

## Morphological analysis

Geert Booij (University of Leiden).

[to appear in Bernd Heine and Heiko Narrog (eds.), *The Oxford Handbook of Grammatical Analysis*. Oxford: Oxford University Press]

### 1. What is morphological analysis?

Morphology is the subdiscipline of linguistics that deals with the internal structure of words. Consider the following sets of English word pairs:

(1)	<i>Verb</i>	<i>Noun</i>
	bake	baker
	eat	eater
	run	runner
	write	writer

In these word pairs we observe a systematic form-meaning correspondence: the presence of *-er* in the words in the right column correlates with the meaning component ‘one who Vs’ where V stand for the meaning of the corresponding verb in the left column. The observation of such patterns is the basis for assigning the words in the right column an internal morphological structure  $[[x]_V -\text{er}]_N$  where the variable x stands for the phonological form of the base verb. We thus consider these nouns to be complex words. The morphological schema that generalizes over these sets of paradigmatically related words may be formalized as follows:

(2)  $[[x]_V -\text{er}]_N$  ‘who Vs’

This schema expresses the systematic form-meaning correspondence found in this set of word pairs. Words are signs with properties at a number of levels of the grammar: they have a phonological form, syntactic properties such as being a noun or a verb, a meaning, and sometimes a particular pragmatic value. Hence, morphology is not a component of the grammar on a par with phonology or syntax. It deals not only with form, unlike what the etymology of the word suggests, but pertains to all levels of the grammar (Jackendoff, 2002). Morphology is the grammar of a natural language at the word level, and calling morphology ‘the grammar of words’ (Booij, 2007) is therefore quite appropriate.

The schema in (2) expresses a generalization based on a number of existing verb-nouns pairs of the relevant type. Such schemas also indicate how new complex words can be made. Indeed, the process of creating deverbal *-er*-nouns is quite productive in English. Morphological schemas are word-based since they express generalizations concerning established complex words. In that sense, morphology is word-based. The language user will learn these abstract schemas gradually, after having been exposed to a sufficient number of words that instantiate those schemas. The acquisition of these schemas does

not imply that the complex words on which they are based are removed from lexical memory once the schemas have been acquired. Schemas co-exist with the complex words that instantiate these schemas (Bybee, 1988; Bybee, 1995). Hence, the grammar exhibits redundancy, which is no problem given the vastness of human memory. The wrong assumption that the existence of a rule excludes listing outputs of that rule is referred to as the rule-list fallacy (Langacker, 1987).

In morphological analysis we make also use of the notion ‘morpheme’, traditionally defined as the minimal meaning-bearing unit of a language. The word *baker*, for instance, might be said to consist of the lexical morpheme *bake* and the bound morpheme *-er*. However, the systematic paradigmatic relationships between words may also be signaled by other means than morpheme concatenation, such as stem alternation, reduplication, stress, and tone patterns. Therefore, the notion ‘morpheme’ is a useful analytic notion for the description of the internal structure of words, but not the starting point of morphological analysis and morphological structure.

The two basic functions of morphological operations are word formation and inflection. Word formation processes create new words, and hence expand the lexicon of a language. Inflection is the grammatical subsystem that deals with the proper form of words in specific syntactic contexts. In Dutch, for instance, the verb *werk* ‘to work’ has three different finite forms, depending on the number and person of the subject of the clause in which this verb occurs:

(3)	<i>werk</i>	present 1 pers.sg
	<i>werk-t</i>	present 2/3 pers.sg
	<i>werk-en</i>	present 1/2/3 pers.pl
	<i>werk-te</i>	past 1/2/3 pers.sg
	<i>werk-te-n</i>	past 1/2/3 pers.pl

We consider these five forms as forms of the same word. The notion ‘word’ in this more abstract sense is usually referred to as ‘lexeme’. Thus, Dutch has a lexeme WERK (lexemes may be indicated by small capitals in order to avoid ambiguity). The stem form of this lexeme is *werk*, and the different inflectional affixes are added to this stem. The 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* has the following morphological structure:

(4)	<i>werk</i>	<i>-er</i>	<i>-s</i>
	<i>work</i>	<i>-AGENT-PL</i>	
			‘workers’

This is a simple example of morphological analysis, 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.

The set of verbal forms in (3) illustrates the well known problem that there is no one-to-one mapping between morphemes and units of (grammatical) meaning, also

referred to as ‘cumulative exponence’. For instance, the *-t* of *werkt* expresses values for the following grammatical categories:

- (5) Person: 2 or 3  
Number: singular  
Tense: present

For this reason, a morpheme like *-t* is traditionally called a ‘portmanteau morpheme’. The Dutch sentence *Jan werkt* will receive the following glossing:

- (6) Jan                werk-t  
John.3SG        work-3SG.PRES  
‘John is working’

Grammatical features that are expressed by the same morpheme are separated by a dot instead of a hyphen. The combination of feature values for person and number is usually given without an internal dot.

Two other notions are important for morphological analysis, the notions ‘root’ and ‘stem’. The stem of a word is the form minus its inflectional markers. The root of a word is the stem minus its word formation morphemes. Hence, in the English word *workers* the stem is *worker*, and the root is *work*. Another example is that in Polish the root noun *kos* ‘scythe’ can be turned into a verbal stem with the meaning ‘to mow’ by adding the verbalizing suffix *-i*. This verbal stem can then be used for deriving verbal forms such as the present participle *košonc* ‘mowing’, the phonetic form of *kos-i-onc*.

## 2. Word formation

Natural languages make use of a number of formal means for the formation of complex lexemes: compounding, affixation, reduplication, conversion, stem alternation (also referred to as internal modification), stress, and tone.

In compounding two or more lexemes are combined into a new one. Cross-linguistically, compounding is one the most common means for word formation, in particular compounding in which one of the constituents is the head (so-called endocentric compounding). The English word *football* is a compound, consisting of the two lexemes *foot* and *ball*, of which the second functions as the head: a football is a particular kind of ball, not a kind of foot. There are also languages with left-headed compounds, such as Maori. The Maori compound *roro-hiko* ‘lit. brain electricity, computer’ denotes a particular kind of brain, namely a computer, not a particular form of electricity. In exocentric compounds such as *pickpocket* there is no constituent that functions as the head: a *pickpocket* is neither a pocket nor a pick. An example of an exocentric compound from Mandarin Chinese, a language with lots of exocentric compounds, is the compound *tian fang* consisting of the verb *tian* ‘to fill’ and the noun *fang* ‘room’ with the meaning ‘second wife (to a widower)’. Besides subordinating compounds, with one of the constituents functioning as the head, there are also coordinating compounds, such as Sanskrit *maa-pio* ‘mother and father’, and English *singer-actor*.

