9 Morphology

ANDREW SPENCER

1 Introduction

Morphology is about the structure of words. All languages have words and in all languages some words, at least, have an internal structure, and consist of one or more *morphemes*. Thus, the form *cats* comprises the root morpheme "cat" to which is added the suffix morpheme "s" indicating plural. Now, for this characterization to mean anything we have to know what a word is. How do we know, for instance, that a string such as *the cat* is two separate words, and that *the* is not a prefix? Conversely, how do we know that the "s" of *cats* isn't a word in its own right. Here we need the help of syntax: *the cat* is a phrase which can be extended by addition of other phrases: *the very black cat*. The form *cats* can never be split up this way, the reason being that the "s" component is an element which can only exist as part of a word, specifically at the end of a noun. In other words, "s" is a suffix and hence a *bound morpheme*. The property of indivisibility exhibited by *cats* is *lexical integrity*. A single word such as *cats* contrasts rather neatly with the fully fledged (but synonymous) phrase *more than one cat*, in which it is clear that *more*, *than*, and *one* are all independent words and can all be separated by other words or phrases.

This chapter will examine the different structures words exhibit and the morphological relationships they bear to each other, and the nature of the morpheme. We begin by clarifying the notion "word" itself.

1.1 The lexeme concept

If we ask how many words are listed in (1) we can give at least two answers

(1) {cat, cats}

In one sense there are obviously two, but in another sense there is only one word, CAT, and only one entry will be found in a dictionary for it. The plural, *cats*, is formed by a completely general rule from the singular form *cat* and there is no need to record the plural form separately. In addition, we can describe *cat* as "the singular form of the word CAT" and *cats* as "the plural form of the word CAT." This gives us another interpretation for the term "word," as becomes clear when we look at the word "sheep." Here the singular form of the word SHEEP has exactly the same shape as the plural form, even though these are distinct linguistic entities. Given the vagaries of English orthography, this identity of shape can be true of the spoken form, the written form, or both (as with "sheep"). Thus, the written shape of the base form of the verb "read" (pronounced like "reed") is identical to that of the past tense, "read" (pronounced like "red") despite the difference in pronunciation, while *the taxes*, *the tax's* ("of the tax") and *the taxes'* ("of the taxes") differ solely in spelling.

It is rather useful to have different terms for these three different senses of the word "word." We will therefore say that there is a *lexeme* CAT which has two *word forms*, *cat* and *cats*. The names of lexemes are conventionally written in small capitals. The grammatical description "the singular / plural of CAT" is a *grammatical word*. Thus, *sheep* is one word form corresponding to one lexeme, but it is two grammatical words (the singular and the plural of SHEEP).

We can think of a lexeme as a complex representation linking a (single) meaning with a set of word forms, or more accurately, linking a meaning with a set of grammatical words, which are then associated with corresponding word forms. From the point of view of the dictionary (or lexicon), this is therefore a *lexical entry*. There is no demand here that the set of forms correspond to only one meaning, or that only one set of forms correspond to a given meaning. If several forms correspond to one meaning we have pure *synonymy*: e.g. {*boat, boats*}, {*ship, ships*}. If a single form corresponds to more than one completely unrelated meaning, as with {*write*, *right*, *rite*}, or {*bank*, *bank*} then we have *homophony* or *homonymy*. We then treat the homophones / homonyms as distinct lexemes which just happen to share the same shape (written and / or spoken). In some cases these meanings are felt as related to each other, and we have a case of *polysemy*. Thus, the word "head" means a body part, the person in charge of an organization, a technical term in linguistics, and so on, and these meanings are associated by some kind of metaphorical extension. In general, polysemy tends to be either ignored (where the meanings are close) or treated like homophony (but see below in section 3.2 on verbs like BREAK).

In linguistics a form-meaning pair is a *sign* and the lexeme is a prototypical example of a sign. The traditional definition of morpheme is "the smallest meaningful component of a word," and this entails that we consider all morphemes as signs. However, this turns out to be very controversial, for some types of morpheme, at least.

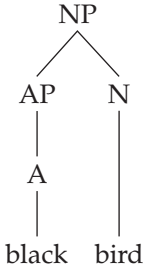
1.2 Inflection, derivation and compounding: preliminaries

In this section I briefly introduce certain important notions which will figure widely later: *inflection*, in which we create word forms of lexemes (such as the plural or past tense), *derivation*, in which we create new lexemes from old lexemes, and the *compound word*, a single word formed by combining two other words. We begin with compounds.

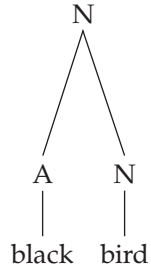
The most straightforward type of compound simply consists of two words concatenated together: *morphology + article = morphology article*; *house + boat = houseboat*. The right-hand member is the *head* of the compound, determining the syntactic category and meaning of the whole (a morphology article is a kind of article, a houseboat is a kind of boat, as compared with a boathouse, which is a kind of house). The left-hand member is the modifier. In transparent cases such as *morphology article* the meaning of the whole is derived from the meanings of the components (though the precise meaning is indeterminate and depends on the context of use).

There is an important distinction in many languages between compounds and phrases. In many cases the difference is obvious. In a hackneyed example such as *blackbird* as opposed to *black bird* the compound has stress on *black*, while the phrase is stressed on *bird* (in neutral contexts at least). Moreover, a black bird is necessarily black, while a blackbird is a particular species of bird whatever its colour (female blackbirds are brown, for instance). This means that the semantics of this compound is *non-compositional*, i.e. we can't determine the meaning of the whole just from the meanings of the parts. The semantics of phrases (idioms apart) is compositional. The difference can be illustrated syntactically as in (2, 3) (making very conservative assumptions about syntactic structure):

(2)



(3)



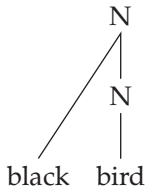
This is the standard story, though there are interesting subtleties. For instance, there is no way of determining the syntactic category of the modifier in *blackbird*, because it is fixed as part of the compound and can't be subjected to any of the morphological or syntactic manipulations that real adjectives can. Thus, compare (4) and (5):

(4) a. a very black bird
b. a blacker bird
c. a bird, black as coal, flew overhead

(5) a. *a very blackbird
b. *a blackerbird
c. *a black-as-coal-bird

Moreover, *black* doesn't mean "black" in *blackbird* (because a blackbird doesn't actually have to be black). Thus, the modifier *black* has neither category nor meaning; it just has a bare morphophonological shape. Therefore, (3) should be rewritten as (6):

(6)



The point is that *blackbird* is a lexicalized compound whose internal structure is only of historical significance, unlike a non-lexicalized coinage such as *morphology article*. In time, with changes in pronunciation, even this historical structure becomes opaque. Thus, *husband* is derived etymologically from (modern) "house" and "bond," but it isn't recognized as a compound by anyone except students of Middle English.

