

Four kinds of lexical items: Words, lexemes, inventorial items, and mental items

Quatre types d'éléments lexicaux : mots, lexèmes, éléments inventoriaux et éléments mentaux

Martin Haspelmath

Max Planck Institute for Evolutionary Anthropology

martin.haspelmath@eva.mpg.de

Abstract

This paper gives an overview of four senses of the terms “lexical (item/entity)” and “lexicon”, as well as several senses of the term “lexicalization”. That these terms are used in different senses in the literature has been discussed before, and it has been noted that this polysemy is sometimes confusing, but here I provide not only concrete definitions of *word(-form)* and *lexeme* and succinct discussion of the relevant issues, but I also propose two new terms: INVENTORIUM (the unpredictable elements of a language) and MENTALICON (the elements that a speaker stores in memory). The latter two are crucially different because all speakers store many predictable elements. The four different senses can thus be distinguished clearly by using the four terms *word-form*, *lexeme*, *inventorial item*, and *mental item*. In the final section of the paper, I note that the term *lexicalization* also has multiple senses, but its most important sense is *inventorization*.

Keywords: terminology, word-form, lexicon, lexeme, lexicalization

Résumé

Cet article donne un aperçu des quatre sens des termes « (élément) lexical » et « lexique », ainsi que de plusieurs sens du terme « lexicalisation ». Le fait que ces termes soient utilisés dans des sens différents dans la littérature a déjà été discuté, et il a été noté que cette polysémie est parfois déroutante, mais ici je fournis non seulement des définitions concrètes de *mot(-forme)* et de *lexème* et une discussion succincte des termes pertinents problématiques, mais je propose également deux nouveaux termes : INVENTORIUM (les éléments imprévisibles d'un langage) et MENTALIQUE (les éléments qu'un locuteur stocke en mémoire). Ces deux derniers sont fondamentalement différents car tous les locuteurs stockent de nombreux éléments prévisibles. Les quatre sens différents peuvent ainsi être clairement distingués en utilisant les quatre termes *forme de mot*, *lexème*, *élément inventorial*, et *élément mental*. Dans la dernière section de l'article, je note que le terme *lexicalisation* a également de multiples sens, mais son sens le plus important est celui d'« inventorisat[i]on ».

Mots-clés: terminologie, mot-forme, lexique/lexicon, lexème, lexicalisation

1. Overview

This paper points out the multiple ambiguity of the terms *word* or *lexical item* (or *lexical entity*) in linguistics and psycholinguistics, and proposes clear terms for four widely needed senses:

- lexical entity as WORD-FORM (or simply WORD) (Section 3)
- lexical item as LEXEME (an abstract element based on a root) (Section 4)
- lexical item as INVENTORIAL ITEM (an item of the *inventorium*) (Section 5)
- lexical item as MENTAL ITEM (an item of the “mental lexicon” or *mentalicon*) (Section 6)

This fourfold polysemy also extends partially to the term *lexicalization*, as will be discussed in Section 7.

It is clear that all four concepts are needed in linguistics, so that they are complementary, not in competition as might be thought. Similarly, the term *lexicon* has been used in four different senses, corresponding to the four kinds of lexical entities:

- lexicon as the component of grammar that creates word-forms
- lexicon as the component of grammar that stores and creates lexemes
- lexicon as the INVENTORIUM, i.e., the set of unpredictable elements of a language
- lexicon as the MENTALICON, i.e., the set of memorized elements (the “mental lexicon”)

We often read that there are different views or conceptions of the lexicon, as in the following quotation:

“The nature, structure, and role of the lexicon in the grammar of natural languages has been a subject of debate for at least the last 50 years.” (Davis & Koenig, 2021, p. 125)

However, in view of the polysemy of the terms *lexicon*, *lexical (item/entity)*, and *word*, it appears that these debates have often been based on misunderstandings rather than on incompatible views or competing theories. In many cases, linguists have simply talked past each other rather than disagreed in substance.

In the present paper, I provide discussions of the above four senses of *word*, *lexical*, and *lexicon*, and I suggest the two new terms INVENTORIUM (for the set of unpredictable elements of a language) and MENTALICON (for the “mental lexicon”). I will not make any controversial statements about the nature of the four kinds of lexical entities or on the nature of the lexicon in any of the four senses. It seems to me that all the concepts are widely agreed upon, though different linguists are of course interested in them to different degrees. It is true that some linguists might argue that one or more of them is an artificial concept that is not really needed because it does not directly mirror the reality of languages and language. So, when I say that the four concepts are uncontroversial, what I mean is that they have all been used by linguists

(and will probably be used in the future by many), and that it is very largely clear what they imply, even if linguists may disagree about the importance of the different concepts.

That the terms *lexicon* and *lexical* have several rather different meanings has been noted for quite some time, most prominently by Aronoff (1988; 1994) and Jackendoff (1995; 2002; 2013). The current paper thus does not claim originality, except perhaps for the last section: In Section 7, I will discuss the term *lexicalization*, which is associated with even more than four senses, because it has been used both for a synchronic concept (*lexification*) and for several diachronic concepts (especially *lexemization* and *inventorization*). Again, the substance of what I will say is not controversial, as far as I can see, so that the main contribution of this paper consists in clarifying the concepts that linguists use. Aronoff (1988, p. 10) feels that “we have been fooled by our own terminology”, and Cappelle (2022, p. 185) even thinks that as a result of linguists’ different understanding of their basic concepts, “linguistic theory has been in a state of turmoil in the last half-century”. While our deep questions about the nature of the language system will not find a definitive answer anytime soon, we may perhaps make a little progress by becoming more aware of terminological polysemy and by enriching our technical vocabulary.

2. The “lexical” stereotype

The term *lexical* can be said to be associated with the four stereotypical properties in (1), deriving from the way dictionaries have been designed in Western linguistics.¹ A lexical item as found in a dictionary can be expected (a) to be written solid between spaces (i.e., an orthographic word), (b) to be contentful (denoting an object, action or property), (c) to constitute unpredictable information, and (d) to be memorized by language users. If all these conditions are fulfilled, an expression is definitely expected to be included in a dictionary.

- | | |
|--|----------------------------------|
| (1) a. solid between spaces | (= word-forms, Section 3) |
| b. contentful (vs. semantically poor or empty) | (= lexemes, Section 4) |
| c. idiosyncratic (vs. predictable) | (= inventorial items, Section 5) |
| d. memorized (vs. constructed online) | (= mental items, Section 6) |

But these four properties do not always go together: Semantically poor elements may be orthographic words (e.g., prepositions like *at* or *by*); contentful words may be predictable (e.g., *driver*, predictable from *drive* and *-er*); idiosyncratic expressions may have internal spaces rather than being written solid (e.g., idioms like *take part*); and predictable combinations may be memorized (e.g., high-frequency phrases such as *that’s a problem*).

¹ The word *lexicon* is the Greek counterpart of Latin *dictionarium*, and it was used by 19th century classicists for Latin and Greek dictionaries. The term *lexicon* as used by linguists as a technical term (following Bloomfield, 1933) thus derives from a special (particularly prestigious) expression for “dictionary”.

As Aronoff (1988) noted, Bloomfield's (1933, p. 162) original notion of *lexicon* included only the morphs of the language, leaving aside unpredictable morph combinations, and ignoring the distinction between idiosyncratic and memorized expressions. But subsequently, complex idiosyncratic expressions were added (e.g., idiomatic compounds like *honey-moon*), and since the 1970s, even regular derived forms were often said to be “lexical”, so that the lexicon came to be thought of as including (some) morphological rules (e.g., Jackendoff, 1975; Bybee, 1995; Scalise & Guevara, 2005). The idea of “lexical rules” may seem to be an odd development, but after Bloomfield's idea of morphs as the basic building blocks was given up, the stereotypical syntactic tree contained “words” as its leaves (its ultimate constituents, added to the tree by “lexical insertion”), and these were often regularly derived.

