

# Forward-formation, back-formation, and the nature of the inventorium

MARTIN HASPELMATH

*Max Planck Institute for Evolutionary Anthropology (Leipzig)*

draft, October 2024

This paper discusses back-formation from the perspective of a view of morphology in which there are productive and unproductive declarative schemas, but no abstract processes, so that the issue of the direction of derivation does not arise. This paradigmatic perspective does not make a strict distinction between rules and listed expressions either, and everything that speakers must remember is said to be in the inventorium (a new term with a meaning similar to Jackendoff's extended lexicon). Moreover, no distinction is made between "potential" and "existing words": Just like sentences, regular complex words are potential, and complex words are inventorized if they are in some way idiosyncratic. On this view, back-formation is just as normal as forward-formation, though when it comes to word coinage, it is clear that back-coinage is much less common.

## 1. Some traditional assumptions

Many (or most) discussions of word formation assume that the normal way of "forming words" is by "attaching" an element to a base (as in 1a), or perhaps by applying an "operation" to a base (as in 1b), or by compounding two elements (as in 1c), which is also a kind of "attachment" or "operation".

- (1) English
- |                               |                      |
|-------------------------------|----------------------|
| a. <i>help</i> + <i>er</i>    | → <i>helper</i>      |
| b. <i>shelf</i> (+ voicing)   | → (to) <i>shelve</i> |
| c. <i>flower</i> + <i>pot</i> | → <i>flowerpot</i>   |

Against these background assumptions, back-formations as in (2)-(3) are unexpected and create conceptual problems. It has often been said that they cannot be synchronic word-formation processes, but must be of diachronic relevance only.

- (2) English
- |                      |                       |
|----------------------|-----------------------|
| a. <i>editor</i>     | → (to) <i>edit</i>    |
| b. <i>babysitter</i> | → (to) <i>babysit</i> |
| c. <i>injury</i>     | → (to) <i>injure</i>  |
- (3) German
- |                                    |  |
|------------------------------------|--|
| a. <i>blödsinnig</i> 'idiotic'     | → <i>Blödsinn</i> 'nonsense'                   |
| b. <i>abkehren</i> 'turn away'     | → <i>Abkehr</i> 'turning away'                 |
| c. <i>Notlandung</i> 'em. landing' | → <i>notlanden</i> 'make an emergency landing' |

In this paper, I question a number of assumptions that are often implicit and sometimes explicit in discussions of word formation. Many linguists have said that word-formation consists in adding "new words" to the "lexicon", and that back-formations like *edit* are

based on “existing words” like *editor*. Instead of the traditional vague notion of the “lexicon” (which is often understood as a mental entity), I argue that we need the more specific notion of *inventorium*, i.e. the set of all conventional forms and constructions that make up a language system (Haspelmath 2024b). In particular, I argue that the following four assumptions, which are commonly made, are in fact not necessary:

- (4) unnecessary assumptions:
  - a. derivational “processes” with directionality (§3)
  - b. a distinction between lexicon and morphosyntax (§5)
  - c. “word-making”: a distinction between potential and existing words (§6)
  - d. “institutionalization”: a distinction between “system” and “norm” (§6)

But of course, many traditional concepts are crucial, and I will rely on the following:

- (5) necessary concepts:
  - a. schemas with a generative function and a relational function (§2)
  - b. idiomatization and conventionalization
  - c. word coinage (§7)
  - d. productivity and unproductivity of a schema (§2)

In the traditional domain of “word formation”, “forward-formations” as in (1) are surely more frequent than back-formations as in (2)-(3), but I will argue that the reason for this is not that there is a problem with the “operation” or “process” of back-formation. In one sense of the word, back-formation is very common in inflectional morphology (§4.1), and I suggest that the reason why it is perceived to be rare in derivational morphology and compounding is that there are far fewer opportunities for back-formations to occur.

I start by discussing morphosyntactic schemas as an alternative to derivations with directionality (§2-3), and then I point out that forward-formation is just one possibility, and not necessarily a more “natural” one (§4). I question the “word-making” stereotype and the distinction between grammar and “lexicon” and introduce the notion of *inventorium* (§5). Next I discuss and reject the idea that there are “potential words” and “existing words” (§6), pointing out that while many complex words are idiosyncratic (not fully predictable) and thus inventorized, there are also many complex lexemes that are completely regular and thus just as “potential” as complex sentences. Finally, I introduce the notion of word coinage, or instantaneous inventorization (§7), before getting back to back-formation (§8).

## 2. Morphosyntactic schemas and their consequences

Colloquially, we often talk about word formation in terms of processes or operations (such as attaching an affix or applying a phonological rule), but it has been widely observed by theoretically oriented morphologists that morphological “formations” are better thought of as happening in a paradigmatic way, based on analogies or established schemas (e.g. Becker 1990; 1993; Bochner 1993; Blevins 2006; Booij 2010; Jackendoff & Audring 2020ab), and this is in fact true of all of morphosyntax. For example, a novel word such as *ambiclitic* (a technical term of linguistics) can be formed analogically on the basis of the earlier word pair *preposition/ambiposition*, the earlier word *proclitic*, and an analogical schema, as shown in (6).

- (6) *preposition* : *ambiposition*  
 = *proclitic* : X  
 (X = *ambiclitic*)

Here we see a four-part analogy (e.g. Blevins & Blevins 2009: 3) with the variable X, which can be replaced by the novel form in a process of analogical creation.

On the paradigmatic view of morphosyntax, all structure arises from “paradigms”, i.e. sets of forms that are related both semantically and in terms of their shape (like *preposition* and *ambiposition*). For example, morphological structure arises from sets of related forms as in (7), and syntactic structure arises in exactly the same way from sets of related forms as in (8).

- (7) *warmer*            ‘warm’ + ‘more’  
*clearer*            ‘clear’ + ‘more’  
 ...
- (8) *this tree*            ‘this’ + ‘tree’  
*this lake*            ‘this’ + ‘lake’  
 ...

Thus, the fundamental notions of paradigmatic morphosyntax are not “constituent” and “combination” (or “merge”), but *shape* and *meaning*.<sup>1</sup> Paradigms such as (7) and (8) can give rise to general schemas, and such schemas are often productive. If they are productive, they must be part of a grammatical description, but if they are unproductive, they can still be relevant for the behaviour of speakers.

There are many ways in which schemas can be notated. As a simple illustration, we can write the schema for the English *-er* comparative (*warmer*, *clearer*, ...) as in (9) (inspired by Jackendoff & Audring 2020ab).

- (9) semantics:            ‘more A’  
 morphosyntax:        [A AFFIX]<sub>A</sub>  
 phonology:            [... əɾ]

This schema is linked to the schema for (monosyllabic) adjectives in such a way that regular relationships such as *warm/warmer*, *clear/clearer* (etc.) are captured. Thus, if we know the adjective *nice*, we can infer the existence of the comparative *nicer*, and if we know the form *taller*, we can infer the existence of the form *tall*. There is no particular directionality inherent in this conceptualization of grammatical structure. Of course, one can still make use of terms such as *root*, *affix*, *compound*, and so on, but they are not atoms of the system and do not have any special status.

