

# Lexicalization in Morphology

Martin Hilpert, Institut de langue et littérature anglaises, Université de Neuchâtel

<https://doi.org/10.1093/acrefore/9780199384655.013.622>

**Published online:** 25 January 2019

## Summary

The term *lexicalization* describes the addition of new open-class elements to a repository of holistically processed linguistic units. At the basis of lexicalization are word-formation processes such as affixation, compounding, or borrowing, which are a necessary precondition for lexicalization. Still, lexicalization goes beyond word formation in important respects. First, lexicalization also involves multi-word expressions and set phrases; second, it includes a range of processes that follow the coinage of a new element. These processes conjointly lead to holistic processing, that is, the cognitive treatment of a linguistic element as a unified whole. Holistic processing contrasts with analytic processing, which is the cognitive treatment of a linguistic unit as a complex whole that is composed of several parts. Lexicalization is usefully contrasted with grammaticalization, that is, the emergence of new linguistic units that fulfill grammatical functions. Finally, lexicalization is also a concept that lends itself to the study of cross-linguistic differences in the types of meaning that are lexicalized in specific domains such as, for example, motion.

**Keywords:** closed-class, derivation, grammaticalization, holistic processing, inflection, lexeme, morphology, open-class, satellite-framed, typology, verb-framed, word formation

**Subjects:** Morphology

## 1. What Is Lexicalization?

---

Defined in the simplest of terms, lexicalization is the process by which new words are added to a language. Elements such as *emoji*, *crowdsourcing*, or *BFF* are relatively recent examples of lexicalization in English; the entire vocabulary of a language can be seen as the result of this process. A closer look at the phenomenon does however necessitate a definition that is more complex than simply the addition of new words.

First of all, lexicalization concerns not only new words, but also multi-word units such as fixed expressions (*cutting edge*), idioms (*let the cat out of the bag*), and set phrases (*They don't make 'em like that anymore*). These types of units are part of what is typically considered the lexicon of a language, that is, the repository of elements that carry specific meanings and that are stored and processed as holistic units in the minds of speakers (cf. Lehmann, 2002, p. 1). A more appropriate definition of lexicalization would, instead of making reference only to new words, refer more generally to the addition of new linguistic elements. Yet even this definition needs to be specified further, because not every new element that emerges in language use is an instance of lexicalization.

As will be discussed in section 4, it is useful to contrast lexicalization with grammaticalization, that is, the process by which new grammatical constructions come into existence (Hopper & Traugott, 2003). Lexicalization, as the term already suggests, is concerned with new lexical elements, while grammaticalization is concerned with new grammatical structures. To take an example, the emergence of a new passive construction such as the English *get*-passive (cf. Wanner, 2009), as in *You get paid at the end of the month*, is an example of grammaticalization, not lexicalization. In order to differentiate between grammaticalization and lexicalization, it can be stated that the latter is concerned with the addition of new open-class elements to the lexicon of a language. The term *open-class* refers to categories of linguistic elements to which new members can be added freely and instantaneously. A shorter alternative for the term *open-class element* is the term *lexeme*, which simultaneously makes clear that the item in question is stored in the lexicon. Central open-class categories of English are nouns, verbs, adjectives, and adverbs. Speakers who are asked to come up with a number of prototypical words will typically produce examples that draw on these categories. They contrast with so-called closed-class categories, such as pronouns, determiners, and auxiliaries, which are not targets of lexicalization, but of grammaticalization. In light of these considerations, a more refined definition of lexicalization would make clear that the process adds new open-class elements to a language.

While this definition already captures a lot of useful information, it still leaves an important aspect unaddressed, namely that linguistic elements undergo a number of types of formal and semantic changes once they have become lexicalized. An example of this can be seen in the English word *holiday*, which historically emerged as the collocation of *holy* and *day* (cf. the OED entry for *holiday*). To a present-day speaker of English, this historical connection, when pointed out, will not come as a great surprise. Still, with regard to meaning and pronunciation the word *holiday* has diverged from its erstwhile source to such a degree that it no longer triggers an automatic association with the word *holy*. At some point along the line, speakers ceased to process *holy day* by breaking the expression down into its component parts (i.e., analytic processing), and instead adopted a holistic processing strategy, in which *holiday* is understood as a single item with an idiosyncratic meaning. Lehmann (2002, p. 3) views holistic processing as a defining characteristic of lexicalized units. Integrating this point into the definition that was developed up to now, the term *lexicalization* will be used thusly for the purposes of this article:

Lexicalization is the process of adding new open-class elements to a repository of holistically processed linguistic units.

This definition sounds a good deal more technical than the one we started out with, and it raises a number of new questions. Elevating holistic processing to a definitional criterion provokes the question of whether newly formed complex words, which are at that point still fully analyzable, would not be seen as instances of lexicalization. A discussion of that point will touch on issues relating to cognition, because the notion of holistic processing presupposes a linguistic framework that makes reference to cognitive processes, which demands some further clarification. The term *open-class* suggests a basic distinction between lexical and grammatical elements.