Nonetheless, noun + noun compounding is a fully *productive* process in English. Simplifying somewhat, we can say that a process is productive if it applies freely in principle to all the lexemes of the language of the relevant type, allowing new forms to be created at will even if they have never been used before. Such processes therefore have to be semantically regular, without any lexicalized idiosyncrasy of meaning, otherwise, hearers would have no way of knowing what a new coining was supposed to mean (see Aronoff and Anshen 1998, for more detailed discussion). The meaning of such compounds is admittedly vague: a morphology article is an article which has some connection with morphology. On the other hand, adjective + noun compounds aren't productive and there are virtually no verb + noun compounds (there is a tiny handful of exceptions like *swearword* and *drawbridge*).

A variety of types of productive compounding are known in the languages of the world. A particularly interesting type, which has been the subject of some debate in recent years, is that known as noun incorporation (see Mithun 1984). In noun incorporation we see an alternation in which the direct object of a verb may form a compound with that verb. In (7) we see two examples from Chukchee (a member of a small language group spoken in northeast Siberia):

(7) a. Gəmnan tə -piri -gʔen pojgə-n
I.ERG 1sgSUBJ-take-3sgOBJ spear-ABS

b. Gəm tə -pojgə-pere-gʔak
I.ABS 1sgSUBJ-spear-take-1sgSUBJ
"I took the spear"

In (7a) the subject pronoun is in the ergative case (the case used to mark the subject of a transitive sentence), while the object is in the absolutive case. Being transitive, the verb agrees with both the subject and the object. In (7b) the root of the object noun has formed a compound with the verb root. This renders the verb intransitive, so it agrees solely with the 1st person subject. The subject pronoun is now in the absolutive case, the case used for intransitive subjects. Finally, notice that the 1sg prefix comes to the left of the incorporated noun root and the vowels of the root have changed. This is due to vowel harmony, under which the "weak" vowel /i/ is changed to /e/ when there is a "strong" vowel elsewhere in the word (e.g. the /o/ of *pojgə*). Vowel harmony only operates within a word, and this helps us identify the incorporative complex as a single word form morphologically. Examples (7a, b) differ slightly in emphasis but are otherwise synonymous. Thus, it is clear that *pojgə* still realizes the "spear" lexeme even when it is compounded. Noun incorporation is completely productive in Chukchee, with very few restrictions.

Turning to derivation, the nouns *writer*, *painter*, *walker* are clearly related to the verbs *write*, *paint*, *walk*, meaning roughly "person who writes, paints, walks," by suffixation of *-er*. I shall call these *subject nominals*. It is customary to treat *write* and *writer* as distinct lexemes related by derivation, rather than word forms of a single lexeme. For instance, *writer* is a noun, while *write* is a verb. The morphological operations which realize derivation (such as *-er* affixation) may or may not be regular and productive. Thus, *apply* has a subject nominal *applic-ant*, with irregular suffix *-ant* added to an irregular form of the root, *applic-*. I discuss derivation in more detail in section 3.1.

As a verb lexeme, WRITE has its own set of grammatical words expressed by the forms *write*, *writes*, *writing*, *wrote*, *written*. Similarly, WRITER has its own set of forms: *writer*, *writers*. These grammatical words are the *inflected* forms of the lexeme and the process of constructing inflected forms is known as inflection ("inflectional morphology"). The meanings of the inflected form are predictable (plural of noun, past tense of verb, or whatever), while the shape of inflected forms is generally determined by affixation to the *stem* form of the lexeme. The stem consists of the root and any derivational affixes. In morphologically complex languages a given lexeme might have several stems for different types of inflection (for example, all verbs may have separate present tense and past tense stems). Irregularity, either in the stem or the affix, is not uncommon. Thus, *knife* has the irregular stem form *knive-* in the plural (*knives*), while *ox* has the irregular suffix *-en* (*oxen*). Irregularity of form can be complete as in *total suppletion*, when one inflected form bears no shape relation to the rest of the paradigm (e.g. *went* as the past tense of GO). Where there is still some overlap we talk of partial suppletion (as in *brought ~ bring*, where the first two consonants are identical). Even where the shapes are irregular, the past tense meaning is exactly the same as it is for any other verb, whether

irregular (such as *write ~ wrote*, *bring ~ brought*, *go ~ went*) or regular (e.g. *scribble ~ scribbled*).

Inflections express grammatical or *functional categories*. The inflectional system organizes the forms of words into systematic groupings or *paradigms*. There are essentially two sorts of function subserved by inflection. Many inflections signal an aspect of meaning which is *grammaticalized*, such as number (singular vs. plural) or tense. This means that the words of a given class obligatorily signal the grammatical distinction: thus, all verbs in English have to have a past tense (even if these are not actually distinct forms, as in *put*). Booij (1994) refers to this as *inherent inflection*.

One typical inherent inflection for nouns is *case*, in which the grammatical or semantic role of a noun in a sentence is shown by its form. In Russian a noun generally has distinct forms for the subject, direct object or indirect object:

(8) Len -a dala Ir -e knig -u
Lena-NOMINATIVE gave Ira-DATIVE book-ACCUSATIVE
"Lena gave Ira a book."

Lena, *Ire*, *knigu* in (8) are case-inflected forms of the lexemes LENA, IRA, KNIGA.

Verbs exhibit much greater variety in their inflectional systems. Two common inherent inflections are *tense* and *aspect*. Tense refers to anchoring in time, as with English *wrote* (past) as opposed to *writes* (non-past – present or future reference). A given language may distinguish a number of different tenses (such as recent vs. remote past) or no tense at all. Aspect refers to the manner in which an event unfolds over time. A very common aspectual distinction is that between completed (perfective) and non-completed (imperfective) events. In Slavonic languages most verbs have separate perfective and imperfective paradigms, e.g. *op'isat'* (perf.) ~ *op'isivat'* (impf.) "describe" (see also section 3.2). Many languages have very rich aspectual markings modifying the meaning of the base verb in very subtle ways. Below is just a small selection of the fifteen aspectual affixes described for Chukchee by Skorik (1977: 179–202):

(9) -lʔet prolonged continuous action:
ʔəttʔe ninepiŋku-*lʔet*-qin . . . ottəlgən
dog jump-*ASP*-3/3 stick
"The dog jumped over the stick *over and over again*."

(10) -cir prolonged interrupted action:
. . . ŋinqejmuri nʔejŋew-*cir*-muri jaralʔa
us.children called-*ASP*-1plOBJ people.at.home
"The people at home *kept* calling us children."