The second widespread process used in word formation is affixation, whereby an affix is prefixed, suffixed, infix, or circumfixed to some input form. Each of the four options is illustrated in (7):

- (7) prefixation: *un-happy* from *happy* (English)  
 suffixation: *happi-ness* from *happy* (English)  
 infixation: *s-m-ka:t* ‘to roughen’ from *ska:t* ‘rough’ (Khmu, a language spoken in Laos)  
 circumfixation: *ge-been-te* ‘bones’ from *been* ‘bone’ (Dutch)

Compounding and affixation are referred to as concatenative morphology since their mode of operation is that of concatenating roots, stems, and affixes. A special form of concatenative morphology is reduplication. This is the process in which a stem or part thereof is copied and prefixed or suffixed to that stem, as in Javanese *baita-baita* ‘various ships (full reduplication) and *tə-tamu* ‘to visit’ from *tamu* ‘guest’ (partial reduplication with copying of the initial consonant and insertion of a default vowel).

In non-concatenative morphology, other formal means are involved in the creation of new morphological forms. In the case of internal modification a stem with a different form is created, for instance, by replacing a vowel pattern or a consonant pattern (or both) with another one. Vowel alternations are characteristic of a number of Indo-European languages; in Semitic languages verbal roots may appear in a number of different ‘binyanim’, templates with specific patterns of consonants and vowels, sometimes in combination with a prefix:

- (8) *stem alternation* (Dutch, Indo-European)  
*sluit* ‘to close’ *slot* ‘lock’  
*bind* ‘to bind’ *band* ‘bond’

*binyan system* (Modern Hebrew, Semitic)  
*katav* (pattern CaCaC) ‘wrote’  
*ni-ktav* (pattern ni-CCcaC) ‘was written’  
*kitev* (pattern CiCeC) ‘inscribed’ (intensive meaning)

Other non-concatenative means for marking morphological operations are stress (as in the English word pair *to review* (verb) versus *réview* (noun) and the use of tone to mark specific morphological categories. In some cases the tonal marking can be analysed as a case of concatenative morphology. An example of the latter from the African language Noni is the following set of pairs of singular and plural nouns (Hyman, 2000) p. 590:

- (9) *singular: LH tone pattern*      *plural: H tone pattern*  
*bwě*                                      *bwé*                                      ‘dog’  
*jĩn*                                        *jín*                                        ‘maggot’

Since the roots involved may be assumed to have a lexical High tone, Hyman qualifies the Low tone that is part of the singular tone pattern as a tonal prefix that marks the singular. Hence, affixes can also consist of suprasegmental units.

New lexemes may also be created without overt formal marking, which is referred to as conversion. A well known case is the conversion of nouns into verbs, a very productive process in languages like English and Dutch; the following examples are from Dutch (the nouns are recent English loans except *contact*):

(10)	<i>noun</i>	<i>verb</i>
	contact ‘id.’	contact ‘to make contact with’
	computer	computer ‘to make us of a computer’
	skype	skype ‘to communicate by means of Skype’
	sms	sms ‘to send an SMS message’

These recently coined examples show how productive this way of creating lexemes is in Dutch. The morphological structure of such verbs can be represented as  $[[x]_N]_V$ , and the corresponding meaning as ‘to perform an act in which N is involved’. In this way we avoid the assumption of arbitrary zero-morphemes that are sometimes used in morphological analyses to account for conversion.

The common denominator for all word formation processes except compounding is derivation. Hence, the notion ‘word formation’ comprises both compounding and derivation.

### 3. The lexicon

The term ‘lexicon’ refers to the component of the grammar that minimally contains a specification of the lexical units of a particular language. The set of lexical units is larger than the set of words. Idiomatic expressions such as *to kick the bucket* ‘to die’ that are phrasal in nature need to be listed as well. This also applies to the many noun phrases that are used as classifying terms such as *blue cheese* and *yellow pages* that form established ways of denoting certain entities. The distinction between the notions ‘word’ and ‘lexical unit’ is important for a proper understanding of the relationship between morphology and syntax, as we will see below.

The set of words that are to be registered in the lexicon is the set of established words, that is, the set of words that is used by more than one native speakers and on more than one occasion. Thus, the lexicon is part of the language norm since it specifies the lexical conventions of a language. This norm can be changed by adding new words to the lexicon.

New complex words, once established, will be added to the lexicon. From a diachronic perspective there are other means of extending the set of complex words as well, as will be discussed below, in the section on diachrony.

In the lexicon, established complex words coexist with the schemas according to which they are formed. The schemas express generalizations about sets of established words, and indicate how new words can be formed. The relation between the schemas and their instantiations can be conceived of as a hierarchical lexicon in which a schema forms a node that dominates its instantiations. All properties shared by a set of words are specified in the schema, and the individual words inherit these properties from the node that dominates them.

#### 4. Inflection

So far, the focus of this chapter has been on lexeme formation. In this section, we will consider the analytical challenges posed by the other domain of morphology, the study of inflectional systems. In a language with inflection, lexemes may have more than one inflectional form. The set of inflectional forms of a lexeme is traditionally represented in the form of a paradigm in which each cell contains a form that expresses a particular array of grammatical features. Inflection has two basic functions. The first is that of creating forms of lexemes with a formal marking for certain grammatical categories such as number for nouns, and tense and aspect for verbs. This kind of inflection is referred to as inherent inflection (Booij, 1993; Booij, 1996) because the choice of the inflectional form is not governed by syntactic context, but by semantic considerations; it reflects the choice of the language user as to what information (s)he wants to convey. The other kind of inflection is contextual inflection, inflection that is determined by the syntactic context in which a lexeme occurs. In a language with a case system, for instance, nouns must have a particular case form because they are the head of an NP in a certain syntactic position that requires case marking. In many languages, finite forms of verbs have to agree with person and number properties of the subject NP. Adjectives may have to agree with respect to certain properties (such as gender, case, and (in)definiteness) with the nominal head that they modify. Verbs and prepositions may require specific case markings on their NP-arguments, as is the case for German. The following sentence illustrates both inherent and contextual inflection for Latin. The cases of pure contextual inflection are in bold print:

- (11) Te                    semper,            ut    **omn-ibus**        pate-t,  
      You.ACC.2SG always,        as    all.DAT.PL        is.clear-PRES.3SG
- immoderat-o**                    amor-e  
      unfettered-MASC.SG.ABL        love- MASC.SG.ABL
- complex-a**    sum  
      embrace.PART.PERF-FEM.SG.NOM        be.PRES.1SG

(DAT = dative, ABL = ablative). This sentence (meaning ‘I have always embraced you with unfettered love, as everyone knows’) is from a famous medieval love letter by Heloïse to Abelard (Janson, 2004), p. 139. The verb *patere* ‘to be clear’ requires its non-subject argument to be marked by the dative case. The ablative marking on *amor* ‘love’ is a case of inherent inflection (‘semantic case’), chosen to express a circumstantial meaning. The corresponding marking on the adjective *immoderato*, on the other hand, is a case of contextual inflection, required by the rule of agreement between modifying adjective and nominal head. The word sequence *complexa sum* is the perfective form of a so-called deponent verb, a verb with active meaning but a passive form. In the perfect tense, such deponent verbs have a periphrastic form, consisting of two words, a perfect/past participle and a form of the verb *esse* ‘to be’. The participle has inflection for feminine gender since the writer of this sentence, Heloïse, is a female. The subject itself, however, is not expressed by a separate noun. We may still consider this contextual

inflection if we take the notion ‘context’ to include the pragmatic context. The ending *-a* expresses both feminine gender and nominative case. An overt subject for the finite form *patet* is missing as well. In fact, the form of the verb enables us to reconstruct the subject as a 3<sup>rd</sup> person singular entity. The verb thus agrees in person and number with an abstract, non-overt subject. Hence, the suffix *-t* of *patet* expresses both inherent inflection (tense) and contextual inflection.

