Chapter 7
Morphology: the structure of words

Geert Booij

Universiteit Leiden

Morphology deals with the systematic correspondence between the form and meaning of words. The study of these regularities comprises the domains of inflection and word formation. Inflection concerns the expression of morphosyntactic properties, sometimes required by a specific syntactic context. Word formation deals with the creation of new (complex) words by various morphological mechanisms such as compounding, affixation, truncation, and segmental and tonal alternations.

The role of morphology in the grammar of natural languages is subject to theoretical debate. First, there are various ideas about the format in which morphological regularities should be expressed (rules or schemas?) Second, there are various models of the position of morphology in the architecture of grammar which will be discussed. Important issues are the interface between morphology and phonology (how does morphological structure influence the pronunciation of complex words?), and the interface between morphology and syntax (demarcation of word versus phrase, lexical integrity, phrases as building blocks of words).

Morphology plays an important role in theories of the acquisition of language and in theories of language change. Hence, language acquisition and language change will also be discussed in this chapter.

Morphology is also very relevant for linguistic typology, which is partially morphology-based.. Therefore, this chapter will also discuss the morphological classification of languages (analytic, poly-synthetic, agglutinative languages, etc.).

Keywords: word structure, inflection, word formation, language change, language acquisition, language processing, linguistic typology

1. Introduction

Words are the basic building blocks of sentences.[1] Most words are a pairing of sound and meaning, and the meaning of a sentence is computed on the basis of the meanings of the

---

[1]   The text of this section and of section 2 is partially based on the text of section 1 of Booij (to appear).

constituent words, and the way in which they are combined. The relation between sound and meaning may be arbitrary. The meaning of the verb *sing*, for instance, cannot be read off its sound form, and the relation between sound and meaning in this word is therefore arbitrary. However, the relation between sound and meaning of a word may be (partially or completely) non-arbitrary, or motivated. This is the case for complex words, words with 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, usually defined as '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 classified as a lexical morpheme, as it can occur as a word of its own, whereas the morpheme *–er,* which serves to evoke the meaning 'agent of the action' when combined with verbs, is a bound morpheme of a particular type, a suffix, that is attached at the right edge of a base word.

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: $[V–er]_N$. 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 either, 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.

However, the distinction between simplex and complex word is not always that straightforward. In a word like *solips-ism*, we recognize a suffix *–ism*, even though there is no base word *solips* in English. The reason is that the suffix *–ism* is used systematically in English to coin, among others, nouns denoting philosophical theories such as *Marxism*.

Another example is the series of English verbs that contain the Latin root *–duce*, as in *deduce, induce, produce,* and *seduce*. There is no word *duce*, yet, these verbs are similar in that the presence of *–duce* implies that their nominalization always ends in *–duction*: *deduction, induction, production, seduction*. These latter roots, without a clear meaning, are referred to as submorphemic units. A similar observation can be made for Dutch verbs such as *begin* 'to begin' and *vergeet* 'to forget'. These verbs behave as if they begin with a prefix *be–* and *ver–* respectively, even though there are no verbs *gin* and *geet*. Dutch past participles normally have a prefix *ge–*, but this prefix has to be absent if the verbal stem begins with a prefix. Hence, the past particple of the verb *be-treur* 'to deplore' derived from the base verb *treur* 'be sad' is *betreurd*, not **gebetreurd*. Similarly, the past participles of *begin* and *vergeet* are *begonnen* and *vergeten*, without the prefix *ge–*.

One might also assume a certain internal complexity for words with phonaesthemes, (sequences of) sounds such as the initial consonants of *glass, gleam, glow, glare*, and *glimpse* that all express the idea of 'light, shine' (Marchand 1969: 411). It is not possible to divide a word like *glow* into two meaningful morphemes, *gl–* and *–ow*, yet language users are able to observe a certain systematicity here.

A second example of complex words are the following plural nouns in English: *apples, books, pages*, which all end in the plural morpheme *–s* (a morpheme with different phonetic realizations: [z], [s], [ɪz]). 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 of a word, and is usually represented with small capitals.