## 2. Processes That Give Rise to Lexicalization

---

Where do new words come from? The area of linguistics that has been chiefly concerned with this question is the study of word formation (Štekauer & Lieber, 2005; Müller, Ohnheiser, Olsen, & Rainer, 2016; Plag, 2003), which has identified different processes by which speakers can coin new lexical elements. A discussion of these word formation processes first necessitates the introduction of a few basic terms. Up to this point, the term *word* has been used interchangeably with the more technical expressions of an open-class element or a lexeme. The need for a more differentiated terminology becomes clear when we consider forms such as English *defend*, *defends*, *defending*, and *defended*. These forms differ from each other in sound and meaning, and from a lay perspective they could be said to be different words. At the same time, they are sufficiently similar in sound, meaning, and not least in their function as verbs to make the argument that they instantiate the same lexeme, namely the verb *defend*. By contrast, the forms *defendant*, *defense*, or *defensible* are more different from one another. While they do share substantial aspects of their form and meaning, they instantiate different word classes than the verb *defend*, and their meanings significantly go beyond the meaning of that verb. These observations support the idea that *defendant*, *defense*, and *defensible* are different lexemes.

The terms that allow us to pinpoint the difference between the two sets of forms are inflectional morphology and derivational morphology. The first set exemplifies a paradigm of forms that differ with respect to their inflectional markers, in this case the English verbal inflections *-s*, *-ing*, and *-ed*. Inflectional morphology is concerned with the relationship between different word forms that belong to the same lexeme (Haspelmath & Sims, 2010, p. 18). Derivational morphology, by contrast, is concerned with the formation of new lexemes. Derivational affixes such as English *-ant* or *-able* (as in *defend-ant*, *defend-able*) can be combined with existing lexemes. The element to which an affix is added is called the host or the base. A combination of a base and a derivational affix yields a new element that is added to the lexicon.

Importantly, the differences between inflectional and derivational morphology go beyond the issue of word formation. A first basic difference between the two is that inflectional affixes do not change the word class of the base to which they are attached. Derivational affixes, by contrast, may do so. The base *defend* is verbal; the lexeme *defendant* is a noun. In terms of meaning, the main function of inflectional morphology is to encode grammatical categories. These include tense, aspect, mood, person, and voice, which are typically expressed on verbs, and number, gender, and case, which are often found on nouns. Naturally, languages differ with regard to the extent that they express grammatical categories by means of inflectional markers and the elements that act as morphological hosts for inflectional markers. The functions and meanings of derivational affixes are cross-linguistically very diverse, but a few typical functions can be identified. Many languages have derivational affixes that function like English *-er*, which allows the expression of an agent who carries out an activity, as in *teacher*, *runner*, or *writer*. Similarly common are derivational affixes that attach to a nominal host and yield lexemes that express qualities. English lexemes such as *childish*, *foolish*, or *hellish* illustrate this function. Inflectional and derivational morphology further differ with regard to word structure, such that derivational affixes have a tendency to be situated relatively close to the morphological base, whereas inflectional affixes tend to be found at the outer edges of

words. In an English word form such as *nationalizations*, which can be broken down into the stem *nation* and the affixes *-al*, *-ize*, *-ation*, and *-s*, it is only the final plural affix that represents an inflectional category. This tendency does not hold only for English but is in fact cross-linguistically valid. It is in line with the observation that derivational morphemes allow the formation of lexicalized elements, which thus have a high degree of internal unity, whereas inflectional morphemes only allow the formation of word forms, which speakers tend to process analytically by breaking them down into their component parts. This last point, however, needs to be qualified. Psycholinguistic evidence (Katz, Rexer, & Lukatela, 1991; Sereno & Jongman, 1997) demonstrates that speakers use holistic processing for highly frequent inflected word forms such as *cats* or *walked*. These findings suggest that even fully transparent complex forms can be stored as lexemes in the lexicon, as long as they are frequent enough. The term *redundant representation* is used for this phenomenon. If a word that could be processed analytically is nonetheless represented as a whole word in the lexicon, that representation is, strictly speaking, redundant.

With a basic distinction of inflectional and derivational morphology in place, we can now proceed to a more detailed discussion of different processes that create the preconditions for lexicalization, some of which actually go beyond derivational morphology.