Thus, it appears that the diverse senses of the terms *lexicon* and *lexical* are due to the different properties associated with stereotypical dictionary entries. If the four properties seen in (1a-d) are kept apart carefully, we can distinguish clearly between four different types of “lexical” elements, as we will see in the following sections. We begin with *word(-form)* in Section 3, and then move on to *lexeme* (Section 4), *inventorial item* (Section 5), and *mental item* (Section 6).

3. Word (or word-form)

Perhaps the most basic sense of *word* is “string of characters” in written texts of European languages like Latin or English, and linguists sometimes use *grammatical word* (or *morphosyntactic word*) as a general concept that largely corresponds to this.² For greater precision, I use WORD-FORM, a term that has become more widespread in morphology over the last two decades (Haspelmath, 2002 p. 13; Booij, 2012, p. 3; Spencer, 2013, p. 1).³ So in a sentence like *My dog likes other dogs*, there are five words in the sense of word-forms, among them two different word-forms (*dog, dogs*) that both belong to the lexeme DOG.

Linguists typically make a distinction between syntactic rules and morphological rules, and they commonly treat word-forms as the leaves of the syntactic tree. Word-forms are generally thought of as “syntactic atoms” (Di Sciullo & Williams, 1987, p. 1), and they are treated as the normal outcomes of rules of inflection. Plural forms (such as *dog-s*), past-tense forms (such

² The term *grammatical word* is used particularly in contexts where it is contrasted with the “phonological word” (e.g., Hall, 1999; Aikhenvald et al., 2020). I do not discuss phonological words here, although one might say that they represent yet another type of “lexical” element (see approaches such as “Lexical Phonology”, e.g., Mohanan, 1986). Other terms equivalent to *grammatical word* that have sometimes been used are *morphological word* and *syntactic word* (Haspelmath, 2011, p. 37).

³ It seems that this term was ultimately adopted from Russian linguistics (Russian *slovoforma*); see, e.g., Zaliznjak (1967, p. 2), Mel'čuk (2006, p. 20). An alternative spelling is *wordform* (Gebhardt, 2023, p. 82). Note that Matthews (1991, p. 30) also talks about “word-forms”, but in a different sense: By this term, he means the shape of a grammatical word or word-form.

as *like-d*) or genitive-case forms (such as *my*) are normally word-forms. In situations where a grammatical meaning of the inflectional type is expressed “not within a word but by a syntactic phrase” (Chumakina, 2013, p. 1), we use the special term *periphrasis*. For example, the English Present Perfect may be said to express an “inflectional meaning”, but forms like *has done* are periphrastic and consist of two word-forms.

Word-forms have often been thought to be “lexical” in the sense that they are created “in the lexicon” (as opposed to “in the syntax”), and they have been said to exhibit “lexical integrity” (inaccessibility of words to syntactic rules). For example, according to Bresnan (2001, p. 92), the principle of lexical integrity means that “morphologically complete words are leaves of the c[onstituent]-structure tree and each leaf corresponds to one and only one c-structure node” (see also O’Neill, 2016, p. 240 for some historical discussion of lexical integrity).

Of course, the spelling system of a language does not determine what its word-forms are, so we need an independent way of identifying word-forms (or grammatical words). Many linguists seem to think that no definition with sharp boundaries can be given, and that “word-form” can only be defined as a prototype. However, the method of using a battery of symptoms or diagnostics cannot be applied if we do not know that the “word-form” has an underlying reality as a natural kind (as an innate building block of grammar). If the word category is not innately given, the method of using test batteries will perhaps only identify a stereotype that is based on our spelling habits.

Thus, in Haspelmath (2023a), I proposed a new definition of *word(-form)*, given in (2). The forms singled out by this definition corresponds fairly closely to what we intuitively regard as a word-form.

(2) word (or word-form)

A word is (i) a free morph, or (ii) a clitic, or (iii) a root or compound, possibly augmented by nonrequired affixes and augmented by required affixes if there are any.

This definition is very complex, making it clear that ‘word’ is an artificial concept that is unlikely to reflect the reality of languages. Still, it is useful to have this definition because linguists very frequently use the term *word(-form)* in the sense given in (2). The idea here is that it is much better to have an explicit definition than to think about words in terms of a vague stereotype, because it would be all too easy to take the ‘word’ notion as an established scientific concept on which further theorizing can rely (e.g., about morphology and syntax as two different levels or modules). The complex and unnatural definition in (2) suggests that it may not be advisable to build theories (such as lexicalism) on the notion of ‘word-form’.

In (3a-e) and (4a-b), I give examples of different types of English and Italian expressions which are words according to the definition in (2).

(3) English

- a. *hello* (a free morph)
- b. *at* (a clitic preposition)
- c. *tree* (a root with no required affixes)
- d. *sad-ness* (a root plus nonrequired affix *-ness* with no required affixes)
- e. *peach tree* (a compound with no required affixes)

(4) Italian

- a. *alber-o* ‘tree’ (a root with required singular affix *-o*)
- b. *alber-i* ‘trees’ (a root with required plural affix *-i*)

In English, there are no (or very few) required affixes, so words are limited to free morphs, clitics, and roots or compounds possibly augmented by nonrequired affixes (as in *tree-s*, or *sad-ness*; these suffixes are not required in that they can be simply absent in free forms such as *sad* or *a tree*). Italian and many other languages have required affixes which cannot be simply absent in a free form (*alber-* always occurs with a suffix).

The adjective *lexical* often has the sense ‘relating to a word-form’, e.g., when we talk about “lexical stress” (i.e., word stress), or “lexical rules” in phonology (as opposed to postlexical rules, which apply across word-form boundaries). Word-forms are thus lexical entities in some sense, though linguists do not call them “lexical items” (this latter expression is used for lexemes, inventorial items or mental items).

4. Lexeme (or content word)

While a word-form can be informally characterized as a “text word”, a lexeme can be said to be a “dictionary word”. In morphology textbooks, this is often made clear by using a small-caps notation, as in the first paragraph of the preceding section: *dog* and *dogs* are two word-forms of the lexeme DOG. Thus, a lexeme is an abstract entity and is best seen as the set of word-forms that form an inflectional paradigm (see Haspelmath, 2024a).

Moreover, lexemes are generally thought of as belonging to the major (or contentful) word classes noun, verb or adjective. Non-inflecting function words such as English *if*, *and*, *to*, *not* are not normally called “lexemes”, even though dictionaries would list them. Function words are considered to be grammatical items and they are often called FUNCTIONAL CATEGORIES (e.g., Muysken, 2008), contrasting with LEXICAL CATEGORIES such as *dog*, *run* and *white* (e.g., Baker & Croft, 2017).⁴ Since *lexical* has multiple senses, more precise term for the latter would be “contentful categories”, or “lexemic categories”.

⁴ The term *lexical category* for the major word classes seems to be fairly young; until the 1980s, they were often called *major lexical categories* in the generative tradition, as opposed to *minor (lexical) categories* (= functional categories) (e.g., Gazdar et al., 1985, pp. 23-29).