(11) -cit / -cet alternating action:
. . . natcə-*cet*-qenat . . .
hide-*ASP*-3plSUBJ
"They played at hide-and-seek"

(12) -skəcet accelerated action:
qənwer ŋəto-*sqəcat*-gʔe gəmnin təletumgin
at last come.out-*ASP*-3sgSUBJ my companion
"At last my companion sprang out"

More than one of these can be combined:

(13) mət-ra-təla-tenmawə-plətko-ŋŋo-gʔa
1pl-FUT-GRADUALLY-prepare-FINISH-BEGIN-FUT
"we will *begin to gradually finish* the preparations"

Other types of verb inflection include mood (whether a statement is presented as fact, possibility, hypothetical situation and so on) such as the subjunctive mood of Romance languages, the optative expressing a wish (e.g. Ancient Greek), imperative for issuing commands, and interrogative, a special set of verb forms used for asking questions (e.g. the Eskimo languages). Many language groups signal polarity (negation) inflectionally (Bantu, Turkic, Athapaskan, and others). It is very common for a given inflectional morpheme to signal a complex mixture of tense, aspect, mood, and polarity.

Any of the above functional categories can be expressed syntactically, by word order or by function words such as the English aspectual auxiliaries (*has been reading*). One purely morphological type of inherent inflection is *inflectional class*: *declensions* for nouns and adjectives and *conjugations* for verbs. Which noun or verb goes in which class is in general arbitrary. Russian nouns can be put into four main declensions depending on the inflections they take (though different descriptive traditions distinguish different numbers of declensions):

(14) Russian noun classes

	Class I inanimate	Class I animate	Class II	Class III	Class IV
	"law"	"boy"	"room"	"bone"	"place"
Singular					
Nominative	zakon	mal'čik	komnata	kost'	mesto
Accusative	zakon	mal'čika	komnatu	kost'	mesto
Genitive	zakona	mal'čika	komnati	kost'i	mesta
Dative	zakonu	mal'čiku	komnate	kost'i	mestu
Instrumental	zakonom	mal'čikom	komnatoj	kost'ju	mestom
Prepositional	zakone	mal'čike	komnate	kost'i	meste
Plural					
Nominative	zakoni	mal'čiki	komnati	kost'i	mesta
Accusative	zakoni	mal'čikov	komnati	kost'i	mesta
Genitive	zakonov	mal'čikov	komnat	kostej	mest
Dative	zakonam	mal'čikam	komnatam	kost'am	mestam
Instrumental	zakonam'i	mal'čikam'i	komnatam'i	kost'am'i	mestam'i
Prepositional	zakonax	mal'čikax	komnatax	kost'ax	mestax

(The symbol ′ represents palatalization. Consonants are always palatalized before /e/. The case names are traditional and represent a variety of syntactic functions.) I have given two subtypes of class I nouns, one animate the other inanimate. In the inanimates the accusative case is always the same as the nominative, while in the animates the accusative takes the form of the genitive. This type of situation, in which parts of a paradigm are systematically identical, is known as *syncretism*. There are other syncretisms here, too. For instance, the dative, instrumental and prepositional plural endings are the same for all classes, that is, the class feature is neutralized and there is effectively a single set of endings for the whole of the class "noun." On the other hand, the behavior of pairs such as "law" and "boy" require us to set up a covert category of animacy for Russian, which never has any direct expression (there is no form which has a suffix identifiable as the "animacy" suffix) but which is nonetheless part of the inflectional system. Note that it is the property "animacy" which is covert, not the accusative case. We know this because class II nouns have a separate accusative, in the singular at least (see Corbett and Fraser 1993, for more detailed discussion of the implications of these data).

Russian verbs inflect so as to indicate the person and number of their subject (see below on "agreement") as well as for tense and occur in two main conjugations (together with a plethora of minor variations on each of these classes):

(15) Principal Russian verb classes

Class I verb *uznat′* "to recognize", class II verb *govor′it′* "to speak"

	Class I	Class II
Non-past tense		
1sg	uzn-aj-u	govor′-u
2sg	uzn-aj-o-š	govor′-i-š
3sg	uzn-aj-o-t	govor′-i-t
1pl	uzn-aj-o-m	govor′-i-m
2pl	uzn-aj-o-te	govor′-i-te
3pl	uzn-aj-ut	govor′-at

As can be seen, the endmost suffixes are common to both classes, except in 1sg, 3pl forms. Both types have a special stem forming suffix, *-aj-* and *-i-* respectively, and class I has in addition a "linking vowel" *-o-*. The *-aj/-i* formatives are found throughout the inflectional system of the verbs.

The other role of inflection is to realize the syntactic functions of *agreement* and *government*. This is what Booij (1994) calls *contextual inflection*, because it is determined by the syntactic context in which the lexeme is used. In many languages a verb must *agree with* its subject and / or object, by cross-referencing various of their properties. This occurs marginally in English for third person non-past verb forms: *Harriet writes* vs. *the girls write*. In Chukchee transitive verbs agree with both the subject and the object, in rather complex ways. The system for one of the six tense forms in the indicative mood is shown in

(16) (see Muravyova 1998; empty cells represent non-existent forms in which the subject and object would have the same person features):

(16) *pela-* "to leave (someone, something)" Simple past

Subj	Obj		Subj	Obj	
1sg	1sg	----------	1pl	1sg	----------
	2sg	tə-pela-gət		2sg	mət-pela-gət
	3sg	tə-pela-gʔan		3sg	mət-pela-gʔan
	1pl	----------		1pl	----------
	2pl	tə-pela-tək		2pl	mət-pela-tək
	3pl	tə-pela-nat		3pl	mət-pela-nat
2sg	1sg	ena-pela-gʔe	2pl	1sg	ena-pela-tək
	2sg	----------		2sg	----------
	3sg	pela-gʔan		3sg	pela-tkə
	1pl	pela-tko-gʔe		1pl	pela-tko-tək
	2pl	----------		2pl	----------
	3pl	pela-nat		3pl	pela-tkə
3sg	1sg	ena-pela-gʔe	3pl	1sg	na-pela-gəm
	2sg	na-pela-gət		2sg	na-pela-gət
	3sg	pela-nen		3sg	na-pela-gʔan
	1pl	na-pela-mək		1pl	na-pela-mək
	2pl	na-pela-tək		2pl	na-pela-tək
	3pl	pela-nenat		3pl	na-pela-nat

The verb references the person and number both of the subject and of the object, though there is no simple relationship between many of the affixes and their functions. Thus, although the prefixes *tə-* and *mət-* clearly meaning "1sg/1pl subject" respectively, the prefix *na-* seems to mean "3pl subject" or "3sg subject with 2nd person object or 1pl object" and the suffix *-nen* seems to mean "3sg object but only if the subject is 3sg." One consequence of this is that some forms correspond to more than one subject–object pairing, e.g. *napelagət*, which means either "3sg leaves 2sg (s/he leaves thee)" or "3pl leaves 2sg (they leave thee)." The system proves to be even more complex than this when the full set of tenses, moods, and voices is taken into account. Patterns such as this are typical of languages with rich agreement systems, and such data have been instrumental in changing the views of linguists about the nature of the morpheme.

