[to appear in John Taylor ed. *The Oxford Handbook of the Word*. Oxford: Oxford University Press]

[draft version, August 2012]

Chapter 11. The structure of words

Geert Booij, Leiden University Centre of Linguistics

Abstract

This chapter discusses the internal morphological structure of words. Complex words are created by processes of word formation and inflection. In concatenative morphology, complex words are created by the combination of words and affixes. In non-concatenative morphology use is made of vowel and consonant alternations and of tonal patterns. Complex words may also be formed by means of paradigmatic morphology, such as the replacement of one affix in a complex word by another.

Normally, the meaning of a complex word is a function of the meaning of its constituents and the meaning of the morphological construction as a whole, in accordance with the principle of compositionality. Morphological structures may also have holistic meaning properties. In some complex words there is a mismatch between the formal structure and the semantics.

The internal morphological structure plays an important role in the interface with phonology: the phonetic realization of morphemes is partially determined by their position in morphological structure.

Complex words, once formed, may lose (part of) their motivation; hence, their morphological structure may become opaque. This may also lead to changes in the phonological forms of these words.

Complex words may also arise as an effect of univerbation, the process by which word combinations become words. Grammaticalization may lead to the expansion of the set of complex grammatical words such as prepositions and conjunctions.

1. Introduction

Words may have an internal structure. For instance, the English word *singer* can be divided into two constituents, *sing* and *-er*. Both constituents contribute to the meaning of the word as a whole. These constituents are referred to as morphemes, the minimal meaning-bearing units of a language. The word *singer* is therefore a complex word, as opposed to the word *sing*, which has no internal morphological structure and is therefore a simplex word. The morpheme *sing* is a lexical morpheme, as it can also occur as a word of its own, whereas the morpheme *-er,* which expresses the meaning 'agent of the action', is a bound morpheme.

How do we know that *singer* is a complex word, whereas other words that also end in *-er* are not considered complex, such as *border, father,* and *order*? The reason is that for a word to be considered complex we expect a systematic correspondence between its form and meaning. The internal structure of the noun *singer* is determined on the basis of a comparison of sets of words such as the following:

(1)

verb	*agent noun in* -er
bake	baker
speak	speaker
dance	dancer
use	user

The two sets of words stand in a systematic form-meaning relationship, and on the basis of this relationship we can assign an internal morphological structure to nouns in *-er* with a verbal subconstituent. We call these nouns deverbal, as they are derived from base words that are verbs. In the case of *border, father,* and *order* there is no base word to be found, and there is no agentive meaning, and hence we consider these words as simplex. This also makes it clear why we want to assign internal morphological structure to words like *singer*: the meaning of this word is not completely arbitrary, but motivated, namely, by its constituents and their arrangement.

A second example of complex words are the following plural nouns in English: *apples, books, pages*, which all end in the plural morpheme *-s*. These words are also complex since they show a systematic form-meaning correspondence with the words *apple, book*, and *page*. The difference with the agent nouns is that this is not a case of word formation, but of inflection. Whereas *sing* and *singer* are two different words, with their own entry in a dictionary, this is not the case for *apples*, which is an inflectional form of the lexeme APPLE, as is the singular form *apple*. A lexeme is the abstract unit that stands for the set of inflectional forms, and is usually represented with small capitals.

The two basic functions of morphological operations that create complex words are word formation and inflection. Word formation processes create new lexemes and hence expand the lexicon of a language. Once the speaker of English has discovered the pattern exemplified in (1), (s)he may hit on a word formation schema which we can characterize, informally, as follows:

(2) $[[x]_V \text{ er}]_N$ 'one who Vs'

The speaker may then use this schema to create new words with the approrpriate form and meaning, such as *skyper*, derived from the verb *(to) skype*, or *texter*, derived from the verb *(to) text*. The new words may not only have the meaning predicted by the schema, but also additional idiosyncratic properties. For instance, the *Urban Dictionary*[1] defines a *texter* as 'a person who prefers to send text messages instead of picking up the phone'.

Inflection is the grammatical subsystem that deals with the proper form of lexemes, often in relation to specific syntactic contexts. In Dutch, for instance, the verb *werk* 'to work' has five different finite forms. The selection of a present or past form depends on the information the speaker wants to convey, and this is called inherent inflection. The choice of a particular present or past form depends on the number and person of the subject of the clause in which the verb occurs, and this is called contextual inflection since the choice of a form depends on the syntactic context:

(3)	werk	present 1st person singular
	werk-t	present 2nd/3rd person singular
	werk-en	present 1st/2nd/3rd person plural
	werk-te	past 1st/2nd/3rd person singular
	werk-te-n	past 1st/2nd/3rd person plural

We consider these five forms as forms of the same lexeme. Thus, Dutch has a lexeme WERK. The stem form of this lexeme is *werk*, and the different inflectional affixes are added to this stem. The Dutch word *werker* 'worker' is a different lexeme than the word *werk* 'to work' (so Dutch has the lexemes WERK and WERKER). The plural form of this noun *werkers* 'workers' has the following morphological structure:

(4)	werk	-er	-s
	work	-AGENT	-PLURAL
	'workers'		

This is a simple example of the morphological analysis of a complex word, and presented in a form that follows the conventions of interlinear morphemic glossing (Lehmann 2004). The first line presents the internal constituency of the complex word. The second line provides a morpheme by morpheme glossing, and the third line gives a paraphrase of the meaning of the linguistic unit.

