the pronoun *my* always refers to the speaker. If Heather were to say: *John said that this is my book*, then John has asserted that the book is Heather’s. Heather is the speaker: the author of the entire utterance, even if that utterance describes what John said. Regardless of the context, the pronoun *my* refers to the speaker of the entire utterance. The only exception is internal to a direct quotation. In (10b) and examples like it, the pronoun systematically fails to have an indexical interpretation. If Heather says: *John has a keep-your-nose-out-of-my-business attitude*, then the personal pronouns *your* and *my* do not have indexical readings: it is not Heather’s business that is being talked about, but Johns’.

Finally, examples like (10d) show that the left-hand member need not be a part of language at all, but can be any sound that can be repeated, and hence quoted. For example, the sound often referred to as the “raspberry cheer”, formed by sticking the tongue out between closed lips, and then blowing, is not a part of any language, yet it may occur as the left member of a quotative compound, precisely because what is in the quotes is not analysed from the point of view of the grammatical structure of the compound.

3.1.1 Semantic considerations in detecting compounds

Semantic considerations can sometimes also be brought to bear on identifying compounds. Compounds quite often take on a meaning which is not simply the sum of their parts. This is a guide or clue, not a foolproof test; however, it does converse nicely with the syntactic test in the case at hand. Thus, a *black board* is simply a board that is black, but a *blackboard* is more than that: it is a slate for writing on in chalk often found at the front of a classroom. *Blackboards* need not be black at all; often they are green. On the other hand, a *black board* might not be suitable as a *blackboard*. I don’t recommend teaching using a two-by-four painted with black paint, for example. Similarly, there is a green house on my street that is

not a greenhouse, and I have been in greenhouses that are not green at all (although hopefully the plants inside them are). And if a billionaire leaves his mansion to his favorite pet Fido, it does not thereby become a doghouse in the usual sense. Often, the meaning of a compound is narrower and more specialized than that of a noun phrase made of the same morphemes, and it includes culturally specific information. The head rule still applies: a blackboard is a kind of board, and greenhouses and doghouses are special kinds of houses. However, meaning of the non-head often becomes vague and is related to the meaning of the head in a way that may not be predictable.

3.1.2 Phonological criteria for compounds

A third way to tell compounds from phrases in English involves phonological properties, in particular metrical prominence, which may be realised as stress or pitch. Thus, the following compounds are pronounced slightly differently from the corresponding phrases (in many dialects of English).

(11)	bláckbòard	blác̀k bóard
	gréenhòuse	gréen hóuse
	blúebèrry	blúe bérry
	sáles manager	yóung mánager
	fórest ranger	júnior ránger
	túrkey sandwhich	rótten sándwhich

The compounds typically have a prominent stress on the first member of the compound, and secondary stress on the second member. The phrases, on the other hand, always have the highest stress on the head noun, the preceding adjective bearing secondary stress (except perhaps in metalinguistic or contrastive situations). The right hand column of examples are instances of a very general stress rule in English phrasal phonology:

(xx) The Nuclear Stress Rule.

Assign highest stress to the last word in a phrase.

This rule also applies to verb phrases like Chris [eats rice] and prepositional phrases like [under the table], explaining why (in the absence of contrastive focus) rice is pronounce louder than eat and table is pronounced louder than under. However, compounds are subject to a special rule (subject to dialect variation—this is one point which frequently separates Canadian from American varieties of North American English, for example):

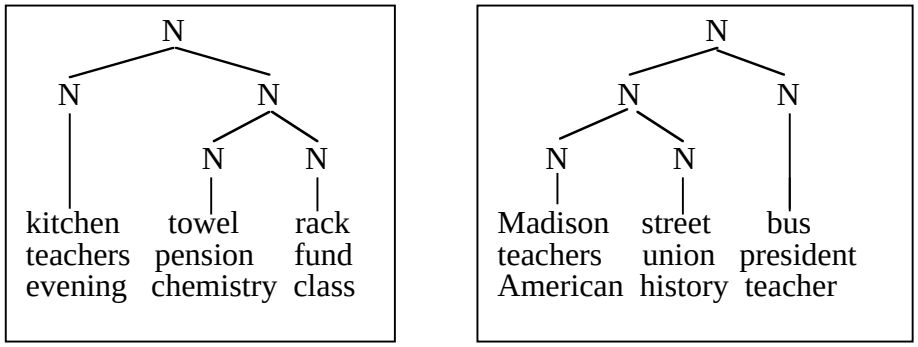
(xx) The compound stress rule:

> In a compound structure, assign highest stress to the second element if it is internally branching; otherwise assign highest stress to the first element.