### 2.1 Affixation

Affixation as a word formation process involves the addition of a derivational morpheme to an existing lexeme. Affixation typically exhibits restrictions, so that an affix can be combined with only a particular kind of base, but not with others. For example, the English derivational suffix *-ness* can be freely attached to adjectival bases (*weakness*, *darkness*, *sadness*, *selfishness*, etc.), but not to verbal bases (with rare exceptions such as *forgiveness*). The derivational suffix *-ee*, as in *employee*, is restrictive not only with respect to the word class of the base, but even with regard to the meaning of that base: the suffix combines only with English transitive verbs that refer to a sentient human being as their direct object; *trainee*, *divorcee*, or *nominee* illustrate this pattern. Further restrictions pertain to complex words in which two derivational suffixes follow one another. The suffix *-ous* can precede the suffix *-ness* (*joyousness*, *graciousness*, etc.) but the reverse sequence of *-ness* and *-ous* is not possible. A structural restriction on the suffix *-ous* is that it cannot be combined with bases that already contain another derivational suffix. A final characteristic of affixation that needs to be mentioned here is that the combination of a base with an affix may result in formal changes to either of the two elements. To illustrate, the English adjective *gracious* is the result of combining the noun *grace* with the derivational suffix *-ous*. Yet, it is evident that the lexicalized element differs from the mere juxtaposition of noun and suffix. Most important, the pronunciation is different. The alveolar fricative /s/ is replaced by the palatal fricative /ʃ/. Also, the spelling is different, given that the grapheme <e> is rendered as an <i>. Affixation can also alter the stress pattern of the base, as can be seen in lexemes such as *parasitic*, *Chinese*, or *accidental*.

## 2.2 Compounding

Another word formation process that involves the concatenation of linguistic units is called compounding. Whereas affixation combines a base with an affix, which by definition cannot appear on its own, compounds are complex lexemes that combine two or more lexemes. English nouns such as *lightbulb*, *easy chair*, or *software installation manual* illustrate the general principle. Compared to derivational affixation, compounding in English is relatively unconstrained (Libben, 2008, p. 73). If there is a semantic motivation for speakers to combine different lexemes into a compound, they are generally free to do so. An important characteristic of compounds is their headedness. The head of a compound can be defined in terms of form and meaning. Formally, the head of a compound is the lexeme that lends its part of speech category to the entire compound. To illustrate, the compound *easy chair* has all the structural characteristics of an English noun: it can take an article, a modifying adjective, or a plural suffix, which allows the conclusion that *chair* is its head. With regard to meaning, the status of *chair* as the head is even more obvious, because an *easy chair* is a kind of chair, not a kind of easy. Compounds in which a head can be identified in this way are called endocentric compounds. English lexemes such as *part time*, *underway*, or *meanwhile* resist an analysis of this kind. The compound *part time* combines two nouns, and yet it functions either as an adjective (as in *John is a part time photographer*) or as an adverb (*John works part time*). Compounds whose internal elements cannot be identified as a head are called exocentric compounds. Certain lexemes represent borderline cases between compounding and affixation. In forms such as *neoliberal* or *physiotherapy*, the initial element is neither a lexeme nor a clear case of an affix: The element *physio-* in *physiotherapy* is not an autonomous element, which makes it difficult to analyze it as a lexeme, and simultaneously it goes beyond the typical role of a derivational affix, in contributing highly specific meaning to the combined form.

## 2.3 Conversion

In many languages, new lexemes can be created out of already existing ones by recruiting them into a different part of speech category. The verbal lexeme *swim* has thus given rise to a nominal lexeme, as in *Let's go for a swim*. This word formation process is known as conversion. Many examples of conversion do not show any formal differences between its source, that is, the element that forms the basis for the formation of a new word, and its target, that is, the resulting element. There are however cases in English in which the stress pattern in the target is altered. To illustrate, the verb *record* has a different stress pattern than the noun *record*. With regard to meaning, the targets of conversion can convey a meaning that is more specific than that of their sources. The nominal lexeme *drink*, as in *He had another drink*, evokes the idea of an alcoholic beverage. The verbal lexeme *bread*, as in *He breaded the chicken*, refers to a method of food preparation that involves breadcrumbs. Typically, both the source and the target of conversion instantiate open-class elements. Cases in which closed-class elements are involved as sources (e.g., *up the hill* > *We upped the budget by 36%*) are sometimes discussed as potential cases of degrammaticalization (Janda, 2001, p. 299; Norde, 2001, p. 235).

## 2.4 Reductive Word Formation Processes

Clippings, acronyms, and initialisms are created by reductive word formation processes, which allow speakers to produce phonologically shortened variants of existing lexemes. These shortened forms may ultimately conventionalize and through that become lexicalized units in their own right. As the name suggests, clippings represent variants of a lexeme from which some of its phonological substance is clipped off. A *refrigerator* can thus be referred to as a *fridge* or a *ref* in different varieties of English. The emancipation of an acronym such as *laser* or an initialism such as DNA is evident in the fact that these forms are familiar to speakers who do not know what the letters in these expressions stand for (respectively, *light amplification by stimulated emission of radiation* and *deoxyribonucleic acid*).