Cross-linguistically, the most common form of word formation is compounding, the combination of two or more lexemes into a complex word, such as the English word *songbook* composed of the nouns *song* and *book*. Many languages make also use of derivation, the process in which bound morphemes (affixes) such as *-er* are attached to a base word. These two mechanisms are instances of concatenative morphology, in which complex words are created by means of the concatenation, or stringing together, of morphemes.

However, these are not the only means of making complex words. In reduplication, a complete or partial copy of a word is added to that base word. Thus, reduplication may be seen as a specific form of compounding (total reduplication) or

[1] http://www.urbandictionary.com/

affixation (partial reduplication). Here are some examples from the Austronesian language Begak spoken in Malaysia (Goudswaard 2005: 52-56):

(5) *total reduplication*

suran 'story'	suran-suran 'many stories'
puti 'white'	puti-puti 'very white'
panow 'to go'	panow-panow 'to go a little bit'

partial reduplication

bua 'fruit'	bə-bua 'various types of fruit'
bunu' 'to kill'	bə-bunu' 'to kill each other'
satu 'one'	sə-satu 'only one'

In the case of partial reduplication, a copy of the first consonant forms a prefix together with a fix vowel [ə].

Other, non-concatenative, mechanisms are the use of specific vowel and/or consonantal patterns (as in *sing-sang, bring-brought*), and tonal patterns. The use of vowel alternations is a characteristic of Indo-European languages. The Semitic languages are well known for their system of creating related verbal lexemes by combining a consonantal root with a specific pattern of C and V positions, and a vowel melody. In addition, prefixes may be used. This type of morphology is called root-and-pattern morphology. A set of verbal lexemes with the same morphological pattern is called a *binyan* **(**'building', plural *binyanim*). The root *qtl* 'to kill' as used in Biblical Hebrew has the following five binyan forms with active meaning (Aronoff 1994: 124), of which the 3SG.MASC.PERF forms are illustrated in (6)[2]:

(6)

Binyan	3SG.M.PF	CV-PATTERN	V-PATTERN	*gloss*
Qal	qâtal	CVCVC	â-a	to kill
Nif'al	ni-qtal	ni-CCVC	a	to kill oneself
Pi'el	qittel	CVC_iC_iVC	i-e	to massacre
Hif'il	hi-qtil	hi-CCVC	i	to cause to kill
Hitpa'el	hit-qattel	hit-CVC_iC_iVC	a-e	to kill oneself

In addition to these five binyanim, there are two binyanim with a passive meaning, the Pu'al as a passive variant of the Pi'el, and the Hof'al as the passive of the Hif'il.

The use of a tone pattern as a morphological marker can be found in many tone languages. Here is an example from Ngiti, a central-Sudanic language of Congo (Kutsch Lojenga 1994: 135):

(7)

singular	*plural*	
àba-du	abá-du	my father(s)
abhu-du	abhú-du	my grandfather(s)
akpà-du	akpá-di	my husband(s)
andà-du	andá-du	my uncle(s)

[2] The apostrophe ' indicates a glottal stop, and the symbol *â* stands for an open /o/; C_iC_i stands for two identical consonants.

In the plural possessive form, the prefinal syllable always carries high tone (indicated by the acute accent on the vowel letter, e.g. *á*), whereas the other syllables carry mid tone (indicated by the lack of an accent symbol). In the singular, some vowels carry low tone (indicated as *à*). Thus, plurality is systematically marked by a specific tone pattern.

Finally, there are operations like blending (*smog* < *smoke* + *fog*), the formation of acronyms (*N.A.T.O*, spoken as [neɪtəʊ] < *North Atlantic Treaty Organization*), clippings (*mike* < *microphone*), and truncations (*commie* < *communist*, *Becky* < *Rebecca*) in which only parts of words appear in the derived word (see Chapter 13).

In the domain of concatenative morphology, the structure of a complex word can be represented by means of labeled bracketing of a concatenation of morphemes. For instance, the English compound *songbook*, the derived word *singer,* and the plural noun *books* can be represented as follows:

(8) $[[song]_N\ [book]_N]_N$
$[[sing]_V\ er]_N$
$[[book]_N\ s]_N$

It is obvious that in the case of non-concatenative morphology it is not possible to represent the structure of a word in terms of a sequence of morphemes. For instance, the structure of *smog* can be represented linearly as *sm-og*, but these two constituents are not morphemes. The structure of this word can be represented by a non-linear representation of the following type:

(9)
```
s  m ou k
|  |
C C V  C
     |  |
   f o  g
```

This representation indicates how the abstract CCVC skeleton of the word *smog* is ‘fleshed out’ by sound segments of the words *smoke* and *fog*.

2. Morphological classification

Languages may be classified according to the role and nature of their morphology (Comrie 1981; Haspelmath 2009). A first dimension is the index of synthesis: languages that do not make use of morphology are called analytic or isolating, languages with a lot of morphology are called synthetic. Hence, languages may be ranked on an index of synthesis. Traditionally, Chinese is referred to as an isolating language because it has no, or almost no inflection. Yet, there is no doubt that word formation, in particular compounding, is very productive in this language (Packard 2000). Hence, Chinese is not analytic in an absolute sense.