The syntactic schema that arises from the paradigm in (8) can be notated as in (10), which is quite analogous to (9).<sup>2</sup>

<sup>1</sup> A form is a pairing of a shape and a meaning. (I try to avoid the term *form* in the other sense that it often has (‘shape’), though this is difficult. Linguists often contrast “semantic” and “formal” aspects of forms, where they really mean aspects of meaning and aspects of shape.)

<sup>2</sup> The paradigmatic view of morphosyntax is sometimes called “word-based” (e.g. Blevins 2006), but there is no reason to restrict it to the “word” domain (and thus to “morphology”). Becker (1993: 4) says: “Whereas syntax describes how atomic signs connect to form complex signs, morphology describes the regular relations between atomic signs.” But in actual fact, both “morphology” and “syntax” necessarily deal with the structure of complex signs, because “atomic signs” do not have any (semiotic) structure. The notion of “word” does not seem to be a natural one (Haspelmath 2023a), so it is good to have an approach where it plays no role.

(10) semantics:	‘this N’
morphosyntax:	[DET N] <sub>NP</sub>
phonology:	[ðis ...]

The schema-based approach can also capture relationships between two forms where both contain an affix and the base does not occur on its own, as in *hedonism/hedonist*, *deism/deist* (there is no \**hedon* or \**de*). In (11), we see the two linked schemas in the notation of Jackendoff & Audring (2020ab) (the symbol “≈” is taken from Booij 2010).

(11) semantics:	‘ideology X’	≈	‘representative of X’
morphosyntax:	[X AFFIX]		[X AFFIX]
phonology:	[... ɪst]		[... ɪzm]

Such pairs of linked schemas (also called *sister schemas*) may have both a generative role and a relational role. Knowing the schema in (11) is fully compatible with having many word pairs of the type *hedonist/hedonism* in one’s mental lexicon, because schemas and regular forms can coexist. Such linked schemas may be unproductive, i.e. they may not have a generative role but may be limited to their relational role (expressing the relationships between the words over which they generalize). Or they may be productive without restriction, or productive in a limited way. If one adopts this view of morphosyntax, two important consequences follow directly: (i) morphosyntax is conceived of in declarative terms (see §3), and (ii) forward-formation is just one of the possibilities (see §4).

Process metaphors permeate our ways of talking (and sometimes also our ways of thinking) about grammatical patterns. But it has long been known that talking in terms of processes can be problematic. Hockett (1954) was worried about diachronic-sounding process metaphors, and his alternative was an “item-and-arrangement” view (a kind of declarative approach). But *arrangements* require *items*, so they suggest “morpheme-based” modes of description, which are also known to be problematic (e.g. Anderson 2015). On the paradigmatic view, the “morpheme” problem disappears, because schemas are more general and can readily accommodate also endophonic patterns (such as German umlaut Plurals, e.g. *Garten/Gärten* ‘garden(s)’, *Mutter/Mütter* ‘mother(s)’).

Moreover, on the paradigmatic view there is no separate “lexicon component”, distinct from a “grammatical rules” component. A language system consists of forms and schemas or constructions of different degrees of generality, and they must all be stored by speakers in their “mental lexicon” (or *mentalicon*). But a language user’s mentalicon must contain many fully regular forms that can also be created from a productive schema, and different speakers have different mentalicons (because what is stored in one’s memory depends on one’s life experience).

To describe a language (as a system of social conventions), we need the additional notion of *inventorium* (§5-6), i.e. the set of all forms and constructions that one MUST know in order to use a language (what Jackendoff 2013 calls “extended lexicon”; however, he does not emphasize its social and conventional nature). The *inventorium* is very large and contains a lot of information that we do not often think of as idiosyncratic. For example, the not only the composite terms *stroll-er* and *pacifi-er* are conventional (in British English, one is more likely to say *pram* and *dummy*, respectively), but also fairly transparent compounds such as *high chair* and *bathtub* – after all, one cannot predict that the correct terms are not “baby chair” or “bathing tub”. The underappreciated richness of the *inventorium* will be important later (§5-7).

### 3. Morphosyntax is conceived of in declarative terms

As noted in the introduction, one reason why back-formation has sometimes seemed paradoxical or problematic is that “word formation” is typically conceived of in processual terms. Linguists talk about “operations” of word formation, of “affixation” and “compounding”, and of processes of “lexicalization” and “enriching” the lexicon. This kind of “item and process” language may be harmless when it is purely metaphorical, but linguists have often formalized these regularities in terms of clearly asymmetric rules with an input and an output, and thus with a clear directionality. However, as Gaeta & Montermini (2022) noted in their description of the Budapest workshop,

“recent research trends in morphology in various theoretical frameworks have shifted the focus from purely derivational rules to lexical / derivational networks or paradigms. As a consequence, the very role of directionality in word-formation (and more generally in linguistics) has been challenged.”

For example, Plag’s (2003: 187) “schema-based model” (and also Haspelmath’s 2002) is explicitly non-directional, and this also applies to Booij’s (2010) *Construction Morphology* (cf. Štekauer 2015: §2.4-5), as well as to Jackendoff & Audring’s (2020ab) *Relational Morphology*.

If one takes a declarative view, the question of directionality does not arise in the same way. At least since Marchand (1964), linguists have often asked which of the elements in “conversion” pairs such as those in (12) is basic and which is derived.

- |      |                 |                  |
|------|-----------------|------------------|
| (12) | <i>a cook</i>   | <i>to cook</i>   |
|      | <i>a hammer</i> | <i>to hammer</i> |
|      | <i>a tape</i>   | <i>to tape</i>   |
|      | <i>a dance</i>  | <i>to dance</i>  |
|      | <i>a rain</i>   | <i>to rain</i>   |

These debates do not seem to show any sign of subsiding (e.g. Grestenberger & Kastner 2022), but if a declarative conception is adopted, there is no immediate need to decide on a directionality. The examples in (12) can all be captured by two sister schemas as shown in (13), although eventually more needs to be said (because the pairs in (12) are by no means uniform).

- |                 |                      |   |                       |
|-----------------|----------------------|---|-----------------------|
| (13) semantics: | ‘thing related to Y’ | ≈ | ‘action related to Y’ |
| morphosyntax:   | [Y] <sub>N</sub>     |   | [Y] <sub>V</sub>      |
| phonology       | [...]                |   | [...]                 |

It has often been said that the direction of derivation can be inferred from the semantic relationship between two related words (e.g. Cetnarowska 1993; Iacobini 2000; Valera 2015), but this cannot be a general solution because the semantic relationships of morphologically related words are too diverse. Moreover, there are also non-conversion pairs, such as those in (14), where the “semantic direction of derivation” does not correspond to the “formal direction of derivation”, and like those in (15), where it is not fully clear what the semantic direction of derivation is.