Another related use of the adjective *lexical* is in the term “lexical noun phrase”, which is sometimes used in the sense of *full nominal*, contrasting with (*personal*) *pronoun* (e.g., *the teacher* vs. *she*). For example, Du Bois (1987) distinguished between “lexical A-arguments” and “pronominal A-arguments”.⁵ Again, this use of “lexical” may be unexpected as pronouns are listed in dictionaries, too. But like function words, they have primarily grammatical functions and are “non-lexical” in this way. The truly typical “lexemes” or “lexical items” are nouns, verbs, and adjectives, i.e., the kinds of elements that are most clearly associated with inflectional paradigms in the Indo-European languages, and that are always by convention discursively primary (Boye, 2023).

In a simplified way, we can thus define a *lexeme* as in (5).

(5) **lexeme**

A lexeme is the set of word-forms that contain the same noun root, verb root or adjective root and may only contain inflectional affixes in addition.

Eventually, we want the definition of *lexeme* to be somewhat broader, to include not only sets of root-based inflected forms (such as DOG, containing the forms *dog* and *dog-s*), but also sets of compound-based forms (e.g., FLOWER-POT, containing the forms *flower-pot* and *flower-pot-s*) and sets of forms based on a combination of a root with derivational affixes (e.g., REAL-IZE, containing the forms *real-ize*, *real-ize-s*, *real-ize-d*, and *real-iz-ing*). For details, I refer the reader to Haspelmath (2024a), where I also provide a definition of the key notion of “inflection”. This definition is artificial in a way similar to the definition of *word* in the previous section, but it can be applied uniformly across languages and thus allows us to arrive at a non-stereotypical understanding of *lexeme*.

The term *lexical* is often used in reference to lexemes, as when we talk about “lexical meanings” (as opposed to sentence meanings or grammatical meanings; e.g., Geeraerts, 2010), or “lexical case” (assigned by a specific verb lexeme; e.g., Woolford, 2006). In such contexts, it would be more precise to talk about “lexemic meanings” (studied in “lexemic semantics”), and “lexemic case”. If we wanted to refer to the set of lexemes in a language, as opposed to its grammatical rules and grammatical markers, we could call it the *lexemicon*.

However, the terms *lexicon* and *lexical* have also been used in a more technical sense that is not closely related to the notion of a lexeme: In many versions of generative grammar, beginning with Chomsky (1965), the syntactic rule component is distinct from another component called “lexicon”, whose elements are inserted into syntactic trees (“lexical insertion”). These elements are most often lexemes (e.g., Chomsky, 1970; Anderson, 1992),

⁵ In their detailed discussion of Du Bois’s claims, Haig & Schnell (2016, p. 594) note that the term *full nominal* is more transparent than Du Bois’s “lexical NP”.

though in some versions, inflected forms are thought to originate in the “lexicon”, too.⁶ But even if only simple and derived lexemes are thought to originate in the “lexicon”, this component is not merely a list but also includes “lexical rules” (e.g., Briscoe & Copestake, 1999). The notion of “lexicalism” refers to the idea that “lexical rules” are distinct (in some sense) from syntactic rules (e.g., Bresnan & Mchombo, 1995; Williams, 2007). When the term *lexical* is used in this way, referring to a presumed component of a formal grammar, it cannot be replaced by “lexemic”, even though the component is thought to be primarily responsible for lexemes.

In addition to the narrow (but widely adopted) sense of *lexeme* as defined in (5), many linguists use the term in a broader sense, to refer to all kinds of elements that are part of the inventorium. For example, Masini (2009) discusses composite expressions such as Italian *casa di cura* ‘care home’, or *carta di credito* ‘credit card’, which are neither single words nor compounds (see Haspelmath, 2024b), and she calls them “phrasal lexemes” (see also Pepper, 2023 for the notion of “binominal lexeme”). These kinds of forms are similar to (compound) lexemes in that the patterns are typically productive but the forms are often semantically idiosyncratic. Thus, while they are not lexemes in the sense given in (5), they are part of the “lexicon” in the sense of *inventorium*, as introduced in the next section.

5. Inventorial item (or element of the inventorium)

If the “lexicon” is “a list of basic irregularities” (Bloomfield, 1933, p. 274), then it cannot be limited to containing the morphs of a language. All languages contain many non-compositional and otherwise unpredictable forms that must be learned in addition to the productive rules of the language, e.g., idiomatic compounds (like *honeymoon*), idiomatic phrases (like *take part*) and other fixed expressions (e.g., famous quotations such as *may the force be with you*, from *Star Wars*). Those elements that must thus be “listed”, i.e., the morphs and the complex unpredictable forms, have been called *listemes* (Di Sciullo & Williams, 1987, p. 3; Harley, 2006, pp. 11-12), and while this term has gained some notoriety among morphologists, it has not been widely adopted. A much more commonly used term for listemes is “lexical item” (e.g., Jackendoff, 1995).

Here I suggest two new terms, INVENTORIUM (for the inventory of morphs and fixed expressions) and INVENTORIAL ITEM (for the elements of the inventorium). As there is no clear boundary between phrasal expressions, partially filled constructions, and fully schematic

⁶ It should be noted that the term *lexeme* comes from a European tradition (e.g., Lyons, 1968, p. 197; Matthews, 1972), and that it has rarely been used in the American generative tradition. *Lexeme* does not appear in Aronoff (1976), Anderson (1992) or Embick (2015) (though Aronoff, 1994, Section 1.1 admitted that he should have used *lexeme* instead of *word* in his 1976 book).

constructions (see, e.g., Croft & Cruse, 2004, p. 255; Culicover et al., 2017), the inventorium also includes the constructions of a language.

(6) inventorium

The inventorium of a language is the set of forms and constructions that cannot be predicted from other forms or constructions.

(7) inventorial item

An inventorial item is a form or construction of a language that cannot be predicted from other forms or constructions.

An inventorial item is thus exactly what Goldberg (1995, p. 4) called a *construction*: a pairing of meaning and form that is not predictable from anything else.⁷ The inventorium is thus the set of meaning-related conventions of a language that speakers or signers must know in order to use the language. If one wants to have a single-word term for elements of the inventorium, one can call them *inventoremes* in the terminological framework proposed here. An inventoreme is an element that must be listed and is thus perhaps the same as Di Sciullo & Williams's *listeme*.⁸

In their insightful discussion of the term *compound*, Gaeta & Ricca (2009) note that compounds have either been identified as being morphologically well-formed lexemes, or as belonging to the inventorium. In a stereotypical compound (e.g., English *flat-iron*), these two properties come together, but they can be dissociated: Languages may have “compound-like” non-lexemic inventorial items (e.g., Italian *ferro da stiro* [iron for flattening]), or “compound-like” non-inventorial lexemes such as German *Entscheidungs-ort* ‘decision place’ (which has no idiosyncratic meaning). Thus, it is not sufficient to use the term “compound” in a stereotypical meaning, and the term “lexical” (used by Gaeta & Ricca) is not sufficient to distinguish the relevant senses. As we saw in Section 4, regular complex lexemic patterns have been called “lexical”, too.

⁷ However, in later work, Goldberg changed her definition of *construction*; and as discussed in Haspelmath (2023b), it is better to reserve the term *construction* for inventorial items that include an open slot, i.e., not to treat morphs as constructions.

⁸ Goldberg (1995) observed that her notion of *construction* corresponds closely to Di Sciullo & Williams's notion of *listeme*. It should be noted, however, that as generative linguists, Di Sciullo & Williams are not interested in language systems as sets of social conventions, so they do not distinguish between the inventorium and the mentalicon (see Section 6). On the one hand, they introduce listemes as the kinds of elements that “must be memorized” (i.e., that are part of the conventions of a language), but on the other hand, they say that listemes are “memorized objects” (Di Sciullo & Williams, 1987, p. 3). They thus simply ignore the fact that all speakers memorize many forms that need not be memorized because they are not idiosyncratic and could alternatively be constructed from other conventions of the language.