Adjectives often agree with the nouns they modify. This is extremely marginal in English, only being found for *this* and *that* (*this / that cat* vs. *these / those cats*). In Russian, however, an adjective agrees with its noun in number and case:

(17) a. bol'šoj mal'čik — Masculine nominative singular
big boy
b. bol'šogo mal'čika — Masculine genitive singular
c. bol'šim mal'čikam — Masculine dative plural

(18) a. bol'šaja devuška — Feminine nominative singular
big girl
b. bol'šoj devuški — Feminine genitive singular
c. bol'šim devuškam — Feminine dative plural

It might be thought that the adjective agrees in declension, but this is wrong. All nouns in Russian have one of three *genders*, masculine, feminine, or neuter. Male and female humans are masculine and feminine respectively and for other nouns gender depends largely on declensional class. Members of class I are masculine, those of classes II, III are feminine and those of class IV are neuter. However, there are certain exceptions. Thus, the word *mužčina* "man" belongs to class II, yet it is masculine: *bol'šoj mužčina* "big man." As is stressed by Aronoff (1994), gender is an essentially syntactic property, which governs agreement. Declension class is a purely morphological property which the syntax has no direct access to. Aronoff points out that the existence of arbitrary inflectional classes is one of the prime motivations for treating morphology as an autonomous linguistic module.

We have seen that a direct object in Russian is in the accusative case. This can be thought of as an instance of *government*: a transitive verb *governs* the accusative. Likewise, prepositions in Russian have to take specific cases, as shown in (19):

(19) a. okolo dom-a
near house-GENITIVE
"near the house"
b. v dom
in house.ACCUSATIVE
"into the house"
c. v dom-e
in house-PREPOSITIONAL
"in the house"

Notice how "motion towards" as opposed to "location at" is signaled solely by case choice in (19b, c), otherwise, it is an arbitrary matter which preposition governs which case.

One of the perennial theoretical problems in morphology is whether there is a clear-cut distinction between inflection and derivation and if so how to draw it. Inflection is often thought to be "of relevance to syntax," which is clearly true of contextual inflection, but not so obvious with inherent inflection. Yet we don't want to say that plurals or past tenses are derivational and hence create new lexemes. Booij's contextual / inherent distinction is designed to ameliorate this problem (though we are now left with the task of distinguishing inherent inflection from derivation). A typical borderline case is that of the aspectual forms of Chukchee given above. Chukchee has a set of six tense-aspect forms in which aspect (roughly perfective vs. imperfective) is

grammaticalized and expressed as part of the obligatory conjugation system. However, the affixes illustrated in (9–13) are not like this. Rather, they are optional elements which are added to modify the overall meaning of the verb. Does this make them derivational, then? Do we wish to say that "to *verb* in a prolonged interrupted fashion" is a new lexeme related to *verb* (derivation) or a form of the word *verb* (inherent inflection)? Cases like this are quite common and promise to provide fertile ground for future research into the problem.

2 The Morpheme Concept and Agglutinating Morphology

2.1 Item-and-arrangement morphology

If we return to the example of *writer* we can easily segment it into two component forms or *morphs*, a verb *base*, *write*, and a derivational suffix *-er*. (I use "base" as a catch-all meaning anything to which an affix is added, whether derivational or inflectional). It is usually claimed that the suffix as well as the base has a meaning and that the meaning of the derived word is obtained by combining the meanings of the two component morphs as shown in (20):

(20) [[WRITE] PERSON WHO]

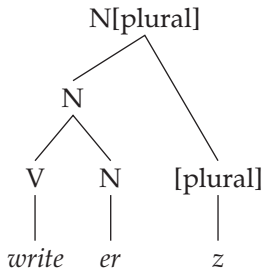
write er

On this basis both of the morphemes are a pairing of a pronunciation (or shape, the morph) and a meaning. They are thus signs and hence are both lexemes, making the combination essentially a compound, like *houseboat*. Admittedly, *-er* is a bound morpheme, but in many languages lexemes can be compounded in the form of a bound stem. Thus, the form *pojgə* in the Chukchee noun incorporation example (7b) is in fact a bound stem form (the word for "spear" itself always surfaces with a case suffix), and even in English one might argue that there are compounds consisting solely of bound roots, the so-called neoclassical compounds such as *gram-o-phone* (or *phon-o-gram*). The traditional account of plural morphology treats the plural suffix in the same way, a type of sign with a phonology and a semantics, as shown in (21):

(21) *-z* = </z/, [plural]>

This way of looking at things immediately leads us to the conclusion that words have a hierarchical structure which can be represented as a tree diagram. A possible structure for *writers* is shown in (22):

(22)



In (22) the grammatical property [plural] is said to *percolate* from the suffix to the top of the tree, ensuring that the entire word is interpreted as a plural form.

The set of assumptions underlying representations such as (22) derive from what is generally called the *item-and-arrangement* theory (IA for short): morphemes are "things" which are arranged in a particular way ("morphotactics") and which contribute their meaning to the meaning of the whole word. In an "ideal" morphological system each morpheme contributes one meaning and each meaning is associated with just one morpheme ("one form – one function"). Such a morphological ideal is often called *agglutination* (and morphologists still sometimes speak of "agglutinating" languages where this type of morphology predominates).

It should be obvious that this approach is at odds with the lexeme concept: the plural form *cats* would not, after all, be a word form belonging to an abstract lexeme, CAT, rather it would be a compound form, in which the meaning of the suffix (or perhaps we should say the head of the compound?), PLURAL, is grammaticalized. Where inflection is concerned this has proved impossible to maintain, for three main reasons. First, it is not always possible to identify a single segmentable morph for the putative morpheme; for instance, where is the plural morpheme in *men* (see section 2.2.2)? Second, there are significant deviations from the form-meaning pairing in affixation and these undermine the assumption that inflections are signs. Third, for such a theory to work we must be able to explain in a satisfactory way how complex words are constructed, and in particular how the morphemes get strung out in the right order. For complex inflectional systems this turns out to be very tricky.

2.2 Deviations from agglutination

The "ideal" type of morphology, then, is often seen as the addition of a semantically transparent affix to a base, so-called *concatenative* morphology. There are several ways in which morphological systems present deviations from the agglutinating ideal of one form – one function. The first of these is caused by the fact that a given morpheme may have more than one shape (allomorphy). Beyond this, we find that there are operations which can't easily be analyzed as the addition of a meaningful element but rather take the form of a phonological process, often called *non-concatenative* morphology. Languages abound

in such operations and there have been a number of ingenious ways of dealing with them. I shall mention just three particularly salient cases here (introductory discussion of different types of operations can be found in Bauer 1988, Spencer 1991, 1998).