- |                                |   |                              |
|--------------------------------|---|------------------------------|
| (14) a. Icelandic              |   |                              |
| <i>kvelja</i>                  | – | <i>kvelja-st</i>             |
| ‘torment, make someone suffer’ |   | ‘suffer’ (see Anderson 2020) |

- b. English  
*linguist* – *linguist-ics*  
 ‘practitioner of linguistics’ ‘the science of language(s)’
- c. English  
*assassin* – *assassin-ate*  
 ‘murderer of a politician’ ‘murder a politician’
- (15) English  
*beauty* – *beauti-ful*  
 ? ‘state of being beautiful’ ? ‘exhibiting beauty’
- warm* – *warm-er (than X)*  
 ? ‘higher in temperature than average’ ? ‘higher in temperature (than X)’

It is true, of course, that the productivity of such schemas is often quite restricted, but they exist, and in a paradigmatic, schema-based approach, they can be modelled easily. For example, Jackendoff & Audring (2020b: 6) describe the relationship between English *assassin* and *assassinate* as in (16).

- (16) a. *assassin*:  
 semantics: [PERSON WHO [MURDERS POLITICIAN]<sub>11</sub>]<sub>10</sub>  
 morphosyntax: N<sub>10</sub>  
 phonology: /əsæsən/
- b. *assassinate*:  
 semantics: [MURDER POLITICIAN]<sub>1</sub> 1  
 morphosyntax: [V N<sub>10</sub> aff<sub>12</sub>]<sub>5</sub>  
 phonology: /əsæsən10 ejt<sub>12/11</sub>

Thus, the schema-based approach puts few restrictions on the way in which two forms may be related formally and semantically. In the next section, we will see that this is the right approach, as the constraints come from elsewhere (§4.3).

## 4. Forward-formation is just one possibility

Back-formation is typically regarded as something that is unusual and that does not quite fit into the usual categories of morphology. However, on the paradigmatic view, forward-formation and back-formation are symmetric. Here we will see that in addition, there are cross-formations, which are a third possibility offered by the paradigmatic perspective. We will see that there are external reasons for the prominence of forward formation, but they need not have any relation to the way morphological schemas work.

### 4.1. Cross-formation in derivation and inflection

In addition to forward-formation and back-formation, there is also cross-formation (Becker 1993), a phenomenon that was already exemplified in (11) above. Another

example comes from German country names and country adjectives, as illustrated in (17): Country names often have the suffix *-ien* (corresponding to English *-ia*), which is replaced by the suffix *-isch* in the corresponding adjectives (where English adds the suffix *-n*).

(17)	German			(English)
	<i>Belgien</i>	↔	<i>belgisch</i>	( <i>Belgium/Belgi-an</i> )
	<i>Algerien</i>	↔	<i>algerisch</i>	( <i>Algeria/Algeria-n</i> )
	<i>Pamphylien</i>	↔	<i>pamphylich</i>	( <i>Pamphylia/Pamphylia-n</i> )
	<i>Oltenien</i>	↔	<i>oltenisch</i>	( <i>Oltenia/Oltenia-n</i> )

This might be regarded as a combination of subtraction and addition, but such process metaphors are not needed if the schema-based approach is adopted.

The examples in (17) include rare country names in order to illustrate the productivity of the pattern, but productive cross-formations can also be illustrated with inflectional paradigms, such as Latin verbal person inflection, as shown in (18).

(18) Latin verbal inflection (traditional paradigm)

1SG	<i>venio</i>	‘I come’
2SG	<i>venis</i>	‘you come’
3SG	<i>venit</i>	‘s/he comes’

Such triples are not very similar to the *help/help-er* pair that we saw above in (1) because there is no basic form *\*ven(i)* from which all the others could be derived. Instead, inflected forms can be seen as cross-formations similar to (11) and (17):

(19) Latin inflectional cross-formations

<i>venio</i>	↔	<i>venis</i>
<i>venit</i>	↔	<i>venio</i>
<i>venis</i>	↔	<i>venit</i>

The dual-direction arrow here is meant to express the fact that (at least in principle) one can get from any inflected form in the paradigm to any other form. The dual-direction perspective can also be applied to English singular/plural pairs as in (20).

(20) English inflectional formations

<i>book</i>	↔	<i>books</i>
<i>table</i>	↔	<i>tables</i>

If a Latin speaker knows the form *venis*, they can infer the existence of forms such as *venit* and *venio*, and similarly for all other inflected forms. Likewise, if an English speaker knows the plural forms *books* and *tables*, they can infer the existence of the singular forms *book* and *table* – they can “back-form” singulars from plurals. From this perspective, we can thus identify “back-formation” in inflection, too:

(21) “inflectional back-formation”

<i>books</i>	→	<i>book</i>
<i>tables</i>	→	<i>table</i>

Since inflection and derivation are not all that different (Haspelmath 2024a), back-formation in inflection is comparable to back-formation in derivation (as seen above in

(1)-(3)).<sup>3</sup> Whatever differences exist must be due to special circumstances, and we will see what these are below.

#### 4.2. Is forward-formation more natural?

Štekauer (2015: 340) suggests that forward-formation may be the normal case because it is “natural”. He cites works such as Mayerthaler (1981) and Dressler (2005) and summarizes their view as follows:

“According to the theory of natural morphology the most natural process in inflectional and derivational morphology is captured by the principle of diagrammaticity, i.e. *constructional iconicity*. This principle requires that a new meaning be represented by a corresponding form.”

Another author who attributed formal complexity to diagrammatic iconicity was Givón (1991: 87), who proposed the principle in (22):

(22) **The quantity principle A** (Givón 1991: 87)

“A larger chunk of information will be given a larger chunk of code.”

e.g. act ⇒ act-ive ⇒ act-iv-ate ⇒ act-iv-at-ion

According to this principle, we expect languages to show forward-formation, but not back-formation, unless it somehow “subtracts” meaning. Something along these lines is also presupposed by Iacobini (2000: 866) in his handbook article on determining the direction of derivation:

“The direction of derivation can ... be identified with a semantic and morpho-phonological growth which acts upon a base to form a derived word. The derivative is distinguished from the base because of its greater semantic specificity and more complex morphological structure, which usually reveal themselves in the addition of a morphological element (typically an affix)...”

Of course, all this makes intuitive sense, and the diagrammaticity principle has rarely been questioned (but see Haspelmath 2008). Thus, what one might call the diagrammatic stereotype (summarized in (23)) seems to be fairly deeply entrenched in the discipline.

(23) **The diagrammatic stereotype**

Adding a formative implies adding a meaning, and new words are created by creating composite forms.