The word *sum* illustrates another morphological phenomenon, that of suppletion. This is the situation of different lexical roots playing a role in filling the cells of the paradigm of a lexeme. Two additional formal complications in inflection are the role of inflectional classes (declinations for nouns and conjugations for verbs), and stem selection. In example (11), the word *amor* ‘love’ belongs to the ‘third declination’ of nouns which implies that the ablative singular case is expressed by the ending *-e*, whereas the participle *immoderat-us* is inflected according to the default declination class for adjectives, hence the corresponding ending *-o*. In verbal paradigms the stem to which the inflectional endings are attached may have more than one form. The verb *patet* is a form of the verb *patere* ‘to be clear’ that belongs to the second conjugation, the verbs with the ‘thematic vowel’ *e* in between the root and the inflectional endings. The stem form of the participle *complexa* is /pleks/, whereas the form /plekt/ is used in the finite forms present and past (as in *com-plect-or* ‘I embrace’).

This small piece of morphological analysis, which includes the use of notions like suppletion, deponent verb, stem allomorphy, and inflectional class shows how complex the relation between form and meaning can be in inflectional systems (Aronoff, 1994).

The contextual inflectional marking on the Latin adjective *immoderato* is a case of dependent marking since the adjective is dependent on the noun, the head of the noun phrase *immoderato amore*. There are also languages that mark the head rather than the dependent (Nichols, 1986). For instance, Hungarian exhibits head marking, as in:

- (12) az ember ház-a  
 the man house-3SG  
 ‘the man’s house’

Where the noun *ház* ‘house’ is the head.

The functional distinction between inherent and contextual inflection made above can be used for predicting the order in which the relevant inflectional elements occur in complex words: contextual inflection tends to be peripheral to inherent inflection. For instance, the ablative singular form of the Finnish word for ‘cat’ is *kisso-i-lta*, with the ablative suffix *-lta* ordered after the plural suffix *-i*. Inflection in its turn is peripheral to word formation. When inflectional systems erode, it is usually the contextual inflection that gets lost first. For instance, most Romance languages have lost their case system while preserving the morphological expression of number on nouns, a case of inherent inflection.

The existence of a rich inflectional system implies that a lexeme may have quite a number of forms. Unlike what is the case for lexeme formation, it therefore is not very realistic to assume that all inflected forms of the established lexemes of a language are stored in the lexical memory of speakers, certainly not for languages with rich inflection.

Storage of inflected forms is most probable for irregular forms, and for regular forms with a certain frequency of occurrence (Booij, 1999).

## 5. Interfaces

### 5.1. The interface with phonology

As we saw above, the basic levels of morphological analysis are the phonological level, the level of morphosyntactic structure, and the semantic level. The interface between these three types of information is subject to certain general principles.

As to the interface between phonology and morphology, the morphological structure of a complex word co-determines the phonological form of the complex word, in particular its prosodic structure. In the default case, a word in the morphological sense corresponds to a word in the phonological sense, referred to as the phonological word or the prosodic word. For instance, the English word *baker* is one prosodic word. The prosodic word is the domain of syllabification (that is, the division of the phonological string of a word into syllables). The syllabification of the word *baker* is as follows:

(13) (be:.kər)<sub>ω</sub>

(the dot indicates a syllable boundary, the  $\omega$  stands for ‘prosodic word’). This shows that morphological structure and prosodic structure are not necessarily isomorphic: there is no prosodic boundary that corresponds to the word-internal morphological boundary of this word. Since the suffix *-er* forms one prosodic word with the stem, we call it a cohering affix. Affixes may also form a prosodic word of their own, however, and are then qualified as non-cohering. Prefixes in Germanic languages are often non-cohering, and a number of suffixes as well. For instances, in careful speech the English word *un-able* has a syllable boundary before the first vowel of *able* even though normally a consonant belongs to the same syllable as the next vowel. Hence, *un-* is a non-cohering prefix (Booij and Rubach, 1984), and forms a prosodic word of its own. Therefore, *unable* is a prosodic compound. Suffixes may also be non-cohering. An example is the Dutch suffix *-achtig* ‘-like’, as in *rood-achtig* ‘red-like, reddish’. The prosodic structure of this word is (ro:d)<sub>ω</sub> (ɑx.təγ)<sub>ω</sub>. Consequently, the final /d/ of the first constituent *rood* with the underlying phonological form /ro:d/ is devoiced since Dutch obstruents are voiceless at the end of a syllable, and the phonetic form of this word is [ro:t.ɑx.təx]. Thus, the phonetic form of this adjective contrasts with that of the synonymous adjective *rod-ig* [ro:dəx] with the cohering suffix *-ig* /əγ/ in which the voicedness of the underlying /d/ is preserved because it does not occur in syllable-final position. In many languages the lexeme constituents of compounds form separate prosodic words, with the effect that the compound-internal morphological boundary coincides with a syllable boundary. Thus we get audible minimal pairs of the following type (examples from Dutch):

(14)	loods-pet ‘pilot cap’ [lo:ts.pət]	lood-spēt ‘lead drop’ [lo:t.spēt]
	balk-anker ‘beam brace’ [balk.ɑŋ.kər]	bal-kanker ‘testicle cancer’ [bal.kɑŋ.kər]

A second example of the influence of morphological structure on the phonetic realisation of complex words is that of word stress, one of the most intensively studied aspects of phonology as far as English is concerned. Some English suffixes are stress-neutral, that is, they do not change the stress pattern of their stem, whereas other are stress-shifting, and shift the main stress of the stem rightward. The English suffix *-er*, for instance, is stress-neutral, unlike the suffix *-ee*, as can be seen in a pair like *emplóy-er* versus *employ-ée* (cf. the classical study of (Chomsky and Halle, 1968), and for a more recent discussion (Hammond, 1999)). The interaction between phonology and morphology is the main focus of the theory of Lexical Phonology (cf. (Booij, 2000) for a survey), and also an important domain of research in Optimality Theory (Kager, 1999).

## 5.2. The interface with semantics

The basic principle that governs the interface between the formal structure of complex words and their meaning is the principle of compositionality: the meaning of a complex word is a compositional function of the meaning of its constituents and the morphological structure. For instance, the meaning of *football* is a compositional function of the meanings of *foot* and *ball*, and the meaning contribution of the compound structure that can be circumscribed as follows for English compounds, which are right-headed:

(15) [X Y]<sub>Y</sub> ‘Y that has some relation R to X’

Hence, a football is a ball which has something to do with feet. The exact relation between *foot* and *ball* does not belong to the domain of linguistic knowledge, but is part of our knowledge of the world (of games). We know that the foot is the device used for hitting the ball. This makes clear that the specific semantic interpretation of individual words is underdetermined by the linguistic system as such.

