

ENCYCLOPEDIA OF COGNITIVE SCIENCE

©Macmillan Reference, Ltd.

Morphology

Stephen R. Anderson

Yale University, New Haven, Connecticut, USA

Contents:

Introduction
Inflection
Word Formation
Representation of Morphological Knowledge

Morphology, in linguistics, is the study of the forms of words, and the ways in which words are related to other words of the same language. Formal differences among words serve a variety of purposes, from the creation of new lexical items to the indication of grammatical structure.

Introduction

If you ask most non-linguists what the primary thing is that has to be learned if one is to ‘know’ a language, the answer is likely to be “the words of the language”. Learning vocabulary is a major focus of language instruction, and while everyone knows that there is a certain amount of ‘grammar’ that characterizes a language as well, even this is often treated as a kind of annotation to the set of words—the ‘uses of the Accusative’, etc. But what is it that is involved in knowing the words of a language?

Obviously, a good deal of this is a matter of learning that *cat*, pronounced [k^hæt], is a word of English, a noun that refers to a “feline mammal usually having thick soft fur and being unable to roar”. The notion that the word is a combination of sound and meaning—indeed, *the* unit in which the two are united—was the basis of the theory of the linguistic *sign* developed by Ferdinand deSaussure at the beginning of the 20th century. But if words like *cat* were all there were in language, the only thing that would matter about the form of a word would be the fact that it differs from the forms of other words (i.e., *cat* is pronounced differently from *mat*, *cap*, *dog*, etc.). Clearly there is no more specific connection between the parts of the sound of *cat* and the parts of its meaning: the initial [k^h], for example, does not refer to the fur. The connection between sound and meaning is irreducible here.

But of course *cat* and words like it are not the end of the story. Another word of English is *cats*, a single word in pronunciation but one that can be seen to be made up of a part *cat* and another part *-s*, with the meaning of the whole made up of the meaning of *cat* and the meaning of *-s* (‘plural’). *Cattish* behavior is that which is similar to that of a cat; and while a *catbird* is not itself a kind of cat, its name comes from the fact that it sometimes sounds like one. All of these words are clearly connected with *cat*, but on the other hand they are also all words in their own right.

We might, of course, simply have memorized *cats*, *cattish* and *catbird* along with *cat*, even though the words seem to have some sort of relation to one another. But suppose we learn about a new animal, a *wug*, say ‘a large, hairy bovine mammal known for being aggressive and braying’. We do not need to learn independently that two of these are *wugs*, or that *wuggish* behaviour is likely to involve attacking one’s fellows, or that a *wugbird* (if there were such a thing) might be a bird with a braying call. All of these things follow from the knowledge we have not just of the specific words of our language, but of their relations to one another, in form and meaning. The latter is our knowledge of the morphology of our language.

In some languages, the use of morphology to pack complex meanings into a single word is much more elaborate than in English. In West Greenlandic, for example, *tusaanngitsuusaartuaannarsiinnaanngivipputit* is a single word meaning ‘you simply cannot pretend not to be hearing all the time’. Other languages do much less of this sort of thing: Chinese and Vietnamese are often cited in this connection, though Chinese does have rather exuberant use of compounding (structures like *catbird* made up of two exiting items). Despite this variation, however, morphology is an aspect of the grammar of all languages, and in some it rivals syntax in the expressive power it permits.

Inflection

Traditionally, morphology is divided into several types, depending on the role played in grammar by a given formation. The most basic division is between inflection and word formation: the latter is easy enough to characterize as ‘morphology that creates new words’ (*wuggish*, *wug-like*, *wugbird*), but inflection (e.g., *wugs*) is rather harder to define. Often, inflection is defined by example: categories like number (e.g., ‘plural’), gender (e.g., masculine, feminine and neuter in Latin), tense (‘past’), aspect (e.g., the difference between the *imparfait* and the *passé simple* in French), case (‘accusative’), person (1st vs. 2nd vs. 3rd), and perhaps a few others are inflectional while everything else is word formation. But this approach is inadequate, because the same category may be inflectional in some languages, and not in others. In Fula (a West Atlantic language), for example, the category ‘diminutive’ is fully integrated into the grammar of agreement in the language, just as much so as person, number, and gender. Verbs whose subjects are diminutive indicate this with an agreement marker, as do adjectives modifying diminutive nouns, etc. In English, in contrast, diminutives appear in forms like *piglet*, but these are clearly cases of word formation. On the other hand, while number is clearly involved in important parts of English grammar (verbs agree with their subjects in number), other languages, like K^Wak^Wala (or ‘Kwakiutl’) treat the category of plural as something that can optionally be added to nouns, or to verbs, as an elaboration of meaning that has no further grammatical consequence.