This accounts for the fact that the nouns in the first column of (xx) have main stress on the first part rather than the second. In a sense the compound stress rule in English can be seen largely as a device for (partially) marking the difference between compounds and phrases.

The Compound Stress Rule as we gave it is more complex than it needs to be for the simple compounds in (xx). For these the simpler rule "assign highest stress to the first part of a compound" would have done just as well. However, the full form of the rule has the effect that it not only distinguishes compounds from phrases, but can reveal the internal structure of three-root compounds. Back in section C.2 we saw that three-noun compounds can have either of two structures, as follows:

(xx)



Interestingly, these two types of compounds also have different characteristic stress patterns in some varieties of English. In the first type, the highest stress is on the middle root of the compound (kithchen TOWEL rack), whereas in the second type, the highest stress is on the first root of the compound (MADISON street bus). The fancy version of the Compound Stress Rule can capture this difference. First we form two-member compounds like *towel rack* and *Madison street*. The Compound Stress Rule correctly expresses the fact that these compounds consistently have stress on the first root: we say *TOWEL rack* and *MADISON street*. Then one forms a larger compound by adding a third root to this unit. If one adds the third root at the end, as in *Madison street bus*, the compound stress rule applies in the same way, saying that the first unit—Madison street—should bear the larger stress. Since we already know that Madison is more stressed than street inside this compound, Madison ends up with the strongest stress in the word as a whole. Suppose, however, that one adds the third root at the beginning, as in kitchen towel rack. This time the head of the compound is itself a compound, and hence is internally branching. Intuitively, this makes it heavier than the nonhead, overriding the general tendency for the nonhead to receive higher stress. In this case, the stress goes on towel rack, of which towel is the most stress item. This shows how phonological rules can sometimes be used to determine not only that a certain group of morphemes is a compound, but also what its word structure is.

Practice Exercise C-X. Consider the following four-root compounds. The first has its primary stress on the very first root, the second has primary stress on the third root:

MADISON avenue bus schedule.

Law school ENTRANCE exam.

Is this difference explained by the compound stress rule? Explain. Does the compound stress rule predict that highest stress can appear on the second root of a four root compound in English? If so, can you find or construct a relevant example?

For students in Montreal: the compound stress rule predicts a subtle difference in meaning between the two strings in (12). Can you explain the difference clearly, identifying which route or routes correspond to which strings?

(12) a. PINE Avenue bus schedule

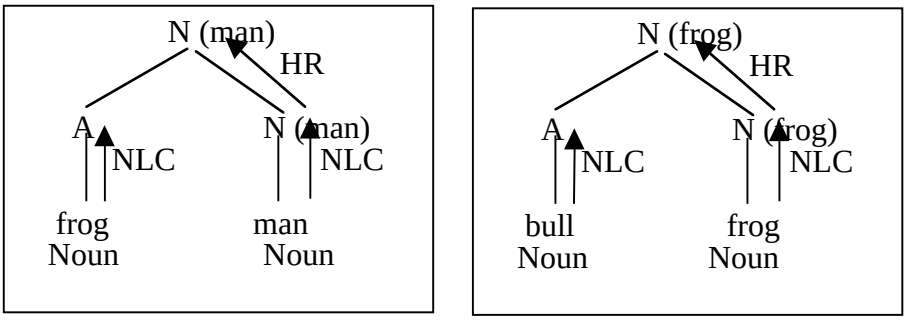
b. Pine Avenue BUS schedule

3.6 Varieties of Compounding

So far we have concentrated on the most important and common kind of compounds, the kind in which the Head Rule applies, in either its left-headed or right-headed version. However, there are a few other types of compounds as well. Moreover, languages do not necessarily allow all the same types of compounding. In this section, we will survey some of the important dimensions of variation in compounding.