The two basic functions of morphological operations that create complex words are (i) word formation and (ii) inflection. Word formation processes create new words, 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]$_V$ er]$_N$ 'Agent of action V'

This schema states that nouns may consist of a verbal base of whatever phonological composition, followed by the suffix –*er*. The meaning correlation of this structure is a compositional function of the meaning of the base verb, and the agent interpretation is a property of this morphological configuration as a whole (the affix –*er* does not have a meaning by itself in isolation). The speaker may then use this schema to create new words with the appropriate 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* (www.urbandictionary.com) defines a *texter* not just as someone who texts, but as 'a person who prefers to send text messages instead of picking up the phone'.

Instead of agent nouns, deverbal nouns in –*er* might also denote instruments, as in *cooker*, *fastener*, and *vacuum-cleaner*., and other meanings such as location (as in *diner,* an American type of restaurant (Booij and Lieber 2004)). This is a case of polysemy of word formation patterns, a pervasive phenomenon in the languages of the world. It has to do with semantic extension mechanisms: language users can easily jump from one semantic concept to a related one. In this case, instruments can be seen as a kind of impersonal agents (Jurafsky 1996).

Inflection is the subsystem of the grammar of natural languages that deals with the proper form of words, 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 which kind of information the speaker wants to convey, and this is called inherent inflection. The choice of a particular present or past form, on the other hand, depends on the number and person of the subject of the clause in which the verb occurs, and is therefore a case of contextual inflection (Booij 1996):

(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 word, and this more abstract notion of word is usually referred to as '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. 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 expression. The rules for glossing are codified as the Leipzig glossing rules (www.eva.mpg.de/lingua/resources/glossing-rules.php).

Morphology is a subdiscipline of linguistics with its own textbooks, reference works, and journals. Two recent textbooks are Haspelmath and Sims (2010) and Booij (2012). Reference works are Spencer and Zwicky (1998), and Booij, Lehmann et al. (2000), Booij, Lehmann et al. (2004). A recent reference work for English morphology is Bauer, Lieber et al. (2013). Two journals devoted to morphology are *Morphology* (1988– , published by Springer), and *Word Structure* (2008–, published by Edinburgh University Press).

2. Types of morphological operation

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 affixes and words. In addition, languages may also allow for deriving new words from existing ones by only changing the syntactic category of a word. This is referred to as conversion. Examples are the English verbs *to skype* and *to text* mentioned above which are verbs derived from the nouns *skype* and *text* respectively, without this derivational relationship being marked overly

by means of an affix. In English, the conversion of nouns to verbs is a very productive process. Conversely, nouns may be derived from verbs in this way, as is illustrated by noun such as *fall* and *help*.

Word formation by means of affixation means that an affix is added to a base from. The affix can appear before the base word (prefixation), after the base word (suffixation), or, far more rarely, within the base word. Examples of English prefixes are *un–* and *re–*. The negative prefix *un–* can attach to verbs (*un-do*), adjectives (*un-kind*) and nouns (*un-belief*). The suffix *re–* attaches to verbs (and may also appear in nouns and adjective derived from such verbs), as in *re-do* and *re-use*. An example of suffixation is the addition of the suffix *–able* to a verb for making adjectives, as in *do-able* and *re-us-able*. An example of infixation is the use of *fucking* and *goddamn* as pejorative infixes in words like *un-fucking-believable* and *in-goddamn-consistent*.

Compounding is an intensively studied word formation process, as can be seen from a recent handbook on compounding (Lieber and Stekauer 2009). Its defining property is that two or more words (in the sense of lexemes) are concatenated. The form of the lexeme used is often the stem form. In English there is no phonological difference between a concrete word and its stem: a word like *dog* is both a concrete word and a stem, as in *dog food*. However, in many languages stems do not occur as words by themselves, and must be supplemented by an overt inflectional ending for their use of concrete words in a sentence.. For example, in Modern Greek the compound for 'deer hunter' is *elaf-o-kinig-os*. The first constituent is the stem form of *elafi* 'deer', without its inflectional ending *–i*. This stem is followed by a linking element *–o–*, a meaningless constituent, and is then followed by the head constituent *kinig-os*. The inflectional ending on the second stem *kinig–*, *kinigos* is the inflectional ending for the whole compound.