## 2.5 Blending

The word formation process of blending can be viewed as a combination of clipping and compounding. It is similar to compounding in that it combines several autonomous lexemes into a new form. Both of these lexemes may undergo clipping in the process. In examples such as *brunch* (from *breakfast* and *lunch*) or *spork* (from *spoon* and *fork*) this is exactly what happens. In many other cases, however, one of the combined words may be retained in its entirety, as forms such as *staycation*, *manscaping*, or *foodgasm* illustrate. While blending has been regarded as a largely unsystematic word formation process, Gries (2004) shows that it is actually subject to a number of regularities. When two lexemes are combined into a blend, the source words tend to be broken up in a way that preserves their syllable structure (cf. Plag, 2003, p. 124). Beyond that, the second source word is typically longer than the first. The longer source word also typically lends its overall length and stress pattern to the resulting blend. The first word tends to be more frequent and semantically more basic. All of these characteristics apply to *staycation*: *stay* is shorter and more frequent than *vacation*, which is identical to the blend in length and stress pattern. Gries (2004, p. 654) further demonstrates that shorter source words contribute a relatively greater portion of their phonological substance to the blend. In *foodgasm*, the shorter source word *food* contributes all of its phonological substance to the blend, while the longer source word *orgasm* contributes only four out of its six phonemes. Lastly, speakers tend to produce blends that exploit phonological overlap between the two source words. This can be observed in *manscaping*: the two phonemes /æɪn/ of the source word *man* are also present in the first syllable of the second source word, *landscaping*. The blend *bromance* even retains both of its source words, *bro* and *romance*, in their entirety.

## 2.6 Coinage

The processes discussed allow speakers to express new ideas on the basis of already existing items and structures that are combined in new ways. Yet new lexemes can also come into being through coinage, which is a term that describes the invention of new words that do not draw on any preexisting morphological structures. While coinage is relatively uncommon in ordinary, conversational language use, it does occur with some frequency in advertising and in some genres of fiction. For example, J. K. Rowling's Harry Potter series features the fictional sport *Quidditch*.

While this coinage is made up of existing phonemes of English and combines them in regular ways that conform to the phonotactics of English, it does not incorporate established lexical material in the way that blends or compounds do.

### 2.7 Borrowing

Other sections have been concerned with processes that give rise to the formation of new words, which may subsequently enter the lexicon of a language. The process of borrowing is an alternative way of adding new material to the lexicon. Speakers regularly adopt words from other languages into their vocabulary, rather than using an equivalent lexeme that already exists in their first language. Such an equivalent lexeme may also happen to be simply unavailable in their native vocabulary. The language that provides the new element is called the donor language; the language into which the new element is incorporated is called the recipient language. Borrowing can proceed in different ways. In the most straightforward case, a word is borrowed in its existing form from the donor language and used as such in the recipient language. The English word *internet* has become a globally used lexeme in this way. The integration of a borrowed item into the lexicon of the recipient language is evident when that item is used with inflectional affixes of the recipient language. For example, in the German phrase *die Entwicklung des Internets* ‘the development of the internet’, the borrowed item is inflected with a German genitive case marker and assigned neutral grammatical gender. Borrowed items may further undergo an adaptation that makes their pronunciation conform to the phonology of the recipient language. In Hawaiian, whose phonemic inventory is smaller than that of English and whose phonotactics do not allow consonant clusters or coda consonants, the borrowed collocation *Merry Christmas* is thus famously rendered as *Meli Kalikimaka* (Akmajian, Farmer, & Bickmore, 2017, p. 97). Borrowing involves the recipient language to an even stronger degree in what is called loan translation. The difference between simple borrowing and loan translation can be seen in the English lexemes *kindergarten* and *beer garden*, for both of which German is the donor language. The German compound *Biergarten*, which refers to an ‘open air pub’, has been rendered into English through a literal translation of its component parts. Not all elements of a language are equally likely to be borrowed into a recipient language. Thomason and Kaufman (1988, p. 74) arrange different types of linguistic elements on a scale that correlates the intensity of contact between two languages with increasing pervasiveness of borrowing. The elements that are most easily borrowed are lexemes with non-basic meanings, such as *internet* or *Biergarten*. Greater language contact can lead to the borrowing of basic lexical vocabulary and even of function words such as conjunctions or particles. The higher the intensity of language contact, and thus the pervasiveness of bilingualism in the community of speakers, the greater the likelihood becomes of what Thomason and Kaufman call structural borrowing. Structural borrowing pertains to phenomena that describe changes in the grammatical organization of the recipient language, including new word order patterns, new constraints on syllable structure, or the adoption of inflectional affixes. Heine and Kuteva (2005) discuss an example of structural borrowing in Pennsylvania German documented by Enninger (1980, p. 344). Under the influence of American English, the perfect auxiliary *havve* ‘have’ has been generalized in Pennsylvania German to contexts in which the perfect would originally have been formed with *sei* ‘be’. This yields English-like

structures such as *was geschehne hott* ‘what happened has’, instead of *was geschehne iss* ‘what happened is’. For the purposes of the present article, it can be stated that even casual language contact is sufficient to trigger borrowing and the lexicalization of material from a contact language.

### 2.8 Is Word Formation the Same as Lexicalization?

The processes of word formation that were discussed in this section, alongside the process of borrowing, are necessary prerequisites for lexicalization insofar as they provide the material that can be added to the lexicon of a language. Following Brinton and Traugott (2005, p. 34), this article will maintain a terminological distinction between word formation and lexicalization in order to differentiate between the creation of new open-class elements and the processes by which they become units that are holistically processed and permanently represented in the lexicon. Section 3 discusses why it is useful to distinguish between these two aspects.