2.2.1 *Allomorphy*

The regular past tense ending appears as three different morphs depending on the final sound of the verb stem: *walk-ed* (/t/), *jogg-ed* (/d/), *trott-ed* (/əd/, where /ə/ is the schwa or reduced vowel). This variation is *allomorphy*, and we say that (/t, d, əd/) are the three *allomorphs* of the past tense morpheme. In this case the allomorphy is conditioned solely by the phonology of the stem: /əd/ after /t, d/, /t/ after a voiceless sound, /d/ elsewhere. Other cases of allomorphy may be irregular. For instance, while *mend* and *pen* have regular pasts, *mended*, *penned*, the verb *bend* takes an unexpected *-t* ending and adds this to an irregular stem form lacking the final *-d*: *ben-t*. Thus, both stem and suffix show irregular allomorphy. Where a given morpheme is realized by more than one allomorph we have a (mild) deviation from the agglutinative ideal.

2.2.2 *Processual morphology*

Certain types of irregular verb in English form their past tense by taking the basic root, *sing*, *run*, *drive*, *write* and changing its vowel: *sang*, *ran*, *drove*, *wrote*. This kind of process is called *ablaut* or *apophony*. In a number of languages, most famously Semitic languages such as Arabic and Hebrew, apophony is regular and widespread throughout the grammar. It is very difficult to represent this in terms of the addition of an affix to a base (though see McCarthy 1982, for the classic item-and-arrangement analysis of Semitic). Another well attested phenomenon is *reduplication*, illustrated by the Tagalog examples in (23):

(23)	a.	sulat	"writing"	su-sulat	"will write"
	b.	basa	"reading"		
		mambasa	infinitive	mam-ba-basa	nominalization
	c.	magpa-sulat	causative	magpa-pa-sulat	"will cause to write"

Here, morphological categories are signaled by a kind of prefix, which consists of a copy of certain of the segments of the stem. Any analysis of this phenomenon has to recognize that there is a process involved at some level (see McCarthy and Prince 1998, for a summary of some recent proposals).

A particularly drastic type of non-affixal morphology is so-called *subtractive* morphology in which a morphological category is signaled by loss of a portion of the base. Anderson (1992: 64–6) lists a number of inflectional processes which, apparently, have to be so analyzed, such as the example in (24) from the Muskogean group:

(24)

a.	balaa-ka "lie down (sg.)"	bal-ka "lie down (pl.)"	(Alabama)
b.	bonot-li "roll up (sg. Obj.)"	bon-li "roll up (pl. Obj.)"	(Choctaw)
c.	atakaa-li "hang (sg.)"	atak-li "hang (pl.)"	(Koasati)

Here, the plural or plural object form of the verb is derived from the singular form by removing the rhyme of the final syllable of the stem: *bal<aa>*, *bon<ot>*, *atak<aa>*.

2.2.3 Form: meaning deviations

In this subsection we examine the idealization that one form corresponds to one meaning / function and vice versa. We already know of two types of deviation: synonymy (many forms – one meaning) and homonymy (one form – many meanings). However, four additional types of deviation can be distinguished when we look at the meanings or functions of morphemes within a single word.

2.2.3.1 One morph, two meanings

The Russian case system shown in (14) clearly has a grammatical category of "plural" but no single identifiable morpheme signaling number. Thus, *-am* means "dative" and "plural" simultaneously. Note that this is not homonymy, because the suffix simultaneously conveys both meanings within the same word form and these meanings are inseparable. We say that the morph shows *fusion* or *cumulation* of two separate meanings.

2.2.3.2 One meaning, two morphs

One and the same function can be signaled (redundantly) by different morphs in a given word. A simple example is found in Latin:

(25) Latin verbs: *amo* "I love / I have loved"

P/N (sg.)	Present	Perfect
1st	am-o	am-a-v-i
2nd	am-a-s	am-a-v-isti
3rd	am-a-t	am-a-v-it

The *-v-* morph realizes perfect tense, and has no other function, so we can say that *-v-* is the *principal exponent* of (perfect) tense. However, the 1sg endings also differ with tense, and thus serve as secondary exponents of this category. This means that the meaning of "perfect tense" extends over both *-v-* and *-i* in *amavi*. This is often referred to as *extended exponence*.

2.2.3.3 One meaning, no morph: null morphemes

Notice that there is no ending in the genitive plural of Russian class II and IV nouns in (14). In a morpheme-based theory we must say that this property

set, "genitive plural, class II/IV," is signaled (cumulatively) by a *null* or *zero morpheme*: *komnat-Ø*. Similarly, in derivation we often find cases of *conversion*, in which a word belonging basically to one category (such as the noun *chair* or the verb *run*) is used in another (the verb *to chair*, the noun *a run*). Given agglutination, this, again, would have to be handled by assuming a null morpheme.

2.2.3.4 One morph, no meaning

A traditional type of meaningless morpheme is the famous *cranberry morph*. Words such as *blueberry*, *blackberry*, *cloudberry*, *cranberry* etc. are clearly compounds of *berry* and refer to types of berry, but what does "cran" mean? More subtly, we saw that the *black* of *blackbird* doesn't have any meaning, strictly speaking. Aronoff (1976) argues in detail for English that there are cranberry morphs which have morphological properties (show allomorphy) and which therefore have to be regarded as morphemes. Thus, a verb such as *understand* is derived morphologically from the prefix *under-* (as in *underwrite*, *undertake*, *undermine*, . . .) and *stand* (as in *withstand*). This is clear because they have the same irregularity in the past tense as the base verb (*understood*, *withstood*). However, neither the prefix nor the base preserves its meaning, or any meaning. I return to such cases in section 3.2.

Cranberries are the examples of meaningless morphs most often cited, but the phenomenon is actually more widespread and more subtle. Thus, the adjectives in (26) illustrate a case in which a morpheme can be said to be meaningful only by stretching the meaning of "meaning" rather uncomfortably:

(26)

Noun	Adjective	
morphology	morphological	morphological theory
navy	naval	naval uniform
poetry	poetic	poetic license
nerve	nervous	nervous system

These are different from normal adjectives ending in the same suffixes such as *topical*, *sympathetic*, or *adventurous* in that they don't express qualities or properties. Thus, we can say *very topical article*, *unsympathetic remark*, *highly adventurous project* but the adjectives in (26) can't be modified in this way: **very morphological theory*. The reason is that the adjectives in (26) are really no different from the basic nouns but used in the syntactic contexts where an adjective is needed, i.e. to modify a noun. Indeed, in a number of cases we can idiomatically replace such phrases with compounds: *morphology theory*, *navy uniform*, or marginally *nerve system*. Thus, the derivational morphology which creates the adjectives changes the syntactic category of the word but doesn't add any element of meaning and thus, strictly speaking, is a kind of cranberry suffix. This type of category-shifting morphology is often referred to as *transposition*.

