2

Morphological analysis

2.1 The atoms of words

Words can be chopped into smaller pieces. At the phonological level, words can be divided into syllables or segments, and segments into their constituent phonological features. At the morphological level, words may consist of more than one unit as well, which we may call the morphological atoms of a word: pieces that are no further divisible into morphological subparts. Just as there are different kinds of atom in chemistry, there are different kinds of atom in morphology, and it is quite useful for morphological analysis to be acquainted with their classification. A good classification is an important analytic instrument, developed in order to get a better understanding of the structure and formation of words.

As we saw in Chapter 1, the Polish lexeme KOT "cat" has a paradigm of case forms; compare this to the case forms of the noun KOBIETA "woman" in (1). Each cell of the paradigm of Polish nouns is occupied by a **grammatical**

(1)

	SINGULAR		PLURAL	
NOMINATIVE	kot	kobiet-a	kot-y	kobiet-y
GENITIVE	kot-a	kobiet-y	kot-ów	kobiet
DATIVE	kot-u	kobieci-e	kot-om	kobiet-om
ACCUSATIVE	kot-a	kobiet-ę	kot-y	kobiet-y
INSTRUMENTAL	kot-em	kobiet-ą	kot-ami	kobiet-ami
LOCATIVE	koci-e	kobieci-e	kot-ach	kobiet-ach
VOCATIVE	koci-e	kobiet-o	kot-y	kobiet-y

word, i.e. a form of a lexeme with a particular property for the grammatical categories number and case. Grammatical words may share the same word form. For instance, both the GEN.SG and the ACC.SG form of KOT have the form *kot-a*. The phenomenon that two or more grammatical words have the same word form is called **syncretism**. The distinction between lexeme, grammatical word, and word form shows that the general notion 'word' subsumes a number of different notions. In most cases it is clear which interpretation of 'word' is intended, but sometimes it will be necessary to use the more specific notions.

Each of the word forms of KOT consists of a stem and an **inflectional ending** (or **desinence**). The **stem** of a word is the word form minus its inflectional affixes, in this example *kot-*. It is the stem that forms the basis for word-formation, not the whole word form. This might not be so clear for the Polish noun KOT, because the NOM.SG word form *kot* of this word happens to have no overt ending. However, the noun KOBIETA does have an overt ending. For that reason, one may speak of a **zero-ending** for the NOM.SG. form of KOT, and likewise for the GEN.PL form of KOBIETA. The following example from Italian also illustrates the role of the stem. The singular form of *macchina* "machine" has the inflectional ending *-a*, and the plural ending is *-e*:

(**2**) macchin-a "machine" macchin-e "machines" macchin-ista "machinist"

It is the stem *macchin-* that is used as the basis for word-formation, as shown by *macchinista*. In English, the form of the stem is identical to that of the SG word form, and this is why English morphology is sometimes qualified as word-based morphology, in contrast to the stem-based morphology of, for instance, most Romance and Slavic languages. This is a superficial difference: these languages all have lexeme-based morphology, they only differ in that the stem-forms of lexemes do not always correspond to word forms.

Stems can be either simplex or complex. If they are simplex they are called **roots**. Roots may be turned into stems by the addition of a morpheme, as the following examples from Polish (Szymanek 1989: 87) illustrate:

(**3**)

a.	b.
butelk-a "bottle"	butelk-owa-ć "to bottle"
filtr "filter"	filtr-owa-ć "to filter"
bial-y "white"	biel-i-ć "to whiten"
głuch-y "deaf"	głuch-ną-ć "to become deaf"

The verbs in (3b) are given here in their citation form, the infinitive. The **citation form** is the form in which a word is mentioned when we talk about it, and the form in which it is listed in a dictionary. In many languages, the infinitive is the citation form of a verb. In languages with case, the NOM.SG form is the citation form of nouns. Each of these Polish infinitives consists of a root, followed by a verbalizing morpheme that turns the root into a stem, and is followed by the infinitival ending *-ć*. It is the stem-forms that are used when new words are derived from these verbs.

Stem-forming suffixes play an important role in many Indo-European languages. Italian verbs, for instance, have a **thematic vowel** after the root morpheme, and this thematic vowel recurs in words derived from these verbs:

(**4**) larg-o "wide" al-larg-a-re "to widen"
profond-o "deep" ap-profond-i-re "to deepen"
al-larg-a-ment-o "widening"
ap-profond-i-ment-o "deepening"

The thematic vowel is not a part of the root, as it does not occur in the roots *larg-* and *profond-*. On the other hand, it cannot be seen as part of the infinitival suffix, because we do not want to miss the generalization that all infinitives end in *-re*. Hence, the vowels preceding the ending *-re* must be assigned a morphological status of their own. Consequently, the noun *allargamento* contains five morphemes: a prefix *al-*, a root *larg*, a thematic vowel *-a-*, the derivational morpheme *-ment*, and the inflectional ending *-o*. So this word has five morphological atoms, which cannot be decomposed further into smaller morphological constituents. Each of these five atoms has a different name because they have different functions in the make-up of this word.