A second form of interface in which semantics is involved pertains to argument structure. The creation of complex verbs may have predictable consequences for the syntactic valency of these verbs. For instance, when we create causative verbs from adjectives, with the meaning ‘to cause something to become A’, the event expressed by the causative verb presupposes an Agent and a Patient. Hence, such causative verbs will have at least two arguments, and hence be transitive. Thus, the semantics of a class of complex words may have predictable consequences for their syntactic valency. Such rules that predict the relationship between argument structure and syntactic valency are referred to as ‘linking rules’.

The principle of compositionality takes a syntagmatic perspective on the semantics of complex words. There are clear cases of word formation, however, where we need a paradigmatic perspective in order to account for the meaning of a complex word. The meaning of the Dutch compound *huisman* ‘househusband’, for instance, can only be understood when seen as part of the following equation:

(16) vrouw ‘woman’ : man ‘man’ = huisvrouw ‘housewife’ : huisman ‘househusband’

A *huisvrouw* is a woman without a paid job who stays at home to take care of the household, a *huisman* is the male counterpart of such a woman.

The idea that paradigmatic relationships play a role in the semantics of complex words can also be seen in the interpretation of Dutch words ending in the noun *boer* ‘farmer’. This word functions as the head of a compound like *groenteboer* ‘green-grocer’, reflecting a time when the farmer was both the grower and the retailer of vegetables. In present-day Dutch, the constituent *boer* has developed the more general meaning retailer, witness compounds like *sigarenboer* ‘cigar seller, tobacconist’ and *tijdschriftenboer* ‘magazines salesman’. Thus, Dutch developed a particular subpattern of NN compounds of the following type:

(17)  $[[x]_N [boer]_N]_N$  ‘seller of x’

which may be qualified as a ‘constructional idiom’ (Jackendoff 2002) at the word level: a construction of which one position is lexically filled, and another one still a variable.

The same observation can be made for the left constituents of compounds. For instance, in most Dutch compounds that begin with the word *hoofd* ‘head’, the meaning of that constituent is ‘main’, as in *hoofd-ingang* ‘main entrance’ and *hoofd-bezwaar* ‘main objection’. In Maale, an Omotic language spoken in southern Ethiopia, the noun *nayi* ‘child’ has developed into a word with the general meaning of agent, as in *bayi nayi* ‘lit. cattle child, one who brings cattle to grazing area’. This reflects the fact that cattle herding is typically a children’s task in that society (Amha, 2001), p. 78. Thus, such compound constituents may develop into affix-like elements (referred to as affixoids).

This implies again that we conceive of the lexicon as a hierarchy of levels: at the bottom the individual coined complex words, at the top the abstract schemas according to which these words have been formed, and intermediate generalizations like (17). In the case of NN compounds, we thus get (at least) three levels:

(18)  $[[x]_N [y]_N]_N$  ‘y with some relation R to x’  
 $\quad \quad \quad |$   
 $[[x]_N [boer]_N]_N$  ‘seller of N’  
 $\quad \quad \quad |$   
 $[[sigaren]_N [boer]_N]_N$  ‘seller of cigars’

At each level a construct(ion) instantiates the constructional schema by which it is dominated.

### 5.3. The relation between morphology and syntax

Our approach so far can be qualified as ‘lexicalism’. This term denotes the set of theories in which morphology is separated from syntax in the sense that the structure of complex words is not dealt with by the syntax, but by lexical rules that express generalizations about established and potential complex words. This does not mean that syntax and morphology do not interact, but that syntactic rules cannot manipulate parts of words. This principle is referred to as the principle of Lexical Integrity:

(19) *Principle of Lexical Integrity*

“The syntax neither manipulates nor has access to the internal structure of words”  
 (Anderson, 1992) p. 84)

As argued in (Booij, 2008), this principle is too strong, since one should not exclude the possibility that syntactic rules may have access to the internal morphological structure even though manipulation must be excluded. An example that shows that access of syntactic or semantic rules to word-internal structure cannot be completely ruled out comes from Georgian. In Georgian we find expressions such as the following (Harris, 2006):

- (20) *sam*                    *tit-moč’r-il-i*                    (*k’aci*)  
 Three.OBL      finger-cut.off-PTCP-NOM      man.NOM  
 ‘(a man) with three fingers cut off’

The first word *sam* has to appear in the oblique form, because it modifies the word *tit* ‘finger’ which is part of the second word. That is, both for the purpose of case assignment (to the independent word *sam* only) and semantic interpretation, *sam* and *tit* form a unit. As Harris argues, the word *sam* cannot be considered a part of the next word even though its form is indeterminate since it could also be a stem form, because recursive modification is not allowed within Georgian compounds. Hence, it should be interpreted as the oblique form of an independent word. This case assignment thus requires access to the internal morphological structure of the second word in (20). The construction in (20) may be compared to that in (21) where the first word bears nominative case, and you get a different interpretation:

- (21) *sam-i*                    *tit-moč’r-il-i*  
 Three-NOM      finger-cut.off-PTCP-NOM  
 ‘three (men, people, statues) with fingers cut off’

In (21), the word form *sami* agrees in case marking with the second word as a whole, and hence it is a modifier of the whole word. Note, however, that the word *tit*, being part of a compound, does not receive case marking itself.

The need to access the internal structure of complex words is also shown by scope phenomena: in some cases a modifier may have scope over a sub-constituent of a complex word. This is illustrated in (22) with Dutch phrases in which the pre-nominal adjective modifies only the first noun constituent of the NN compounds. I use Dutch examples here even though the English glosses have the same properties. For Dutch we can be certain that these linguistic units are phrases because the adjectives are inflected witness the inflectional ending *-e*:

- (22) [A [NN]<sub>N</sub>]<sub>NP</sub>  
*visuel-e*    *informatie-verwerking*  
 visual-NONNEUTER.SG.INDEF    information-processing  
 ‘visual information processing’  
  
*intellectuel-e*                                    *eigendoms-rechten*

intellectual-NEUTER.PL.INDEF property-rights  
'intellectual property rights'

The principle of Lexical Integrity can be used to determine the status of lexical units such as preverb + verb combinations that are found in several Indo-European languages, and also in Uralic languages like Hungarian (Kiefer and Honti, 2003). The latter language has lexical units such as *tévét nez* 'be engaged in television watching'. The two parts of this lexical unit can be split in certain syntactic contexts, for instance by the negative word *nem* 'not'. The splittability of these units is evidence for their not being words. This is confirmed by the fact that the noun constituent *tévét* in this example is marked with accusative case by the suffix *-t*. This assignment of structural case to *tévé* shows that it must be an independent word. Given the principle of Lexical Integrity, one does not expect structural case assignment to a sub-constituent of a word. Thus, we can make a principled distinction between morphological and syntactic constructs.

This does not imply that syntactic constructs cannot form part of words. In the following Dutch examples, AN phrases are used in the non-head position of complex words. These complex words as such are morphological constructs, but one of their constituents is formed in accordance with the rules of syntax:

- (23) [[[oude]<sub>A</sub> [mannen]<sub>N</sub>]<sub>NP</sub> [huis]<sub>N</sub>]<sub>N</sub> 'old men's home'  
[[[vierde]<sub>A</sub> [klas]<sub>N</sub>]<sub>NP</sub> -er]<sub>N</sub> 'fourth class pupil'

These cases of word formation do not invalidate lexicalism, since the distinction between syntactic and morphological constructs is maintained.