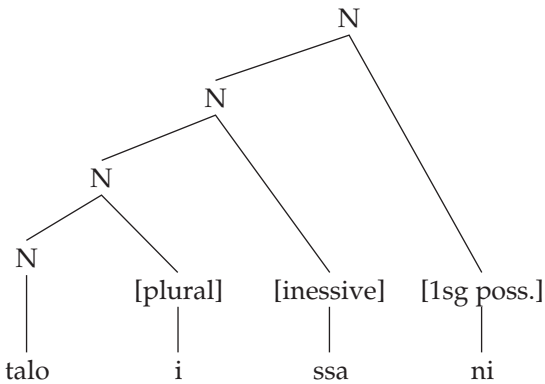
2.3 Morpheme order

The order of morphemes in a word is usually strictly determined, even in languages with very free word order and linguistic theory has to have some set of mechanisms for guaranteeing this order. A simple example of polymorphemic inflection is provided by nouns in Finnish, a typical "agglutinative" language. It has number, case, and possessor inflection on nouns, with a separate formative for each function. Some examples of inflected forms of the word *talo* "house" are given in (27) (data from Karlsson 1987):

(27) talo-ssa-ni "in my house"
talo-lle-mme "onto our house"
talo-i-sta-si "out of thy house"
talo-i-lta-nne "off of your houses"

On a morphemic account a word form such as *taloissani* "in my houses" would have the form (28):

(28)



How do we ensure that the morphemes come in this order and not, say, **talossaini*, or **talonissai*?

Lieber (1992) argues that we can make use of the syntactician's notion of *subcategorization* or *selectional frame*. Each ending is given a frame in its lexical entry stating what kind of stem it can occur to the right / left of. For possessor suffixes the frame stipulates that they occur next to a stem marked for case, while case suffixes are marked to occur next to Number-marked stems. A possible entry for the inessive suffix *-ssa* is shown in (29):

(29) <*ssa*, [CASE:Iness], . . . >: [$_{N}$ [NUMBER:{Sg, Pl}]_____]

Notice that we have to allow for the Number specification to be either singular or plural. (An alternative would be to invent some notation meaning "any value of the feature NUMBER.")

Finnish, however, presents a problem for an approach of this sort. The point of the subcategorization approach is that the addition of a suffix is dependent on the structure of the stem as built up so far. The nominative case ending is zero in the singular and *-i* in the plural, so the nominative plural of *talo* should be **taloi*. However, the real form is *talo-t*. The first problem is interpreting the meaning of the *-t* suffix: is it a plural marker found only in the nominative (30a) or a nominative marker found only in the plural (30b)?

(30)		root	NUMBER	CASE
	a.	talo	t_{Pl}	$Ø_{Nom}$
	b.	talo	$Ø_{Pl}$	t_{Nom}

In analysis (30a) how does the grammar know that *-t* and not *-i* must be inserted after the stem? In analysis (30b) how does the grammar know that *-Ø* and not *-i* must be inserted after the stem?

The only way around this problem is to reject the assumptions of agglutination and say that the *-t* formative appears next to the stem but cumulates case and number, as in (31):

(31)

root	NUMBER / CASE
talo	$t_{Nom,Pl}$

However, this will lead to considerable complications because we will now have a stem marked for plural to which other case markers could attach. We must therefore impose some principle saying that once a form is marked for case it can't be marked again. Indeed, Lieber (1992) introduces essentially such a scheme by means of the "categorial signature," a significant departure from the item-and-arrangement model for inflection which makes it to all intents and purposes just a variant of the model to be discussed below.

2.4 Rule function morphology

The morpheme concept thus serves even agglutinative languages like Finnish rather badly. An alternative conception has been argued for by many morphologists (see Anderson 1992, Aronoff 1994, for examples and surveys of the literature), under which affixation is just one of a set of phonological operations which can be performed on a base, triggered by the inflectional properties of the whole word. This is a variant of the classical *item-and-process* approach to morphology but I shall refer to it as the *rule function* approach (since processes are stated as rules which are usually thought of as something like functions in the mathematical sense).

To handle Finnish nouns we would set up a battery of rules applying in three blocks, one essentially for number marking, one for case and one for

possession. We start with a complete inflectional characterization of the word, say, {plural, inessive, 1sgpossessed}. This triggers three rules which take the current stem and add the appropriate suffixes:

(32) a. {Plural}(X) = X + i
b. {Inessive}(X) = X + ssa
c. {1sgpossessed}(X) = X + ni

A rule such as {plural}(X) is intended to be read "the plural form of (the base form of) X is constructed by adding *-i* to X." Taking the base *talo* these build up the required form in the obvious way. Note that there is no need for subcategorization since morpheme order is reflected in the ordering of the rule functions. The "zero affixes" of the singular and the nominative are handled very naturally: there is no rule corresponding to these properties, therefore nothing is done to the base form. There is thus no need for dubious constructs such as strings of null morphemes.

The problem with nominative plural forms is handled by having a more specific (a) rule:

(33) a. (i) {Plural, Nom}(X) = X + t
(ii) {Plural}(X) = X + i

The rules in (33) are regulated by a very general principle (often called the *elsewhere condition*): if two rules can apply to the same base, it is the more specific which wins out. Confronted with the need to construct the nominative plural form, rules (33a(i) and (ii)) are both applicable, but only (33a(i)) applies because it is the more specific. Rule (33a(ii)) is the rule which applies "everywhere else" or in the *default* case. The use of the notion of default in such systems has become extremely important in recent research and some form of it is even accepted by protagonists of a (very highly modified) type of item-and-arrangement theory (Halle and Marantz 1993, Wunderlich 1996).

The rules of (32) are affixations but could just as easily be any phonological process, such as vowel ablaut or reduplication. The rule function approach rejects the idea that affixes are lexemes compounded with base lexemes. Rule functions can also handle syncretisms very neatly. There can be a problem for morpheme-based theories because they are often defined over parts of the paradigm independently of the actual affixes. A simple example of this is the relationship between the perfect participle (*has cooked*) and the passive participle (*was cooked*). These are identical for all verbs despite variation in form (*has / was written*, *has / was sung*, etc.). It is hard to see how the generalization can be stated over such diverse "lexemes" as *-ed*, *-en*, *u*-ablaut and so on. This can be captured in the morphology by stating a *rule of referral* for constructing passive participles from perfect participles as in (34):

(34) {Passive participle}(X) = {Perfect participle}(X)

As Stump (1993) argues, (34) is just another type of rule function and can interact with other rule functions in a variety of ways. To achieve the same effect, morpheme-based theories have to undermine the morpheme concept fairly drastically, for example by reconstructing a rule of referral over the meaning content of morphemes independently of their form (as in the rules of impoverishment of Halle and Marantz 1993).