The general term for bound morphemes that are added to roots and stems is affix. If an affix appears before the root/stem, it is a **prefix**, if it appears after the root/stem, it is a **suffix**. So *al-* and *ap-* are prefixes, whereas *-a*, *-ment*, and *-o* are suffixes. Two other types of affixation are illustrated in (5):

(**5**) **infix** (within a root): Khmu (Laos) *s-m-ka:t* "roughen" < *ska:t* "rough"; Alabama (Stump 2001: 131) *ho-chi-fna* "smell, 2SG" < *hofna* "to smell", *chifip-as-ka* "poke, 2PL" < *chifipka* "to poke";
circumfix (combination of prefix and suffix): Dutch *ge-fiets-t* "cycled, PAST PARTICIPLE" < *fiets* "to cycle"; German *Ge-sing-e* "singing" < *sing* "to sing"

Infixation and circumfixation are much rarer than prefixation and suffixation.

Affixes are bound morphemes, but not all bound morphemes are affixes. There are many roots from Greek and Latin that are used in so called **neo-classical compounds** but do not occur as words by themselves. These compounds are called 'neo-classical' because they consist of constituents from the classical languages Greek and Latin that were combined into compounds long after these languages ceased to be 'living languages'. In such compounds either one or both constituents are not lexemes:

(**6**) micro-: micro-scope, micro-phone, micro-gram, micro-wave
tele-: tele-phone, tele-vision, tele-communication
-graph: di-graph, sono-graph, photo-graph, tele-graph
-scope: micro-scope, tele-scope, cine-scope, spectro-scope

Neo-classical roots such as *scope* and *graph* can also be used nowadays as words, but in that case they have a more specific meaning than in these compounds. Such non-lexical roots are called **combining forms** since they only occur in combination with other morphemes. These bound roots cannot be considered affixes since that would imply that words such as *necrology* would consist of affixes only. This goes against the idea that each word has at least one stem. Thus, we might adapt our definition of what compounds are, and define them as combinations of lexemes and/or non-affixal roots.

The bound morphemes in neo-classical compounds have an identifiable meaning, but there are also morphemes that have no clear meaning. In the word *cranberry* the part *berry* is identifiable, and this makes us interpret the word *cranberry* as denoting a particular kind of berry. Yet, *cran-* has no particular meaning. Similarly, the Dutch compound *stiefvader* "stepfather" denotes a particular kind of father, and hence can be parsed into *stief* and *vader*. However, the morpheme *stief* does not occur as a word. This phenomenon of **cranberry morphemes** is widespread, and is to be expected since complex words can lexicalize and thus survive, even though one of their constituent morphemes has disappeared from the lexicon. The following examples from Dutch illustrate the same phenomenon for derived words with suffixes that are still used for coining new words (the constituent before the suffix does not occur as a lexeme):

(**7**) arge-loos "naive", beslommer-ing "chore", dier-baar "dear, precious", le-lijk "ugly", moei-zaam "difficult", sprook-je "fairy tale", veil-ig "safe"

These recognizable suffixes determine the syntactic category of the word of which they form a constituent. For example, *-baar* is a suffix that creates adjectives, and hence *dierbaar* is predictably an adjective. This implies that when we have to decompose words into morphemes, not all morphemes have an identifiable lexical or grammatical meaning. Cranberry morphemes like English *cran-* and Dutch *dier-* thus form a problem for an exclusively meaning-based definition of the notion morpheme. This also applies to another kind of non-affixal bound root, the recurrent constituents of words borrowed from Latin such as the following English verbs:

(**8**) conceive, deceive, perceive, receive
adduce, deduce, induce, produce, reduce
admit, permit, remit, transmit

It makes sense to consider these words complex, because of recurrent elements such as *ad-*, *con-*, *de-*, *in-*, *per-*, *pro-*, *re-*, and *trans-* which are prefixes, and bound roots like *-ceive*, *-duce*, and *-mit*. Although these bound roots have no identifiable meaning, they should be recognized as morphemes since they determine the form of corresponding noun: all verbs in *-ceive* have a corresponding noun in *-ception*, those ending in *-duce* one in *-duction*, and verbs in *-mit* one in *-mission*. There is a wealth of such bound morphemes in the non-native part of the English lexicon, as the following examples illustrate:

(**9**) arct-ic, cred-ible, in-del-ible, gradu-al, mor-al, mus-ic, negoti-ate, per-for-ate, per-nic-ious

In lexeme-based morphology these bound roots do not have a lexical entry of their own, they only occur as part of established (listed) complex lexemes. In morpheme-based morphology, on the other hand, they will have to be represented as bound lexical morphemes with their own lexical entry. The advantage of the lexeme-based approach is that it correctly predicts that new combinations of a prefix and a bound root such as *demit* or *perduce* are not to be expected, because we cannot assign a meaning to such new combinations.

Boundness of morphemes is also created through allomorphy. **Allomorphy** is the phenomenon that a morpheme may have more than one shape, corresponds with more than one morph. A **morph** is a particular phonological form of a morpheme. Allomorphy is found in both affixes and root morphemes. In the Italian examples in (4) we saw the prefixes

al- and *ap-*. In fact, these are two allomorphs of the prefix *ad-*, in which the final consonant /d/ has assimilated to the first consonant of the root morpheme. This kind of allomorphy can be accounted for by assuming one common **underlying form** /ad/ for the different allomorphs of this prefix, and a rule of assimilation that derives its different surface forms.