## 6. Morphological classification

Languages may be classified according to the role and nature of morphology in each language (Comrie, 1981; Haspelmath, 2008). A first dimension of classification 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 stems, leading to relatively complex words, as illustrated by the following one-word-sentence of Central Alaskan Yup'ik (Mithun, 2000) p. 923):

- (24) Tuntutuq=gguq  
tuntu-te-u-q=gguq  
moose-catch-INDIC.INTRANSITIVE=HEARSAY  
'He got a moose'

The third dimension of classification is the index of fusion. In fusional languages, one morpheme may express more than one grammatical feature. Above, we saw that Latin is such a language. Such languages can be contrasted with 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, unlike what is the case for Latin:

- (25) çocuk-lar-in  
child-PL-GEN  
'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 (Haspelmath, 2008). For instance, many Indo-European languages are fusional in their inflectional system, but agglutinating in their derivational morphology. Chinese also illustrates this point since, as mentioned above, it is synthetic as far as word formation is concerned, but analytic as far as inflection is concerned.

## 7. Affix ordering

In languages with a reasonably rich morphology affix ordering is an important topic for morphological analysis. The basic question is how we can account for the ordering in which the different types of morphemes have to appear in multiply complex words.

A well known general principle (Greenberg, 1963) is that inflectional morphemes are peripheral to derivational morphemes. Within the domain of inflection, contextual inflection appears to be peripheral to inherent inflection (Booij 1993, 1996).

The ordering of inflectional affixes has also been investigated in detail by Bybee (Bybee, 1985). Bybee proposed a semantic Relevance Hierarchy for the order of affixes: the more semantically relevant an affix is for the stem, the closer it is to the stem. Hence, derivational morphemes, which have obviously a profound effect on the meaning of the stem are closer to the stem than inflectional ones.

As to the inflectional affixes on nouns, it is predicted that case markers and (in)definiteness markers on nouns will be peripheral to gender markers since they do not have a semantic effect on the stem of the noun. Instead, they relate the noun to its syntactic context.

In the case of verbs, the following hierarchy can be observed for languages with markers for voice, aspect, tense, and agreement:

- (26) Voice > Aspect > Tense > Agreement

Voice markers such as Passive have a strong semantic effect, and affect the argument structure of the verb. At the other end of the hierarchy, tense has the deictic role of relating the event expressed by the verb to the moment of speaking. Agreement, a case of contextual inflection, also has an external role in that it relates the verb to its syntactic context.

Another example is the ordering of suffixes in the following Maale verb (Amha, 2001), p. 114:

- (27) *gap-is-é-ne*  
finish-CAUSATIVE-PERFECT-DECLARATIVE  
'finished'

The causative suffix affects the argument structure and hence the syntactic valency of the root *gap*. Hence, it is closest to the root. The declarative suffix on the other hand expresses a property of the whole sentence since it indicates that the sentence is declarative. That is, it does not modify the semantic content of the verbal root as such, and is therefore the most peripheral suffix.

For some languages with complicated sequences of morphemes in words one finds descriptions with templates. A template specifies a number of slots for specific morphemes. This kind of affix ordering is referred to as 'position class morphology' (cf. (Inkelas, 1993) for a discussion of such template morphology in Nimboran, a Papuan language of New Guinea). For Athapaskan (Amerindian) there is a detailed study that relates the order of affixes to their semantic scope properties (Rice, 2000).

The order of affixes in a complex word may also reflect the different historical strata of the vocabulary of a language. The Dutch vocabulary, for instance, has a non-native (Romance) stratum, besides a native (Germanic) stratum. The basic generalization is that non-native suffixes can attach to non-native (simplex or complex) stems only, whereas native suffixes can attach to both non-native and native stems (Booij, 2002b). Hence, the predicted order is that native suffixes will be peripheral to non-native ones, as illustrated in (28):

- (28) *real-iser-ing* 'realisation'  
*controvers-ieel-heid* 'controversialness'

The suffixes *-iseer* and *-ieel* are non-native suffixes borrowed from French, whereas the suffixes *-ing* and *-heid* are native suffixes of Germanic origin. The non-native suffixes first attach to the (non-native) roots. They cannot be added after a native suffix, since the attachment of a native suffix makes the stem native.

As to English, there is a long debate on how to deal with the constraints on suffix ordering (Hay and Plag, 2004; Plag, 1996). This is related to how complex words are processed, discussed in section 10 below. The idea is that affixes that are easily and often recognized as parts of complex words tends to be peripheral to affixes that form part of complex words whose internal structure is not so easily parsed. For instance, the English suffix *-less* is easily parsed out, and attaches freely to all kinds of complex words, whereas *-ity* that occurs in less parsable words has a more restricted distribution. Hence, a word like *\*home-less-ity* is odd though it is semantically well-formed (Hay, 2002).

Prosodic properties may also play a role in stacking up affixes. Dutch suffixes that are non-cohering and thus form prosodic words of their own, can easily be attached to already suffixed words - sometimes even to plural forms of nouns - unlike cohering suffixes. For instance, the productive cohering suffix *-ig /əɣ/* '-ish' cannot attach to adjectives that have a participial form whereas the non-cohering suffix *-heid* '-ness'

freely attaches to such participial adjectives. Hence, the contrast in wellformedness between \**woed-end-ig* ‘slightly furious’ and *woed-end-heid* ‘furiousness’. That is, words can be made longer if the suffix starts a new prosodic word (Booij, 2002a).

## 8. Diachrony

The use of productive word formation patterns is not the only source of complex words in the lexicon. Complex words may also arise through univerbation, the process in which phrases become words. Many nominal compounds in Germanic languages have a phrasal origin. For instance, the Dutch compound *koningskroon* ‘king’s crown’ originated as a phrase in which the noun *konink* was marked as the possessor through the genitive case ending *-s*. The case ending that was thus trapped inside a word was then reinterpreted as a semantically empty linking element or stem extension. The system of linking elements became subsequently part of the compounding system of Dutch.

Word formation processes have the function of expanding the sets of words of lexical categories (nouns, verbs, adjectives, and adverbs). Yet, we also find complex words of non-lexical categories such as prepositions. This is due to the process of grammaticalization, defined as follows by (Hopper and Traugott, 2003), p. xv) “We define grammaticalization as the process whereby lexical items come in certain linguistic contexts to serve grammatical functions, and, once grammaticalized, continue to develop new grammatical functions”. The English preposition *during*, for instance, has the shape of a present participle of the verbal root *dure* that we also find in *en-dure* and *dur-ation*. The participle could be reinterpreted as a preposition in absolute participial constructions like *during the war* ‘while the war lasted’. Thus, the class of English prepositions (prepositions are grammatical morphemes) was expanded with a complex word *during*. In the word *notwithstanding* we see a combination of univerbation and grammaticalization.