Despite the intuitively clear nature of the category of inflection, other efforts to define it explicitly do no better. Inflection is generally more *productive* than other sorts of morphology, for instance: virtually every German noun has an accusative, a plural, etc., while only a few English nouns have a diminutive formation like *piglet*. But in some languages, categories that we would certainly like to call inflectional are quite limited: in Basque, for example, only a few dozen verbs (the number varying from one dialect to another) have forms that show agreement. In English, on the other hand, the process of forming nouns in *-ing* from verbs (as in *Fred’s lonely musings about love*) can take virtually any verb as its basis, despite being intuitively a means of creating new words, not of inflecting old ones. A variety of other attempts that can be found in the literature also fail, either because of ready counter-examples, or because they are insufficiently general: inflectional material is generally found at the word’s periphery, while word formational markers are closer to the stem (cf. *piglets* but not **pigslet*), but this property is only useful in words that contain material of both types, and even then, it does not help us to find the boundary in a word like French *im-mort-al-is-er-ait* ‘would immortalize’.

In fact, the intuition underlying the notion of ‘inflection’ seems to be the following: inflectional categories are those that provide information about grammatical structure (such as the fact that a noun in the accusative is likely to be a direct object), or which are referred to by a grammatical rule operating across words (such as the agreement of verbs with their subjects). The validity of other correlates with inflectional status, then, follows not from the nature of the categories themselves, but rather from the existence of grammatical rules in particular languages that refer to them, and to the freedom with which items of particular word classes can appear in positions where they can serve as the targets of such rules.

For any given word, we can organize a complete set of its inflectional variants into a *paradigm* of the word. Thus, a German noun has a particular gender, and a paradigm consisting of forms for two numbers (singular and plural) and four cases (nominative, genitive, dative, and accusative). German adjectives have paradigms that distinguish not only case and number, but also gender (since they can agree with nouns of any of the three genders), plus another category that distinguishes between ‘strong’ and ‘weak’ declensions (depending on the presence of certain demonstrative words within the same phrase).

All of the word forms that make up a single inflectional paradigm have the same basic meaning. In general, they are all constructed on the basis of a basic shape, or stem, though in many languages with complex inflection, the paradigm of a given word may be built from more than one stem. In French, for example, the verb *pouvoir* ‘to be able to’ shows different stems in *(je) peux* ‘I can’ and *(je) pourrais* ‘I would be able to’.

Certain terminology has become more or less accepted in describing facts of these sorts. We refer to a particular sound shape (e.g. [fawnd]) as a specific *word form*; all of the inflectional forms in a single paradigm are said to make up a single *lexeme* (e.g., *find*). A specific *morphosyntactic form* of a particular lexeme (e.g., the past tense of *find*) is realized by a corresponding word form [fawnd]). These terms are all distinct, in their way: thus, the same morphosyntactic form of a given lexeme may correspond to more than one word form (e.g., the past tense of *dive* can be either [daɪvd] or [dowv]), while the same word form can realize more than one morphosyntactic form (e.g., [hit] can be either the past tense of *hit*, the non-third-person present tense of *hit*, or the singular of the noun *hit*).

Word Formation

Inflection, then, is the morphology that distinguishes the various forms within the paradigm of a single lexeme. Some languages, like ancient Greek or Georgian, have a great deal of inflectional morphology, while others (like English) have

much less, and some (like Vietnamese) have hardly any at all. Regardless of this, however, essentially all languages have ways of constructing new lexemes from existing ones, or patterns of word formation. These fall into two broad classes: *compounding* is the process of combining two or more independently existing lexemes (perhaps with some additional material as ‘glue’) into a single new lexeme (as in *catbird*). *Derivation*, in contrast, is the formation of a new lexeme from an existing one by means of material that does not appear by itself as a word. It is common to refer to such non-independent content as *bound* in contrast with independently occurring or *free* elements.