6. Mental item (or element of the mentalicon)

Finally, inventorial items need to be distinguished from mental items. Jackendoff (2013, p. 74) notes that the term “lexical item” can be understood in three ways: (a) for the words (and morphemes) of the language [= word-forms or lexemes], (b) for the exceptional, idiosyncratic parts of the language that cannot be predicted by rule [= inventoremes], and (c) for the parts of a language that are listed in long-term memory. He goes on to say that the third option is preferred in his approach:

“Criteria (a) and (b) are traditional views of the lexicon, in which it is a component of language altogether distinct from rules of grammar. The Parallel Architecture, however, adopts the more psycholinguistically based definition (c).” (Jackendoff, 2013, p. 74)

However, different linguists may simply have different research interests, and different “approaches” may be fully compatible. One may choose to focus on the morphs of a language (as Bloomfield did originally), or on the words (Section 3) or the lexemes (Section 4) of a language, or on the unpredictable elements of a language (Section 5). Jackendoff’s own interests are primarily in languages as mentally represented knowledge systems, and so it makes good sense for him to focus on elements stored in memory, but this is not necessary. Languages can be viewed not only from a cognitive perspective, but also from a social and cultural perspective, as sets of conventions obeyed by its users. If one adopts Jackendoff’s psycholinguistic point of view, one could use the term *mental item* in (9) as an unambiguous replacement for the vague term “lexical item”. It is defined as an element of a language user’s *mentalicon* as defined in (8), more often known as the “mental lexicon”.

(8) *mentalicon*

A language user’s *mentalicon* of a language is the set of forms and constructions that they have stored in their long-term memory.

(9) *mental(ic) item*

A *mental(ic) item* is an element of a language user’s *mentalicon*.

An important advantage of *mentalicon* over “mental lexicon” is that the *mentalicon* includes not only word-like elements, but also fully schematic and partially filled constructions, as Jackendoff emphasizes. Mental items are perhaps better called *mentalic items*, because people store a lot of nonlinguistic information, so a term like “mental item” is not really specific enough.⁹

⁹ On the analogy to *inventoreme* (an inventorial item), one could coin the term *mentaleme* for a mentalic item. Recall the term *listeme* from Di Sciullo & Williams (1987), which is either an inventoreme or a mentaleme, as observed in the previous footnote.

Jackendoff is by no means alone in taking a psychological or cognitive perspective. Since the 1960s, most theoretical linguists have adopted an implicitly (or explicitly) cognitive perspective, but the issues that psycholinguists have discussed are often somewhat different from the issues discussed by theoretical linguists. Emmorey & Fromkin (1988, p. 124) noted that

“It is an open question whether there is an isomorphism between the units and components of the grammar and those implemented and accessed in the linguistic processing system(s), even though a linguistic performance model is ultimately dependent on the grammar.”

The dependence of psycholinguists on theoretical linguists is also a terminological one: While theoretical linguists have often taken pains to distinguish different terms and concepts (e.g., Bauer, 1983, Chapter 1; Carstairs-McCarthy, 2005), psycholinguists have not developed their own concepts or terms and have relied on the concepts developed by grammarians. When they talk about the “mental lexicon” (e.g., Aitchison, 1987) or “lexical access” (e.g., Levelt et al., 1999), what they seem to mean is mentally represented lexemes or word-forms, or perhaps everything that is stored, i.e., a speaker’s entire mentalicon.

7. Four diachronic senses of “lexicalization”

Now that we have seen four different senses of *lexical* and *lexicon* in synchronic grammar, we can move on to the diachronic dimension and to the widely used polysemous term *lexicalization*. While Brinton & Traugott (2005, p. i) say that “lexicalization has been conceptualized in a variety of ways”, here I suggest that the usage variety does not result from different views (or conceptualizations), but simply from different senses of the term *lexicalization*. Corresponding to the four senses of *lexical*, we can distinguish four different senses for *lexicalization*, as listed in (10).

- (10) “Lexicalization” as ...
- a. univervation: the transition from a word combination to a word (Section 7.1)
 - b. lexemization: the transition from an unrestricted combination to a lexeme (Section 7.2)
 - c. inventorization: the passing of an unrestricted combination into the inventorium (Section 7.3)
 - d. mentalicization: the passing of an unrestricted combination into a language user’s mentalicon (Section 7.4)

Not all of these concepts are equally current, and three of the dynamic labels (*lexemization*, *inventorization*, and *mentalicization*) are new here. But we will see that it is useful to distinguish the four senses also for the dynamic processes that often go by the name *lexicalization*. Most commonly, “lexicalization” is understood in the sense of inventorization, so in Section 7.5 below, I will ask whether the current perspective helps us understand the phenomenon of

inventorization. I will not say more about understanding univerbation and lexemization here, although these changes raise important questions, too.

In addition to the four diachronic senses, there is also a synchronic sense that is likewise fairly prominent in the literature: *lexicalization* can refer to the way in which lexical forms are mapped onto semantic elements, e.g., when we talk about directionality being “lexicalized” together with motion (as in English *enter* or *exit*), or manner together with speaking (as in English *whisper* or *shout*) (see, e.g., Talmy, 1985; Levin & Rappaport Hovav, 2019).¹⁰ For this synchronic sense, it has recently been proposed to replace it by the related term *lexification* (Haspelmath, 2023c), which is more transparent also because it makes the close link to *colexification* (François, 2008) clear.¹¹

7.1. “Lexicalization” as univerbation

For the diachronic change from a word combination to a word-form, there is a traditional term, *univerbation*, discussed recently by Lehmann (2020). The term is defined in (11), and a few examples are given in (12)-(13).

(11) **univerbation**

Univerbation is the diachronic transition from a word combination to a word-form.

- (12) a. Latin *cantare habeo* ‘I have to sing’ > Spanish *cantar-é* ‘I will sing’
 b. Greek *thélō hína hupágō* ‘I want to go’ > *thé na páo* > *tha-páo* ‘I’ll go’
- (13) a. Latin *quo modo* ‘in what way’ > *quomodo* > Spanish *cómo* ‘how’
 b. English *all one* > *alone* (Hopper, 1990, p. 154)
 c. Old English *dæges eage* ‘day’s eye’ > *daisy* (a flower, *Bellis perennis*)

The creation of a new affixal future tense in Spanish and other Romance languages is one of the classic cases of *grammaticalization* (Narrog & Heine, 2021), and this latter term is used much more frequently than *univerbation*. Of course, grammaticalization need not involve univerbation with affixed markers, but can instead lead to clitic markers, as in the case of English articles (e.g., *the* = *world*) or French past tense (e.g., *a* = *changé* ‘has changed’). When

¹⁰ The term *lexicalization* is also used in psycholinguistics in a related (but different) sense (e.g., Takashima et al., 2014).

¹¹ The term *lexicalization* has occasionally also been used for the creation of verbs or nouns out of function words (e.g., English *to up*, *ifs and buts*; Bybee, 2010, p. 113), for the integration of loanwords (Blank, 2001, p. 1606), or even simply for the formation of lexemes via lexeme-formation rules (Brinton & Traugott, 2005, Section 2.1). All these are left aside here, and I focus on univerbation, lexemization, inventorization and mentalicization.

a grammatical marker becomes an affix, the diachronic process is also called *agglutination*, using a term that goes back to the early 19th century (see Stolz, 1991; Haspelmath, 2018).¹²

In (13a-c), the result of the univerbation is not a new affix, but a new non-affix morph, because Spanish *cómo* ‘how’ and English *alone* and *daisy* can no longer be segmented. Since *daisy* is a noun (and thus a lexeme), (13c) is also an example of lexemization, as discussed in the next subsection (see 15d).