A related idea is Koontz-Garboden’s Monotonicity Hypothesis in (24), which is illustrated by triples of forms such as (25a-c).

(24) **The Monotonicity Hypothesis** (Koontz-Garboden 2009; 2012)

“Word formation operations do not remove operators from lexical semantic representations.”

---

<sup>3</sup> And as we saw above, syntax is not all that different from morphology; for example, hearing a nominal such as *this earbud* allows one to infer the possibility of other nominals such as *the earbud*, *my earbud*. These are never called “back-formation”, presumably because they do not involve innovative word-forms.



- (25) a. *red* ‘the state of redness’  
 b. *redd-en* ‘a change into the state of redness’  
 c. *redd-en-ed* ‘a state of redness brought about by a change into the state of redness’

At first glance, these examples seem to be representative of what typically happens in morphological marking, but in the next subsection, we will see that the apparent general preference for forward-formation can be explained differently.

### 4.3. The diagrammatic stereotype is wrong for inflection

While it may intuitively be evident that “more meaning” is expressed by “more form”, this is actually wrong. Grammatical (and also lexical) forms serve to counter expectations on the part of the addressee, not to “convey meanings” (see Haspelmath 2025). This can be seen from contrasts like those in (26)–(31), where the asymmetry of coding can go in both directions. I discussed such examples in Haspelmath (2021) under the heading of “differential-coding pairs”, and I refer readers to that paper for further explanation.

- |      |         |   |   |
|------|---------|---|---|
| (26) |         | INDICATIVE  | IMPERATIVE  |
|      | Spanish | <i>canta</i> ‘(he) sings’<br><i>canta-s</i> ‘you sing’            | <b><i>que cante</i></b> ‘let him sing’<br><i>canta!</i> ‘(you) sing!’ |
| (27) |         | NONCAUSAL   | CAUSAL  |
|      | Russian | <i>kipet</i> ‘boil (intr.)’<br><i>slomat’-sja</i> ‘break (intr.)’ | <b><i>kipja-tit</i></b> ‘make boil’<br><i>slomat</i> ‘break (tr.)’    |
| (28) |         | AGENT (ERG)   | PATIENT (ACC)   |
|      | Dyirbal | <i>ngadya</i> ‘I (ERG)’<br><i>yarra-ŋgu</i> ‘man (ERG)’           | <b><i>ngaygu-na</i></b> ‘me (ACC)’<br><i>yarra</i> ‘man (ACC)’        |
| (29) |         | UNIPLEX   | MULTIPLEX   |
|      | Welsh   | <i>cath</i> ‘cat’<br><i>moron-en</i> ‘carrot’                     | <b><i>cath-od</i></b> ‘cats’<br><i>moron</i> ‘carrots’                |
| (30) |         | ATTRIBUTIVE   | PREDICATIVE   |
|      | English | <i>happy (children)</i><br><i>play-ing</i> (children)             | <i>(they) are happy</i><br><i>(they) play</i>                         |
| (31) |         | UNPOSSESSED   | POSSESSED   |
|      | Koyukon | <i>tel</i> ‘socks’<br><b><i>k’e-tlee</i></b> ‘head’               | <i>(se-)tel-e</i> ‘my socks’<br><i>(se-)tlee</i> ‘my head’            |

In all these cases, the marking depends on which grammatical meaning is less expected (and thus needs to be countered), not on “greater semantic specificity”, or on the presence of “operators in the semantic representation”. Givón was thus quite right to give another version of his quantity principle, as in (32) (Givón 1991: 87):

(32) **The quantity principle B**

“Less predictable information will be given a larger chunk of code.”

We can thus say that grammatical marking is generally explained by efficiency of coding, not by the nature of the rules that languages may have. The constraints are substantive rather than formal.

Thus, when we take a broader picture into account, phenomena that are similar to back-formation are not nearly as unusual as it may seem at first glance. But forward-formation has generally been taken as the default in the past, and linguists have typically adopted a perspective under which back-formation is unexpected or even conceptually anomalous. On the schema-based paradigmatic view of §2, this is much less the case.

But there is another question that we have not discussed yet: Is back-formation of synchronic relevance or is it only diachronically relevant? For pairs like *editor/edit*, let alone for *beggar/beg*, very few English speakers will be aware that the agent noun existed first in the language, so they do not seem to be special from a synchronic perspective. However, pairs like *babysitter/babysit* have a different flavour, and in §7 below, we will see what explains this difference.

## 5. The “word-making” stereotype and the inventorium

One of the main reasons why back-formation is a traditional problem is that “word-formation” is typically conceived of as a way of enriching the “lexicon”. For relatively “new words” such as *to babysit*, speakers often have the feeling that they are novel formations that have been added to the stock of words of their language. But what exactly is meant by “formation”, and what is the “lexicon”?

In many influential conceptions of language systems, these systems consist of two distinct COMPONENTS, a lexicon component and a morphosyntax component: The lexicon contains the set of elements that must be remembered, and the morphosyntax contains the regularities for combination, including the regularities for “forming” words. But as was noted earlier (§2), in the paradigmatic tradition (of works such as Booij 2010 and Jackendoff & Audring 2020a), there is no contrast between a “lexicon” and “grammatical schemas”, because words and constructions have the same basic nature, consisting of information from three domains (semantics, morphosyntax, and phonology). But if lexicon and grammar are non-distinct, and if word-formation schemas are non-directional, how do we express the intuition that the directionality was (or is) different in two cases such as (33a) and (33b)?

- (33) a. *to experience* → *experienc-er*<sup>4</sup> (forward-formation)  
 b. *to edit* ← *edit-or* (back-formation)

One way of putting it might be to say that the verb *to experience* was part of English speakers’ mental lexicons before the noun *experienter* was added to it, and that the noun *editor* was part of their mental lexicons before the verb *to edit* was added to it. However, we do not know what their mental lexicons were in the past, and we do not really know what they are now. Linguists can describe linguistic conventions very accurately, even for the past (when the documentation allows it), but we do not know much about mentalicons. While the conventions are uniform across speakers, their mentalicons may be quite different, reflecting their personal histories (e.g. the order in which they happened to acquire different forms).

Thus, to describe how a word is added to the “stock of words” of a language, we need a distinct notion: the set of words and complex forms (and constructions) which speakers

<sup>4</sup> The word *experienter* for a semantic role in experiential verbs is a fairly new formation in English, having been used only since the 1970s (apparently coined by Chafe 1970).

must remember in order to conform to the conventions: the INVENTORIUM of the language (Haspelmath 2024b). This can also be described as the set of “listemes” (e.g. Harley 2006: Ch. 4), though this term has also been used to describe the members of individual speakers’ stored elements. For a description of language change, however, it is crucial to have terms for the conventions of the language, because language change is change in the social conventions or norms, not (necessarily) change in the speakers’ mental lexicons.