Derivation

A typical derivational relation among lexemes is the formation of adjectives like *inflatable* from verbs (*inflate*). In this case, the meaning of the adjective is quite systematically related to that of the verb: VERB-*able* means ‘capable of being VERB-ed’. It is therefore tempting to say that English contains an element *-able* with that meaning, which can simply be added to verbs to yield adjectives. The facts are a bit more complex than that, though.

For one thing, the related adjective may not always be just what we would get by putting the two pieces together. For instance, *navigate* yields *navigable*, *formulate* yields *formulable*, etc. These are instances of *truncation*, where a part of the base is removed as an aspect of the word formation process. Then there are cases such as *applicable* from *apply*, where we see the same variation (or *allomorphy*) in the shape of the stem as in *application*. These patterns show us that the derivational whole may be more than the simple sum of its parts.

When we consider the class of adjectives in *-able* (or its spelling variant *-ible*), we find a number of forms like *credible*, *eligible*, *potable*, *probable*, ... which seem to have the right meaning for the class (they all mean roughly ‘capable of being [SOMETHING]-ed’), but the language does not happen to contain any verb with right form and meaning to serve as their base. This suggests that derivational patterns have a sort of independent existence: they can serve as (at least partial) motivation for the shape and sense of a given lexeme, even in the absence of the possibility of deriving that lexeme from some other existing lexeme. In some instance, the force of this analysis is so strong that it leads to what is called *back-formation*: thus, the word *editor* was originally derived from Latin *e:dere* ‘to bring forth’ plus *-itor*, but it fit so well into the pattern of English agent nouns in *-er* (e.g., *baker*; *driver*) that a hypothetical underlying verb *edit* actually became part of the language.

We may also notice that some *-able* forms do not mean precisely what we might predict. Thus, *comparable* means ‘roughly equal’, not just ‘able to be compared’. In the world of wine, *drinkable* comes to mean ‘rather good’, not just ‘able to be drunk’, etc. This shows us that even though these words may originally arise through the invocation of derivational patterns, the results are in fact full-fledged words of the language; and as such, they can undergo semantic change independent of the words form which they were derived. This is the same phenomenon we see when the word *transmission*, originally referring to the act or process of transmitting (e.g., energy from the engine to the wheels of a car) comes to refer to a somewhat mysterious apparatus which makes strange noises and costs quite a bit to replace.

Finally, we can note that in some cases it is not at all evident how to establish a ‘direction’ of derivation. In Maasai, for example, there are two main noun classes (‘masculine’ and ‘feminine’), and a derivational pattern consists in taking a noun which is ‘basically’ of one class, and treating it as a member of the other. Thus, *en-kérái* is a feminine noun that refers to any child, of either gender; while *ol-kérái* is a corresponding masculine noun meaning ‘large male child’. Here it looks plausible to take the feminine form as the basis for the derivational relationship; but when we consider *ol-abááni* (masculine) ‘doctor’ vs. *enk-abááni* ‘small or female doctor, quack’ it looks as if the direction of derivation goes the other way. In fact, it looks as if what we have here is a case of a relation between two distinct patterns, where membership in the feminine class may (but need not) imply femaleness and/or relatively small size, as opposed to the masculine class which may imply maleness and/or relatively large size. When a word in either class is used in the other, the result is to bring out the additional meaning associated with the class, but there is no inherent directionality to this relationship. The possibility of back formation discussed above suggests that this interpretation of derivational relationships as fundamentally symmetrical may be applicable even to cases where the formal direction of derivation seems obvious.

Compounding

The other variety of word formation, compounding, seems fairly straightforward, even if the actual facts can be quite complex at times. Compounds are built of two (or more) independent words, and have (at least in their original form) a meaning that involves those of their components. Thus, a *catfish* is a kind of fish sharing some property with a cat (in this case, the whiskers). Like derived forms, compounds are independent lexemes in their own right, and as such quickly take on specialized meanings that are not transparently derived from those of their parts. We need to tell a story to explain why a *hotdog* is called that, why a *blackboard* can be white or green, etc.