3 The Structure of the Lexeme

3.1 Derivational morphology

We saw in section 1.2 that compounds show varying degrees of semantic transparency: *morphology article*, *blackbird*, *husband*. Much the same can be said of derivational morphology. In (35) we see examples of fully transparent, compositional derivation:

(35) a. cat-like
b. elephant-like
c. lion-like
d. ape-like

These all mean roughly "like a typical X" where X = {cat, elephant, lion, ape, . . . }, and we can call such forms *similitudinal adjectives*. This derivation is highly productive, in that "X" can be virtually anything: speakers can understand and use a coinage like *iguana-like* without ever having heard it before (though the exact force of *-like* is rather subtle, e.g. what is the precise semantics of *an iguana-like expression / skin / gait*?). This type of suffixation is so transparent that it resembles compounding – indeed, some might claim that *-like* adjectives are in fact compounds (though of a very rare type in English). Now contrast (35) with the examples in (36):

(36)	a.	cat	catty	cat + y
	b.	elephant	elephantine	elephant + ine
	c.	monster	monstrous	monst(e)r + ous

Although these may also be similitudinal adjectives, they are not compositionally derived from their bases. This means that the base has lost all meaning and functions much like the cranberry morpheme *stand* in *understand*.

Much derivational morphology is similar to that of (36), if not quite so dramatically so. For instance, *hopeless* means "without hope," but this isn't true for all uses. If we say "I wouldn't pick him for the job, he's hopeless" we

are saying that he is extremely bad and unsuitable. He himself is not necessarily without hope in the literal sense (he could be one of life's irrepressible optimists). Such cases are the norm and it turns out that there is a cline of transparency running from *cat-like* to *hopeless* to *catty* to *understand*.

3.2 Four types of lexical relatedness

Transparent derivational morphology defines a network of relatedness amongst lexemes, but it is only one of four types of relatedness, one in which morphological relatedness goes hand in hand with semantic relatedness. The second type of relatedness is that mediated solely through semantics, without any morphological relatedness. For example, there are similitudinal adjectives which mean "like *noun*" but which aren't morphologically derived from *noun* e.g. *infantile* (= / = "like an infant"!), or *puerile* both meaning "child-like" (with additional pejorative overtones). These could be said to end in an affix *-ile*. I cited *writer* in section 2.1 as an example of a subject nominalization, and this represents a very productive formation, but not all verbs permit it. The subject nominal corresponding to the verb *fly*, as in "fly a plane (professionally)," is an entirely different word, *pilot*. Admittedly, the form *flier* exists for other uses of the verb, but we wouldn't say for instance **Tom was the flier of that Boeing 747*. This is a kind of derivational suppletion.

The third type of relatedness is represented by *systematic polysemy*, that is, where we have two different lexemes with the same form. A familiar example is the alternation shown in (37) (referred to variously as *inchoative*, *causative*, or *anti-causative*). Most linguists would probably say that there are two distinct, though related, BREAK lexemes here:

(37) a. Tom broke the vase.
b. The vase broke.

In many languages such usages are conveyed morphologically (by what is usually considered derivational morphology). Notice that the verb retains all its purely morphological properties in both usages, so there is no conversion or affixless derivation in the normal sense here. Rather, we seem to have two closely related lexemes which share all the same word forms.

In section 2.2.3 I pointed out that both the prefix and base of *understand* are cranberries. This is the fourth type of derivation in which there is clear morphological relatedness but no semantic connection (*asemantic relatedness*). In Russian this phenomenon is pervasive. Most Russian verbs are derived by prefixation of a few hundred simplex verb roots. In some cases the prefix and root contribute to the overall meaning compositionally, but in many cases it is impossible to ascribe a clear meaning to either root or prefix, just as with *understand*. Thus, from the verb *p'isat'* "write" we obtain *pr'i-p'isat'* "ascribe,"

o-p'isat' "describe," *pod-p'isat'* "sign," and so on. Each of the prefixes occurs in hundreds or thousands of verbs (sometimes with identifiable meaning). Moreover, each of these formations belongs to the same special subclass of class I that *p'isat'* belongs to (thus, the present tense stem is *p'iš-*, rather than the default stem form for class I, **p'isaj-*), and they all show the same pattern of stress shift as the base verb. Finally, they all behave like morphologically prefixed forms. We can see this because the base verb, *p'isat'*, is imperfective in aspect, but nearly all prefixed verbs in Russian are perfective and form a special secondary imperfective (usually by suffixation of *-iv-*). Sure enough, all the derived verbs from *p'isat'* are perfective and form their secondary imperfective in *-iv-*: *pr'i-p'is-iv-at'*, *o-p'is-iv-at'*, *pod-p'is-iv-at'* etc. Russian verbs thus present a much more convincing demonstration of Aronoff's original point because there can be no doubt that the verbs are prefixed, and because the base exists in nearly all cases as an independent verb with exactly the same morphological properties; the majority of the native verb lexemes in the language are like this (indeed, the majority of monomorphemic, non-prefixed verb stems are loans).

This property of the Russian lexicon is particularly damaging to the classical morpheme theory. Of the 28,500 or so verbs in Zaliznjak's *Grammatical Dictionary of Russian*, roughly 24,000 are prefixed. Of these a large proportion are highly regular aspectual or Aktionsart formations (which could be claimed to be more like inflections than lexeme forming derivations). Given this, I would estimate between a third and a half of these 24,000 are like the derivates of *p'isat'* discussed above. Thus, if we consider cases which are uncontroversially independent lexemes it turns out that the *majority* of Russian verbs consist of a cranberry prefix and a cranberry root. The significance of such cases has been significantly underplayed in the literature.

One might wish to claim that there is a fifth type of relatedness illustrated by denominal verbs in English formed by conversion (though deverbal nouns, such as *a bite* or *a broadcast* would do as well). Verbs such as *to saddle (a horse)*, *to shelve (books)*, *to skin (a rabbit)*, *to paper (a wall)*, and many others are clearly derived from nouns, but without any overt morphology. Presumably we would wish to say that this creates new lexemes (it seems far-fetched to regard *to saddle* as an inflected form of the base noun) and hence constitutes a derivational relation. However, this can either be regarded as a type of derivation which happens not to involve morphology (and hence a subtype of standard derivation) or a subtype of systematic polysemy.

In sum: lexemes can be related to each other by (1) morphology which induces a compositional meaning change; (2) systematic meaning relation which is not matched by any formal relatedness (*suppletive derivation*); (3) systematic meaning relation between different meanings associated with the same form (*systematic polysemy*); and (4) purely in terms of shape, *asemantic relatedness*. These extremes define a space within which word relatedness can vary, so that *catty* could be said to be an example of derivation with respect to the suffix but asemantic relatedness with respect to the base.