Allomorphy is also found in root morphemes. In languages such as Dutch, German, and Polish, **obstruents** (that is, stops and fricative consonants, which are articulated with a high degree of obstruction in the mouth) are voiceless at the end of a word. Hence we get alternations of the following kind in pairs of singular and plural nouns:

(**10**)	Dutch	hoed [hut] "hat"	hoed-en [hudən]
	German	Tag [ta:k] "day"	Tag-e [ta:gə]
	Polish	chleb [xlɛp] "bread"	chleb-y [xlɛbɪ]

The symbols between brackets represent the phonetic forms of these word forms; the phonetic symbols are taken from the International Phonetic Alphabet (IPA). This is the alphabet used in dictionaries and grammars to indicate the phonetic forms of words in an unambiguous way. This is necessary because orthographical conventions differ from language to language. For instance, the vowel [u] is represented as *u* in German, but as *oe* in Dutch, as illustrated by the first example in (10).

Some linguists prefer to restrict the term 'allomorphy' to those cases in which the variation in phonetic shape of a morpheme does not follow from the automatic phonological rules of the language. The alternation between voiced and voiceless stops exemplified in (10) is determined by a phonological constraint that excludes voiced obstruents in syllable-final (Dutch and German) or word-final (Polish) position. Hence, the variation in shape of these morphemes is an automatic effect of the phonology of the language. This is usually accounted for by assuming a common underlying form for the different realizations of the morpheme involved, with a morpheme-final voiced obstruent. In the singular forms that lack an overt ending, a process of syllable-final or word-final devoicing then applies. The plural forms will not undergo this process because in these forms the relevant segments do not occur in final position.

This type of alternation can be contrasted to the alternation between voiceless and voiced obstruents in English, as in the singular–plural word pair *wife–wives*. This alternation applies to a small and closed set of English words only. That is, there are alternations that are restricted to a specific set

of words. Another example is that the Dutch diminutive suffix has five different shapes (*-tje*, *-je*, *-etje*, *-pje*, and *-kje*); the choice of one of these depends on the phonological composition of the stem. For instance, the allomorph *-je* has to be selected after stems ending in an obstruent. The alternations involved are unique to diminutive words, and do not follow from general phonological constraints of Dutch. Therefore, a distinction is made between phonology proper (the variation of the kind mentioned in (10) that is the effect of automatic phonological rules) and **morphophonology**, the domain of phonology in which alternations are restricted to a specific subset of words. The term 'allomorphy' might therefore be reserved for such non-automatic alternations, which can be accounted for in two ways. One option is to assume a common underlying form for the allomorphs, and derive the surface forms by means of one or more morphophonological rules, that is, rules whose application depends on non-phonological properties such as the feature DIMINUTIVE. Alternatively, the allomorphs can be listed individually in their surface form, with a specification of the phonological context in which they occur.

In some cases the non-automatic alternation is unique for one or a few words. For instance, the English adjective *platonic*, related to the noun *Plato*, has the morphological structure *platon-ic*, with the root *platon-* and the suffix *-ic*. The morpheme *platon-*, an allomorph of *Plato*, is a bound morpheme since it does not occur as a word of its own. This kind of allomorphy, a heritage from Greek (in the case of *Plato*) and Latin, increases the set of bound non-affixal morphemes enormously. An example from the Latinate substratum of English is *act*, *act-or* vs *ag-ent* with the bound root *ag-*. Although it has to be listed, the allomorph *platon-* does not require its own entry in the lexicon: it can be specified in the lexical entry for *Plato* as the allomorph to be used for the derivation of words from *Plato* by means of non-native suffixes. The same applies to the bound root *ag-*.

Another, more radical form of formal variation in paradigms is the phenomenon of **suppletion**, where there is no phonological similarity between the different forms of a lexeme. In the English word pair *good–better* we observe the suppletive root *bet* for *good*, followed by the comparative suffix *-er*. Thus, we might say that the lexeme GOOD comprises two different stems, *good* and *bet*. In the pair *bad–worse* the suppletive simplex form *worse* even expresses both the meaning of the stem *bad* and the comparative meaning. Some linguists also use the notion 'suppletion' in the domain of word-formation. In the following examples of inhabitative

names in Italian you can observe a formally regular case of derivation, a case of allomorphy, and a case of suppletion respectively:

(**11**) Milano–Milan-ese, Forlì–Forliv-ese, Chieti–Teat-ino

Although alternations in the phonological shape of a morpheme may not be the effect of the phonology of a language, the choice of a particular allomorph or suppletive root can still be phonologically conditioned (Carstairs 1988; Kiparsky 1994). For instance, the Dutch agentive suffix *-aar* is selected after stems ending in the vowel [ə] + *l, r, n* (that is, in a phonologically defined environment), and the allomorph *-er* elsewhere. The Italian verb *andare* "go" has two suppletive roots: *and-* when the root is not stressed, and *vad-* when the root is stressed in the verbal paradigm; see (12). This example illustrates that the choice between suppletive roots may be phonologically governed as well.

(**12**)

	SINGULAR	PLURAL
1.PERS	vádo	andiámo
2.PERS	vái	andáte
3.PERS	vá	vánno

2.2 Morphological operations

Morphology does not only deal with the analysis of existing words into their constituent pieces. The language user is able to make new words or forms of words, and it is this form of creativity that is the focus of morphology. The key notion involved is that of 'morphological operation'. This term denotes a particular kind of linguistic activity, and invokes a dynamic perspective on morphology. Two types of morphological operations have been discussed so far: compounding and affixation. They are the prototypical cases of **concatenative morphology**, in which morphological constituents are concatenated in a linear fashion. Compounding and affixation are the most widespread types of morphology since they create words with a high degree of **transparency**, that is, words of which the formal morphological structure correlates systematically with their semantic interpretation.