Where it is possible to relate the meaning of a compound to those of its parts, it is often possible to establish a privileged relationship between the semantic ‘type’ of the whole compound and that of one of its pieces. Thus, a *dog house* is a kind of *house* (and certainly not a kind of *dog*), *out-doing* is a kind of *doing*, etc. When such a relation can be discerned, we refer to the ‘privileged’ member of the compound as its head, and speak of the compound itself as *endo-centric*.

By no means all compounds would appear to be endocentric, however: a *pickpocket* is neither a kind of pocket nor a kind of picking, and a *sabre-tooth* is a kind of tiger, not a kind of tooth. Traditional grammar provides a variety of names for different types of such *exo-centric* compounds, some deriving from the Sanskrit grammatical tradition in which these were of particular interest. A *bahuvrihi* compound is one whose elements describe a characteristic property or attribute possessed by the referent (e.g., *sabre-tooth*, *flatfoot*), a *dvandva* compound is built of two (or more) parts, each of which contributes equally to the sense (e.g., an *Arab-Israeli* peace treaty).

In some languages, the decision as to which compounds are endocentric and which are not depends on the importance we give to different possible criteria. For instance, in German, *Blauhemd* ‘(soldier wearing a) blue shirt’ is on the face of it a bahuvrihi compound, exocentric because it does not denote a kind of shirt. On the other hand, the gender of the compound (neuter, in this case) is determined by that of its rightmost element (here, *hemd* ‘shirt’). Semantically, *blauhemd* is exocentric; while grammatically, it could be regarded as endocentric with its head on the right.

Languages can vary quite a bit in the kinds of compound patterns they employ. Thus, English compounds of a verb and its object (like *scarecrow*) are rather rare and unproductive, while this constitutes a basic and quite general pattern in French and other Romance languages. English and German tend to have the head, when there is one, on the right (*dollhouse*), while Italian and other romance languages more often have the head on the left (e.g., *caffelatte* ‘coffee with milk’). Most English compounds consist of two elements (though one of these may itself be a compound, as in *[[high school] teacher]*, leading to structures of great complexity such as German *[[[[Leben]s-versicherung]s-gesellschaft]s-angestellter]* ‘life insurance company employee’), but many dvandva compounds in Chinese consist of three or four components, as in *ting-tai-lou-ge* ‘(pavilions-terraces-upper stories-raised alcoves) elaborate architecture’.

Finally, we should note that although we have defined compounds as built from free elements or independent lexemes, this leaves us with no good way of describing structures such as the names of many chemical compounds and drugs (*dichlorobenzene*, *erythromycin*) and words such as *Italo-American*. On the one hand, we surely do not want to say that there is a process that affects a base such as *American* by prefixing *Italo-*. On the other hand, *Italo-*, *erythro-*, *chloro-*, etc. do not occur on their own, but only in this class of compounds. Even more striking examples occur in other languages. For examples, the Mandarin root *yi* ‘ant’ freely forms compounds such as *yiwang* ‘queen ant’ (literally ant-king), *gongyi* ‘worker ant’, *baiyi* ‘white ant, termite’. But *yi* is clearly not a word: the free word for ‘ant’ in Mandarin is *mayi*. While English *erythro* etc are always prefixes, in the Mandarin cases, the roots in question occur in both head and non-head position, and are therefore like normal compound components in every respect except that they are not free forms. It appears that the very definition of compounding need more thought than was initially evident.

Representation of Morphological Knowledge

To this point, we have talked of morphological relationships as existing between whole lexemes (in the case of word formation), or between word forms (in the case of inflection). Much of the tradition of thought about morphology, however, regards these matters in a somewhat different light. We saw at the beginning of this article that the model of the Saussurean sign as the minimal unit where sound and meaning are connected could not serve as a description of the word, since it is often the case that (proper) parts of words display their own connection between sound and meaning. It was this observation, in fact, that led us to explore the varieties of morphology displayed in natural language. But many have felt that the proper place for the sign relation is not the word, but rather a constituent part of words: the *morpheme*. On that picture, morphology is the study of these units, the morphemes: how they may vary in shape (the *allomorphy* they exhibit) and how they can be combined (*morphotactics*).