Now let us consider a highly regular word-formation type where there are no relevant idiosyncrasies. An example is English deadjectival abstract nouns as exemplified in (34) (Gebhardt 2023: 119).

- (34) *wet*            *wetness*  
       *red*            *redness*  
       *heavy*        *heaviness*  
       *obtuse*       *obtuseness*  
       *gregarious* *gregariousness*

This is a typical example of a word-formation pattern, but is this a way of “enriching the lexicon”? The *-ness* words in the right-hand column are completely predictable, so there is no need to remember them (though many speakers may of course have some of them in their mental lexicons). While *wet*, *red*, *heavy*, *obtuse* and *gregarious* are part of the English inventorium, this is not the case with the *-ness* words.

Thus, derivational formation is not the same as “word-making” in the sense of adding words to the inventorium. If someone heard the word *gregariousness* for the first time without knowing the word *gregarious*, they could infer that it exists, because the two linked sister schemas have no particular directionality. If they actually used *gregarious* without having it heard before, would this be back-formation? How do we distinguish back-formation and forward-formation in such cases? Are there any reasons to think that the relationship between *gregarious* and *gregariousness* is any different than the relationship between *book* and *books* (see (20) above)?

The question can be extended to the word pair *edit/editor* (in (2a) and (33b)): If the word-formation process is productive, then how do we know that the verb *edit* is an innovation and did not exist at some earlier time? After all, we would not say that the word *gregariousness* did not exist at some earlier time: If a word-formation process is productive, the question of “existence” should be irrelevant. If English has a productive word-formation pattern of the type *bake/baker*, *hunt/hunter*, *write/writer*, then as soon as the word *editor* comes into the English inventorium (by borrowing it from Latin),<sup>5</sup> the verb *edit* comes into existence as well.

Thus, the application of a word-formation pattern does not necessarily mean that there is a change in the inventorium. However, morphologists often equate word-formation with “word-making” in the sense of adding “a new word” to the language. The following quotation from Geert Booij’s (2012) textbook is characteristic:

“Most complex words have been derived by one of the available word-formation processes of a language. Indeed, one of the main functions of morphology is to expand the set of available words. Once a complex word has been formed, it may get established as a word of the language. This means that it is used by more than one native speaker, and on different occasions, and that language users will recognize it as a word they have come across before. The set of established words of a language functions as the lexical norm or lexical convention of that language.” (Booij 2012: 17)

<sup>5</sup> I am assuming, of course, that the suffix *-or* is pronounced in the same way as *-er* and is thus not distinguishable from *-er* in *baker*, *hunter*, etc.

The problem is that Booij does not say clearly what he means by “available words” or “established words”. In the cognitively oriented view that most linguists adopt, there is little space for Booij’s “lexical norms”, and this notion is not often discussed by authors working in the Jackendoff-Aronoff tradition.

By contrast, in the tradition of Marchand (1969) and Bauer (1983), the notion of “established words” or “existing words” has often been discussed. This is the topic of the next section.

## 6. Against “institutionalization”: Existing words are always idiosyncratic

If we adopt a traditional view of language systems as consisting of a lexicon and one or more separate rule components (the morphosyntax, or morphology plus syntax), then it makes sense to think of some complex words as “existing” in the lexicon, in addition to word-formation rules that could give rise to words that have not been created yet and thus do not “exist” in the lexicon yet. These are potential words, but not “actual” or “existing” words (Bauer 2001: §3.2). For example, we could say that the noun *ambiclitic* is an existing word (used, e.g., by Haspelmath 2023b), while the derived verb *ambicliticize* and its nominal counterpart *ambicliticization* are only potential words because they have not been used yet and are not part of the English lexicon. The following quotation describes a widespread view among linguists:

“besides the words actually attested in dictionaries or corpora, the word-formation rules of a language usually also define an even greater number of potential complex words which have not yet been attested but could come into being at any moment if they should be deemed useful.” (Rainer 2012: 165)

But how do we know whether a word is an existing word or a potential word that has not yet “come into being”? Attestation in a corpus cannot be a sufficient criterion because all attested cases may have been “on-the-fly” uses, and it is not necessary either because many existing words are very rare. In general, corpora reflect both linguistic conventions and creative language use, so they do not allow us to distinguish easily between conventionalized and unconventional expressions.

The view I take here, in contrast to the traditional view, is that we should not think of language systems as consisting of separate components for words and rules (§2), and that all idiosyncrasies are part of the inventorium (as described in the previous section). Fully regular words like *gregariousness* are not part of the inventorium any more than fully regular inflected forms like *books* or fully regular phrases such as *this tree*. Word-formation does not imply addition to the inventorium. As a result, fully regular words such as *ambicliticization* are not part of the inventorium, and the status of “potential words” is in no way different from the status of “potential phrases”. If someone uses the phrase *this tree*, this does not “bring the phrase into existence” and it does not make the phrase “actual”. Likewise, if someone uses *gregariousness* for the first time, this does not make it “actual” or “existing”.

This view goes against a tradition in which the transition from potential to actual words is thought of as a process of gradual “institutionalization” (Bauer 1983; Lipka 1992; 2005; Lipka et al. 2004; Hohenhaus 2005: §3). According to this tradition, new words are first used as “ad hoc formations” (or “nonce formations”), and are institutionalized in a second step, in a social process of conventionalization. According to Lipka (1992; 2005), one needs to distinguish between a level of LANGUAGE SYSTEM

and a level of LANGUAGE NORM, following Eugenio Coseriu's conceptual distinction (Coseriu 1952; Kabatek 2020). But even though the distinction between *system* and *norm* became well-known (especially in Spanish-speaking and German-speaking countries; e.g. Dressler 1985: 322), it has not really been substantiated. Both language systems and language norms (in Coseriu's sense) are conventional, and which aspects of the conventions that we learn belong to the "system" as opposed to the "norm" remains unclear.

Thus, it is better to say that only those complex words (and other complex expressions) are part of the inventorium that show some idiosyncrasies, and that there are no fully predictable "existing" or "actual" words, just as there are no fully predictable "existing sentences". Linguists have sometimes felt that they can identify "established" words which are not irregular, and that identifying a word as "established/existing/institutionalized" is the best characterization. Booij (2012: 17) gives a concrete example:

"In British English the machine that is used for drawing money from one's bank account is called a *cash dispenser*, and in American English it is called an *automatic teller machine (ATM)*. In fact, it would also have been possible to use the compound *money machine* for this device, but the established words function as a lexical norm, and hence they can block the creation of the compound *money machine*. That is, the lexicon as the set of established lexical units of a language may have a blocking effect on the creation of new words. It does not mean that *money machine* is an ill-formed word, only that its use might not be appropriate."

The words *cash dispenser* and *automatic teller machine* are part of the inventorium (of British and American English, respectively), but they are not idiomatic in the sense that their meanings are non-compositional.<sup>6</sup> However, they are not predictable, as noted by Booij – all three examples are potential words based on the word-formation schemas, and the first two are inventorized as they are not fully predictable from the schema.