The formal operations available in morphology have several functions. Affixation is used both in word-formation and in inflection, and this applies to a number of other morphological operations discussed in this section as well.

For each morphological operation, we have to define the set of **base words** to which it applies. Often, the operation is restricted to base words of a particular syntactic category. This is the input category of the operation. The outputs of an operation also belong to a specific syntactic category. The input category of the English suffix *-able* is V, and the output category is A. Hence, verbs are the base words of the suffix *-able*. Thus, in the case of derivation, the morphological operation may result in words of another syntactic category or subcategory than that of the input words. In that case, we speak of a **category-changing** or **class-changing** operation.

If compounding and affixation were the only kinds of morphological operation, morphology could be said to consist of just one operation—concatenation. In such a view, the elements to be concatenated are lexemes and affixes. Affixes are provided with a subcategorization feature that specifies with which kind of morphological elements it has to combine. For instance, the suffix *-able* will be specified as $[V—]_A$, which means that it takes verbs to form adjectives.

The reason why the term 'morphological operation' is more adequate than the term 'concatenation' is that there are also morphological processes that do not consist exclusively of the attachment of affixes to words. In this section I present a short survey of these operations, which are dealt with in more detail in subsequent chapters on derivation and inflection.

A special kind of affixation is the attachment of a complete or partial copy of the base as a prefix or a suffix. This is called **reduplication**, illustrated by the following examples (Uhlenbeck 1978: 90) from Javanese:

(**13**) a. **full reduplication:**

baita "ship"	baita-baita "various ships"
səsupe "ring"	səsupe-səsupe "various rings"
omaha "house"	omaha-omaha "various houses"

b. **partial reduplication:**

gəni "fire"	gəgəni "to warm oneself by the fire"
jawah "rain"	jəjawah "to play in the rain"
tamu "guest"	tətamu "to visit"

In the examples of partial reduplication, the prefix consists of a copy of the first consonant of the base followed by the vowel schwa [ə]. The doubling effect of full reduplication is often reflected by its meaning contribution: for nouns it may express plurality or distributivity (as in 13a), for verbs a high intensity of the action expressed, and for adjectives a higher degree of the property mentioned by the adjective.

Reduplication is a kind of affixation (or compounding, in the case of full reduplication), and hence to a certain extent a case of concatenative morphology. Yet, it is clear that we cannot list reduplicative affixes with their phonological content in the lexicon since this content depends on the phonological composition of the stem. The obvious analysis is the assumption of an abstract affix RED(UPLICATION) that triggers a phonological operation of copying. The copy is then attached to the copied stem.

A second type of morphological operation is the use of tone patterns. Tone patterns belong to the **suprasegmental** properties of languages. In Ngiti, the plural form of kinship terms is expressed systematically by the tone pattern Mid–High on the stem, whatever the tone pattern of the singular (Kutsch Lojenga 1994: 135):

(**14**)

SINGULAR	PLURAL
àba-du	abá-du "my father(s)"
adhà-du	adhá-du "my co-wife(s)"
andà-du	andá-du "my uncle(s)"

Thus, we may speak of a **tonal morpheme** Mid–High which is superimposed on the segmental material of the stem of these nouns. This is why such a tonal morpheme is sometimes called a **suprafix**. This is a case of non-concatenative morphology since this kind of affix is not linearly ordered with respect to its base.

Many languages make use of **internal modification**. Standard examples are the patterns of vowel alternation in the roots of the so-called strong verbs in Germanic languages, called **ablaut, vowel gradation**, or **apophony**. Such vowel alternations are used in a number of Indo-European languages for different forms of the verb:

(**15**) Classical Greek: l**ei**p-o "I leave"; le-l**oi**p-a "I have left", e-lipon "I left"

The *e* in the first root form alternates with *o* in the second, and zero in the third (the second form also exhibits partial reduplication). This pattern of vowel alternation is reflected in Germanic languages, as the following examples from Dutch illustrate:

(**16**)

geef [ɣe:f] "to give"	gaf [ɣɑf] "gave"	gegeven [ɣəɣe:vən] "given"
help [hɛlp] "to help"	hielp [hilp] "helped"	geholpen [ɣəhɔlpən] "helped"
schiet [sxit] "to shoot"	schoot [sxo:t] "shot"	geschoten [ɣəsxo:tən] "shot"

Vowel alternations also play a role in the derivation of deverbal nouns of such verbs, as shown by the related Dutch deverbal nouns *hulp* "help" and *schot* "shot". They only differ from their verbal bases *help* and *schiet* with respect to the root vowel.

Ablaut is not the only kind of vowel alternation with a morphological function. German exhibits an alternation between back vowels and front vowels in singular–plural noun pairs:

(**17**) Apfel [ɑpfəl] Äpfel [ɛpfəl] "apple(s)"
Bach [bɑx] Bäche [bɛçə] "brook(s)"
Buch [bu:x] Bücher [by:çər] "book(s)"

This kind of alternation is called **umlaut** (also called **vowel mutation** or **metaphony**). Historically it is a case of assimilation: back vowels of roots are fronted before a high front vowel in the following syllable (the plural suffix contained a high vowel originally).