## 3. From Analytic Processing to Holistic Processing

---

The main purpose of this section is to discuss in more detail the role of language processing for lexicalization. The definition of lexicalization that was adopted at the end of section 1 involves the notion of holistic processing as a necessary criterion. In order for a word to be processed holistically, it needs to be represented as a single unit in the lexicon. This is necessarily the case for established monomorphemic words such as *lamp* or *cable*. By contrast, if a hearer encounters a word such as *interdependently* for the first time, the word will have to be processed analytically, through decomposition of the word into its morphological components. For many complex words, both processing strategies are in fact available. A word such as *quickly* can of course be analyzed into the base *quick* and the derivational suffix *-ly*, but at the same time, speakers of English will be highly familiar with the word as a unit. The adverb *quickly* is a well-established word, and it even outranks its base *quick* in terms of frequency in corpora such as the British National Corpus. Schreuder and Baayen (1995) developed a psycholinguistic model that presents the alternative between analytic and holistic processing of complex words as a race of two routes that are activated in parallel. Given the relatively higher frequency of the complex form vis-à-vis its base, the case of *quickly* would be one in which the holistic processing route has an advantage. Brinton and Traugott (2005) use the term *institutionalization* to refer to the process by which words establish themselves as conventional expressions within a community of speakers. They point out that this term “is sometimes regarded as a precursor of lexicalization, sometimes identified with it” (2005, p. 45). A complex institutionalized element such as *quickly* may remain fully amenable to analytic processing as long as its form and meaning are compatible with a productive, transparent word formation process.

A reliable diagnostic of lexicalization is the presence of non-compositional meanings in a complex linguistic unit, as for instance in the word *holiday* that was previously discussed. To add another illustration, the word *hopefully* is conventionally used with the meaning ‘it is to be hoped’, which substantially goes beyond its compositional meaning ‘in a hopeful manner’, which could be derived

from analytic processing. The development of non-compositional meanings can proceed gradually when contextual meanings are increasingly associated with a linguistic unit, until that association is completely conventionalized (cf. Bybee, 2010, p. 205).

Another process that often accompanies the lexicalization of complex words is the obliteration of morpheme boundaries. To a speaker of present-day English without any training in historical linguistics, the word *gospel* is opaque, which means that there are no recognizable morpheme boundaries. If the etymology of the word is traced back to its Old English roots, the expression *gōd spell* ‘good tidings’ is however fully analyzable. Like the development of non-compositional meanings, the gradual loss of morpheme boundaries is a consequence of repeated language use. If a complex unit is used with non-compositional meaning to a sufficient extent, analytic processing is a strategy that yields fewer and fewer communicative successes and is eventually abandoned.

A third diagnostic that yields clear evidence of lexicalization is univerbation, that is, the fusion of formerly independent linguistic units into a single word. The example of *gospel* that was just discussed shows univerbation; words such as *getaway*, *holiday* or *stick-to-your-ribs* illustrate the general principle. Univerbation typically goes along with changes in segmental or suprasegmental pronunciation, which then negatively affect analyzability and thus promote holistic processing.

A fourth sign of lexicalization that is visible in the surroundings of the lexicalized item rather than in the item itself is its spread to new semantic and morpho-syntactic contexts. Himmelmann (2004, p. 36) proposes the term *context expansion* for this phenomenon. To illustrate, the lexicalization of *holiday* shows itself in the fact that it can appear in contexts such as *We’re on holiday*, which exclusively select for the lexicalized meaning, but not the source meaning. The lexicalization of *getaway* is revealed by its acquisition of nominal distributional features, such as the possibility of adjectival modification or pluralization.

These four processes can be observed not only in the development of words, but also of multi-word units. For example, the adjectival expression *stick-to-your-ribs*, which is used to describe foods that are tasty, satisfying, and high in calories, exhibits non-compositional meaning: it does not refer to anything literally sticking to someone’s ribs. Expressions such as *whatchamacallit* ‘what you may call it’ show univerbation and an obliteration of morpheme boundaries. Both collocations have acquired new distributional characteristics, so that for example the phrase *a sticky savory stick-to-your-ribs goodness* shows a formerly verbal structure in a new morphosyntactic context.

What unites the four processes is that all of them increase the idiosyncrasy of the linguistic units that they affect. Items that undergo these processes become less predictable and acquire characteristics that have to be memorized as such by language learners. Memorization and routinization ultimately yield the outcome of holistic processing. Taken together, these processes account for the fact that lexicalized words typically have features of form or meaning that make them more than the mere products of word formation processes. The pervasiveness of these idiosyncrasies motivates the distinction of word formation and lexicalization that Brinton and Traugott (2005) advocate.