Morphemes and Words

The notion that words can be regarded as (exhaustively) composed of smaller sign-like units, or morphemes, is extremely appealing. It leads to a simple and uniform theory of morphology, one based on elementary units that can be regarded as making up a sort of lexicon at a finer level of granularity than that of words. Nonetheless, it seems that this picture of word structure as based on a uniform relation of morpheme concatenation is literally too good to be true.

If morphemes are to serve the purpose for which they were intended, they ought to have some rather specific properties. It ought to be possible, for any given word, to divide its meaning into some small number of sub-parts, to divide its form into a corresponding number of continuous sub-strings of phonetic material, and then to establish a correspondence between the parts of meaning and the parts of form. Of course, it is possible to do exactly that in a great many cases (e.g., *inflatable*):

hence the intuitive appeal of this notion. But in many other instances, such a division of the form is much more laboured or even impossible.

One fairly minor problem is posed by parts of the form that are not continuous. When we analyze words containing circumfixes (e.g., *ke—an* in Indonesian *kebisaan* ‘capability’, from *bias* ‘be able’) or infixes (e.g. *—al-* in Sundanese *ngad^ulahar* ‘to eat several’, from *ngadahar* ‘to eat’) one or the other of the component morphemes is not a continuous string of material.

Other cases are more serious. For instance, we may find no component of meaning to correspond to a given piece of form (an ‘empty morph’ such as the *th* in English *lengthen* ‘make long(er)’) or no component of form that relates to some clear aspect of a word’s meaning (e.g., English *hit* ‘past tense of *hit*’). Sometimes two or more components of meaning are indissolubly linked in a single element of form, as in French *au* ([o]) ‘to the (masc.)’ or the ending *—o:* of Latin *amo:* which represents all of ‘first person singular present indicative’, a collection of categories that are indicated separately in other forms. When we look beyond the simple cases, it appears that the relation between form and meaning in the general case is not one-to-one at the level of the morpheme, but rather many-to-many.

In fact, it seems that even though both the forms and the meanings of words can be divided into components, the relation is still best regarded as holding at the level of the entire word, rather than localized exclusively in the morpheme. We have also seen support for this notion in the fact that entire words, presumably composed of multiple morphemes, develop idiosyncratic aspects of meaning that cannot be attributed to any of their component morphemes individually (e.g., *appreciable* and *considerable* come to mean not ‘capable of being appreciated/considered’, but ‘substantial, relatively large’). On this basis, many linguists have come to believe that morphological relations are based on the word rather than the morpheme. Actually, we need to take into account the fact that in highly inflected languages like Latin or Sanskrit, no existing surface word form may supply just the level of detail we need, since all such words have specific inflectional material added. For such a case, we need to say that it is *stems* (full words minus any inflectional affixation) that serve as the basis of morphological generalizations, in the sense of representing the phonological component of a lexeme.

Items and Processes

A further difficulty for the notion that morphemes are the basis of all morphology comes from the fact that in many cases, some of the information carried by the form of a word is represented in a way that does not lend itself to segmentation. One large group of examples of this sort is supplied by instances in which it is the replacement of one part of the form by another, rather than the addition of a new piece, that carries meaning. Such relations of *apophony* include *umlaut* (*goose/geese, mouse/mice*), *ablaut* (*sing/sang/sung*), and such miscellaneous relations as those found in *food/feed, sell/sale, sing/song, breath/breathe*, and many others. Terms for these relations often refer to their historical origins and do not reflect any particularly natural category in the modern language (e.g., *umlaut* as opposed to *ablaut* in modern English).

Sometimes some information is carried in a word’s form not by the addition of some material (a morpheme), but rather by the deletion of something that we might expect. In the Uto-Aztecan language Tohono O’odham (‘Papago’) for example, the perfective form of a verb can in most instances be found by dropping the last consonant of the imperfective form (whatever that may be): thus, *gatwid* ‘shooting’ yields perfective *gatwi* ‘shot’; *hikck* ‘cutting’ yields *hikc* ‘cut’, etc.