7.2. “Lexicalization” as lexemization

“Lexicalization” can also be used for a process in which a new lexeme-stem arises, thus giving rise to a new lexeme. For this sense, we can create a new term that specifically refers to the development of a lexeme, defined as in (14). A few examples are given in (15).

(14) **lexemization**

Lexemization is the diachronic change from an unrestricted combination to a lexeme-stem.

- (15) a. Latin: *com-ed-ere* ‘eat up’ > Spanish *com-er* ‘eat’
 b. Ancient Greek: *paid-íon* ‘little child’ > Modern Greek *peḗí* ‘child’
 c. Old Russian: *vŭz-ŭm-u* [up-take-1SG] > Russian *voz’m-u* ‘I’ll take’
 d. Old English: *dæges eage* ‘day’s eye’ > English *daisy* (a flower, *Bellis perennis*)

While univerbation is the creation of a (possibly complex) word-form from an earlier syntactic combination, lexemization is the creation of a simple lexeme-stem from an earlier composite form (either a syntactic combination or a composite single-word form). The example of English *daisy* in (15d) (< Old English *dæges eage* ‘day’s eye’) shows that univerbation can simultaneously lead to lexemization, but more typical cases of lexemization involve the absorption of an earlier prefix or suffix into the root (as in 15a-c). Hopper (1990) called such processes “demorphologization” (see also Brinton & Traugott, 2005, Section 2.3.3).

In this paper, I focus on single-root lexemes, but as I noted above (Section 4), compound stems consisting of two roots also count as lexeme-stems, e.g., *flower-pot*. So, one might ask whether such compound lexemes also arise via lexemization. This is an interesting question, but I leave it aside here, because not much is known about the way in which productive compound patterns arise in languages. Most cases of “lexicalized compounds” are semantically idiosyncratic and thus fall under inventorization (Section 7.3). By contrast, even semantically regular compounds are lexemes and may have undergone lexemization, but most were created from productive compounding patterns.

¹² In recent decades, the term *agglutination* has more often been used in a synchronic sense, for the type of grammatical marking found in languages of the “agglutinative type”. However, as the morphological classifications of the 19th century are not well-founded (Haspelmath, 2009), it may be worth reviving the diachronic sense of *agglutination*, which is much better understood.

7.3. “Lexicalization” as inventorization (not “institutionalization”)

Most commonly, the term *lexicalization* is understood in the sense of inventorization, as defined in (16), relying on the term *inventorium* that was introduced in Section 5 above.

(16) **inventorization**

Inventorization is a diachronic change by which an unrestricted combination acquires unpredictable properties so that it must be part of the *inventorium*.

As noted earlier, the term *lexicon* was originally introduced for the *inventorium* by Bloomfield (1933, p. 274), who famously characterized the lexicon as “the list of basic irregularities” (first taken as the set of morphs, but later extended to idiosyncratic combinations). It appears that the use of the term *lexicalization* took this sense as its starting point.

The term *lexicalization* in a diachronic sense is most typically used for the development of various idiosyncrasies in composite expressions that are no longer semantically compositional and phonologically regular and thus must be part of the *inventorium*. The following quotation is quite characteristic for the way the term seems to be used most often:

Lexicalization is the change whereby in certain linguistic contexts speakers use a syntactic construction or word formation as a new contentful form with formal and semantic properties that are not completely derivable or predictable from the constituents of the construction or word formation pattern. (Brinton & Traugott, 2005, p. 96)

The term has been used in this sense since the 1960s. Marchand (1960, p. 80) gave examples such as *black market* ‘clandestine market’ and *carry out* ‘execute’, which are cases of *idiomatization* (development of an idiomatic, unpredictable meaning, even though the form stays regular). A particularly nice example of semantic idiosyncrasy of nominal compounds is the contrast between *sleep tablet* (‘tablet for sleep’) and *headache tablet* (‘tablet against headache’) (Lipka et al., 2004). English examples of phonological irregularity from English are the deadjectival abstract nouns *length* and *width*, which are semantically quite regular (‘quality of being long, of being wide’).

An important aspect of the definition of inventorization is that the *inventorium* is the set of conventions that language users must know, and no claim is made about the way in which this knowledge is mentally represented (recall that the *inventorium* is crucially different from the *mentalicon* in this regard). Inventorization is thus a social and historical process, not a cognitive process (see also Section 7.4 below).

To capture this aspect of “lexicalization”, some authors have posited a special process of *institutionalization*. The idea of a progression from “nonce formation” via institutionalization to lexicalization is found in works such as Brinton & Traugott (2005, Section 2.2) and Bauer et al. (2013, p. 30; see also Hohenhaus, 2005, for an overview):

(17) nonce formation > institutionalization > lexicalization

The claim here is that nonce formations result from applying a word-formation rule by which a “potential word” becomes an “actual word” in a speaker’s individual usage. For example, the English verb *to incentivize* was first recorded in 1989 and may not have been used before that date (Bauer et al., 2013, p. 29), so it was only a potential word before that date. But it is only after a word is adopted by a larger number of language users that one can say that it has become part of the language in general, i.e., that it has become institutionalized. It need not be irregular in any way, but once it develops some idiosyncrasies (semantic or phonological), it becomes definitely part of the inventorium. This proposal relies on the idea of a fundamental distinction between potential words and actual words (or possible words and existing words), and Lipka et al. (2004, p. 3) link this distinction to Coseriu’s proposal that in addition to Saussure’s *parole* (language use) and *langue* (language system), there is a third level of language called *norm* (Kabatek, 2020; see also Lyons, 1977, p. 549 for remarks going in a similar direction). Institutionalization would then constitute the passing from an (unrealized) system option to the reality of the norm.

However, the distinction between potential and actual words is very questionable, so the notion of “institutionalization” is not well-founded. There is no analogous distinction between potential and actual sentences, and the idea of a “norm” as a separate level between language system and language use is obscure. It is true that many linguists feel that by applying an inflectional rule (e.g., by forming the plural *coati-s* from the English singular noun *coati*, a raccoon-like mammal), one does not create a “new word”, while by applying a derivational rule (e.g., *ecstatic-ness* from *ecstatic*) or by making a compound (e.g., *coati hunter*), one does create a “new word”. However, it is unclear what the basis for this distinction might be (on the problems of delimiting inflectional and derivational rules, see Plank, 1994, and Haspelmath, 2024a). Many derivational and compound formations are completely unremarkable and are not recorded in dictionaries, so the question whether they are only “potential words” or also “actual words” cannot really be answered.¹³

One cannot deny, of course, that some kinds of morphological patterns seem to have a strong tendency to be used with idiosyncratic meanings. For example, the verb *incentivize* has the specific meaning ‘provide an incentive’ which is not entirely predictable from its parts *incentive* and *-ize*, and perhaps it had this idiomatic meaning right from the beginning when it was first coined. But this would mean that it became part of the inventorium right away, without necessarily undergoing a diachronic process.¹⁴ I would thus reject the idea of

¹³ See also Świątek (2014) for discussion of the problems with the term “nonce formation”.

¹⁴ Perhaps the notion of *coining* can be said to refer the process of instantaneous idiomatization, for purposes of creating a short expression with a highly specific meaning.

“institutionalization” as a separate stage of a progression as in (17), though the precise ways in which complex expressions with idiosyncrasies arise is still poorly understood.