## 4. Differentiating Between Lexicalization and Grammaticalization

---

Earlier, it was argued that not all new elements that emerge in language constitute cases of lexicalization. The term *grammaticalization* captures a range of phenomena that are commonly contrasted with lexicalization. Hopper and Traugott (2003) define grammaticalization as “the change whereby lexical terms and constructions come in certain linguistic contexts to serve grammatical functions, and, once grammaticalized, continue to develop new grammatical functions.” The central aspect of this definition is the term *grammatical functions*. The definition applies to cases such as the English *be going to* construction, in which the motion verb *go* has acquired the grammatical function of future time reference. Grammatical functions, such as tense, definiteness, or voice, can be described as basic conceptual distinctions. Tense is concerned with the distinction of something happening now, in the past, or in the future. Definiteness affords a contrast between an identifiable referent and generic reference. Voice captures the difference between acting and being acted upon. Very roughly then, grammatical functions are basic, schematic meanings that are conveyed by grammaticalized constructions. Schematicity is understood here as pertaining to an expression’s “level of specificity, that is, the fineness of detail with which something is characterized” (Langacker, 1987, p. 132). Lexicalized units differ from grammaticalized constructions with regard to the specificity of their meaning. While grammaticalization usually leads to more general meanings, lexicalization leads to an outcome of specific meanings. For example, the English plural suffix *-s* in *cats* conveys the idea of ‘more than one’, which is rather schematic. By contrast, the lexical item *pistachio* carries a highly specific meaning. Specificity of meaning is an important difference between lexical and grammatical elements, but it is only one aspect. Lehmann (2002, p. 1) points out further differences that relate to language processing:

The essential difference between grammar and lexicon is the following: The grammar is concerned with those signs which are formed regularly and which are handled analytically, while the lexicon is concerned with those signs which are formed irregularly and which are handled holistically. A sign is lexicalized if it is withdrawn from analytic access and inventorized.

This quotation needs some unpacking. Under Lehmann’s definition, signs that are formed regularly and handled analytically include inflectionally formed constructions such as the regular English past tense, which combines a verbal stem with the *-ed* suffix. They also include periphrastic constructions such as the English perfect, which combines a form of the verb *have* with a past participle of a lexical verb. On a more abstract level, also a word order pattern such as the English *wh*-cleft construction, as in *What I want is a desk and a chair*, is a complex sign that speakers form regularly and handle analytically.

Lehmann’s (2002) distinction between analytic and holistic processing invites the idea of a clear dichotomy between grammatical elements that are processed as configurations with parts and lexical elements that are processed as units. Naturally, the picture is more complex than that. Speakers arguably have holistic representations of many grammatical words, such as *these*, *because*,

*him*, or *could*, which makes these elements part of the lexicon. Conversely, while idioms such as *Make hay while the sun shines* are lexicalized units that are typically processed holistically, speakers are able to discern the parts that these expressions consist of, so that they may spontaneously produce modified variants of the idiom, such as *Republicans strained in their attempts to make political hay while Democrats were obsessive in their defensive blocking* (Barlow, 2000, p. 337).

So far, lexicalization and grammaticalization have been contrasted with regard to degrees of specificity of meaning and with regard to holistic and analytic processing. While these aspects show clear differences between the two, it is important to keep in mind that lexicalization and grammaticalization actually involve some of the same underlying processes. Non-compositional meanings emerge in grammaticalization just as in lexicalization, as is illustrated by the development of future time reference in the grammatical *be going to* construction. Univerbation and the obliteration of morpheme boundaries can be seen in its contracted variant, as in *He's gonna come back*. Changes in distribution and morphosyntactic behavior are common corollaries of grammaticalization, as for example in the development of the English perfect auxiliary *have*, which has entered a syntagmatic relation with a participial verb form (Carey, 1994). Traugott and Trousdale (2013) therefore argue that the distinction of lexicalization and grammaticalization should be re-thought in terms of a perspective that integrates these processes into a joint theoretical framework, namely that of Construction Grammar (Goldberg, 1995, 2003, 2006; Hilpert, 2014). A basic commonality of lexicalization and grammaticalization is that both are concerned with the emergence of new linguistic elements. For this, Traugott and Trousdale propose the term *constructionalization*, which they define as follows (2013, p. 22):

Constructionalization is the creation of form<sub>new</sub>-meaning<sub>new</sub> (combinations of) signs. It forms new type nodes, which have new syntax or morphology and new coded meaning, in the linguistic network of a population of speakers. It is accompanied by changes in degree of schematicity, productivity, and compositionality. The constructionalization of schemas always results from a succession of micro-steps and is therefore gradual. New micro-constructions may likewise be created gradually, but they may also be instantaneous. Gradually created micro-constructions tend to be procedural, and instantaneously created micro-constructions tend to be contentful.