A major division of the class of compounds is that between endocentric compounds and exocentric compounds. In endocentric compounds one of the constituent words is the semantic and syntactic head. In English and other Germanic languages the rightmost constituent is the head. For instance, a post office is a particular type of office, and not a particular kind of post. In languages with gender, such as Dutch, the gender of the compound is also determined by the gender of the right constituent (also referred to as the head). The Dutch gloss for *post office* is *post-kantoor*. Its head *kantoor* is of neuter gender, and hence, the word *postkantoor* is also neuter, as can be seen from the selection of the neuter definite

singular determiner *het* (*het kantoor, het postkantoor*), whereas *post* is non-neuter, and selects the definite singular determiner *de* as in *de kantoorpost* 'the office mail'.

The position of the head in compounds is language-specific. For instance, in Indonesian compounds, the head is on the left. Hence, the Indonesian word for 'post office' is the compound *kantoor-pos* (with *kantoor* and *pos* being loans from Dutch), with the reverse order of the two words).

In the case of exocentric compounds there is no constituent that functions as the head. An example is English *pick-pocket*, where the semantic head is the agent who picks pockets, and neither *pick* nor *pocket* is the head. A famous example of such exocentric compounds are the verb-noun compounds in Romance languages like Italian. The compound *lava-piatti* 'lit. wash-dishes, dish washer' denotes a person or a machine who washes the dishes, but this semantic component is not expressed by one of its building blocks. Another example comes from the class of coordinated compounds in Chuvash, a Turkic language spoken in Russia. The compound *erex-săra* 'lit. vodka-beer' denotes all alcoholic beverages. Hence the compound is semantically a hyperonym of the meaning of its constituents (Wälchli 2005: 141).

However, derivation, conversion, and compounding are not the only means of making complex words. In the morphological process of reduplication, a complete or partial copy of a word is added to that base word. Total reduplication may be seen as a specific form of compounding with two identical constituents , whereas in the case of partial reduplication only part of a word is copied and affixed to the base word. Here are some examples of both types of reduplication from Papua Malay (Angela Kluge, pers. comm.):

(5) *total reduplication*
     ada 'sibling'   ada-ada 'siblings'
     lari 'run'   lari-lari 'keep running'
     partial reduplication
     anak 'child'   an-anak 'children'

In this case of partial reduplication, a copy of the first part of the base word forms a prefix attached to that base word.

The meaning of reduplication patterns is often that of increase: increase in number in the case of nouns, increase in intensity or repetition in the case of verbs, and increase in quality in the case of adjectives. Therefore, reduplication is often qualified as being a case of iconicity, a parallelism between form and meaning because an increase of the size of a word

correlates here with an increase in meaning. However, this parallelism does not always obtain.

Other, non-concatenative, morphological operations are the use of specific vowel and/or consonantal patterns (as in creating English past tense form and past participles: *sing-sang-sung, bring-brought*), and tonal patterns (used a lot in the languages of Africa, for instance). The use of vowel alternations for morphological purposes is a characteristic of Indo-European languages such as English. 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. 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
Nifʿal	ni-qtal	ni-CCVC	a	to kill oneself
Piʿel	qittel	CVC_iC_iVC	i-e	to massacre
Hifʿil	hi-qtil	hi-CCVC	i	to cause to kill
Hitpaʿel	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 Puʿal as a passive variant of the Piʿel, and the Hofʿal as the passive of the Hifʿil. In addition, affixes may be used in Semitic languages.

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

(7)

singular	plural	
màlimò	màlímó	teacher(s)
màlàyikà	màlàyíká	angel(s)
kamà	kámá	chief(s)

In the plural forms, the tones of the last two syllables (Mid-Low) of the singular forms are replaced by High tones (High tones are indicated by an acute accent on the vowel letter, e.g.

---

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

á), Low tone by a grave accent, e.g. à, and Mid tone by absence of an accent). Thus, plurality is systematically marked by a specific tone pattern.

Word formation can also take place by what is called back formation: a word is reinterpreted as a complex word, and its base word is then created afterwards. A classic example is the verb *to babysit*, back formed from the compound noun *babysitter*. Another example is the formation of the verb *to buttle* 'to do what a butler does', from the noun *butler*, by interpreting its ending –*e*r as a suffix –*er*.