The notion of "institutionalization" as intermediate in the progression from ad hoc formations to lexicalization has also been criticized by Blank (2001: 1598). He notes that compounds and derivatives are typically somewhat idiosyncratic in their meanings, e.g. the English compound *wheelchair*.

"strictly speaking, the sense of a word-formation is never completely predictable from the meaning of its components. The simple combination of the senses of two words by compounding would lead to rather abstract meanings: a *wheelchair* then would be just a *chair* that is somehow characterized as having a *wheel* (or maybe several)." (Blank 2001: 1598)

Blank concludes that "the distinction between institutionalization and lexicalization should therefore be dropped" (2001: 1599) (see also Haspelmath 2024b: §7.3). Blank does not take into account the fact that many derived words are indeed fully regular, e.g. derivatives such as *gregariousness* (in (34) above) and ad hoc compounds such as *apple-juice seat* (Downing 1977: 818). Because these fully regular words are "lexical" in the sense that they are lexemes (rather than inflected forms), I would say that we should not even talk about "lexicalization". Instead, the term *inventorization* is a good description of a change from a fully regular composite form to an idiosyncratic (unpredictable, or even non-compositional) form.

Inventorization may happen in a slow and gradual fashion, for example when formal idiosyncrasies develop, or when a word-formation pattern gradually loses productivity.

---

<sup>6</sup> An idiomatic expression is one in which the parts have meanings that are incompatible with the overall meaning, e.g. *honeymoon* (which denotes to a post-wedding vacation and has no direct relation to 'honey' or 'moon'). Many conventional compounds are semantically idiosyncratic, but not idiomatic in this sense.

For example, the English 3<sup>rd</sup> person singular form *says* [sez] developed a short vowel (contrasting with the long vowel in *say*) and is thus inventorized, perhaps through a gradual change. And the English noun *replacement* is inventorized (even though it is semantically and formally regular) because the word-formation pattern in *-ment* is no longer productive, so that its existence cannot be predicted. However, inventorization can also happen instantaneously, as in *word coinage*. This is discussed in the next section.

## 7. Word coinage

Much of the time, “new” words are formed in the same way as “new” sentences are formed, i.e. regularly and productively. This applies not only to inflected forms, but also to many derived forms and compounds (we saw a few derived examples in (34) above). The notion of *productivity* has mostly been discussed in the context of word-formation because it seems to be unproblematic in inflection and in syntax. However, the most famous definition of productivity in word-formation includes the idea that productive word-formation is unintentional and unlimited (Schultink’s definition, discussed e.g. by Bauer 2005: xxx), so if we adopt that definition, the most typical cases of word-formation do not exhibit productivity. For example, cases such as (35)-(37) from English, French and German are show neologisms, which are have properties that are not predictable.

- (35) English  
 N>V                      *to mayonnaise* ‘to cover with mayonnaise’  
 -ee                        *festschriftee*    ‘scholar given a festschrift’
- (36) French (Dal & Namer 2016: 73, 78)  
 -itude                    *corsitude*        “Corsicanness”  
 de-...-isation         *dératisation*    “de-ratting”
- (37) German  
*sich ver-...*            *sich verklicken* ‘click wrongly’  
*Klima, kleben*         *Klimakleber*    ‘gluing climate activists’

Instead, cases such as (35)-(37) can be said to represent *word coinage*, i.e. the creation of an element of the inventorium on the basis of a regularity that is not fully productive or with an outcome whose properties are not fully predictable. Word coinage could also be said to be “instantaneous inventorization”. Such coined words appear novel to language users, and this novelty may last for quite a while. It appears that all nsalient back-formations such as the English verb *to babysit* are word coinages.

Word coinage is very common in modern societies, where new technical terms and new social buzzwords are created all the time (see also Keyes 2021, who discusses word coinage from the perspective of a writer). For example, technical terms such as *cash dispenser* or *automatic teller machine* (mentioned in §6) were presumably coined by the businesses when they first created these technologies. They do not involve the sort of gradual semantic or formal change that tends to be the focus of linguists’ discussions of morphological change. They do not fall under Schultink’s classic definition of productivity, which makes reference to “creating” or “forming” (cited after ten Hacken 2020):

“Under productivity as a morphological phenomenon, we understand the possibility for language users to create new formations unintentionally and in an in principle

not countable number, by means of a morphological procedure that is at the basis of the form-meaning correspondence of some words they know.” (Schultink 1961: 113)

In general, word coinage (or instantaneous inventorization) is surely intentional, although intentionality is not part of its definition (which only requires the lack of full predictability). By contrast, creating or forming regular words (whether inflectional or derivational) is something that we do unintentionally all the time, making use of productive regularities (i.e. sets of sister schemas).

The word “coinage” has rarely been treated as a technical term in morphology, as far as I am aware. On the contrary, “coining a word” has sometimes been used synonymously with “forming a word”, in particular in the most widely quoted translation of Schultink’s classic definition:

“By productivity as a morphological phenomenon we understand the possibility for language users to coin... a number of formations...” (translation by van Marle 1985: 45)

But Schultink was simply talking about “forming new formations” (*nieuwe formaties te vormen*), not about “coining”. The difference in ordinary English is clear, and this distinction should be made more clearly by linguists as well.<sup>7</sup>

## 8. Back to back-formation

Let us now go back to back-formation and ask three questions that already came up earlier: (i) Is back-formation a purely diachronic process or can it be a synchronic word-formation process? (ii) What is the definition of “back-formation”? (iii) Why are back-formations not more frequent than they seem to be? The answers that I propose will make use of the discussion of the last six sections.

### 8.1. Diachronic or synchronic?

Some linguists have said that back-formations are restricted to diachrony and should not be part of synchronic word-formation patterns, e.g.

“Once back-formation has occurred, it becomes invisible to speakers; linguistically naïve contemporary speakers have no reason to think, for example, that *peddle* was derived from *peddler*, rather than the other way around.” (Bauer et al. 2013: 20)

It is well-known that Marchand (1969) and Kiparsky (1982) are among those who have said that back-formation is a purely diachronic process, while Plag (2003: 178) has said that it is also synchronically relevant because back-formation can be productive.

The first point to note here is that the diachronic vs. synchronic question can also be asked of ordinary forward-formations when they are unproductive, e.g.

(38) some English demonyms  
*Spain*                      *Span-iard*

<sup>7</sup> Some linguists have made a distinction between creativity on the one hand, and productivity in a narrow sense on the other (see Bauer 2001: §3.10), and perhaps these linguists would say that creative word formation is the same as what is called coinage here. However, the forming of completely novel regular sentences is typically treated as an aspect of “linguistic creativity” (e.g. Chomsky 1982), so I feel that *word coinage* is best opposed to the “formation/creation of words”.