This definition recasts lexicalization as based on the instantaneous creation of contentful signs (which, as has been discussed in section 3, subsequently undergo a series of gradual processes). By contrast, grammaticalization is presented as the constructionalization of schemas, which emerge gradually and have procedural meanings, that is, meanings that signal to the hearer how the contentful signs in an utterance are related to each other and to the speech situation (Traugott & Trousdale, 2013, p. 12). The definition further mentions changes in degree of schematicity, productivity, and compositionality. Traugott and Trousdale (2013, p. 164) associate the constructionalization of grammatical elements with increases in schematicity and productivity and decreases in compositionality. The constructionalization of lexical elements, by contrast, shows decreasing schematicity, productivity, and compositionality. This view integrates the earlier, meaning-based definition of grammaticalization offered by Hopper and Traugott (2003) with Lehmann's (2002) characterization of grammatical signs as units that are formed regularly (i.e.,

productively) and handled analytically (i.e., through a schema). What Traugott and Trousdale's new definition adds is that the meanings of grammatical signs are not only relatively more abstract, but that they are procedural in nature: grammatical meanings allow hearers to make conceptual connections between the different elements in an utterance.

In summary, lexicalization and grammaticalization can be contrasted with regard to degrees of specificity of meaning, the distinction of contentful and procedural meaning, the degree of regularity with which new expressions can be produced, and the distinction between analytic and holistic processing.

## 5. Typological Differences in Lexicalization Patterns

---

Lexicalized elements do not encode only meanings such as 'big', 'move', or 'now', which are conceptually simple and general (Wierzbicka, 1996). Quite commonly, the targets of lexicalization are elements with meanings that unite several quite specific conceptual components. In compounds, such as English *treehouse*, this is clearly in evidence, but even monomorphemic words such as *jog*, *uncle*, or *blonde* are, upon closer inspection, carriers of multiple components of meaning. Interestingly, cross-linguistic differences can be observed with regard to the lexicalization of semantically complex elements. A classic example of such a difference, which is discussed by Talmy (1985), Slobin (1987), and a sizable literature that further developed their insights (see, e.g., Goschler & Stefanowitsch, 2013), concerns verbs that express motion, such as the previously mentioned verb *jog*. Breaking down the meaning of *jog* into its different components, it becomes apparent that the verb encodes the idea of movement through space, but also the manner in which this movement is executed. Other motion verbs in English show the same lexicalized combination of motion and manner: *run*, *roll*, *slide*, *stumble*, *wade*, *gallop*, *crawl*, and many others illustrate a vocabulary of motion verbs that simultaneously encode manner of movement. Crucially, this lexicalization pattern is not equally common across the world's languages. Talmy (1985) distinguishes languages, such as English, that typically conflate motion and manner, from languages, such as Spanish, where conflation of motion and manner is relatively rare, but where another lexicalization pattern is prevalent, namely the conflation of motion and path, as in *entrar* 'move into', *salir* 'move out of', *pasar* 'move past', *subir* 'move up', and other verbs. For this typological split, Talmy (1991, p. 485) proposes two labels that make reference to the lexicalized form in which the path component of a motion event is expressed. Spanish, other Romance languages, and languages such as Hebrew, Turkish, and Korean tend to lexicalize the path in the verb and are hence called verb-framed languages. When speakers of English want to express the path of a motion event, they typically use an additional element that is external to the verb, which Talmy (1985, p. 57) calls a satellite. The preposition *around* in *We jogged around the campus* is such an element. Languages such as English, and other, non-Romance languages of the Indo-European family are labeled satellite-framed languages. It goes without saying that the distinction between the two types is not absolute. For example, Benincà and Poletto (2005) document the presence and use of different types of phrasal verbs in northern varieties of Italian. Rather than making a crisp, categorical distinction, the typological split describes probabilistic tendencies of lexicalization. Language contact complicates matters even further: English has borrowed multiple Romance

motion verbs in which motion and path are conflated, so that the combination of these aspects of meaning is far from alien to its speakers. The general point is that languages may differ with regard to the components of meaning that are typically lexicalized together in a single element.

### Further Reading

Many of the ideas presented in this article draw on the discussion in Lehmann (2002). Brinton and Traugott (2005) offer a comprehensive overview of research on lexicalization and the different ways in which it has been related to grammaticalization. Their book also covers related concepts such as degrammaticalization that were not treated in this article. Himmelmann (2004) presents another comparison of lexicalization and grammaticalization. Traugott and Trousdale (2013, chapter 4) discuss lexicalization from a constructionist perspective. A general introduction to Construction Grammar is offered by Hilpert (2014). *The Handbook of Morphology* (Spencer & Zwicky, 1998) contains detailed discussions of inflectional morphology (Stump, 1998) and derivational morphology (Beard, 1998). Štekauer and Lieber (2005) and Müller, Ohnheiser, Olsen, and Rainer (2016) offer comprehensive overviews of word formation processes. Filipovic and Ibarretxe-Antunano (2015) present an introduction to lexicalization patterns in the domain of motion.