Compounds that have a single, well-defined head are sometimes called endocentric compounds. (Endo means 'inside' in Latin.) Endocentric compounds get not only their morphological and syntactic features from the designated head, but also their primary meaning. Consider the minimal pair *frogman* and *bullfrog*. Both are singular nouns, because their righthand member is a singular noun. However, there is something else to capture. A bullfrog is a kind of frog, whereas a frogman is not a kind of frog; rather, a frogman is a kind of man (one that looks or acts a bit like a frog). Clearly, when *frog* is the head of a compound, the compound refers to a kind of frog, but when *frog* is not the head it does not. This is true quite generally. Speaking informally, we can assume that an element of meaning is also passed from the righthand member to the compound by the head rule, as shown in (xx).

(13)



Not all kinds of endocentric compounds are equally prevalent in all languages.

English abounds in compounds formed from nouns and adjectives as we have seen: the combinations N+N, N+A, A+N, and A+A are all frequent and easily coined, as are the combinations P+N and P+A. However, other types of compound are much rarer. For example, other than P+V, compounds that have verbs as their heads are unusual in English. A few examples exist:

(xx) a. A+V: dry-clean, sharp-shoot

b. N+V: babysit, bartend, trouble-shoot, air-condition, …

c. V+V: stir-fry, slam-dunk, drop-kick, freeze-dry

But they are relatively infrequent, and it is not so easy to make up new ones. Even as the nonhead, verbs are not as comfortable in compounds as other categories. There are extremely few V+A compounds in English, for example:

(xx) V+A: die-hard

V+N compounds are more common: we listed some examples above, including *scarecrow, cut-throat, drawbridge, swearword, pushcart*, etc. However, it is not clear that even this pattern is really productive. Most V+N compounds have been in the English language for a long time, and it is rare for people to coin new words on this pattern. For example, if someone designates a place to meet, it is not a *meet-place*, but a *meeting place*, using a noun-noun compound rather than a verb-noun compound (compare the German example in (xx) above). Or if someone markets a new tool for opening boxes, they are unlikely to call it an *openbox*, but rather a *box-opener*. However, the V-N pattern is wide-spread in French and the other Romance languages, and the presence of a reasonable number of V-N compounds in English is probably due to extensive contact with French earlier in the history of the language.

There is probably also a kind of historical explanation for the A+V and N+V compounds found in English. Whereas these types of compounds are unusual in the language, it is easy to make verbs into nouns by adding –ing, -er, or other derivational affixes. Then it is easy to form A+N and N+N compounds using these deverbal nouns as heads, as we have seen:

(xx)	Not:	meat-eat	but:	meat-eater
		Dish-wash		dishwasher, dishwashing
		Deer-hunt		deer-hunter, deer-hunting

Now there is a common tendency for speakers of a language to see affixes when they are not really there. For example, there used to be no verb ‘to peddle’ in English, but the noun ‘peddler’. However, this noun happened to end in the sounds /er/, and referred to a type of profession. Therefore, speakers over time naturally perceived the last part of peddler as the very common and productive –er affix. *Peddler* seems to bear the same relationship to a root *peddle* that *baker* does to the verb *bake*, for example. They thus inferred that there was (or should be) a verb root

peddle that this affix was attached to. This process of creating a new word over time by stripping off putative affixes is known as **back formation**. Now it is striking that all of the A+V and N+V compounds we have in English correspond to very familiar and often used deverbal compounds.

(xx) a. A+V: dry-clean cf. dry-cleaner, dry-cleaning

sharp-shoot cf. sharpshooter

b. N+V: babysit cf. baby-sitter, baby-sitting

Bartend cf. bartender, bartending

trouble-shoot cf. trouble-shooter, trouble-shooting

air-condition cf. air-conditioner, air-conditioning

It is plausible then to explain the presence of these examples in English as an instance of back-formation. English speakers hear words like baby-sitter frequently, recognize the –er affix, and posit a verb baby-sit that this word was derived from. In this way, we can explain why there are a reasonable number of N+V compounds in English, even though they are not formed productively, and new ones sound odd, particularly if the related N+Ving or N+Ver compounds are not already familiar to a person. And there is substantial individual variation in what is accepted here. For example, I know people who routinely use a verb to grocery-shop, although I would not (before meeting them anyway). This clearly comes from grocery-shopping, but they have taken the back-formation step and others have not.