In addition to idiosyncratic items, the inventorium contains items that are not semantically or formally idiosyncratic, but are still fixed expressions in that they are conventional ways of rendering certain meanings. In his discussion of lexicalization, Bauer (1983, p. 55) mentions unproductive formations such as English *warm-th* and *assign-ment*, which are not semantically or phonologically idiosyncratic, but which are unpredictable because the derivational pattern is unproductive.

Another, less well-known example is fixed expressions of the type that Mel’čuk (2012; 2015) calls CLICHÉS in his typology of phrasemes (or phraseological units, i.e., fixed expressions). They are not idioms, but they are still part of what one must learn in order to know a language fully and to use it felicitously in all contexts. For example, Russian uses *ostorožno, okrašeno* [caution, painted] on signs where English-language signs would have *wet paint*. Russian has the literal translation *syraja kraska* [raw paint], but would not use them in the context of signs, just as English has *caution, painted!* but would not use this expression in this context. Another example of clichés in different language comes from academic texts: French says *autrement dit* [otherwise said], English has *to put it differently*, and German has *mit anderen Worten* [with other words]. All these expressions are perfectly regular, and speakers would be able express these meanings differently (in a manner corresponding more closely to one of the other languages), but these are the conventional expressions used in this specific context.¹⁵

The notion of inventorization encompasses both the development of unproductive derived lexemes (such as *warm-th*) and the development of clichés (such as *to put it differently*), because there is something unpredictable about these expressions, so that they must be learned. The development of clichés has not been included in “lexicalization” so far, so the notion of inventorization is perhaps a bit broader than the earlier notion of idiosyncrasy-driven lexicalization. However, clichés have not played an important role in the literature so far, and I will not say more about them.

Inventorization has to do with the “public lexicon” (i.e., the conventions of the speech community) as opposed to the “mental lexicon” (Sperber & Wilson, 1998). But there is a corresponding mental process, which is here called *mentalication*, and we now take a closer look at this concept.

¹⁵ Mel’čuk’s *clichés* seem to be similar to what some morphologists have called “nameworthiness” or “naming force” (e.g., Gaeta & Ricca, 2009). These issues have not been widely discussed among linguists.

7.4. “Lexicalization” as mentalicization?

As I said earlier, the inventorium is the collection of conventional elements that must be learned and stored by speakers as they cannot be created on the fly. But linguists are often more interested in “mental languages”, i.e., speakers’ mental representations of their knowledge of the social conventions. (A mental language is often called “I-language” or “competence”, especially in generative grammar contexts.) Given the notion of a mentalicon (Section 6), the counterpart of lexicalization at the level of mental systems can be called *mentalicization*.

(18) **mentalicization**

Mentalicization is a psychological change by which an unrestricted combination comes to be part of a speaker’s mentalicon (or “mental lexicon”).

Even though “lexicalization” is generally thought of as a historical process and the term was first used by historical linguists, some linguists now associate it with a mental or psychological process. The following quotation about an English derivational pattern is typical in this respect:

“many *-ity* derivatives are lexicalized, i.e., they have become permanently incorporated into the mental lexicons of speakers, thereby often adopting idiosyncratic meanings, such as *antiquity* ‘state of being antique’ or ‘ancient time,’ *curiosity* ‘quality of being curious’ and ‘curious thing.’ (Plag, 2003, p. 91)

And according to Hilpert (2019), lexicalization is the process of “adding new open-class elements to a repository of holistically processed linguistic units”. He gives examples such as *crowdsourcing* (a new compound), *emoji* (a new loanword), and *cutting edge* (a new idiom). Similarly, Gebhardt (2023, p. 20) says that “an expression becomes lexicalized when it becomes a word or phrase that is stored as a unit”. These authors thus seem to describe “lexicalization” as mentalicization. And in the same vein, the term *nonce-formation* (see (17) above) has been used in the sense of a form that is novel for a particular language user’s mental lexicon (e.g., Dal & Namer, 2018, p. 204).

However, the evidence for lexicalization (of forms such as *antiquity* and *cutting edge*) does not normally come from the “mental lexicons” of speakers, or from “holistic processing”, or from indications of “storage as a unit”. It comes from the observed idiosyncrasies (or more generally, the observed unpredictability), which make it necessary to assume that such forms are stored and processed holistically. But as I noted above, speakers also store (and process holistically) many forms that are not idiosyncratic or otherwise unpredictable.

As a psychological process, mentalicization happens at the individual level, whereas lexicalization is a historical process which must happen at the social level. Since Hermann Paul’s time, historical linguists have often tried to link diachronic changes to psychological

patterns or processes, but even though many of these suggestions may be plausible, we must keep the social and the psychological processes apart conceptually. Mentalicization over a language user's lifespan may play a role in inventorization (or univerbation, or lexemization) over the lifespan of a language, but this role will in any event be very indirect. And next to nothing seems to be known about the way in which mentalicization might give rise to inventorization.

7.5. Does all this help us understand inventorization?

Now that we have seen several concepts that have been called “lexicalization” in the past, let us ask briefly what might explain such changes. As the most important process is inventorization, I will concentrate on this process here.

It appears to me that inventorization is not a process that we understand well, because we cannot easily identify constraints on it. Of course, mentalicization in some speakers is a necessary prerequisite for idiosyncrasies or unpredictability to arise, and an expression must be sufficiently frequent in language use to be stored holistically. But even mentalicization by all speakers is not a sufficient condition, because there are some regular expressions (e.g., *do you like it?*; *that's a problem*; *I am tired*) are so frequent that they are probably stored holistically by all speakers.

Why do some expressions acquire idiosyncrasies and others don't? It seems that we do not have a general answer to this question, and maybe there is none, because idiosyncrasies develop randomly, once the condition of high frequency and widespread mentalicization is fulfilled.

In the case of grammaticalization, it seems that there is an interesting constraint: While grammaticalization changes are very common, antigrammaticalization is very rare (Haspelmath, 1999; 2004). Is there a corresponding constraint on inventorization, or on one of the other types of “lexicalization”? One might want to say that just as antigrammaticalization is very rare, antilexemization is very rare, as we almost never find cases such as (19), where the root *alcoholic* is broken up into the parts *alc-* and *-oholic*, and new words are formed with the element *-oholic*.

(19) antilexemization

(= “folk etymology”, Lehmann, 2002; Brinton & Traugott, 2005 p. 102), e.g.,

alcoholic > *alc-oholic*,

work-oholic,

choc-oholic, ...

However, while it is surprising (and in need of explanation) that antigrammaticalization is very rare, because there would be many opportunities for it, there is no need for a special explanation of the rarity of antilexemization. As it is possible only with unusually long roots (such as *alcoholic*, a learned loanword), it is necessarily an uncommon phenomenon. Thus, we

do not seem to have good reasons to reject the null hypothesis that inventorization is a random process.

8. Conclusion

The main thrust of this paper is the idea that well-recognized different senses of *lexicon*, *lexical* and *lexicalization* can be better kept apart by distinguishing clearly between word-forms and lexemes, and by introducing a number of new terms, especially *inventorium* (and *inventorization*) as well as *mentalicon* (and *mentalicization*). While the *inventorium* (the set of meaning-form conventions that must be remembered) and the *mentalicon* (the set of elements that a particular speaker has stored) are without any doubt linguistically significant, this is much less clear for “lexicon”-like notions based on the *word(-form)* or the *lexeme*, because these are defined in artificial ways and may ultimately be based on Western habits of spelling languages (with spaces between words) and describing languages (in dictionaries and grammar books).