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.

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 and complex words. 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]$_N$

Since compounding can be reapplied to a complex word, we can get recursive compounds, with complex structures, as exemplified by the compound [[[[*song book*] *shop*] *assistant*] *salary*].

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 same applies to *brunch*, the combination of the first part of *br-eakfast* and the second part of *br-unch*.

3. The place of morphology in the grammar

The position of morphology in the architecture of the grammar of natural languages has been debated for decades in the linguistic literature, and it is impossible to summarize this debate in detail in this chapter.

In a syntactic approach, morphology is defined as the 'syntax of morphemes' and the way in which morphemes are combined is accounted for by the same syntax that accounts for the construction of sentences from words. The most recent form of this approach is the theoretical framework of Distributed Morphology (Harley and Noyer 1999, Marantz 2013).

In the lexicalist approach to word formation, on the other hand, there is a separate morphological component in the grammar that accounts for word formation, and precedes the syntactic component that serves to combine words into grammatical sentences. This component enriches the lexicon of a language which provides the words that can be used in the construction of sentences. The inflectional forms of word may be generated pre-syntactically, or post-syntactically. In the first option, the proper inflectional form of a word in a particular syntactic context is selected by syntactic rules such as agreement rules, in the second option, the stem of words is selected for lexical insertion into syntactic structure, and this stem is then operated upon by the post-syntactic rules of inflection. For instance, the verb form *sings* in *My father sings wonderfully* is either selected as the proper form (3rd person singular present tense) by the rule of Subject-Verb agreement, or this rule assigns the features [3rd person singular] to the present tense stem *sing*, and this configuration is then spelled out as *sings*.

The view of inflectional rules as spelling out the phonological form of a word with a specific array of morpho-syntactic features is referred to as realizational morphology (Stump 2001). An alternative model of inflection is the Word-and-Paradigm model in which inflectional forms are computed on the basis of other forms in the paradigm of the same word (Matthews 1974 [2nd edition 1991], Blevins 2006, Ackerman, Blevins et al. 2009, Blevins 2012).

In recent theories of the architecture of grammar, the distinction between lexicon and morphological rules or schemas has been relativized. The lexicon is considered to contain both the set of existing (simplex and complex) words of a language, and the abstract schemas of the type (2) that motivate existing complex words and define how new words can be coined (Booij 2010). Moreover, it cannot be the case that word formation always precedes syntax, since syntactic phrases can be form part of complex words, as illustrated in (9) for English:

(9)   a one-of-a-kind vacation
       a run-of-the-mill blockbuster

all-you-can-eat buffet

a get-it-done-attitude

These facts imply that syntax may define the wellformedness of parts of words, and hence can stand in a feeding relationship to morphology.

The lexicon of a language does not only contain its simplex and (established) complex words, but also larger lexical units that are used to denote concepts or entities. For example, the English phrase *yellow fever* (with main stress on its second constituent) functions as a lexical unit, and denotes a particular disease. In other words, all words are lexical units, but not all lexical units are words.

The demarcation of syntactic units and morphological units (= complex words) follows from the principle of Lexical Integrity: complex words are islands for syntactic rules. The Modern Greek phrase *psichros polemos* 'cold war', for instance, is considered a phrase, even though it is a lexical unit, as the adjective *psichros* agrees in number and gender with the noun *polemos*. Other examples of phrasal lexical units are the particle verbs of Germanic languages, such as English *to look up* and the Dutch particle verb *aan-vallen* 'lit. at-fall, to attack'. The particle verb *aanvallen* is clearly a lexically stored unit, with an unpredictable meaning. Yet, it is phrasal in nature, as its constituents *aan* and *vallen* are split in main clauses:

(10) De insecten vallen ons aan

The insects fall us at

'The insects attack us'

The coinage of new particle verbs can be characterized by schemas similar to those for affixation as given in (2) (Los, Blom et al. 2012).