*Cyprus*      *Cypr-iot*  
*Italy*        *Ital-ian*

Diachronically, one can probably say that *Spain*, *Cyprus* and *Italy* existed first, before the demonyms came in to existence, and these patterns are no longer synchronically productive, so they may only be relevant diachronically.

However, even unproductive patterns may have a relational role (see §2 above), so they may be synchronically relevant, too. But there does not seem to be a difference between forward-formations and back-formations in this regard: Both may be synchronically relevant, and both may be primarily diachronic phenomena in the sense that we only know about the first attestation from the historical records.

## 8.2. Defining forward-formation, back-formation and back-coining

But how exactly is back-formation defined? If the formation of a word simply means that a schema is applied, then we can say the following (for affixal formation; compounding and other types are left aside here for simplicity):

- (39) a. Forward-formation is the formation of a word with an affix on the basis of a schema or word lacking an affix in this position.  
 b. Back-formation is the formation of a word lacking an affix in some position on the basis of a schema or word having an affix in this position.

For example, forming *books* on the basis of *book* (or a general singular noun schema) is forward-formation according to this definition, while forming *book* on the basis of *books* (or a general plural noun schema) is back-formation.

But this does not correspond to the way in which the term “back-formation” is generally used in the literature. Instead, “back-formation” is most typically used for a subtype of word coining as described in §7. We can call this “back-coining”, thus coining a novel term (via a kind of cross-formation).

- (40) **back-coining**  
 Back-coining is the formation of a word with unpredictable properties on the basis of a longer word that can be treated as complex making use of an unproductive regularity or an isolated analogy.

For example, the verb to *babysit* was coined on the basis of the longer word *babysitter*, which can be treated as complex (*babysitt-er*). The regularity is not productive, as it is not possible to form such verbs from all nouns ending in *-er/-or* (*\*to auth* from *author*, *\*to trendset* from *trendsetter*, *\*to booklove* from *booklover*, etc.). Thus, it is a kind of word coinage. It has also been noted that in many cases, word coinage need not be regular formally either, and this is also found in some back-formations, such as to *enthuse* (from *enthusiastic*) or to *surveille* (from *surveillance*) (see Nagano 2006: 260).

Thus, in the present perspective, back-formation is not more special than forward-formation, but back-coining is of course much rarer than “forward-coining”, as we will see in the next subsection.

## 8.3. Why is back-formation (or rather back-coining) not more common?

Finally, let us ask why back-formation, or rather back-coining, has a special feel. Verbs such as to *ghostwrite* (from *ghostwriter*) sound unusual and are not common. It happens



occasionally that a verb such as to edit becomes perfectly normal and unremarkable, but forward-coining is much more common.

We already saw that Štekauer (2015: 340) discussed the rarity of back-formation in the context of morphological naturalness, but we also saw that inflectional back-formations are perfectly normal. What does seem to be rare is back-coinage, but why should this be so?

It seems that the answer is fairly straightforward: There simply are not very many opportunities for novel back-formation to arise. Complex words that are in principle analyzable (such as *editor*) do not often come into a language, so there are not a whole lot of opportunities for word coinage, or for the development of a productive regularity. Forward-coinage has the great advantage that it can operate on the basis of shorter words, of which all languages have a lot. Back-coinage must be based on longer words, of which languages have far fewer.

Thus, if we adopt a paradigmatic perspective on morphology where there is no distinction between morphosyntax and lexicon (§2), and no distinction between “potential” and “existing” expressions (§6), back-formation and forward-formation are not really asymmetric, and it is fairly easy to understand why back-coinage is much rarer than forward-coinage.

## References

- Anderson, Stephen R. 2015. The morpheme: Its nature and use. In Baerman, Matthew (ed.), *The Oxford handbook of inflection*, 11–33. Oxford: Oxford University Press.
- Anderson, Stephen R. 2020. Semantically subtractive morphology. In Körtvélyessy, Lívía & Štekauer, Pavol (eds.), *Complex words: Advances in morphology*, 36–54. Cambridge: Cambridge University Press.
- Audring, Jenny. 2022. Advances in morphological theory: Construction Morphology and Relational Morphology. *Annual Review of Linguistics* 8. 39–58. (doi:<https://doi.org/10.1146/annurev-linguistics-031120-115118>)
- Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.
- Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.
- Bauer, Laurie. 2022. Thinking about back-formation (informal workshop contribution).
- Bauer, Laurie & Lieber, Rochelle & Plag, Ingo. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
- Becker, Thomas. 1990. *Analogie und morphologische Theorie*. München: Fink.
- Becker, Thomas. 1993. Back-formation, cross-formation, and ‘bracketing paradoxes’ in paradigmatic morphology. In Booij, Geert & van Marle, Jaap (eds.), *Yearbook of Morphology 1993*, 1–25. Dordrecht: Springer. (DOI: 10.1007/978-94-017-3712-8\_1)
- Blank, Andreas. 2001. Pathways of lexicalization. In Haspelmath, Martin & König, Ekkehard & Oesterreicher, Wulf & Raible, Wolfgang (eds.), *Language typology and language universals: An international handbook (volume 2)*, 1596–1608. Berlin: Walter de Gruyter. (doi:10.1515/9783110194265-049)
- Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42(3). 531–573. (doi:10.1017/S0022226706004191)
- Blevins, James P. & Blevins, Juliette. 2009. Introduction: Analogy in grammar. In Blevins, James P. & Blevins, Juliette (eds.), *Analogy in grammar: Form and acquisition*, 1–12. Oxford: Oxford University Press.
- Bloomfield, Leonard. 1933. *Language*. New York: H. Holt and Company.
- Bochner, Harry. 1993. *Simplicity in generative morphology*. Berlin: Mouton de Gruyter.
- Booij, Geert E. 2010. *Construction morphology*. Oxford: Oxford University Press.
- Booij, Geert E. 2012. *The grammar of words: An introduction to linguistic morphology (3rd edition)*. Oxford: Oxford University Press.