If we only take the first example of (17) into consideration, we might conclude that plural formation in German is a case of non-concatenative morphology: the plural is created by the replacement of the back root vowel by its front counterpart. However, an alternative analysis in terms of affixation is also possible. Given the three examples in (17), we might conclude that there are at least three different plural suffixes in German: ø (zero), *-e*, and *-er*. In addition, the plural nouns may exhibit stem allomorphy, a vowel alternation triggered by the attachment of the plural suffix. Such morphologically conditioned alternations may also affect consonants (Lieber 1987, 2000). English has cases of consonant modification as well, for instance *defend–defence, offend–offence, belief–believe*, and *proof–prove*.

An interesting kind of non-concatenative morphology is found in, among others, Semitic languages: **root-and-pattern morphology**. The basis of each lexeme is a skeleton of consonants, in most cases three, which functions as the root of the lexeme. The abstract pattern of consonants is combined with one or more vowels which are intertwined with the sequence of consonants. In addition, the lexeme may contain a prefix and a suffix. In the words of Modern Hebrew in (18) (Clark and Berman 1984: 545) the roots *g-d-l* "grow" and *k-t-b* "write" have been used (the *k* and *b* may surface as *x* [x] and *v* respectively). The vowel patterns that are intercalated with the consonantal skeletons are called **transfixes** since they are spread across the consonantal sequence.

(**18**)	*Pattern*	*Root* g-d-l	*Root* k-t-b
	CaCaC	gadal "grow, get bigger"	katav "write"
	hiCCiC	higdil "enlarge"	hixtiv "dictate"
	CCiCa	gdila "growth"	ktiva "writing"
	miCCaC	migdal "tower"	mixtav "letter, missive"
	haCCaCa	hagdala "enlargement"	haxtava "dictation"

The morphological structure of the words in (18) can be represented as the linking between three different morphemes. Each of these morphemes forms a phonological tier of its own: (i) the skeletal tier that consists of a pattern of consonantal and vocalic slots that is characteristic of a particular morphological category, (ii) the sequence of consonants that represents the lexeme, and (iii) the vowels that fill the vocalic slots of the skeletal tier. The words *gadal* and *gdila* in (18) can be represented as in Figure 2.1. The consonants of the lexical root, and the vowel pattern (*a-a* for the base verb and *i-a* for the nominalization) are both linked to the central skeletal CV tier. These three tiers are then conflated into one sequence of sounds at the phonetic level of the grammar, where the phonetic forms of words are specified.

Fig. 2.1 Three-tiered representations of words

The morphological operations discussed so far all have the effect that the phonological form of the input word is changed somehow. Conversion, on the other hand, consists of a change in syntactic (sub)category only. The conversion of nouns to verbs is quite common in European languages; see (19). The verbs are given here in their citation form, the infinitive. The conversion from noun to verb is not indicated directly by means of an affix, and is therefore also called **implicit transposition**, as opposed to **explicit transposition**, which denotes cases of category-changing word-formation in which the change is marked through the addition of an affix. Note that conversion does have indirect morphological effects: the verbs in (19) are recognized as such by their verbal inflectional endings, the infinitival

suffixes (except the English verb since there is no overt infinitival ending in English). The category change may also have an effect on the stress pattern, as in the English pair *convért* (V)—*cónvert* (N), where the noun is derived from the verb, with concomitant stress shift from the last to the first syllable.

(**19**)

	Noun	*Verb*
Dutch	fiets "cycle"	fiets-en "cycle"
English	chain	(to) chain
French	guide "guide"	guid-er "guide"
Latin	corona "crown"	coron-a-re "crown"

If one wants to treat conversion as a kind of affixation, one is forced to assume a **zero-morpheme** that is added to the input word. However, there is no independent evidence for such a zero-affix, and we do not even know if the zero-morpheme should be taken to be a prefix or a suffix. Therefore, conversion as exemplified in (19) is better analysed in terms of the following morphological rule:

(**20**) $[x]_N \rightarrow [[x]_N]_V$

The verbs in (19) therefore have the structure $[[X]_N]_V$.

A defining property of the notion 'conversion' is that it has a direction: in the examples above, the verb has been derived from the noun. This phenomenon must therefore be distinguished from **multifunctionality**, the situation in which words can be used for different syntactic categories without a particular direction in the relation between these different uses of words. In Maori, for example, the word *waiata* can be used as a verb "to sing", as a noun "song", and as a participle "singing" (Bauer 1993: 510). In Sranan, a creole language of Surinam, the word *hebi* functions as an adjective "heavy", a noun "weight", an intransitive verb "to be heavy", and a transitive verb "to make heavy" (Voorhoeve 1979: 43).

Change of category without overt morphological marking is also found in the case of **middle verbs**, which are intransitive and denote a property, whereas the corresponding activity verb denotes an activity (examples from Dutch):

(**21**) Deze aardappelen schillen gemakkelijk
These potatoes peel easily
"These potatoes are easy to peel"
Mars hapt zo heerlijk weg

Mars eats so nicely away
''Mars is so pleasant to eat''

(Mars is a kind of candy bar.) This kind of change from one subcategory of verbs to another subcategory may be subsumed under conversion because there is a clear direction in the relation between the verbs involved: the middle verb is derived from the activity verb.