4. Interfaces

The term 'interface' means that different levels of representation in the grammar (phonological form, morpho-syntactic form, meaning) are related in a systematic fashion. For instance, morphological word structure plays a role in phonology since the proper phonetic realization of morpheme sequences may depend, directly or indirectly, on their morphological structure. For instance, in many English compounds, the main stress of the word is located on its first constituent, as in *kite runner*, where the monosyllabic constituent [kajt] carries the main stress of the compound word. In phrases, the main stress is always on

the right constituents. Hence, the proper application of phonological rules of stress assignment to words requires access to their morphological structure.

Another example of the morphology-phonology interface is that morphological constituents may form domains of phonological rules. This can be illustrated by means of Hungarian. Hungarian is well known for its having vowel harmony, the requirement that within a word all vowels are either front vowels or back vowels. This is why Hungarian suffixes may have more than one phonetic form, depending on the vowels of the stem to which they are attached. For instance, we find *Buda-nál* 'in Buda' versus *Pest-nél* 'in Pest' because *Buda* contains back vowels and hence selects an allomorph of the adhesive suffix 'at, in', with a back vowel, *–nál*, whereas *Pest* contains a front vowel, and hence selects an allomorph with a front vowel, *–nél*. The name of these two cities combined is the compound *Budapest*, a city which is indeed composed of two cities, *Buda* on the right bank of the river Danube, and *Pest* on the left bank. This compound does not violate the constraint of vowel harmony, even though it has both back vowels and front vowels, because the domain of this constraint is the phonological word rather than the morphological word. In many languages, including Hungarian, each word component of compound forms a phonological word of its own, and therefore, *Budapest* consists of two phonological words.

Morphological structure is also revealed by the phonotactics of a word. For instance, if a Dutch word contains the sequence /fp/, you know for certain that there must be a morphological boundary between /f/ and /p/, as this consonant cluster never occurs within one and the same phonological word; hence, this cluster must have been created by a morphological operation, the concatenation of two words, as in the compound *straf-pleiter* 'criminal lawyer'.

Morphological structure may also determine the syntactic properties of a complex word, in particular its word class and its syntactic valence. Many affixes determine the syntactic category of a word. For instance, English words with the adjective *–less* are adjectives. Syntactic valence is the term for the combinatorial properties of a word in sentences. Verbs may be either intransitive or transitive. In the first case, they do not select a direct object, in the second case they do. Examples are *to die* and *to kill* respectively. An example of morphology affecting syntactic valence is that the addition of the suffix *–ize* turns the noun *alphabet* into the transitive verb *alphabetize*. Hence, there may be a systematic correlation between the morphological structure of a word, and its morpho-syntactic properties. This also

applies to gender properties: as we saw above, the gender of a Dutch compound is identical to the gender of its head.

The semantic interpretation of compounds provides a simple illustration of the interface between morphological structure and meaning: the formal head of a compound in Germanic languages is also the semantic head. That is why a song book is a kind of book, and not a kind of song. A morphological operation may also be motivated by pragmatic meaning. A famous case is the use of diminutive forms of nouns for the expression of endearment. The Dutch diminutive word *mann-etje* 'man-diminutive' derived from the noun *man* 'man' is not only used for referring to men of small size but also to express one's affection for a boy or a man, without a small size of the person being implied.

5. Morphological classification

Languages may be classified according to the role and nature of their morphology (Comrie 1981), (Haspelmath 2009).[3] 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 analytic or isolating language because it has 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):

(11) 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 of evidentiality that indicates the nature of the information source. (A clitic is a word that is phonologically dependent on an adjacent host word, '=' is the symbol for the link between a host word and a clitic.)

---

[3]   The text of this section is partially based on the text of section 2 of Booij (to appear).

The third dimension of morphological classification of languages 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:

(12) 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.

6. Storage and computation

Morphology plays an important role in linguistic research of language processing. An important empirical question is the division of labour between storage and computation. To what extent are complex words retrieved ready-made from lexical memory, and to what extent are complex words coined during the process of language production? In the domain of word formation it is quite clear that lots of complex words are memorized and retrieved from memory. This can be concluded from frequency effects. Words with a higher frequency use will have a higher level of activation, and hence more easily recognized as correct words in lexical decision tasks. In the domain of inflection this may be different, as a word may have quite a number of different forms, in particular in languages with rich inflectional systems. On the other hand, there is solid evidence that the plural forms of nouns in Dutch and Italian are often memorized when they have a sufficiently high frequency, even though they could have been computed on line quite easily (Baayen, Burani et al. 1997, Baayen, McQueen et al. 2003). These findings go against the hypothesis defended in Pinker (1999) that regular inflectional forms are computed on line, and only irregular forms such as past tense forms of English verbs with vowel alternation of the type *sing-sang* are retrieved directly from memory.