### References

- Akmajian, A., Farmer, A. K., & Bickmore, L. (2017). *Linguistics: An introduction to language and communication* (7th ed.). Cambridge, MA: MIT Press.
- Barlow, M. (2000). Usage, blends and grammar. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 315–345). Stanford, CA: CSLI.
- Bauer, L. (1983). *English word formation*. Cambridge, UK: Cambridge University Press.
- Beard, R. (1998). Derivation. In A. Spencer & A. M. Zwicky (Eds.), *The handbook of morphology* (pp. 44–65). Oxford, UK: Blackwell.
- Benincà, P., & Poletto, C. (2005). Phrasal verbs in Venetian and regional Italian. In F. Hinskens (Ed.), *Language variation – European perspectives: Selected papers from the third international conference on language variation in Europe (ICLaVE 3)* (pp. 9–22). Amsterdam, The Netherlands: John Benjamins.
- Brinton, L. J., & Traugott, E. C. (2005). *Lexicalization and language change*. Cambridge, UK: Cambridge University Press.
- Bybee, J. L. (2010). *Language change*. Cambridge, UK: Cambridge University Press.
- Carey, K. (1994). The grammaticalization of the perfect in Old English: An account based on pragmatics and metaphor. In W. Pagliuca (Ed.), *Perspectives on grammaticalization* (pp. 103–117). Amsterdam, The Netherlands: John Benjamins.
- Enninger, W. (1980). Syntactic convergence in a stable triglossia plus trilingualism situation in Kent County, Delaware, U.S. *Zeitschrift für Dialektologie und Linguistik, Beiheft*, 32, 343–350.

- Filipovic, L., & Ibarretxe-Antunano, I. (2015). Motion. In E. Dabrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 527–546). Berlin, Germany: De Gruyter.
- Goldberg, A. E. (1995). *Constructions. A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.
- Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5), 219–224.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford, UK: Oxford University Press.
- Goschler, J., & Stefanowitsch, A. (Eds.). (2013). *Variation and change in the encoding of motion events*. Amsterdam, The Netherlands: John Benjamins.
- Gries, S. T. (2004). Shouldn't it be breakfunch? A quantitative analysis of the structure of blends. *Linguistics*, 42(3), 639–667.
- Haspelmath, M., & Sims, A. (2010). *Understanding morphology*. London, UK: Hodder.
- Heine, B. & Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge, UK: Cambridge University Press.
- Hilpert, M. (2014). *Construction grammar and its application to English*. Edinburgh, U.K.: Edinburgh University Press.
- Himmelmann, N. (2004). Lexicalization and grammaticization: opposite or orthogonal? In W. Bisang, N. P. Himmelmann, & B. Wiemer (Eds.), *What makes grammaticalization—A look from its components and its fringes* (pp. 21–42). Berlin, Germany: Mouton de Gruyter.
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Janda, R. D. (2001). Beyond 'pathways' and 'unidirectionality': On the discontinuity of language transmission and the counterability of grammaticalization. *Language Sciences*, 23, 265–340.
- Katz, L., Rexer, K., & Lukatela, G. (1991). The processing of inflected words. *Psychological Research*, 53(1), 25–32.
- Langacker, R. W. (1987). *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Lehmann, C. (2002). New reflections on grammaticalization and lexicalization. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 1–18). Amsterdam, The Netherlands: John Benjamins.
- Libben, G. (2008). How do we parse compound words? In N. Srinivasan, A. K. Gupta, & J. Pandey (Eds.), *Advances in cognitive science* (pp. 154–172). New Delhi, India: SAGE.
- Müller, P. O., Ohnheiser, I., Olsen, S., & Rainer, F. (Eds.). (2016). *Word-formation: An international handbook of the languages of Europe*. Berlin, Germany: De Gruyter.
- Norde, M. (2001). Deflexion as a counterdirectional factor in grammatical change. *Language Sciences*, 23, 231–264.
- Plag, I. (2003). *Word-formation in English*. Cambridge, UK: Cambridge University Press.

- Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale, NJ: Lawrence Erlbaum.
- Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory and Cognition*, 25, 425–437.
- Slobin, D. I. (1987). Thinking for speaking. *Berkeley Linguistic Society*, 13, 435–445.
- Spencer, A., & Zwicky, A. M. (Eds.). (1998). *The handbook of morphology*. Oxford, UK: Blackwell.
- Štekauer, P., & Lieber, R. (Eds.). (2005). *Handbook of word-formation*. Dordrecht, The Netherlands: Springer.
- Stump, G. (1998). Inflection. In A. Spencer & A. M. Zwicky (Eds.), *The handbook of morphology* (pp. 13–43). Oxford, UK: Blackwell.
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description 3: Grammatical categories and the lexicon* (pp. 57–149). Cambridge, UK: Cambridge University Press.
- Talmy, L. (1991). Path to realization—Via aspect and result. *Berkeley Linguistics Society*, 17, 480–519.
- Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Traugott, E. C., & Trousdale, G. (2013). *Constructionalization and constructional changes*. Oxford, UK: Oxford University Press.
- Wanner, A. (2009). *Deconstructing the English passive*. Berlin, Germany: Mouton de Gruyter.
- Wierzbicka, A. (1996). *Semantics: Primes and universals*. Oxford, UK: Oxford University Press.