The second index on which languages can be ranked is that of polysynthesis: some languages allow the incorporation of lexical morphemes, leading to relatively complex words, as illustrated by the following one-word-sentence of Central Alaskan

Yup'ik in which the lexical morpheme *tuntu* 'moose' is incorporated in the verb *te* 'to catch' (Mithun 2000: 923):

(10) tuntutuq=gguq
tuntu-te-u-q=gguq
moose-catch-INDICATIVE.INTRANSITIVE-3SINGULAR=HEARSAY
'He got a moose'

In this example, the complex word is followed by a clitic morpheme ('=' is the symbol for the link between a host word and a clitic) of evidentiality that indicates the nature of the information source.

The third dimension of classification is the index of fusion. In fusional languages, one morpheme may express more than one grammatical feature. Dutch is an example such a language. For instance, in (3), the inflectional suffix *-te* expresses both 'past tense' and 'singular number'. Fusional languages differ from agglutinating languages in which each bound morpheme corresponds with one grammatical feature. Turkish is the textbook example of an agglutinating language. For instance, case and number in Turkish are expressed by different suffixes:

(11) çocuk-lar-ın
child-PLURAL-GENITIVE
'of the children'

These three indices of morphological complexity are useful in given a global characterization of the morphology of a language. One should be aware, however, that languages are not homogeneous with respect to these indices. For instance, many Indo-European languages are fusional in their inflectional system, but agglutinating in their derivational morphology. Chinese also illustrates this point since it is synthetic as far as word formation is concerned, but analytic as far as inflection is concerned, as it has no inflection.

3. Morpheme order and hierarchical structure

A basic question in the analysis of the structure of complex words is what determines the order of its constituent morphemes.

A first principle is that the order of morphemes reflects the order of morphological derivation and its semantics. In the word *acceptability*, the suffix *-ity* is at the right edge because this word means 'the property of being acceptable'. This suffix takes adjectives as base words, and forms nouns from them. In its turn, *acceptable* is formed by adding the suffix *-able* to *accept*, with the meaning 'can be accepted'. The suffix *-able* takes verbs as bases to create adjectives. Thus, the following hierarchical word structure can be assumed, which also implies a specific linear order of the morphemes involved:

(12) $[[[\text{accept}]_V \text{ able}]_A \text{ ity}]_N$

At the same time, this hierarchical structure will determine the way in which this complex word is assigned its meaning.

A second principle that governs the order of morphemes is that inflectional morphemes are usually peripheral to derivational morphemes, this being one of the universals proposed by Greenberg (Greenberg 1963). This is exemplified by the word *workers*, in which the derivational morpheme *-er* precedes the plural suffix *-s*.

Within the domain of inflection, there are also principles for the order of morphemes. Consider the following inflected form of the Italian verb *temere* 'to fear'

(13) tem-e-va-no 'they feared'

In this verb form, the stem *tem* 'to fear' is followed by a so-called thematic vowel *-e-*, then by the suffix *-va-* which expresses past imperfective , and finally by the suffix *-no* which expresses third person plural. This form conforms to the following general schema:

(14) [base – thematic vowel]$_{\text{stem}}$ – inherent inflection – contextual inflection

The stem of a word is what remains of a word after removal of the inflectional elements. Contextual inflection is the type of inflection dictated by syntax, such as case marking on nouns and adjectives in specific syntactic configurations, and person and number marking on verbs in languages with subject-verb agreement. In example (13), the morpheme *-va-* expresses inherent inflection (past imperfective), and the morpheme *-no* is a case of contextual inflection, as it may be required by the presence of a third person plural subject. The base may consist of a root followed by a thematic vowel, as is the case in *temevano*, but it may also be complex itself. That is, the stem may contain derivational prefixes and/or suffixes.

The schema in (14) is an instantiation of the more general tendency of contextual inflection to be peripheral to inherent inflection (Booij 1996), which also applies to nouns. For instance, in inflected nouns case marking (a case of contextual inflection) is peripheral to plural marking (a case of inherent inflection), as in the inflected Turkish noun in (11) where the plural suffix indeed precedes the case suffix.

Languages may also exhibit 'position class morphology'. This means that they have complex words in which the order of the morphemes is determined by a template with a number of slots. Morphemes are then specified for appearing in a particular slot. For instance, in Nimboran, a Papuan language of New Guinea, each finite verb contains a number of morphemes, including the root. There are 8 positions, with the root appearing in position 0. The morpheme for 'plural subject' has to appear in slot 1, right after the root (slot 0), whereas the morpheme for 'person of subject' has to appear in slot 8. preceded by the morpheme for tense in slot 7 (Inkelas 1993).

In Kimatuumbi, a Bantu language spoken in Tanzania, the structure of the verb may be assigned a hierarchical structure of the following type (Odden 1996: 71) which specifies structural positions for specific morphemes:

(15) [(relative) [subject [object [[root – suffix]$_{\text{derivational stem}}$ final vowel]$_{\text{inflected stem}}$]]]]

This structure is exemplified by the following verb form:

(16) cha-ba-nį-telek-ey-a
what-they-me-cook-CAUSATIVE-FINAL VOWEL
'what they make me cook'

In compounds the order of the lexemes depends on which lexeme is the head, if there is one (these are so-called endocentric compounds). In Germanic languages, the head of a compound is its rightmost constituent, a regularity often referred to as the Right-hand Head Rule (Williams 1981). It is this constituent that determines word class and subclasses such as gender, etc. Compare the following minimal pairs of compounds in Dutch:

(17) a koor 'choir, neuter gender', zang 'song, common gender'
koor-zang 'lit. choir-song, choral singing' (common gender)
zang-koor 'lit. song-choir, choir' (neuter gender)
b water 'water, neuter gender', tafel 'table, common gender'
water-tafel 'water table, common gender'
tafel-water 'table-water, neuter gender'

These data show that in these Dutch compounds the head is on the right. For instance, *tafel-water* is a type of water, not a type of table, and it has neuter gender, just like its right constituent word *water*.