3.3 Mixed categories

We end this section by looking at a set of cases which occupy a borderline position in some sense, and which are currently the focus of a good deal of research effort. A very common form of verb-to-adjective transposition is illustrated by *participles*. These are adjectival forms associated with verb lexemes, often expressing verbal tenses, aspects or voices, but not adding any lexical meanings and hence usually considered inflectional forms. Examples in English would be the present and past / passive participles of *running water* or *a snatched / stolen kiss*. In many languages it is particularly obvious that the participles are adjectives; for instance, they not only modify nouns but also agree with them in number, gender or case (something verbs don't normally do).

Participles illustrate an intriguing problem, illustrated by the Russian examples in (38):

(38) a. Devuška čitaet gazetu
girl.NOM reads newspaper.ACC
"The girl is reading the newspaper"
b. devuška čitajuščaja gazetu
girl.NOM reading.FEM.NOM.SG newspaper.ACC
"the girl reading the paper"

In (38a) we see that the transitive verb "read" takes a direct object in the accusative case. The participle in (38b) takes the same direct object marked in the same way but corresponding to the subject in (38a) is the noun modified by the participle, "girl," with which the participle agrees just like an adjective would (see (18)). This shows that the participle is not like a normal adjective because Russian adjectives do not take complements (especially not in the accusative case!). Participles, however, take exactly the same set of complements as their base verb, and mark them in exactly the same way as the verb. Forms with this Janus-like behavior are often referred to as *mixed categories*. Deverbal nominalizations provide further instances. Thus, in *Tom's writing the letter (would be surprising)* the nominal *writing* expresses an object in the manner of a verb (*Tom wrote the letter*) not a noun (cf. *Tom's writing of the letter*) but expresses its subject in the manner of a noun (cf. *Tom's letter*). This type of morphology, changing, so to speak, only half a category, raises a variety of theoretical questions and deverbal nominalizations in particular have been the subject of intense study in recent years by morphologists and syntacticians.

3.4 Complex predicates

We began the chapter with a discussion of lexical integrity and a good deal of recent research has been devoted to clarifying this notion, and hence the

notion of word. Considerable research effort has been directed in recent years to cases in which there is a mismatch between the number of lexemes and the number of syntactically realized word forms. Such constructions are often referred to as *complex predicates*, a term which is used broadly of two types of phenomenon: (1) a single phonological, syntactic word form corresponds to two lexemes and (2) two phonological, syntactic word forms correspond to one lexeme. We saw one example of the type (1) complex predicate when we discussed noun incorporation (see examples (7), section 1.2). In this subsection I briefly mention two type (2) cases.

A simple example of a type (2) complex predicate is provided by an English phrasal verb such as *turn . . . off*. In *Tom turned the light off* we have a single verb lexeme *turn off* with the meaning "extinguish," but the two components can be separated by the verb's object. In the general case, we cannot predict the meaning of the phrasal verb from its components (compare for instance *Low temperatures will slow the process up / down* or *They ran a huge bill up*). A similar phenomenon is found in Hungarian, but with preposed particles ("preverbs"). Thus, *megérkez-* "arrive" has the preverb *meg-*. In (39) we see the preverb (PV) functioning as a prefix to the verb (the prefix receives the initial word stress, for instance; the accent in Hungarian orthography indicates vowel length, not stress):

(39) Meg-érkezett
MEG-arrived
"S/he arrived"

However, in certain morphosyntactic circumstances (negation, questions, focussing) it can appear separated to the right of the verb ((40), Szíj 1981: 209):

(40) Nem érkezett meg
NEG arrived MEG
"S/he didn't arrive"

In (41) we see forms of the verb *meg-néz-* "to watch, look at" as the complement of the verb *akar-* "want," where it remains prefixed to the verb:

(41) Nem akarom meg-nézni ezt a filmet
NEG I.want MEG-watch this the film
"I don't want to watch this film"

However, when the main clause is neither interrogative nor negative, as in (42) we find that the preverb appears to the left of the main verb:

(42) Meg akarom nézni ezt a filmet
MEG I.want watch this the film
"I want to watch this film"

There is good reason to regard such phrasal verbs as single lexical items, i.e. lexemes: the meaning is often (though not always) idiosyncratic and, in Hungarian, processes which derive nouns or adjectives from verbs often apply equally well to the phrasal verbs. This is illustrated below where a simple verb (43) is compared with a particle verb (44) (Ackerman and LeSourd 1997: 89):

(43) a. old-ani
"dissolve"
b. old-ás
"(chemical) solution"
c. old-ható anyag
"dissolvable substance"
d. old-hatatlan anyag
"insoluble substance"

(44) a. meg-old-ani
"solve"
b. meg-old-ás
"solution (to problem)"
c. meg-old-ható feladat
"solvable task"
d. meg-old-hatatlan feladat
"unsolvable task"

Thus, in Hungarian, a single lexeme, *meg-oldani*, can be systematically realized as more than one word in the syntax. Ackerman and LeSourd argue that this calls for a more sophisticated concept of lexical integrity: word forms such as "meg," "oldani," "turn" and "off" are single indivisible words, they cannot be split up once they appear in sentences and thus they exhibit lexical integrity. However, a given lexeme may be realized by a combination of such words, (*meg* = *oldani*, *turn* = *off*) and these may be separated in the syntax, so that, as lexemes, they do not exhibit lexical integrity. In other words, lexical integrity is a property of word forms but not necessarily of lexemes.

4 Conclusions

The notion "word" covers several distinct linguistic concepts, including: lexeme, word form, grammatical word. Not all the properties of words can be explained in terms of syntax or phonology, in particular, the existence of arbitrary inflectional classes demonstrates the need to treat morphology as an autonomous component of grammar. The classical sign-based concept of the morpheme has been extremely influential in thinking about the internal structure of words, but this has been largely abandoned, at least for inflection, where morphologists increasingly appeal to the notion of rule functions and defaults to capture the structure of paradigms and the order of elements, and to account for deviations from the "ideal" of agglutinating morphology.

We surveyed four types of derivational relatedness, showing that words can be related to each other in four main ways: in terms solely of semantics, with no morphological relationship, in terms purely of morphology, with no semantic relationship, in terms of polysemy, in which there is a semantic relationship but the word forms remain the same, and, the standard case, in terms of a semantic relationship mediated by morphology. We also looked at important

cases of mismatch between form and function, the mixed categories and complex predicates.

There are several important phenomena which I have had to pass over, in particular, the question of clitics, structural mismatches between word structure and syntactic or semantic structure, morphology and the semantic representations of words (especially verbs and argument structure) and questions of productivity and regularity and the storing of words in the mental lexicon. The reader should consult some of the references cited (for instance, the chapters of Spencer and Zwicky 1998) for overviews of these and other areas. However, enough has been said to illustrate that the structure of words, their organization into inflectional paradigms and their derivational relationships to each other is extremely rich and an important part of contemporary linguistic theory.

NOTE

Parts of this chapter are based on work conducted as part of research funded by the Economic and Social Research Council (Project Reference R000236115), to whom I express my gratitude. I am grateful to Mark Aronoff for helpful suggestions for improvement.