In Chapter 1 you were introduced to the notion of paradigmatic word-formation, in which a morphological constituent of a word is replaced with another one. A typical case of this kind of word-formation is **affix substitution**, the replacement of one affix with another. In Dutch, female counterparts of agent nouns can be formed by replacing *-er* with *-ster* (Booij 2002*a*: 6):

(**22**)	aanvoerd-er ''captain''	aanvoerd-ster
	betwet-er ''lit. better knower, pedant''	betweet-ster
	rederijk-er ''rhetorician''	rederijk-ster
	reizig-er ''traveller''	reizig-ster
	oproerkraai-er ''ring leader''	oproerkraai-ster

The operation of substitution as a viable way of making new words has developed from systematic relationships between words derived from the same base. In this case, both *-er* and *-ster* can be added to Dutch verbs to form agent nouns. Thus a pattern $[\text{X-}er]_N$: $[\text{X-}ster]_N$ could be observed, which was then extended to other nouns in *-er* without a straighforward verbal base. For instance, there is no Dutch verb *reizig* ''to travel'', and yet, the agent noun *reiziger* has a female counterpart in *-ster*. The presence of the /d/ in *aanvoerdster* also betrays that this word is derived from *aanvoerder*. The /d/ does not belong to the verbal stem *aanvoer* ''to lead'', but is part of the allomorph *-der* that is used after stems ending in /r/. Since it is *-er* that is replaced, the /d/ shows up in the female agent noun as well.

A prototypical case of paradigmatic word-formation is **back formation** in which the direction of derivation is inverted: the less complex word is derived from the more complex word by omitting something. Well-known examples from English are *to sculpt* from *sculptor*, and *to babysit* from *babysitter*. The noun *sculptor* is a borrowing from Latin. Because English has word pairs of the type $\text{V}-[\text{V}+or]_N$, (*terminate–terminator*, etc.), the verb *sculpt* could be reconstructed from the noun *sculptor* by reinterpreting this word as having the structure $[[sculpt]_V or]_N$. The paradigmatic dimension involved here is that a word ending in *-or* is assigned an internal

morphological structure with a verbal base on the basis of existing verb–noun pairs such as *terminate–terminator*.

The emergence of the verb *to babysit* can be reconstructed as follows. The word *babysitter* is a regular compound consisting of two nouns, *baby* and the deverbal noun *sitter*. However, there is no general process of N + V compounding in English. The exceptional NV compound *babysit* could therefore only arise through back formation. In the same way, the Dutch NV compound *stofzuig* "to vacuum-clean" arose through back formation from the regular NN compound *stofzuiger* "lit. dust sucker, vacuum cleaner". In the cases of *to babysit* and *stofzuig*, the structure $[[N][V\text{-}er]_N]_N$ has been reinterpreted as $[[N\ V]_V\text{-}er]_N$, and subsequent back formation led to the rise of these N + V compounds.

2.3 Morphological typology

The catalogue of morphological operations presented in section 2.2 raises the question to what extent the languages of the world make use of these possibilities. First, we can locate each language on a scale of degree of **synthesis**, the average number of morphemes in a word. On one end of the scale we find **isolating languages** that do not make use of morphology at all. A classical example of such a language is Vietnamese (which, however, is said to have compounds). At the other end of the scale we find **polysynthetic languages** such as Greenlandic and Alaskan Yup'ik, languages in which words may contain a considerable number of suffixes after the root.

Before we have a look at some relevant examples, I will first give a short clarification of the notational conventions used in **interlinear morphemic translation**. These conventions are of considerable importance for our understanding of the structure of sentences and words. A space marks the boundary between two words, and a hyphen represents a boundary between two morphemes within one word. Lexical morphemes are represented by lower case letters, and grammatical categories by small capitals. If one morpheme on the first line represents more than one piece of lexical or grammatical information in the morphemic gloss, the categories are separated by a dot, as in the following Latin example (Lehmann 1982: 205):

(**23**) Manu-s manu-m lava-t
hand-NOM.SG hand-ACC.SG wash-3SG
"One hand washes the other"

The only exception to this use of the dot is its absence in combinations of the category PERSON and the category NUMBER, as in 3SG.

Let us now have a look at some examples of polysynthetic words in which the conventions just discussed are also exemplified—first Greenlandic (Fortescue 1984: 273) and then Alaska Yup'ik (Mithun 1999: 28):

(**24**) tuqu-riikatap-puq
die-long.ago–3SG.INDIC
"He died long ago"

anglani-tu-llru-u-nga caknek
enjoy-customarily-PAST-IND.INTR-ISG very.much
"I used to enjoy myself very much"

The first example is a sentence of one word only, the second one contains two words. When we compare this with the number of words in the English glosses (four and seven respectively), we get some idea of what it means for a language to be polysynthetic.

The second scale on which we may rank languages as to their morphological properties is that of **fusion**. In some languages, a word is easily segmentable into its constituent morphemes. An example is Turkish, a language that is therefore characterized as **agglutinative:** the stem of a word is followed by one or more suffixes, each with their own meaning:

(**25**) çocuk-lar-nız-dan
child-PL-your.PL-ABL "from your children"

Most Indo-European languages are **fusional** in their inflectional system since different inflectional properties are often expressed by one and the same morpheme. In the Polish word form *koty* "cat, NOM.PL" the ending *-y* expresses simultaneously the properties NOMINATIVE and PLURAL. In the English word *walks*, the ending *-s* expresses three properties: PRESENT TENSE, SINGULAR, and 3.PERSON. Such units that serve to express more than one morphological property are called **portmanteau morphs**.