It remains true, however, that there is a strong correlation between frequency and exceptionality. The English irregular verbs that create past tense and past participle forms by means of vowel alternation such as *to sing* appear to belong to the top frequency class of English verbs. High frequency means a high degree of entrenchment, and thus these verb do not easily regularize, as their vowel alternation patterns are so well entrenched (Bauer, Lieber et al. 2013: 67).

7. Acquisition of morphology

How are morphological schemas or rules for word formation and inflection acquired by children who are in the process of acquiring their native language? First, they will acquire a small set of complex words that instantiate a particular word formation schema. After sufficient exposure (stage 1) they will be able to recognize an abstract schema as in (2) that summarizes their knowledge of a subset of complex word (stage 2). Thus, they will be able to coin new complex words themselves. In coining new words, they might not be hampered by knowledge of exceptions and lexical conventions. Hence, they might overgeneralize, and create the regular plural form *foots* instead of *feet*, or the compound *war man* for soldier. At stage 3, they will also have learned exceptions like *feet*, and thus command the whole system, both the rules and the exceptions (Berko 1958/2004).

The number of compounds is very high in Germanic languages like Dutch and English, and hence, children acquire the competence to coin new compounds at an early age, in particular the class of noun-noun compounds. Hence these children make a lot of new compounds, also because they are not hampered by the conventions of the language community to which they are going to belong (Clark 2003 [2009, 2nd edn]). For example, one of my daughters made the compound *oorlog-s-man* 'war man' to denote soldiers. The conventional Dutch word for 'soldier' is *soldaat* but this word was not yet known to her. These newly coined compounds will normally not survive, because children will learn the conventional word for a certain concept in due course.

8. Morphological change

The morphology of language has interesting historical dimensions. A first question is how the morphological system of a language may have arisen in the course of time, and secondly, we may ask how such a system can change. Change does not only occur on the level of the language system, but also at the level of the individual complex word: complex words, once

formed, may lose their transparency through a number of factors. In this section, more will be said about these different aspects of morphological change.

Let us first focus on the rise of morphology. A classical idea is that 'today's morphology is yesterday's syntax' (Givón 1971). For example, phrase structures may develop into compounds. An example is the change of noun phrases of the type [N$_{Genitive}$ N], with a first noun case-marked as genitive preceding the head noun of the phrase, into a NN compound, in which the old genitive case marker is reinterpreted as a linking morpheme. For instance, the Middle Dutch noun phrase *coninc-s sone* 'king's son' was reinterpreted as the compound [N-s-N], with the *s* as a linking element. Hence, in present day Dutch we finds lots of NN compounds with the semantically empty linking element *–s*.

A famous example from Romance languages is the development of the adverbial suffix *–mente* from the Latin word *mens* 'mind', embedded in phrases with ablative case such as *clara mente* 'with a clear mind'. This phrase was reinterpreted as one word, with *mente* being reinterpreted as a suffix that denotes the manner of an action. The English suffix *–ment* as in *government* is a reflex of this historical development, as English borrowed the French form of *–mente, ment*.

This change of a phrase into a word is called univerbation. Many words, some still felt as complex, arose this way. For instance, the preposition *beside* derives from the Old English phrase *be sidan* 'by side', and the reflexive pronoun *himself* derives from the phrase *him self*. This shows that the source of a complex word is not always a morphological process, it may also have originated as an effect of univerbation (Brinton and Traugott 2005).