In compound schemas, the order of the elements is part of the definition of the schema, and thus defines the order of the compound's constituents. For instance, Romance languages such as Italian have compounds for agents and instruments in which the verb (V) precedes the noun (N):

(18) lava-piatti 'lit. wash-dishes, dish washer'
spazza-camino 'lit. sweep-chimney, chimney sweeper'

It is the order V-N that is essential for evoking this agent/instrument interpretation.

4. Meaningless elements in word structure

Complex words may contain building blocks that do not contribute to the meaning of the complex words, but are just there for creating a particular morphological form. An example is the occurrence of thematic vowels in verbal forms, exemplified in (9) above. The thematic vowel *e* of the verb *temere* 'to fear' has no meaning contribution of its own. It does define a conjugation class, as not all verbs of Italian have the same thematic vowel. Most verbs have *a*, others have *e* or *i*.

In compounds, the constituents may be linked by an element without a meaning of its own. For instance, Greek compounds have a linking element *-o-*:

(19) gloss-o-logos 'language-LINK-student, linguist''
pag-o-vuno 'ice-LINK-mountain, iceberg'

Such linking elements are also referred to as empty morphs or interfixes. The Slavic language Polish makes use of such interfixes both in compounding and derivation. In compounds, the default linking element is the vowel *o*, as in the following.

(20) gwiazd-o-zbiór ‘star-LINK-collection, star constellation’
star-o-druk ‘old-LINK-print, antique book’

In the domain of derivation, the suffix *-ski* that is used to derive adjectives from nouns (*student* ‘student’- *student*-ski ‘student-like’), is often preceded by an interfix *-ow-*, as in the following denominal adjectives with the meaning ‘related to N’ (Szymanek 2010: 253):

(21)	kat ‘hangman’	kat-ow-ski
	szpieg ‘spy’	szpieg-ow-ski
	tchórz ‘coward’	tchórz-ow-ski

These data show that word structure may comprise building blocks with a purely formal role.

Words may also show an internal formal structure without a correlating semantic structure. English has many prefixed verbs with a Latinate root, such as:

(22) con-ceive, de-ceive, per-ceive, re-ceive,
de-duce, in-duce, re-duce, se-duce

The roots *ceive* and *duce* do not have a meaning of their own. Yet these roots have specific properties. In particular, the nominalized form of all verbs in *-ceive* ends in *-ception*, and the nominalized form of all verbs in *-duce* ends in *-duction*. These generalizations can only be made if the formal internal structure of these verbs is recognized.

Complex words may lose their internal semantic structure due to the loss of the base word. For instance, Dutch has many verbs with a prefix followed by a root that once was a verb but has since disappeared as a verb on its own:

(23)	*prefixed verb*	*past participle*
	be-gin ‘to begin’	be-gonnen / *ge-be-gonn-en
	ge-niet ‘to enjoy’	ge-noten / *ge-ge-not-en
	ont-beer ‘to miss’	ont-beer-d / *ge-ont-beer-d
	ver-geet ‘to forget’	ver-get-en / *ge-ver-get-en

We have to assign internal morphological structure to these Dutch verbs since they behave as prefixed verbs. In Dutch, past participles are formed from the stem by means of prefixation with *ge-* and simultaneous suffixation with *-t/-d* or *-en*, as in *ge-werk-t* ‘worked’ and *ge-slap-en* ‘slept’. However, when the verb is prefixed, the prefix *ge-* has to be omitted:

(24) *simplex verb* *participle* *prefixed verb* *participle*

slaap ‘sleep’	ge-slap-en	be-slaap ‘sleep on’	be-slap-en
		ver-slaap ‘oversleep’	ver-slap-en
neem ‘take’	ge-nom-en	ont-neem ‘take away’	ont-nom-en
		ver-neem ‘hear’	ver-nomen

These facts also show that word-internal morphological structure may be relevant for the application of morphological rules. If the syllable *ver-* is a prefix, there is no prefix *g-*, but if *ver-* is just a syllable, as in the Dutch verb *verbaliseer* ‘to fine’, the past participle is *ge-verbaliseer-d*, not **verbaliseer-d*.

In the case of verbs such as those listed in (22-23) we speak of ‘formally complex words’. Words may also be partially motivated from a semantic point of view. In the English word *cranberry* we recognize the word *berry*, but we do not know what *cran* means. However, because a cranberry is a type of berry we have to consider *cranberry* a compound. Many words end in a recognizable suffix but the part before the suffix has no meaning of its own. This applies to many words in *-ism*:

(25) altru-ism, aut-ism, bapt-ism, pacif-ism, solips-ism

There are no base words *altru, aut, bapt, pacif*, and *solips*. However, we can tell from the fact that the words in (25) end in *-ism* that they are nouns with meanings such as ‘predisposition, activity, ideology’. This internal structure can also be seen when we derive corresponding nouns in *-ist* which denote a certain type of person: the part *-ism* is replaced with the suffix *-ist*:

(26) altru-ist, aut-ist, bapt-ist, pacif-ist, solips-ist

These facts show that words may exhibit a formal internal structure without a completely corresponding semantic structure.