Grammaticalization can lead to the rise of word formation processes, since lexemes can become bound morphemes, prefixes or suffixes, which belong to the class of grammatical morphemes. In French, for instance, the French preposition *sur* has a prefixal counterpart *sur-* with the meaning ‘over’, as in *surexposition* ‘overexposure’ (Amiot, 2005). The English suffix *-wise* as used in *money-wise* ‘as far as money is concerned’ originates from the noun *wise* ‘manner’. Thus, univerbation and grammaticalization are mechanisms of language change that lead to the expansion of non-lexical categories, and to the rise of new derivational processes (Booij, 2005a; Heine and König, 2005).

Complex words can also be subject to the process of lexicalization, and thus lose their morphological transparency. The Dutch word *aardappel* ‘potato’, for instance, is historically a compound, consisting of the stems *aard* ‘earth’ and *appel* ‘apple’. Yet, it is no longer perceived as a kind of apple, and it is syllabified as a simplex word, without the word-internal morphological boundary coinciding with a syllable boundary: *aar.dap.pel*, not *aard.ap.pel*.

Language contact is another source of word formation processes. Dutch, English and German have borrowed many complex words from French in the course of time, for instance deadjectival nouns ending in *-ité*, with some phonological adaptation

(Dutch *-iteit*, English *-ity*, German *-ität*). Speakers of these languages could abstract a word formation schema on the basis of a number of such loans and the corresponding adjectives, and use these productively. Thus, Dutch now has a number of nouns in *-iteit* for which there is no French counterpart since they have been coined in Dutch on the basis of a Dutch adjective, such as *stomm-iteit* ‘stupidity’ derived from *stom* ‘stupid’. Similarly, the English deverbal suffix *-ee* derives from the French passive participle ending *-é*, but has now gone its own ways, and combines with English verbs, as in *standee* derived from the verb *stand*.

Word formation processes can also be lost. In the course of time a word formation process can lose its productivity, with the effect that no more words of that type are coined. This means that we might still find a number of instantiations of the relevant word formation schema, but no extension of that class. The Dutch suffix *-lijk* for instance, which is similar in function to the English suffix *-able* in coining deverbal adjectives, lost its productivity, and the role of coining new deverbal adjectives is taken over by the productive suffix *-baar* ‘-able’.

As mentioned above, inflectional systems may be subject to strong erosion. Most Romance and Germanic languages have lost their case system for nouns and agreeing adjectives. Verbal inflection may also change considerably. Afrikaans, a daughter language of Dutch spoken in South-Africa lost most of its verbal morphology, with the effect that there is only one verb form left. Past tense in this language is expressed periphrastically, as illustrated in (29):

- (29) *Ek het geloop*  
I have walked  
‘I walked’

This kind of inflectional erosion is due to the effect of contact between speakers of Dutch and speakers of African languages and Malay in South-Africa. The simplification of the inflection of nouns in English (loss of case and gender marking) may also be due to the effect of language contact between Anglo-Saxon speakers and Viking invaders.

Language change may lead to the rise of constructions with specific morphology. Consider the following English phrases, and their labeling (Rosenbach, 2006; Rosenbach, 2007):

- (30) Determiner genitive: John’s book, the young girl’s eyes  
Descriptive (classifying) genitive: a men’s suit, the lawyer’s fees

The morpheme *-s* in these constructions derives historically from a genitive case ending. After the loss of the case system, this use of *-s* for the marking of specifiers in certain types of noun phrases persisted. Hence, this use of the *-s* is a case of construction-dependent morphology (Booij, 2005b). Another example is the use of the old dative suffix *-en* in Dutch collective constructions such as

- (31) *met zijn vier-en*  
with his four-*en*  
‘the four of us’

na en-en  
after one-*en*  
'after one o'clock'

This use of the suffix *-en* does not follow from synchronic case-marking, but is the residue of case-marking in a earlier stage of Dutch. Morphological elements can be 'trapped' in a construction, and thus become dependent on that construction.

As mentioned above, individual complex words may lose their internal morphological structure in the course of time, a form of lexicalization. The Dutch denominal adjective *natuur-lijk* 'natural' has acquired the additional meaning 'of course' when used as an adverb, just like English *naturally*. The word has become opaque, and speakers may no longer feel a relationship with the base noun *natuur* 'nature'. This independence of the highly frequent adverb *natuurlijk* /na:ty:rlək/ manifests itself in the fact that its phonetic form is often reduced to forms like [ty:rlək] or even [ty:k]. This contrasts with the use of this word as an adjective with the regular meaning 'natural'. Used with that meaning, the word's phonetic form cannot be reduced, and it must be pronounced as [na:ty:rlək].

Words may also lose their morphological transparency because the base word was lost in the course of time. In the English verb *begin*, for instance, we might still recognize a prefix *be-*, but the root *gin* is no longer used as a verb. Hence, we might conclude that this verb has become a simplex one. In some cases there is reason to speak of words that are only formally complex. Such formal complexity may be the result of the loss of the base word, or of borrowing, as happened in English through the influx of Latinate verbs. English verbs in *-ceive* such as *conceive*, *perceive* and *receive* share the property that their corresponding nominal ends in *-ception*, and their corresponding adjective in *-tive*. For that reason, we would like to assume that these verbs consist of two constituents, a prefix and a root. Yet, there is no systematic form-meaning correspondence involved in such sets of similar words.

## 9. Processing complex words

The processing of complex words is an important domain of psycholinguistic research. The main debate concerns the balance between computation and storage of complex word forms. One position, which has been most eloquently defended by Steve Pinker (Pinker, 1999) is that complex words that are irregular are stored in the lexicon, but that regular forms are computed on-line when the utterance is being processed. For instance, the inflectional forms of English regular verbs are assumed not to be stored, unlike those of the irregular verbs. The different patterns of irregular verb forms may however be recognized and stored in a kind of associative memory.

This view of the balance between storage and computation has been challenged by many psycholinguistic research results. It appears that fully regular forms may induce frequency effects. If a word form has a relatively high frequency of use, this will speed up the processing of that word in processing tasks such as lexical decision tasks. In such tasks subjects have to decide whether a letter sequence shown on a screen is a correct word of the language. The idea behind the frequency effect is that the frequent use of a

word heightens its resting activation level (or ‘lexical strength’), and thus the word is faster recognized. A frequency effect is only possible if the relevant words or word forms are stored, and hence can be subject to frequency effects. Thus, we can conclude that regular word forms may be stored.

These findings can be modeled in a dual route model with competition. When the meaning of a complex word has to be understood, there are two routes available: the complex word is either retrieved directly from lexical memory, or first decomposed into its morphological constituents, on the basis of which the meaning is then computed. Since both routes are available, they will compete. If a complex word is an established one, with a high frequency, and thus a high resting activation level, the direct route will win. If the word to be understood is a completely new one that is not stored, or has a low frequency of use, the decompositional route will be the fastest (Baayen et al., 1997)

At a more fundamental level, the issue is whether we are justified to make a distinction between symbolic rules or schemas on the one hand, and representations on the other. The distinction between rule and representation is denied in connectionist approaches to language structure. These issues are too complex for being discussed in this chapter, however (cf. (Bybee and McClelland, 2005) for a recent general discussion of the issues involved).