Linguists have long been aware of the problems with their terminology, but these problems have persisted over the last decades. That the term *word* has multiple senses was recognized by Lyons (1968, p. 196) among others, and that *lexicon* can mean at least three different things was pointed out by Hoeksema (1985, Chapter 1) and Jackendoff (2002, Chapter 6) among others. Thus, the present paper does not add much in terms of substance, but it makes concrete proposals for remedying the situation: The different senses can be distinguished by actually using different terms when precision is needed.¹⁶

Linguists will no doubt continue to use the terms *word* and *lexical*, and in many cases, these do not present any problems because the context makes this meaning clear. However, it is useful to be aware that a stereotypical word is associated with a number of properties that need not go together (as we saw in Section 2), so that in many technical contexts, it makes good sense to use more precise terminology.

Finally, I discussed the term *lexicalization* (Section 7) and showed that it can likewise be understood in four different senses: univerbation, lexemization, inventorization, and mentalicization. Again, this discussion does not add any new insights, and I am currently pessimistic about understanding the phenomenon of inventorization (Section 7.5), but the

¹⁶ There has been one previous suggestion for introducing a new term, by Aronoff (1988), who emphasized the conceptual contrast between “idiosyncratic-lexical” (= *inventorial* in current terms) and “categorical-lexical” (= *lexemic* in current terms), proposing the new term *umlical* for the latter (from “uninflected member of a major lexical category”). However, Aronoff did not follow up on this suggestion and did not use this term in his subsequent papers, perhaps because the term did not look very plausible. Hopefully the terms *inventorium* and *mentalicon* will be found to be more memorable and better connected to the earlier literature.

conceptual distinctions that are highlighted here may help us avoid talking past each other in the future.

References

- Aikhenvald, A. Y., Dixon, R. M. W., & White, N. M. (Eds.). (2020). *Phonological word and grammatical word: A cross-linguistic typology*. Oxford University Press.
- Aitchison, J. (1987). *Words in the mind: An introduction to the mental lexicon*. Blackwell.
- Anderson, S. R. (1992). *A-morphous morphology*. Cambridge University Press.
- Aronoff, M. (1976). *Word formation in generative grammar*. MIT Press.
- Aronoff, M. (1988). Two senses of lexical. In *ESCOL (Eastern States Conference on Linguistics)*, '88, 1–11.
- Aronoff, M. (1994). *Morphology by itself: Stems and inflectional classes*. MIT Press.
- Baker, M. C., & Croft, W. (2017). Lexical categories: Legacy, lacuna, and opportunity for functionalists and formalists. *Annual Review of Linguistics*, 3, 179–197.
- Bauer, L., Lieber, R., & Plag, I. (2013). *The Oxford reference guide to English morphology*. Oxford University Press.
- Bauer, L. (1983). *English word-formation*. Cambridge University Press.
- Blank, A. (2001). Pathways of lexicalization. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (Eds.), *Language typology and language universals: An international handbook* (vol. 2) (pp. 1596–1608). Walter de Gruyter. <https://doi.org/10.1515/9783110194265-049>
- Bloomfield, L. (1933). *Language*. H. Holt and Company.
- Booij, G. E. (2012). *The grammar of words: An introduction to linguistic morphology* (3rd ed.). Oxford University Press.
- Boye, K. (2023). Grammaticalization as conventionalization of discursively secondary status: Deconstructing the lexical–grammatical continuum. *Transactions of the Philological Society*, 121(2), 270–292. <https://doi.org/10.1111/1467-968X.12265>
- Bresnan, J., & Mchombo, S. A. (1995). The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory*, 13(2), 181–254.
- Bresnan, J. (2001). *Lexical-functional syntax*. Blackwell.
- Brinton, L. J., & Traugott E. C. (2005). *Lexicalization and language change*. Cambridge University Press.
- Briscoe, T. J., & Copestake, A. (1999). Lexical rules in constraint-based grammar. *Computational Linguistics*, 25(4), 487–526.

- Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5), 425–455. <https://doi.org/10.1080/01690969508407111>
- Bybee, J. L. (2010). *Language, usage and cognition*. Cambridge University Press.
- Cappelle, B. (2022). Lexical Integrity: A mere construct or more a construction? *Yearbook of the German Cognitive Linguistics Association*, 10(1), 183–216. <https://doi.org/10.1515/gcla-2022-0009>
- Carstairs-McCarthy, A. (2005). Basic terminology. In P. Štekauer & R. Lieber (Eds.), *Handbook of word-formation* (pp. 5–23). Springer. https://doi.org/10.1007/1-4020-3596-9_1
- Chomsky, N. A. (1965). *Aspects of the theory of syntax*. MIT Press.
- Chomsky, N. A. (1970). Remarks on nominalization. In R. A. Jacobs & P. S. Rosenbaum (Eds.), *Readings in English transformational grammar* (pp. 184–221). Ginn.
- Chumakina, M. (2013). Introduction. In M. Chumakina & G. Corbett (Eds.), *Periphrasis: The role of syntax and morphology in paradigms* (pp. 1–23). British Academy.
- Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge University Press.
- Culicover, P. W., Jackendoff, R., & Audring, J. (2017). Multiword constructions in the grammar. *Topics in Cognitive Science*, 9(3), 552–568. <https://doi.org/10.1111/tops.12255>
- Dal, G., & Namer, F. (2018). Playful nonce-formations in French: Creativity and productivity. In S. Arndt-Lappe, A. Braun & C. Moulin (Eds.), *Expanding the lexicon: Linguistic innovation, morphological productivity, and ludicity* (pp. 203–228). De Gruyter. <https://doi.org/10.1515/9783110501933>
- Davis, A., & Koenig, J.-P. (2021). The lexicon in HPSG. In S. Müller, A. Abeillé, R. D. Borsley & J.-P. Koenig (Eds.), *Head-Driven Phrase Structure Grammar: The handbook* (pp. 125–176). Language Science Press.
- Di Sciullo, A.-M., & Williams, E. (1987). *On the definition of word*. MIT Press.
- Du Bois, J. W. (1987). The discourse basis of ergativity. *Language*, 63, 805–855.
- Embick, D. (2015). *The morpheme: A theoretical introduction*. De Gruyter Mouton.
- Emmorey, K. D., & Fromkin, V. A. (1988). The mental lexicon. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey* (vol. 3) (pp. 124–149). Cambridge University Press.
- François, A. (2008). Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In M. Vanhove (Ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations* (pp. 163–216). Benjamins.

- Gaeta, L., & Ricca, D. (2009). Composita solvantur: Compounds as lexical units or morphological objects? *Rivista di Linguistica*, 21(1), 35–70. https://www.italian-journal-linguistics.com/app/uploads/2021/05/03.gaeta_.pdf
- Gazdar, G., Klein, E., Pullum, G. K., & Sag, I. A. (1985). *Generalized Phrase Structure Grammar*. Harvard University Press.
- Gebhardt, L. (2023). *The study of words: An introduction*. Routledge.
- Geeraerts, D. (2010). *Theories of lexical semantics: An introduction to the history and current state of theories of word meanings*. Oxford University Press.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. The University of Chicago Press.
- Haig, G., & Schnell, S. (2016). The discourse basis of ergativity revisited. *Language*, 92(3), 591–618. <https://doi.org/10.1353/lan.2016.0049>
- Hall, T. A. (1999). The phonological word: A review. In T. A. Hall & U. Kleinhenz (Eds.), *Studies on the phonological word* (pp. 1–22). Benjamins.
- Harley, H. (2006). *English words: A linguistic introduction*. Blackwell.
- Haspelmath, M. (1999). Why is grammaticalization irreversible? *Linguistics*, 37(6), 1043–1068.
- Haspelmath, M. (2002). *Understanding morphology*. Arnold. <https://zenodo.org/record/1236482>
- Haspelmath, M. (2004). On directionality in language change with particular reference to grammaticalization. In O. Fischer, M. Norde & H. Perridon (Eds.), *Up and down the cline: The nature of grammaticalization* (pp. 17–44). Benjamins.
- Haspelmath, M. (2009). An empirical test of the Agglutination Hypothesis. In S. Scalise, E. Magni & A. Bisetto (Eds.), *Universals of language today* (pp. 13–29). Springer.
- Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica*, 45(1), 31–80. <https://doi.org/10.1515/flin-2017-1005>
- Haspelmath, M. (2018). Revisiting the anasyntetic spiral. In B. Heine & H. Narrog (Eds.), *Grammaticalization from a typological perspective* (pp. 97–115). Oxford University Press.
- Haspelmath, M. (2023a). Defining the word. *WORD*, 69(3), 283–297. <https://doi.org/10.1080/00437956.2023.2237272>
- Haspelmath, M. (2023b). On what a construction is. *Constructions*, 15(1). <https://doi.org/10.24338/cons-539>