5. Word structure and semantic interpretation

The classic principle of compositionality states that the meaning of a complex linguistic expression is a compositional function of the meaning of its constituent words and the manner of their combination. The interpretation of the word *acceptability* mentioned in section 3, ‘the property of being acceptable’, is a clear illustration of this principle. In the case of the right-headed compounds of Dutch, also discussed in section 3, we saw the role of hierarchical structure: it is the head that determines primarily which concept is denoted by the compound as a whole.

The Italian VN compounds in (18) are called exocentric, because they do not have a head. The constituent lexemes contribute to the meaning, but the agent or instrument meaning is not expressed by a morpheme. This meaning component is a property of the morphological construction as a whole.

Holistic properties of complex words can also be observed for the class of so-called co-compounds, in which two words are coordinated into a compound. In some cases, the meaning of the compound is a hyperonym of the meanings of the constituent lexemes. Here are some examples from Lezgian (Haspelmath 1993: 108):

(27) biba-dide ‘father-mother, parents’
xeb-mal ‘sheep-cattle, domestic animals’

It is the coordination construction that evokes the hyperonym interpretation.

Reduplication is interesting from the semantic point of view because it is the prototypical example of the role of iconicity (meaning as a direct reflection of form) at the level of the word: the formal repetition of the base word may indicate a repetition of its meaning, that is ‘increase’. Hence, a reduplicated noun may indicate plurality, a reduplicated verb that the action denoted by the base verb takes place intensively or iteratively, and a reduplicated adjective may indicate a high degree of the property expressed by the base adjective (see the examples 5). This illustrates once more that meaning components of complex words may arise from the morphological construction as such. Note also that the meaning of reduplication is not always that of plurality or intensity. In the reduplicated form of the Begak verb *panow* ‘to go’, the verb *panow-panow* ‘to go a little bit’ in (5), the meaning contribution of the reduplication pattern is that of diminishment. Hence, other types of meaning than ‘increase’ may be linked to the reduplication construction.

A second challenge for the compositionality principle as the sole principle for constructing the meanings of complex words is the phenomenon of conversion, the change of words to another word class without the addition of overt morphological material. In English, nouns can be converted to verbs without adding an affix, as is illustrated by the following sentences in which the recently coined verbs *to virus*, *to transition*, and *to friend*, and are used (data from *The Boston Globe*, 24 May, 2010):

(28) I keep getting a message saying my computer is virused
He is transitioning to a new job.
Do you get offended when someone won’t friend you on Facebook?

Although it is easy to reconstruct the meaning of these verbs on the basis of the meaning of the corresponding nouns, there is an additional meaning component ‘to perform an action related to the base noun’ which is not marked by a morpheme, but only by the change of word class. Moreover, the exact nature of the relation between the noun and the kind of action has to be determined on the basis of context or conceptual knowledge.

Conversion may be qualified as a type of paradigmatic word formation, since the meaning of the verb is to be defined with reference to the meaning of the corresponding noun (the term ‘paradigm’ is used to denote systematic relationships between words in the lexicon of a language). One may ‘translate’ this paradigmatic relation into a syntagmatic structure, by assuming conversion constructions such as the following for English (Booij 2010):

(29) $[[x]_{Ni}]_{Vj}$ ‘action V_j related to N_i’

In other cases of paradigmatic word formation, such a structural interpretation is not available, and we then have to make explicit reference to the paradigmatic

relationship between the two sets of words in order to express the relationship between form and meaning. Consider the following Dutch word pairs (Booij 2010: 34):

(30)	*verb*	*noun*
	alloc-eer ‘to allocate’	alloc-atie ‘allocation’
	communic-eer ‘to communicate’	communic-atie ‘communication’
	reden-eer ‘to reason’	reden-atie ‘reasoning’
	stabil-is-eer ‘to stabilize’	stabilis-atie ‘stabilization’

The nouns in *-atie* ‘-ation’ are semantically the nominalizations of the verbs on the left, just like their English counterparts. Yet, the part before *-atie* is not identical to the corresponding verb, which has an additional morpheme *-eer*. This mismatch between form and meaning is due to the historical fact that the inflectional ending *-er* of French infinitives has been reinterpreted in Dutch as a derivational suffix *-eer*. In combination with the massive borrowing of the corresponding nouns in *-atie*, this led to the following productive paradigmatic pattern, where the symbol ≈ stands for ‘paradigmatically related to’, the angled brackets demarcate a morphological schema, and the double arrow stands for the relationship between form and meaning (SEM):

(31) $< [\text{x-eer}]_{Vi} \leftrightarrow [\text{SEM}]_i > \approx <[\text{x-atie}]_{Nj} \leftrightarrow [\text{action of SEM}_i]_j >$

New Dutch nouns in *-atie* can readily be formed on the basis of verbs in *-eer*. For instance, now that the verb *implement-eer* ‘to implement’ has been coined, the corresponding noun *implement-atie* ‘implementation’ is its derived nominal.

Mismatch between form and meaning may also be due to the fact that a constituent of a complex word has a different semantic scope from what we would expect given its structure. This is illustrated by the following complex adjectives of Dutch

(32)			
	a	tak ‘branch’	ge-tak-t ‘branched’
		spits ‘point’	ge-spits-t ‘pointed’
	b	rok ‘skirt’	kort-ge-rok-t ‘short-skirted’
		jas ‘coat’	wit-ge-jas-t ‘white-coated’
		schouder ‘shoulder’	breed-ge-schouder-d ‘broad-shouldered’

Denominal adjectives can be formed with a prefix *-ge* and the simultaneous attachment of the suffix *t/d*, as shown in (32a). These complex adjectives can function as the head of an adjectival compound of the type Adjective + Adjective. However, the modifier adjective does not modify its adjectival head as a whole, but only its nominal base part. For instance, *kortgerokt* does not mean ‘skirted in a short manner’ but ‘dressed in a short skirt’. Hence, the semantic scope of the modifier is more restricted than the structure of the AA compounds would make us to expect. This restricted semantic scope is a property of this type of AA-compounds as a whole. Similar examples from English are the lexical units *criminal lawyer* and *atomic physicist* in which the adjectives do no modify the meaning of the whole noun, but only of a subpart if it.