When a language tends to be more agglutinative, it will tend to have more morphemes per word than a fusional language, and hence it will be higher on the scale of synthesis as well. The average number of morphemes per word in Turkish is estimated to be four times higher than that in English (Csató and Johanson 1998: 208).

The kind of typological classification discussed so far mainly has a descriptive and orientational function: by locating a language on a number

of scales, we know roughly what kind of morphological system we may expect. But it does not provide fine-grained classifications. Germanic languages are fusional in their inflectional systems, but agglutinative in their system of derivational word-formation. Moreover, for an adequate descriptive classification other parameters are also relevant, for instance the parameter of reduplication: languages of the Austronesian family make wide use of reduplication patterns, whereas this does not apply to most Indo-European languages of Europe. Languages may also differ in the extent to which they make use of prefixation or suffixation.

Morphological typology becomes theoretically interesting if it enables us to predict certain properties of a language on the basis of other properties. For instance, the following **morphological universal** has been proposed by Greenberg (1963: 95):

(**26**) If a language has the category of Gender, it always has the category of Number.

This universal has the form of an implication, and hence it predicts that of the four following logically possible languages only the first three exist:

(**27**) a. Languages with Gender and Number
b. Languages with Number only
c. Languages without Gender or Number
d. Languages with Gender only

Thus, an **implicational universal** is a restriction on the class of possible natural languages, and hence contributes to the definition of the notion 'possible natural language'.

Some implicational universals pertain to markedness phenomena. **Markedness** is the asymmetrical distribution of properties. An example of a **markedness universal** is that there are many languages in which the singular is not expressed by a morpheme, but only the plural, whereas there are no languages where only the singular is expressed by a morpheme. For example, the asymmetric distributional pattern of singular and plural morphemes given in Table 2.1 has been found. This table shows that languages with a singular morpheme only must be excluded in order to restrict the degree of variation in natural language. Hence, we might formulate the following implication universal for this markedness pattern: 'if singular number is expressed by a morpheme, then plural number as well'.

Table 2.1. *Distribution of Number morphemes*

	Absence of SG *morpheme*	*Presence of* SG *morpheme*
Presence of PL *morpheme*	English, Dutch	Latvian, Italian
Absence of PL *morpheme*	Chinese, Maori	

Source: Croft 1990: 69.

This generalization concerning the expression of number has to be amended slightly, however. There are languages where, for those entities that always occur in pairs or in groups, the plural form of the noun has no overt suffix, and the singular form ends in a **singulative** suffix. This is the case for Turkana (Dimmendaal 1983: 224–8), a language of Kenya with the singulative suffixes *-a* and *-it* (the prefixes are gender markers):

(**29**)	*singulative*	*plural*
	ɛ-sɪkɪn-a "breast"	ŋ-sɪkɪn "breasts"
	e-turkàna-ɪt "Turkana person"	ŋ-tùrkanà "Turkana people"

This reversal of the markedness pattern concerning singular–plural in a special domain is called **local markedness** (Tiersma 1982).

The use of **hierarchies** in morphological typology is illustrated by the following hierarchy for the different values for the category number:

(**30**) singular > plural > dual

This hierarchy ranks singular above plural, and plural above dual. It expresses that singular forms are less marked than plurals, and plurals are less marked than duals. This means that if a language has a **dual** (that is, a word form with 2 as the value for the category number), it has also a plural, and if a language has a plural, it also has a singular. Hence, this hierarchy restricts the variation space of natural language: certain types of logically possible languages are excluded, such as a language with singular and dual only.

Not all the typological universals are absolute ones; some are statistical tendencies only. For instance, there are many more languages that only use suffixes (Turkish is an example) than there are languages that only use prefixes. Hence, there is a suffixing preference in natural languages. Yet, there are languages that are exclusively prefixing, so there is no absolute universal involved here. Many of the universals discussed in Greenberg (1963) are of this statistical nature.

Summary

Words can be divided into different kinds of morphemes such as roots and affixes, the morphological atoms of language. These morphemes may vary in shape (allomorphy), a variation that does not always follow from the phonological system of the language. In the case of suppletion, different stems co-occur in the paradigm of one lexeme.

The set of morphological operations available to human languages comprises more than concatenation: conversion, reduplication (concatenation plus copying), different types of phonetic modification, root-and-pattern morphology, and paradigmatic word-formation also play a role.

Languages do not all make the same use of the available morphological operations, and can be classified according to the indices of synthesis and fusion. In addition to this purely classificatory typology, morphologists make cross-linguistic comparisons of morphological systems in order to find constraints on the degree of morphological variation of natural language.

Questions

1. Identify the bound constituents of the following English words: *disagreeable, acceptability, ungrammaticality, discriminatory, permafrost, fascination, protolanguage, versification, intolerance, productivity, unidirectionality.*
2. Consider the sets of morphologically related words in French in the table that exhibit variation in the underlined vowel of their base word (Dell and Selkirk 1978).