- Haspelmath, M. (2023c). Coexpression and synexpression patterns across languages: Comparative concepts and possible explanations. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1236853>
- Haspelmath, M. (2024a). Inflection and derivation as traditional comparative concepts. *Linguistics*, 62(1). <https://doi.org/10.1515/ling-2022-0086>
- Haspelmath, M. (2024b). Compound and incorporation constructions as combinations of unexpandable roots. Manuscript, MPI-EVA. <https://zenodo.org/record/8137251>
- Hilpert, M. (2019). Lexicalization in morphology. *Oxford Research Encyclopedia of Linguistics*, 2019. <https://doi.org/10.1093/acrefore/9780199384655.013.622>
- Hoeksema, J. (1985). *Categorial morphology*. Garland.
- Hohenhaus, P. (2005). Lexicalization and institutionalization. In P. Štekauer & R. Lieber (Eds.), *Handbook of word-formation* (pp. 353–373). Springer. doi:10.1007/1-4020-3596-9
- Hopper, P. (1990). Where do words come from? In W. Croft, KM. Denning & S. Kemmer (Eds.), *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday* (pp. 151–160). Benjamins.
- Jackendoff, R. (1975). Morphological and semantic regularities in the lexicon. *Language*, 51(3), 639–671.
- Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E.-J. van der Linden, A. Schenk & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives* (pp. 133–165). Taylor and Francis.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.
- Jackendoff, R. (2013). Constructions in the Parallel Architecture. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of Construction Grammar* (pp. 70–92). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195396683.013.0005>
- Kabatek, J. (2020). Linguistic norm in the linguistic theory of Eugenio Coseriu. In F. Lebsanft & F. Tacke (Eds.), *Manual of standardization in the Romance languages*. De Gruyter.
- Lehmann, C. (2002). New reflections on grammaticalization and lexicalization. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 1–18). Benjamins.
- Lehmann, C. (2020). Univerbation. *Folia Linguistica*, 54(s41–s1), 205–252. <https://doi.org/10.1515/flih-2020-0007>
- Levelt, W. J. M., Roelofs, A. & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38. <https://doi.org/10.1017/S0140525X99001776>

- Levin, B., & Rappaport Hovav, M. (2019). Lexicalization patterns. In R. Truswell (Ed.), *Oxford handbook of event structure* (pp. 395–425). Oxford University Press.
- Lipka, L., Handl, S., & Falkner, W. (2004). Lexicalization and institutionalization: The state of the art in 2004. *SKASE Journal of Theoretical Linguistics*, 1, 2–19.
- Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge University Press.
- Lyons, J. (1977). *Semantics*. Cambridge University Press.
- Marchand, H. (1960). *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. Harrassowitz.
- Masini, F. (2009). Phrasal lexemes, compounds and phrases: A constructionist perspective. *Word Structure*, 2(2), 254–271. <https://doi.org/10.3366/E1750124509000440>
- Matthews, P. H. (1972). *Inflectional morphology*. Cambridge University Press.
- Matthews, P. H. (1991). *Morphology* [2nd ed.]. Cambridge University Press.
- Mel'čuk, I. (2006). *Aspects of the theory of morphology*. De Gruyter.
- Mel'čuk, I. (2012). Phraseology in the language, in the dictionary, and in the computer. *Yearbook of Phraseology*, 3(1), 31–56. <https://doi.org/10.1515/phras-2012-0003>
- Mel'čuk, I. (2015). Clichés, an understudied subclass of phrasemes. *Yearbook of Phraseology*, 6(1), 55–86. <https://doi.org/10.1515/phras-2015-0005>
- Mohanan, K.P. (1986). *The theory of lexical phonology*. Reidel.
- Muysken, P. (2008). *Functional categories*. Cambridge University Press.
- Narrog, H., & Heine, B. (2021). *Grammaticalization*. Oxford University Press.
- O'Neill, P. (2016). Lexicalism, the principle of morphology-free syntax and the principle of syntax-free morphology. In A. Hippisley & G. Stump (Eds.), *The Cambridge handbook of morphology* (pp. 237–271). Cambridge University Press. <https://doi.org/10.1017/9781139814720.010>
- Pepper, S. (2023). Defining and typologizing binominal lexemes. In S. Pepper, F. Masini & S. Mattioli (Eds.), *Binominal lexemes in cross-linguistic perspective: Towards a typology of complex lexemes* (pp. 23–72). De Gruyter Mouton.
- Plag, I. (2003). *Word-formation in English*. Cambridge University Press.
- Plank, F. (1994). Inflection and derivation. In R.E. Asher (Ed.), *Encyclopedia of Language and Linguistics* (pp. 1671–1678). Pergamon. doi:10.5281/zenodo.5817771

Scalise, S., & Guevara, E. (2005). The lexicalist approach to word-formation and the notion of the lexicon. In P. Štekauer & R. Lieber (Eds.), *Handbook of word-formation* (pp. 147–187). Springer. https://doi.org/10.1007/1-4020-3596-9_1

Spencer, A. (2013). *Lexical relatedness*. Oxford University Press.

Sperber, D., & Wilson, D. (1998). The mapping between the mental and the public lexicon. In P. Carruthers & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 184–200). Cambridge University Press.

Stolz, T. (1991). Agglutinationstheorie und Grammatikalisierungsforschung: Einige alte und neue Gedanken zur Entstehung von gebundener Morphologie. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 44(3), 325–338.

Świątek, D. (2014). The notion of “nonce formation” revisited. *Prace Naukowe Akademii im. Jana Długosza w Częstochowie: Studia Neofilologiczne*, 10, 207–221.

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *NeuroImage*, 84, 265–278. <https://doi.org/10.1016/j.neuroimage.2013.08.023>

Talmy, L. (1985). Lexicalization patterns. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. III) (pp. 57–149). Cambridge University Press.

Williams, E. (2007). Dumping lexicalism. In G. Ramchand & C. Reiss (Eds.), *The Oxford handbook of linguistic interfaces* (pp. 352–382). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199247455.013.0012>

Woolford, E. (2006). Lexical case, inherent case, and argument structure. *Linguistic Inquiry*, 37(1), 111–130. <https://doi.org/10.1162/002438906775321175>

Zaliznjak, A. A. (1967). *Russkoe imennoe slovoizmenenie (Русское именное словоизменение)*. Nauka. https://inslav.ru/images/stories/pdf/2002_Zalizniak_RIS_i_statji.pdf