In sum, even though the structure of complex words is necessary for computing their meaning, there are a number of complications in the relationship between the

morphological structure and the semantic structure of complex words, which require us to specify holistic semantic properties of complex words.

6. Word structure and the interface with phonology

The structure of complex words is not only important for computing their meaning, but also for computing their phonological structure and properties. The systematic relationship between two different levels of linguistic structure is referred to as interface. Whereas section 5 discussed the interface between morphological form and semantic structure, this section focuses on the way that morphological structure determines the way in which the constituent morphemes of a complex word are realized phonetically. This type of interface is an important topic of research in phonology.

A first illustration of this interface is the assignment of stress patterns to complex words. In Dutch compounds, main stress falls on the first constituent, and secondary stress falls on the second constituent:

(33) a kóor-zàng ‘lit. choir-song, choral singing’
záng-kòor ‘lit. song-choir, choir
b. wáter-tàfel ‘water table’
táfel-wàter ‘table-water’

A similar rule applies to English (the so-called Compound Stress Rule), but in some types of English compounds, the second constituent carries main stress – compare *Bráttle Street* with *Massachusetts Ávenue*. In the words in (33a), two consecutive syllables carry stress (a stress clash), whereas Dutch simplex words exhibit a regular rhythmic alternation between stressed and unstressed syllables. This stress clash is due to the fact that each compound constituent is a phonological word of its own (cf. Chapter 7), and hence a separate domain of stress assignment. In other words, the presence of morphological structure is essential in explaining why two consecutive syllables of a word can be stressed.

Affixes may also play a crucial role in computing the location of the main stress of a word. For instance, English nouns in *-ity* require the main stress of the word on the last syllable before the suffix:

(34) superfícial superficiálity
fúnctional functionálity

Languages differ in the extent to which morphological structure plays a role in stress assignment. For instance, in Polish the presence of suffixes has no influence on the location of the main stress: both simplex and suffixed words carry main stress on the penultimate syllable. On the other hand, in Polish compound words, each part functions as a separate stress domain.

Word structure also plays a role in the way in which a word is divided into syllables. In Germanic languages, the morphological boundary between the constituents of a compound usually coincides with a syllable boundary, because each of these constituents corresponds with a separate phonological word. On the other hand, the

morphological boundary before vowel-initial suffixes is ignored in syllabification. The contrast is illustrated by the syllabification of the suffixed words *walker* and *walking* versus the compound word *walk-in* (as in *walk-in wardrobe*) (where σ denotes a syllable):

(35) walk-er (wa:)$_{\sigma}$(kər)$_{\sigma}$
walk-ing (wa:)$_{\sigma}$(kər)$_{\sigma}$
walk-in (wa:k)$_{\sigma}$(in)$_{\sigma}$

In this respect, prefixes often differ from suffixes. In English, the morphological boundary after *un-* coincides with a syllable boundary. For instance, in *un-able*, the syllabification in careful speech is (*un*)$_{\sigma}$(*a*)$_{\sigma}$(*ble*)$_{\sigma}$ rather than (*u*)$_{\sigma}$(*na*)$_{\sigma}$(*ble*)$_{\sigma}$, even though in simplex words the C in a VCV sequence always forms one syllable with the following V. In *un-natural* the prefix *un-* also preserves word-like properties in that *unnatural* may be spoken with a lengthened, geminate [n:], which we do not find in simplex words.

A third domain of phonology where word structure plays a role is that of allomorphy: morphemes may have different phonetic shapes co-varying with their appearance as independent words or as parts of complex words. Consider the following set of related words:

(36) sane sanity
opaque opacity

The letter *a* stands for the tense vowel [eI] in the adjectives, and for the lax vowel [ɛ] in the corresponding de-adjectival nouns. This alternation does not apply, however, to all words with this suffix witness a noun like *obesity* where the vowel of the second syllable is the tense vowel [i:].

Allomorphy may be triggered by specific morphemes. In English, the Latinate prefixes *in-* and *con-* undergo a specific assimilation process of the nasal consonant

(37) in-adequate, im-potent, in-tact, i[ŋ]-competent, il-logical, ir-rational
com-petent, con-tact, co[ŋ]-gress, col-league, cor-relation

This assimilation process does not apply to other morphemes ending in /n/. For instance, the nasal consonant of *un-* in *unless* is not pronounced as [l].

A different type of allomorphy is the alternation between voiced and voiceless stops and fricatives at the end of morphemes in Dutch and German. The rule is that these consonants are realized as voiceless at the end of a syllable. Here are some examples from German:

(38) Bund [bunt] ‘union’ Bund-e [bundə] ‘unions’ syllables: bun.də
Sarg [sɑrk] ‘coffin’ Särg-e [sɛrgə] ‘coffins’ syllables: sɛr.gə

In this case, morphological structure has an indirect effect only. Due to the addition of the plural morpheme *-e* [ə] the final consonant of the stem morpheme appears in the onset of the second syllable. Hence, it is no longer syllable-final and will therefore not be devoiced.