Parts of compounds may develop into derivational affixes. This is exemplified by the English suffixes *–dom*, *–hood*, and *–ship* as in *kingdom, brotherhood, friendship*, that were once words and appeared as the right part of compounds (Trips 2009). For instance, *–hood* derived from the Old English word *had* 'state, rank, order'. These words became very productive as suffixes, witness the recent coinings *gurudom, guruhood*, and *guruship*. Such words acquired more general meanings when used as parts of compounds, and ended up as derivational affixes. In some cases, the lexical origin is still visible, as in words beginning with *great* such as *great grandfather* and *great grandchild*. The adjective *great* has acquired a specialized use for denoting family relationships, which is only available when this word forms part of a complex word. Similarly, a number of English words turned into prefixes with a more specific meaning than the corresponding preposition, as *out* in *out-perform*, *over* in *over-do*, and *under* in *under-line* (Booij and Hüning 2013).

A second dimension of morphological change is that complex words, once formed, may loose their semantic transparency, and might even become simplex words. An example of semantic opacity is the compound *breakfast*, with the original meaning of breaking the overnight fasting. For most English speakers, this word will no longer be perceived as a compound word. This may also affect the pronunciation. In a compound such as *postman*, the second part is experienced as a kind of affix for 'agent', and hence, the vowel of *man* is usually realized as [ə]. Similarly, the vowel of the second syllable in *breakfast* is also pronounced as [ə], and not with the original full vowel of *fast*.

Morphological systems may also change through language contact. This is clearly visible in English that has a lot of non-native suffixes due to its contact with French, Greek and Latin. For instance, besides the suffix *–ness* (of Germanic origin) we have *–ity* as a suffix for nouns, as in *obesity* and *opacity*. Most of these non-native suffixes prefer to be attached to a non-native base word. For the adjective *good*, the noun *goodness* is chosen, and *goodity* is odd, whereas both *denseness* and *density* are grammatical, since native suffixes can usually be attached to both native and non-native base words. There are also lots of non-native prefixes such as *super–* and *re–*. These non-native prefixes are often less choosy about their base words, and also attach to native words, witness *super-man* and *re-send*.

One may wonder how affixes, that is, bound morphemes, can be borrowed from another language, as they are not independent entities. The answer is that in borrowing a number of words of a certain morphological type may be borrowed, for instance a number of words in *–ity* (the English adaptation of French *–ité*), and on the basis of such a set of morphologically similar words, the language user may discover a pattern, for instance [A-*ity*]$_N$. Hence, there will be English words in *–ity* without a correspondent word in *–ité* in French, for example *oddity*.

Even one borrowed word with sufficient impact may lead to the borrowing of an affix. This is the case for the prefix *über,* which was borrowed in English from German on the basis of the philosophical notion of *Über-mensch* from Nietzsche's philosophical work. Nowadays, we find lots of English words with this prefix, as in *über-burger, über-guru, über-diva, über-style*, and *über-station wagon.* In all these cases the meaning of *über* is something like 'exhibiting the characteristic properties to an extreme extent'.

Languages may also lose parts of their morphological system. A well-known example is the loss of case morphology in almost all Germanic languages, and most Romance languages. The function of case markers has been taken over, at least partially, by other formal means

such as word order, or the use of prepositions and postpositions. Remnants of the old case system may still be seen in lexicalized expressions (as in Dutch *te-gelijk-er-tiid* 'at the same time, simultaneously', with the case ending *–er* on the adjective *gelijk* 'same'), and in the recycled form of the genitive case marker of English, now a possessor marker (as in *John's book* and *the king of England's hat*).

References

Ackerman, F., et al. (2009) Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins and J. Blevins (eds) *Analogy in grammar*, pp. 54-82. Oxford: Oxford University Press.

Aronoff, M. (1994) *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT Press.

Baayen, R. H., et al. (1997) Effects of semantic markedness in the processing of regular nominal singulars and plurals in Italian. In G. Booij and J. Van Marle (eds) *Yearbook of Morphology 1996*, pp. 13-34. Dordrecht: Kluwer.

Baayen, R. H., et al. (2003) Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R.H. Baayen and R. Schreuder (eds) *Morphological structure in language processing,* pp. 355-390. Berlin: Mouton de Gruyter**.**

Bauer, L., et al. (2013) *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.

Berko, J. G. (1958/2004) The child's learning of English morphology. In B. Lust and C. Foley (eds) *First language acquisition: The essential readings*, pp. 253-273. Malden, MA: Blackwell.

Blevins, J. P. (2006) Word-based morphology.*Journal of Linguistics* 42: 531-573.