- Cetnarowska, Bożena. 1993. *The syntax, semantics and derivation of bare nominalisations in English* (Prace Naukowe Uniwersytetu Śląskiego w Katowicach). Katowice: Uniwersytet Śląski.
- Chafe, Wallace L. 1970. *Meaning and the structure of language*. Chicago: The University of Chicago Press.
- Chomsky, Noam. 1982. A note on the creative aspect of language use. *The Philosophical Review* 91(3). 423–434. (doi:10.2307/2184692)
- Dal, Georgette & Namer, Fiammetta. 2016. Productivity. In Hippisley, Andrew & Stump, Gregory (eds.), *The Cambridge handbook of morphology*, 70–89. Cambridge: Cambridge University Press.
- Diessel, Holger. 2019. *The grammar network*. Cambridge: Cambridge University Press.
- Dressler, Wolfgang U. 1985. On the predictiveness of Natural Morphology. *Journal of Linguistics* 21(2). 321–337. (doi:10.1017/S002222670001029X)
- Dressler, Wolfgang U. 2005. Word formation in Natural Morphology. In Štekauer, Pavol & Lieber, Rochelle (eds.), *Handbook of word-formation* (Studies in Natural Language and Linguistic Theory 64), 267–284. Dordrecht: Springer. (doi:10.1007/1-4020-3596-9\_1)
- Gaeta, Livio & Montermini, Fabio. 2022. Back-formation in a new theoretical universe (workshop description, IMM Budapest). (<http://www.nytud.hu/imm20/>)
- Gebhardt, Lewis. 2023. *The study of words: An introduction*. London: Routledge. (doi:10.4324/9781003030188)
- Givón, T. 1991. Markedness in grammar: Distributional, communicative and cognitive correlates of syntactic structure. *Studies in Language* 15(2). 335–370. (doi:10.1075/sl.15.2.05giv)
- Grestenberger, Laura & Kastner, Itamar. 2022. Directionality in cross-categorial derivations. *Glossa: a journal of general linguistics*. Open Library of Humanities 7(1). (doi:10.16995/glossa.8710) (<https://www.glossa-journal.org/article/id/8710/>) (Accessed August 31, 2022.)
- Harley, Heidi. 2006. *English words: A linguistic introduction*. Malden, MA: Blackwell.
- Haspelmath, Martin. 2002. *Understanding morphology*. London: Arnold. (<https://zenodo.org/record/1236482>)
- Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1). 1–33.
- Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics* 57(3). 605–633. (doi:10.1017/S0022226720000535)
- Haspelmath, Martin. 2022. Against lexicalization (and what to replace it with). London: UCL. (doi:<https://zenodo.org/record/6408756>)
- Haspelmath, Martin. 2023a. Defining the word. *WORD* 69(3). 283–297. (doi:10.1080/00437956.2023.2237272)
- Haspelmath, Martin. 2023b. Types of clitics in the world’s languages. *Linguistic Typology at the Crossroads* 3(2). 1–59. (doi:10.6092/issn.2785-0943/16057)
- Haspelmath, Martin. 2024a. Inflection and derivation as traditional comparative concepts. *Linguistics* 62(1). 43–77. (doi:10.1515/ling-2022-0086)
- Haspelmath, Martin. 2024b. Four kinds of lexical items: Words, lexemes, inventorial items, and mental items. *Lexique: Revue en Sciences du Langage* (34). 71–95. (doi:10.54563/lexique.1737)
- Haspelmath, Martin. 2025. Ambiguity avoidance vs. expectation sensitivity as functional factors in grammatical change (to appear). In Gvozdanović, Jadranka (ed.), (*A volume with papers from ICHL Heidelberg*).
- Hockett, Charles F. 1954. Two models of grammatical description. *WORD* 10(2–3). 210–234.
- Hohenhaus, Peter. 2005. Lexicalization and institutionalization. In Štekauer, Pavol & Lieber, Rochelle (eds.), *Handbook of word-formation*, 353–373. Dordrecht: Springer. (doi:10.1007/1-4020-3596-9)
- Iacobini, Claudio. 2000. Base and direction of derivation. In Booij, Geert E. & Lehmann, Christian & Mugdan, Joachim (eds.), *Morphology: An international handbook on inflection and word-formation (Volume 1)*, 865–876. Berlin: Walter de Gruyter.

- Jackendoff, Ray. 2013. Constructions in the Parallel Architecture. In Hoffmann, Thomas & Trousdale, Graeme (eds.), *The Oxford handbook of Construction Grammar*, 70–92. Oxford: Oxford University Press.  
(<https://doi.org/10.1093/oxfordhb/9780195396683.013.0005>)
- Jackendoff, Ray & Audring, Jenny. 2020a. Relational Morphology: A cousin of Construction Grammar. *Frontiers in Psychology* 11.  
(<https://www.frontiersin.org/articles/10.3389/fpsyg.2020.02241>)
- Jackendoff, Ray & Audring, Jenny. 2020b. *The texture of the lexicon: Relational Morphology and the Parallel Architecture*. Oxford: Oxford University Press.
- Kabatek, Johannes. 2020. Linguistic norm in the linguistic theory of Eugenio Coseriu. In Lebsanft, Franz & Tacke, Felix (eds.), *Manual of standardization in the Romance languages*, 127–144. Berlin: De Gruyter.
- Kiparsky, Paul. 1982. Lexical Morphology and Phonology. In Yang, In-Seok (ed.), *Linguistics in the Morning Calm: Selected Papers from SICOL-1981*, 3–91. Seoul: Hanshin.
- Koontz-Garboden, Andrew. 2009. Anticausativization. *Natural Language & Linguistic Theory* 27. 77–138. (doi:10.1007/s11049-008-9058-9)
- Koontz-Garboden, Andrew. 2012. The Monotonicity Hypothesis. In McNally, Louise & Demonte, Violeta (eds.), *Telicity, change, and state: A cross-categorical view of event structure*, 139–161. Oxford: Oxford University Press.
- Lipka, Leonhard. 1992. Lexicalization and institutionalisation in English and German. *Linguistica Pragensia* 1(1–13).
- Lipka, Leonhard. 2005. Lexicalization and institutionalization: Revisited and extended. *SKASE Journal of Theoretical Linguistics* 2(2). 40–42.
- Lipka, Leonhard & Handl, Susanne & Falkner, Wolfgang. 2004. Lexicalization and institutionalization: The state of the art in 2004. *SKASE Journal of Theoretical Linguistics* 1. 2–19.
- Marchand, Hans. 1964. A set of criteria for the establishing of derivational relationship between words unmarked by derivational morphemes. *Indogermanische Forschungen* 69. 10–19.
- Marchand, Hans. 1969. *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. Munich: C.H. Beck.
- Mayerthaler, Willi. 1981. *Morphologische Natürlichkeit*. Wiesbaden: Athenaion.
- Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.
- Rainer, Franz. 2012. Morphological metaphysics: virtual, potential, and actual words. *Word Structure* 5(2). 165–182. (doi:10.3366/word.2012.0025)
- Štekauer, Pavol. 2015. Backformation. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation*, 340–352. Berlin: De Gruyter Mouton.  
(doi:10.1515/9783110246254-016)
- ten Hacken, Pius. 2020. Productivity and anticipation in language processing. *SKASE Journal of Theoretical Linguistics* 17(4). 1–17.
- Valera, Salvador. 2015. Conversion. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe (Volume II)*, 322–339. Berlin: De Gruyter Mouton.  
(doi:10.1515/9783110246278)

Martin Haspelmath  
 Max Planck Institute for Evolutionary Anthropology  
 Deutscher Platz 6  
 04109 Leipzig  
 Germany  
 martin\_haspelmath@eva.mpg.de