7. Loss of word-internal structure

Complex words, once formed and stored in lexical memory, may lose their semantic transparency, because the meaning of the word as a whole has become idiosyncratic and is no longer fully derivable from the meanings of its constituents. This may also lead to phonological change. Classic examples are the faded compounds *lord* and *lady* that derive historically from the old English compounds *hlāf-weard* 'bread-warden' and *hlāf-dige* 'bread-kneader'.

One cause of this loss of motivation is that one of the constituent parts is no longer a word by itself. This can be seen in the names of the days such as *Tuesday* and *Wednesday*, where only the part *day* is recognizable. *Tues*- derives from *Tiwe*s, the genitive form of *Tiw*, the name of a god of war and law, and *Wednes*- derives from the genitive form of the Germanic god *Wodan*. And a *holi-day* is no longer a *holy-day*. In the Dutch word *barn-steen* and its German equivalent *Bern-stein*, both meaning 'amber', the second part is recognized as the regular word for 'stone', but there is no recognizable morpheme *barn/bern*. This constituent derives historically from the verb *bern* 'to burn' which no longer exists in these languages.

The phonological form of a demotivated compound may change into the canonical form of a simplex word. For instance, *gospel* (< *god-spell* 'good message') has lost the /d/, since /ds/ is not a regular word-medial consonant cluster in English simplex words, and the second syllable of a simplex word has [ə] as its canonical vowel. In certain varieties of Dutch, the still existing standard Dutch compound *boom-gaard* 'lit. tree-garden, orchard' has acquired the canonical phonetic form of a simplex word (two syllables, the first stressed, the second with [ə]). Hence the word *boomgaard* [bó:mɣà:rt] changed into *bongerd* [bóŋərt].

Loss of transparency can also be seen in the way in which a word is syllabified. For instance, although Dutch still has the words *voort* 'forth' and *aan* 'on', the compound *voortaan* has the meaning 'from now on', in which the two constituents are no longer recognized. Hence, this word is syllabified as $(vo{:}r)_\sigma$ $(ta{:}n)_\sigma$, not as $(vo{:}rt)_\sigma$ $(a{:}n)_\sigma$, as would be expected if it were still recognized as a compound.

In sum, complex words, once formed, may be subject to a process of demotivation and lose their morphological structure. The cause of this loss is semantic in origin, but it may have effects on phonological form as well, since demotivation moves the phonetic forms of words into the direction of the canonical phonological forms of simplex words.

8. Univerbation and grammaticalization

Morphological processes such as compounding and derivation are the main sources of complex words in a language. However, there is a second source, the historical process by which a sequence of words, often a phrase, becomes tighter, and behaves (and is often written) as one word. This process is referred to as univerbation. An example from English is the word *notwithstanding*, which derives historically from the word *not* and the participle *withstanding*. In modern English it occupies the position of a single word, the preposition in a prepositional phrase, as in *notwithstanding his request* (compare the ungrammatical *withstanding his request*). Another example is the conjunction *because*

from Middle-English *bi + cause* ‘by cause of’, parallel to French *par cause* ‘by cause of’, and the compound *handful* from *hand + full*, as in *a handful of eggs*, in which *handful* refers to a quantity, not to what is (literally) contained in a hand or a spoon.

When a univerbated word sequence becomes the model for a class of words, univerbation leads to the rise of a morphological process. A famous example is the rise of the adverbial suffix *-ment*(*e*) in various Romance languages. It derives from the ablative form of the Latin noun *mens* ‘mind’. For instance, *clara mente* (two words) meant ‘with a clear mind’; interpreted as one word, it became the adverb *claramente* ‘in a clear manner’. This and similar words could then function as models for new deadjectival adverbs in *-mente*. The historical origin can still be observed in the fact that *-mente* takes the feminine form of adjectives as stem form since Latin *mens* is a feminine noun, as in Italian *chiar-a-mente* and French *clair-e-ment* in which the stem forms *chiar-a* and *clair-e* have a feminine ending.

An important difference between word formation and univerbation is that the output words of word formation always belong to the class of lexical words: nouns, verbs, adjectives, and adverbs, whereas univerbation may lead to the expansion of classes of grammatical (or function) words. For instance, *notwithstanding* is a preposition, and *because* is a conjunction.

A second source of grammatical words with word-internal structure is the process of grammaticalization. This is the process in which words of lexical classes become grammatical words within specific syntactic constructions, or grammatical words develop into even more grammatical words (Hopper and Traugott 2003). An example is the reinterpretation of past participles in participial clauses as prepositions, as illustrated by the use of *given* in *given this situation*. Dutch has many prepositions with the form of a present or past participles that now function as prepositions:

(39) *present participles > prepositions*
ge-dur-end-e ‘lit. lasting, during’
staan-d-e ‘’lit. standing, during’
niet-tegen-staan-de ‘not-with-stand-ing

past participles > prepositions
ge-gev-en ‘given’
ge-zien ‘lit. seen, because of’
uit-ge-zonder-d ‘out-PREF-separate-SUFF, with the exception of’

In conclusion, univerbation may lead to new complex words, including words belonging to grammatical classes. The process of grammaticalization may lead to the expansion of sets of grammatical words with word-internal structure. Therefore, we should not conceive of morphological processes as the only source of complex words.