Other important results of psycholinguistic research are its findings concerning the structure of the lexicon. It is clear that the lexicon is not a list, but rather a network of relationships between words, relations along different dimensions such as phonology, semantics, and morphological structure. The lexicon is a web of words. This means that the paradigmatic relationships between words (either simplex or complex) are essential for understanding how morphology works.

An example of the role of paradigmatic structure is the family size effect (Baayen and Schreuder, 2003): a simplex word is more easily processed the larger the number of words derived from it (its morphological family) is. This effect presupposes that a word is linked to its derivatives in the lexicon. Another paradigmatic effect is that the choice of a linking element in Dutch compounds can be very well predicted on the basis of the constituent family of the left and the right constituent. For instance, subjects tend to choose the linking element *-en* for a new compound that begins with the constituent *rat* ‘rat’ since *-en* is the preferred linking element in established compounds that begin with *rat*.

The processing of complex words can also be investigated through naturalistic data such as slips of the tongue, which give a clue as to how complex words may be represented in the mental lexicon. When a complex word is stored, it might be stored including its internal morphological structure. In the following slips of the tongue, two morphemes have been exchanged, which suggests that the internal structure of such words is indeed available (Cohen, 1987):

- (32) Ik heb vrijdag maan < Ik heb maan-dag vrij  
I have Friday moon I have Mon-day free  
‘I am free on Monday
- een vloeibare drinkstof < een drink-bare vloeistof  
a liquid drink-stuff a drink-able flow-stuff

‘a drink-able liquid’

Thus, speech errors can provide linguistic evidence for the way in which complex words are stored in lexical memory.

## 10. Morphological productivity

A much discussed notion in morphological research is the notion of productivity. Morphological processes differ in the extent to which they are used for coining new words or word forms. The classical example from the domain of inflection is the difference between the regular formation of past tense forms of English, which is productive and applies to newly coined verbs, and the class of irregular verbs, which form their past tense by means of vowel change and occasionally consonantal changes as well, a case of internal modification (*sing-sang, bring-brought, etc.*). In the domain of word formation, one may observe that new deadjectival nouns ending in *-th* are hardly ever formed by speakers of English, whereas new deadjectival nouns ending in *-ness* can be made readily: *coolness* is more readily coined than *coolth*.

Quantitative measures of productivity make use of the notion ‘type frequency’. A productive process will result in a large number of different types. For instance, in English the number of types of past tense forms of regular verbs is much higher than the number of types of irregular past tense forms, and the number of types of nouns in *-ness* is much higher than that in *-th*. Baayen makes use of the notion ‘hapax’ in measuring quantitative productivity: the number of types of a particular type of complex word that occur only once in a given corpus (that is, the number of hapaxes), is a good measure of the productivity of the morphological process involved (Baayen, 1992).

One approach to this phenomenon of differences in productivity is to consider it as a property of the language system: some morphological processes are unproductive (hence, the relevant set of words cannot be expanded), whereas other processes are productive, and may lead to new words or word forms. The actual, quantitative productivity of a productive process is then determined by two types of factors: the number of potential bases to which the process can apply (system-internal factors), and pragmatic, system-external factors such as the need for a particular word. This is the position taken in (Bauer, 2001). The number of potential bases depends on the number of linguistic constraints on the morphological process involved: the more constraints there are, the less chances the process can apply and create new forms.

However, Baayen (Baayen, 1992; Baayen, 2008) and Bybee (Bybee, 1995) have argued that productivity is inherently a gradual notion. Even in the case of processes with a very low productivity, the relevant class of words can occasionally be extended. For instance, one may come across the word *coolth* in language corpora.

As to the non-systemic factors, the productivity of word formation processes may depend on factors such as written versus spoken language, specific registers, and speech communities. Certain types of word formation are used productively in written language only. The suffix *-ity* is typically used productively in scientific and technical discourse (Baayen 2008).

Processing factors also play a role in productivity. The output of a word formation process is morphologically more transparent and will be readily decomposed in

processing if the frequency of the derived word is lower than that of the base word. Decomposition of complex words will strengthen the accessibility of the corresponding morphological schema, and thus increase productivity (Hay and Baayen, 2002).

## 11. Tools for morphological research

A primary source of information about morphology is formed by the descriptive grammars of individual languages which usually give a description of inflection and word formation. The availability of such grammars is essential for typological research. An excellent manual for morphological description is (Payne, 1997).

Typological databases are mainly based on such descriptive grammars. Among others, the following databases on morphological typology can be found on the internet:

Surrey Morphology Group: <http://www.surrey.ac.uk/LIS/SMG>

Universals Archive of the University of Konstanz: <http://typo.uni-konstanz.de/archive/intro>

The Morbo database on compounding, University of Bologna: <http://morbo.lingue.unibo.it>

It is important that in morphological analysis, linguists use the same conventions, and the same glossing rules. The Leipzig glossing rules are used as a standard these days. They can be found on <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>. Tools for language description are available on [www.eva.mpg.de/lingua](http://www.eva.mpg.de/lingua). Other tools including the Ethnologue survey of languages can be found on [www.sil.org](http://www.sil.org).

Sources for the outputs of word formation are the dictionaries of individual languages. Searching for morphological information has been made much easier thanks to electronic dictionaries, many of them on line. However, the role of dictionaries in morphological research is now strongly being reduced in favour of corpora of language use. Corpora do not suffer from the restrictions of dictionaries that the data are filtered by the lexicographer, and are always lagging behind as to what happens in actual language use. In fact, present-day good dictionaries are based on corpora as well. Moreover, corpora provide the possibility to investigate how the productive use of morphological processes correlates with factors of language use and properties of language users. Therefore, corpus-based linguistic research has become indispensable for adequate morphological research (Baayen 2008).

## References

- AMHA, AZEB. 2001. The Maale language. Leiden: University of Leiden, Research School of Asian, African, and Amerindian Studies.
- AMIOT, DANY. 2005. Between compounding and derivation: elements of word formation corresponding to prepositions. Morphology and its demarcations, ed. by

- Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer and Franz Rainer, 81-96. Amsterdam / Philadelphia: Benjamins.
- ANDERSON, STEPHEN. 1992. A-morphous morphology. Cambridge UK: Cambridge University Press.
- ARONOFF, MARK. 1994. Morphology by itself: Stems and inflectional classes. Cambridge Mass.: MIT Press.
- BAAYEN, R. HARALD. 1992. Quantitative aspects of morphological productivity. Yearbook of Morphology 1991, ed. by Geert Booij and Jaap van Marle. Dordrecht: Kluwer.
- . 2008. Corpus linguistics in morphology: morphological productivity. Handbook of corpus linguistics, ed. by Anke Lüdeling, M. Kytö and T. McEnery. Berlin: De Gruyter.
- BAAYEN, R. HARALD and SCHREUDER, ROBERT (eds.) 2003. Morphological structure in language processing. Berlin: Mouton de Gruyter.
- BAAYEN, R. HARALD, DIJKSTRA, TON and SCHREUDER, ROBERT. 1997. Singulars and plurals in Dutch. Evidence for a parallel dual route model. Journal of Memory and Language, 36.94-117.
- BAUER, LAURIE. 2001. Morphological productivity. Cambridge: Cambridge University Press.
- BOOIJ, GEERT. 1993. Against split morphology. Yearbook of Morphology 1993, ed. by Geert Booij and Jaap van Marle, 27-49. Dordrecht / Boston: Kluwer.
- . 1996. Inherent versus contextual inflection and the split morphology hypothesis. Yearbook of Morphology 1995, ed. by Geert Booij and Jaap van Marle, 1-16. Dordrecht / Boston: Kluwer.
- . 1999. Lexical storage and regular processes. Brain and Behavioural Sciences, 22.1016.
- . 2000. The phonology-morphology interface. The first Glot International state-of-the-article book. The latest in linguistics, ed. by Lisa Cheng and Rint Sybesma, 287-306. Berlin: Mouton de Gruyter.
- . 2002a. Prosodic restrictions on affixation in Dutch. Yearbook of Morphology 2001, ed. by Geert Booij and Jaap van Marle, 183-202. Dordrecht: Kluwer.
- . 2002b. The morphology of Dutch. Oxford: Oxford University Press.
- . 2005a. Compounding and derivation: evidence for construction morphology. Morphology and its demarcations, ed. by Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer and Franz Rainer, 109-32. Amsterdam / Philadelphia: John Benjamins.
- . 2005b. Construction-dependent morphology. Lingue e Linguaggio, 4.31-46.
- . 2007. The grammar of words. An introduction to morphology. 2nd edition: Oxford textbooks in linguistics. Oxford: Oxford University Press.
- . 2008. Lexical integrity as a morphological universal, a constructionist view. Universals of language today, ed. by Elisabetta Magni and Sergio Scalise. Dordrecht: Springer.
- BOOIJ, GEERT and RUBACH, JERZY. 1984. Morphological and prosodic domains in Lexical Phonology. Phonology Yearbook, 1.1-27.
- BYBEE, JOAN. 1985. Morphology. A study of the relation between meaning and form. Amsterdam / Philadelphia: Benjamins.

- . 1988. Morphology as lexical organization. *Theoretical morphology*, ed. by Michael Hammond and Michael Noonan, 119-41. San Diego: Academic Press.
- . 1995. Regular morphology and the lexicon. *Language and Cognitive Processes*, 10.425-55.
- BYBEE, JOAN and MCCLELLAND, JAMES. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22.381-410.
- CHOMSKY, NOAM and HALLE, MORRIS. 1968. *The sound pattern of English*. New York: Harper and Row.
- COHEN, ANTHONY. 1987. Morfologie op heterdaad betrapt. *Forum der Letteren*, 28.90-97.
- COMRIE, BERNARD. 1981. *Language universals and linguistic typology*. Oxford UK: Blackwell.
- GREENBERG, JOSEPH. 1963. Some universals of grammar, with particular reference to the order of meaningful elements. *Universals of language*, ed. by Joseph Greenberg, 73-113. Cambridge Mass.: MIT Press.
- HAMMOND, MICHAEL. 1999. *The phonology of English*. Oxford: Oxford University Press.
- HARRIS, ALICE C. 2006. In other words: external modifiers in Georgian. *Morphology*, 16.205-29.
- HASPELMATH, MARTIN. 2008. The agglutination hypothesis. *Universals of language today*, ed. by Elisabetta Magni and Sergio Scalise. Dordrecht: Springer.
- HAY, JENNIFER. 2002. From speech perception to morphology: affix ordering revisited. *Language*, 78.527-55.
- HAY, JENNIFER and BAAYEN, R. HARALD. 2002. Parsing and productivity. *Yearbook of Morphology 2001*, ed. by Geert Booij and Jaap van Marle, 203-36. Dordrecht: Springer.
- HAY, JENNIFER and PLAG, INGO. 2004. What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory*.565-96.
- HEINE, BERND and KÖNIG, CHRISTA. 2005. Grammatical hybrids: Between serialization, compounding and derivation in !Xun. *Morphology and its demarcations*, ed. by Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer and Franz Rainer, 81-96. Amsterdam / Philadelphia: Benjamins.
- HOPPER, PAUL J. and TRAUGOTT, ELISABETH C. . 2003. *Grammaticalization*. Cambridge U.K.: Cambridge University Press.
- HYMAN, LARRY. 2000. Suprasegmental units. *Morphology / Morphologie. A handbook on inflection and word formation / Ein Handbuch zur Flexion und Wortbildung. Volume 1.*, ed. by Geert Booij, Christian Lehmann and Joachim Mugdan. Berlin: De Gruyter.
- INKELAS, SHARON. 1993. Nimboran position class morphology. *Natural Language and Linguistic Theory*, 11.559-624.
- JACKENDOFF, RAY. 2002. *Foundations of language*. Oxford: Oxford University Press.
- JANSON, TORE. 2004. *A natural history of Latin*. Oxford: Oxford University Press.
- KAGER, RENÉ. 1999. *Optimality Theory*. Cambridge UK: Cambridge University Press.
- KIEFER, FERENC and HONTI, LÁSZLÓ. 2003. Verbal 'prefixation' in the Uralic languages. *Acta Linguistica Hungarica*, 50.137-53.

- LANGACKER, RONALD. 1987. *Foundations of Cognitive Grammar: Theoretical prerequisites*. Stanford, California: Stanford University Press.
- LEHMANN, CHRISTIAN. 2004. *Interlinear morphemic glossing. Morphology / Morphologie. A handbook on inflection and word formation / Ein Handbuch zur Flexion und Wortbildung. Volume 2* ed. by Geert Booij, Christian Lehmann and Joachim Mugdan. Berlin: De Gruyter.
- MITHUN, MARIANNE. 2000. *Incorporation. Morphologie / Morphology. Ein internationales Handbuch zur Flexion und Wortbildung / An international handbook on inflection and word formation. Vol. 1* ed. by Geert Booij, Joachim Mugdan and Christian Lehmann, 916-28. Berlin: De Gruyter.
- NICHOLS, JOANNA. 1986. *Head-marking and dependent-marking grammar. Language*, 62.56-119.
- PACKARD, JEROME. 2000. *The morphology of Chinese. A linguistic and cognitive approach*. Cambridge UK: Cambridge University Press.
- PAYNE, THOMAS E. 1997. *Describing morphosyntax. A guide for field linguists*. Cambridge UK: Cambridge University Press. [Reprinted in 1999, 2001 ].
- PINKER, STEVEN. 1999. *Words and rules*. New York: Basic Books.
- PLAG, INGO. 1996. *Selectional restrictions in English suffixation revisited. Linguistics*, 34.769-98.
- RICE, KEREN. 2000. *Morpheme order and semantic scope. word formation in the Atahapaskan verb*. Cambridge: Cambridge University Press.
- ROSENBACH, ANETTE. 2006. *Descriptive genitives in English: a case study on constructional gradience. English Language and Linguistics*, 10.77-118.
- . 2007. *Emerging variation: determiner genitives and noun-modifiers. English Language and Linguistics*, 11.143-89.