Blevins, J. P. (2012) Word-based morphology from Aristotle to modern WP. In K. Allan (ed.) *The Oxford Handbook of the History of Linguistics*, pp. 397-417. Oxford: Oxford University Press**.**

Booij, G. (1996) Inherent versus contextual inflection and the split morphology hypothesis. In G. Booij and J. v. Marle (eds) *Yearbook of Morphology 1995*, pp. 1-16. Dordrecht/ Boston: Kluwer.

Booij, G. (2010) *Construction morphology.* Oxford: Oxford University Press.

Booij, G. (2012) *The grammar of words. An introduction to morphology.* 3rd edn. Oxford: Oxford University Press.

Booij, G. (To appear). The structure of words. Chapter 11 of J. Taylor (ed.) *The Oxford Handbook of the Word*. Oxford: Oxford University Press.

Booij, G. and M. Hüning (2013) Affixoids and constructional idioms. In R. Boogaart, T. Colleman and G. Rutten (eds) *Extending the scope of construction-based grammar*, pp. ??? Berlin: De Gruyter Mouton.

Booij, G., et al. (eds) (2000) *Morphologie / Morphology. Ein internationales Handbuch zur Flexion und Wortbildung / An international handbook on inflection and word formation (in colloboration with Wolfgang Kesselheim and Stavros Skopeteas).* Handbücher zur Sprachwissenschaft. Berlin: De Gruyter.

Booij, G., et al. (eds) (2004) *Morphologie / Morphology. Ein internationales Handbuch zur Flexion und Wortbildung / An International handbook on Inflection and Word Formation (in colloboration with Wolfgang Kessselheim).* Handbücher zur Sprachwissenschaft. Berlin: De Gruyter.

Booij, G. and R. Lieber (2004) On the paradigmatic nature of affixal semantics in English. *Linguistics* 42: 327-357.

Brinton, L. J. and E. C. Traugott (2005) *Lexicalization and language change*. Cambridge: Cambridge University Press.

Clark, E. (2003) *First language acquisition*. [2nd edn 2009.] Cambridge: Cambridge University Press.

Comrie, B. (1981) *Language universals and linguistic typology*. Oxford: Blackwell.

Givón, T. (1971) Historical syntax and synchronic morphology, an archeologist's field trip. *Chicago Linguistics Society* **7**: 394-415.

Harley, H. and R. Noyer (1999) Distributed Morphology. *Glot International* 4(4): 3-9.

Haspelmath, M. (2009) An empirical test of the agglutination hypothesis. In E. Magni, S. Scalise and A. Bisetto (eds) *Universals of language today*, pp. 13-29. Dordrecht: Springer**.**

Haspelmath, M. and A. D. Sims (2010) *Understanding morphology*. London: Hodder Education.

Jurafsky, D. (1996) Universal tendencies in the semantics of the diminutive. *Language* 72: 533-578.

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

Lieber, R. and P. Stekauer (2009) *The Oxford handbook of Compounding*. Oxford: Oxford University Press.

Los, B., et al. (2012) *Morphosyntactic change: a comparative study of particles and prefixes*. Cambridge: Cambridge University Press.

Marantz, A. (2013) No escape from morphemes in morphological processing. *Language and Cognitive Processes* 28: 905-916.

Marchand, H. (1969) *The categories and types of present-day English word formation*. München: Beck.

Matthews, P. H. (1974) *Morphology*. [2nd edn 1991.] Cambridge: Cambridge University Press.

Mithun, M. (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 (in collaboration with Wolfgang Kesselheim and Stavros Skopeteas)*, pp. 916-928. Berlin: De Gruyter**.**

Packard, J. (2000) *The morphology of Chinese. A linguistic and cognitive approach*. Cambridge: Cambridge University Press.

Stump, G. T. (2001) *Inflectional morphology. A theory of paradigm structure*. Cambridge: Cambridge University Press.

Trips, C. (2009) *Lexical semantics and diachronic morphology: the development of the derivational suffixes -hood, -dom, and -ship in the history of English.* Tübingen: Niemeyer.

Wälchli, B. (2005) *Co-compounds and natural coordination*. Oxford: Oxford University Press.