9. Suggestions for further reading

Recent introductions to the study of word structure are Booij (2012) and Haspelmath & Sims (2010). Handbooks of morphology are (Booij et al. 2000/2004) and Spencer &

Zwicky (1998). A handbook on compounding is Lieber & Stekauer (2009) and a typological survey of word formation can be found in Štekauer et al. (2012).

Information on the morphology of the world's languages is provided in Haspelmath et al. (2005), which can also be consulted on-line: http://wals.info/. The standard glossing rules for word-internal structure, the Leipzig glossing rules, are available on http://www.eva.mpg.de/lingua/resources/glossing-rules.php.

The interface between morphology and phonology is the topic of Volume 4 of Van Oostendorp et al. (2011).

The lexicalization and grammaticalization of complex words are discussed in Hopper & Traugott (2003), Brinton & Traugott (2005), and in Narrog & Heine (2011).

References

Aronoff, Mark (1994). *Morphology by Itself: Stems and Inflectional Classes*. Cambridge Mass.: MIT Press.

Booij, Geert (1996). 'Inherent versus Contextual Inflection and the Split Morphology Hypothesis', in G. Booij and J.van Marle (eds.), *Yearbook of Morphology 1995*. Dordrecht / Boston: Kluwer, 1-16.

Booij, Geert (2010). *Construction Morphology*. Oxford: Oxford University Press.

Booij, Geert. 2012. *The Grammar of Words.* 3rd edition Oxford: Oxford University Press.

Booij, Geert, Lehmann, Christian, and Mugdan, Joachim (eds.) 2000/2004. *Morphologie / Morphology. Ein internationales Handbuch zur Flexion und Wortbildung / An International Handbook on Inflection and Word Formation.* (Handbücher zur Sprach- und Kommunikationswissenschaft 17.1/2). Berlin: De Gruyter.

Brinton, Laurel J., and Traugott, Elizabeth C. (2005). *Lexicalization and Language Change*. Cambridge UK: Cambridge University Press.

Comrie, Bernard (1981). *Language Universals and Linguistic Typology*. Oxford UK: Blackwell.

Goudswaard, Nelleke (2005). *The Begak (Ida'an) Language of Sabah*. Utrecht: LOT.

Greenberg, Joseph (1963). 'Some Universals of Grammar, with Particular Reference to the Order of Meaningful Elements'. In J. Greenberg (ed.), *Universals of Language.* Cambridge Mass.: MIT Press, 73-113.

Haspelmath, Martin (1993). *A Grammar of Lezgian*. Berlin: Mouton de Gruyter.

Haspelmath, Martin. 2009. 'An Empirical Test of the Agglutination Hypothesis', in E. Magni, S. Scalise and A. Bisetto (eds.), *Universals of Language Today*. Dordrecht: Springer, 13-29.

Haspelmath, Martin, Dryer, Matthew S., Gil, David ,and Comrie, Bernard (eds.) (2005). *The World Atlas of Language Structures*. Oxford: Oxford University Press.

Haspelmath, Martin, and Sims, Andrea D. 2010. *Understanding Morphology*. London: Hodder Education.

Hopper, Paul J., and Traugott, Elisabeth C. (2003). *Grammaticalization*. Cambridge U.K.: Cambridge University Press.

Inkelas, Sharon (1993). 'Nimboran Position Class Morphology', *Natural Language and Linguistic Theory* 11:559-624.

Kutsch Lojenga, Constance (1994). *Ngiti, A Central-Sudanic language of Zaire.* Köln: Rüdiger Köppe Verlag.

Lehmann, Christian (2004). 'Interlinear Morphemic Glossing', in G. Booij, C. Lehmann, and J. Mugdan (eds.), *Morphology / Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung / A Handbook on Inflection and Word Formation / (*Handbücher zur Sprach- und Kommunikationswissenchaft 17.2). Berlin: De Gruyter, 1834-57.

Lieber, Rochelle and Štekauer, Pavol (2009). *The Oxford Handbook of Compounding*. Oxford: Oxford University Press.

Mithun, Marianne (2000). 'Incorporation', in G. Booij, J. Mugdan, and C. Lehmann (eds.), *Morphologie / Morphology. Ein internationales Handbuch zur Flexion und Wortbildung / An International Handbook on Inflection and Word Formation.*

(Handbücher zur Sprach- und Kommunikationswissenschaft 17.1). Berlin: De Gruyter, 916-28.
Narrog, Heiko, and Heine, Bernd (eds.) (2011). *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press.
Odden, David (1996). *The Phonology and Morphology of Kimatuumbi*. Oxford: Clarendon Press.
Packard, Jerome (2000). *The Morphology of Chinese. A Linguistic and Cognitive Approach*. Cambridge UK: Cambridge University Press.
Spencer, Andrew, and Zwicky, Arnold (eds.) (1998). *The Handbook of Morphology*. Oxford: Blackwell.
Štekauer, Pavol, Valera, Sandor, and Körtvélyessy, Lívia (2012). *Word-formation in the World's Languages*. Cambridge: Cambridge University Press.
Szymanek, Bogdan (2010). *A Panorama of Polish Word-formation.* Lublin: Wydawnictwo KUL.
Van Oostendorp, Marc, Ewen, Colin J., Hume, Elizabeth and Rice, Keren (eds.) (2011). *The Blackwell Companion to Phonology*. Malden Mass. / Oxford: Blackwell.
Williams, Edwin (1981). 'On the Notions 'Lexically Related' and 'head of a Word'', *Linguistic Inquiry* 12:245-74.