Examples like these (and several other sorts which considerations of space prevent us from going into here) suggest that the relations between words that constitute a language’s morphology are best construed as a collection of *processes* relating one class of words to another, rather than as a collection of constituent morphemic *items* that can be concatenated with one another to yield complex words. Of course, the simplest and most straightforward instance of such a process is one that adds material to the form (a prefix at the beginning, a suffix at the end, or an infix within the basic stem), but this is only one of the formal relations we find in the morphologies of natural languages. Others include changes, permutations, deletions, and the like. Linguists set on treating all morphological relations as involving the addition of morphemes have proposed analyses of many of these apparent processes in such terms, but it is possible to ask whether the extensions required in the notion of what constitutes an ‘affix’ do not in the end empty it of its original theoretical significance.

Conclusions

We have seen above that the forms of words can carry complex and highly structured information. Words do not serve simply as minimal signs, arbitrary chunks of sound that bear meaning simply by virtue of being distinct from one another. Some aspects of a word’s form may indicate the relation of its underlying lexeme to others (markers of derivational morphology or of compound structure), while others indicate properties of the grammatical structure within which it is found (markers of inflectional properties). All of these relations seem to be best construed as knowledge about the relations between *words* however: relations between whole lexemes, even when these can be regarded as containing markers of their relations to still other lexemes; and relations between word forms that realize paradigmatic alternatives built on a single

lexeme's basic stem(s) in the case of inflection. These relations connect substantively defined classes in a way that is only partially directional in its essential nature, and the formal connections among these classes are signalled in ways that are best represented as processes relating one shape to another.

Glossary

Allomorphy: The study of the various formal shapes that can be taken by individual meaningful elements ('morphemes'), and the patterns of such variation that characterize the grammar of a particular language.

Apophony: A meaningful relation between two words which is signalled not (only) by the addition of an affix, but also by a change in the quality of a vowel or consonant, a change which is correlated with the meaning difference in question rather than with the phonological shape of the form. For example, English *man* and *men* stand in an apophonic relation, since it is precisely the difference between the vowels of the two words that signals the difference between singular and plural.

Bahuvrihi (compound): Sanskrit term of a compound such as English *tenderfoot* which refers not to a kind of *foot*, but to an individual 'having or characterized by tender feet'. The word *bahuvrihi* is itself a compound of this type: it means literally 'much-rice', and refers to someone '(having) much rice'.

Morpheme: A hypothetical unit in the analysis of words, corresponding closely to the linguistic sign. To the extent it is possible to divide the form of every word exhaustively into a sequence of discrete chunks, to divide its meaning in a similar fashion, and establish a one-to-one correspondence between the components of form and those of meaning, each such combination constitutes a morpheme.

Morphotactics: The study of the patterns according to which minimal meaningful elements ('morphemes') can be combined to form larger units, particularly words.

(Linguistic) Sign: The basic unit in terms of which meaning is represented by form in language. The sign is 'minimal' in the sense that no sub-part of its form can be correlated with some particular sub-part of its meaning. The notion is central to the linguistic theory of Ferdinand deSaussure

Readings

Anderson, SR (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.

Aronoff, M (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Bybee, JL (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins.

Carstairs-McCarthy, A. *Current morphology*. London: Routledge.

Halle, M, & A. Marantz. (1993). Distributed morphology and the pieces of inflection. In K Hale and SJ Keyser [eds.], *The view from building 20*. Cambridge, MA: MIT Press. 111-176. Marchand, H. (1969). The categories and types of present-day English word-formation. Munich: C. H. Beck.

Matthews, PH (1991). *Morphology* (2nd edition). Cambridge: Cambridge University Press.

Pinker, S. (1999) *Words and rules*. New York: Basic Books.

Spencer, A. (1991). *Morphological theory*. Oxford: Blackwell.

Spencer, A., & AM Zwicky [eds.]. *The handbook of morphology*. Cambridge: Blackwell.