Base word	*Derived words*	
[œ]	[œ]	[ɔ]
fleur ''flower''	fleurette ''small flower''	floral ''floral''
seul ''alone''	seulement ''only''	solitude ''solitude''
peuple ''people''	peuplade ''tribe''	populaire ''popular''
[ɛ]	[ɛ]	[a]
vain ''idle''	vainement ''in vain''	vanité ''vanity''
clair ''clear''	éclairer ''to light''	clarifier ''to clarify''
mer ''sea''	amerrir ''to land on the sea''	marin ''sailor''
africain ''African''		Africaniste ''Africanist''
humain ''human''		humanité ''humanity''
similaire ''similar''		similarité ''similarity''

a. Formulate the rule that accounts for these vowel alternations.
b. Is this rule an automatic phonological rule or a morphophonological rule?

3. In the following past tense forms of English verbs, the past tense suffix has three different phonetic shapes: *kept, walked, kissed, hugged, lived, added, fitted, coded*.

a. Which are the three phonetic forms of this suffix?
b. Formulate the rules (or rule) that account for this phonetic variation.

4. Make a morphological analysis of the following words of the Amerindian language Cree, and give the interlinear morphemic glossing of the last word (Cowan and Rakušan 1985: 111):

niwa:pahte:n ''I see (it)''
kiwa:pahte:n ''you see (it)''
niwa:pahte:na:n ''we see (it)''
kiwa:pahte:na:wa:w ''you (plural) see (it)''
nima:čiše:n ''I cut (it)''
kima:čiše:n ''you cut (it)''
nima:čiše:na:n ''we cut (it)''
kima:čiše:wa:w ''you (plural) cut (it)''
nitapin ''I sit''
kitapin ''you sit''
nitapina:n ''we sit''
kitapina:wa:w ''you (plural) sit''

5. Consider the following past tense forms of English: *kept, wept, slept*. Which kinds of operation have been used for making these verbal forms?

6. In written Arabic, verbs can be derived from nouns, as the following examples (Becker 1990*b*: 14) show:

zayt ''oil''	zayyata ''to oil''
bukla ''buckle''	bakkala ''to buckle''
turki: ''Turk''	tarraka ''to Turkify''
tilifu:n ''telephone''	talfana ''to telephone''
tilifiziu:n ''television''	talfaza ''to televise''

a. Give a three-tiered representation of the verbs *bakkala* and *talfana*.
b. Give the consonantal template and the vowel template for the morphological category of denominal verbs.

7. The following are singular and plural nouns in Bulgarian (Cowan and Rakušan 1985: 99):

teátər	teátri	''theater(s)''
bóbər	bóbri	''beaver(s)''
pésen	pésni	''song(s)''
psalóm	psalmí	''psalm(s)''
báncik	báncigi	''band saw(s)''
ízverk	ízvergi	''monster(s)''

Give the (morphological and phonological) rules for plural noun formation for this set of word forms.

8. Which types of morphological processes are involved in the following forms (Melčuk 2000: 528) of the Tagalog verbs *patáy* ''to kill'' and *sulat* ''to write''?

	PAST	PRESENT
ACTIVE	pumatáy	pumápataý
	sumulat	sumusulat
PASSIVE	pinatáy	pinápatáy
	sinulat	sinusulat

9. Consider the following pairs of singular and plural nouns of Agta (Wiltshire and Marantz 2000: 558):

takki	taktakki	''leg(s)''
labáng	lablabáng	''patch(es)''
uffu	ufuffu	''thigh(s)''

Formulate the morphological rule that accounts for the formation of the plural nouns.

10. In Italian, adjectives can be derived from nouns and adjectives through the addition of the suffix *-oso* or *-astro*, as illustrated by the following examples (source Scalise 1984: 59):

fama ''fame''	famoso ''famous''
virtú ''virtue''	virtuoso ''virtuous''
giallo ''yellow''	giallastro ''yellowish''
blu ''blue''	bluastro ''bluish''

Give the stem-forms of the four base words listed here.

Further reading

A survey of the different types of morphological operations can be found in chapter 8 of *BLM*, and in Payne (1997). A detailed analysis of vowel and consonant mutation is given in Lieber (1987). The multi-tier interpretation of root-and-pattern morphology has been proposed by McCarthy (1981). This kind of morphology is also found in a number of non-Semitic languages (Broselow 2000).

Formal theories of morphology as a set of operations on lexemes (processual morphology) have been developed in Anderson (1992) and Stump (2001).

Morphological typology is discussed in Comrie (1981) and Croft (1990). The suffixing preference is dealt with in Cutler *et al*. (1985). The verbs of the Athapaskan language Slave are famous for their morphological complexity

(Rice 1989, 2000), and Amerindian languages in general for their polysynthetic nature (Mithun 1999). Polysynthesis is discussed from a theoretical point of view in Baker (1988, 1996, 2001). Possible explanations for the suffixing preference are discussed in Hawkins and Cutler (1988), and in Hall (1988).

Information on the World Atlas of Language Structures that also includes morphological typology can be found at http://linguistics.buffalo.edu/people/faculty/dryer/dryer/atlas. The Universals Archive of the University of Konstanz may be consulted at http://ling.uni-konstanz.de/pages/proj/sprachbau.htm. The website of the Surrey Morphology Group provides typological information on agreement and syncretism: http://www.surrey. ac.uk/LIS/SMG.

A standard for interlinear morphemic glossing is presented in Lehmann (1982). More recent versions are the set of glossing rules proposed by the Leipzig typologist group, based on Lehmann (1982), which can be found at www.eva.mpg.de/lingua/index.html, and Lehman (2004).