English's love of N-N compounding and reluctance to compound verbs is not universal. The same tendency is seen in the Edo language. Mohawk, on the other hand, productively allows N-V compounding, but does not allow either N-N compounding or V-V compounding. Here are some examples of N-V compounds from that language.

Wa'eksohare'	'She dish-washed.'	(ks 'dish' + ohare 'wash')
Wa'kenaktahninu'	'I bed-bought.'	(nakt 'bed' + a + hninu 'buy')
Wahana'tarakwetare'.	'He bread-cut.'	(na'tar 'bread' + a+ kwetar 'cut')
waha'wahrake'.	'He meat-ate.'	('wahr 'meat' + a + k 'eat')

However, V-V compounding is very common in other languages, including Chinese, Igbo (an African language), and Tagalog (check). Here are some examples from Igbo:

(xx) a. kwà-cí

push-be.shut 'to push something shut'

b. fùn-yú

blow-be.off ‘to blow something out

c. cú-pù

pursue-exit ‘to chase someone away’

d. tí-gbú

hit-kill ‘to beat someone to death’

The Mapuche language of South America has all three of these types.

It is not well-understood why languages differ in this way. It could be simply an idiosyncratic fact that children need to learn about each language. On the other hand, it is conceivable that these differences are related somehow to other differences in the syntax or morphology of the languages in question.

There are also compounds for which the Head Rule does not apply. These are known as exocentric compounds, because they have no head inside them (*exo* means ‘outside’). English has a small number of such compounds, including the following examples:

(xx) A+N sweetheart, redcap, Bigfoot

N+N sabertooth,

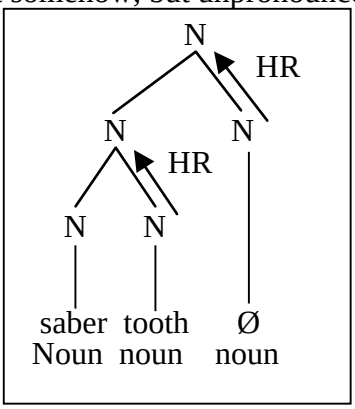
V+N scarecrow, pickpocket, cut-throat, …

At first glance, one might think that these were endocentric compounds. After all, they are all nouns and the right-hand member of the compound is also a noun. However, this time the semantics is not right: a sweetheart is not a kind of heart, for example, nor is a sabertooth (only) a kind of tooth, or a scarecrow a kind of crow. Rather, a sweetheart is someone who has a sweetheart, a sabertooth is an extinct feline that had teeth like sabers, and a scarecrow is a manikin intended to frighten crows. This calls into question whether these examples really have a head in the sense that we have been discussing. Further evidence that they do not comes from patterns of pluralization. We saw that with endocentric compounds, using an irregular plural root as the righthand member of a compound created a plural compound. However, this is not the case with exocentric compounds; rather, the normal plural morpheme –s must be added to the compound as a whole. Thus, we have the following contrasts:

Endocentric		Exocentric	
Clubfoot	clubfeet	bigfoot	bigfoots (*bigfeet)
Bucktooth	buckteeth	sabertooth	sabertooths (*saberteeth)
Maple leaf	maple leaves	Maple leaf	Maple leafs (*Maple leaves)
(things that fall from the maple tree)		(members of Toronto’s hockey team)	

Thus, the plural feature is not inherited from the final N of a compound in exactly those cases in which the meaning is also not inherited from the final N. Thus, these examples are exceptions to the Head Rule for all features. In some cases, one might want to say that these examples have null morphemes that count as their heads. For example, *sabertooth* is a shortened form of *sabertooth tiger*, which is a normal endocentric compound. Perhaps the "tiger" part is still present somehow, but unpronounced. Then the structure of the word would be:

(xx)



This analysis takes the terminology of exocentric seriously, saying that these compounds do have a head, but one that falls outside of the obvious roots that make it up. Similarly, examples like scarecrow could be considered to have an invisible derivational affix that makes verbs into nouns, similar in meaning to the overt affix –er. This would capture the fact that compounds like scarecrow are comparable to the more productive English formation crow-scarer, where the noun-forming affix is visible. This null –er could be considered the head of the compound. This is attractive for Romance languages, where the pattern is more common. Consider the following example from Spanish:

(xx) el lavaplatos

el lava-plato-s

the wash-dish-es

Again, it is clear that the noun *platos* is not the head of this compound in any sense. It is plural, but the noun as a whole is not, nor does the gender of the whole noun come from platos. Finally, a dishwasher is not a kind of dish. Such nouns are always singular and masculine, suggesting that they have a singular masculine head that is neither of the visible roots.

Finally, there is a third kind of compound in which both members of the compound seem in some sense to be heads. The two nouns of such compounds seem to be linked by a kind of co-ordination. A logical name for such compounds would be dicentric compounds, although in fact linguists general call them *dvandva* compounds, a term that comes from the Sanskrit grammatical tradition. Dvandva compounds are not very common in English, but a few examples exist, such as the following:

(xx) a. Austria-Hungary; Bosnia-Hercegovina;

b. mother-daughter (as in 'The mother-daughter relationship is …')

c. player-coach, gentleman-farmer, manservant

Semantically, these are like double-headed compounds. Whereas a frogman is a man, not a frog, and a sabertooth is neither a saber nor a tooth, a player-coach is both a player and a coach. Sometimes these compounds even take double plurals, with both members of the compound switching from a singular form to a plural form.

(xx) a. Menservants (not *manservants)

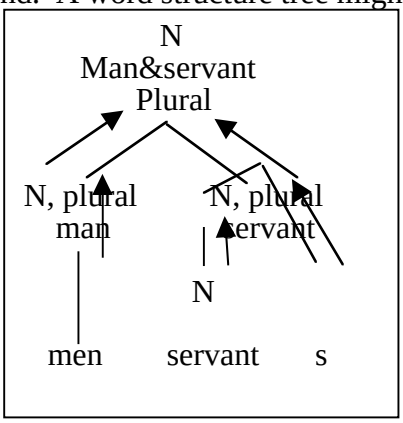
b. Gentlemen-farmers

Not many dvandva compounds in English follow this rule; for most, the plural is formed by adding the regular plural –s to the compound as a whole:

(xx) Player-coaches (not *players-coaches)

Only if the first root of the dvandva compound has an irregular plural root, does the double-plural show up. However the fact that the double plurals are ever possible confirms that these compounds are in some sense dicentric, with features being inherited from both sides of the compound. A word structure tree might look like this:

(xx)



Dvandva compounds are limited in English, but they are much more productive in some other languages. Malayalam, for example, allows them productively: [Fix diacritics]

(xx) a. waḍ^{h}u 'bride' warān 'groom' -maar [plural]
waḍ^{h}uuwarānmaar 'wedding pair'

b. aaṭə 'goat' maaṭə 'cow' -kaḷə [plural]
aaṭəmaaṭəkaḷə 'goats and cows'

c. 'father' 'mother' [plural]
acchanammamaare 'parents'

Again, it is not well understood why some languages permit this more regularly than others, or whether there are interesting regularities. There may be interesting patterns to discover. For example, we noted above that V-V compounds are also rare in English, but common in some other languages. Now, V-V compounds in languages that have them can usually be classified as dvandva compounds. For example, stir-fry something is both to stir it and to fry it; to drop kick something is to both drop it and kick it. The two roots in these compounds must also match in transitivity (both drop and kick are transitive verbs, for example), much as nouns sometimes match in number in dvandva compounds. Thus, it is possible that the real generalization is that English does not permit dvandva compounds, and the fact that it does not permit V-V compounds is simply a special case of this broader generalization. Overall, though, there is still much to be done to discover what typological and implicational universals apply to compounding.