

CLIPP

Christiani Lehmanni inedita, publicanda, publicata

titulus	Univerbation
huius textus situs retis mundialis	christianlehmann.eu/publ/lehmann_univerbation.pdf https://doi.org/10.1515/flih-2020-0007
dies manuscripti postremum modificati	20.10.2020
ocasio orationis habitae	II Coloquio de Lingüística Histórica, Pontificia Universidad Católica de Chile, 15 – 16 noviembre, 2018
volumen publicationem continens	<i>Folia Linguistica Historica</i> 41
annus publicationis	2020
paginae	205-252

Univerbation

Christian Lehmann

University of Erfurt

Abstract

Univerbation is the syntagmatic condensation of a sequence of words recurrent in discourse into one word, as when the Spanish combination *a tras* (to back) becomes *atrás* ‘behind’. It affects both lexemes and grammatical formatives. Unlike processes of word formation, including conversion of a syntactic construction into a word, as in *forget-me-not*, and compounding, as in Spanish *lavaplatos* ‘dishwasher’, univerbation is a spontaneous process.

There are two main types of univerbation: Phrasal univerbation downgrades a phrase to a word, as when Latin *terrae motus* ‘earth’s movement’ becomes Spanish *terremoto* ‘earthquake’. Transgressive univerbation coalesces a string of words which do not form a syntagma into a word, as when French *par ce que* becomes *parce que*.

A set of univerbations may share structural features and may therefore evolve into a pattern of compounding. Thus, *blackbird* originated by univerbation, but may now provide a pattern of compounding. As a consequence, univerbation and compounding are not always easily distinguishable.

The discussion uses empirical evidence adduced in earlier work, mostly from Romance and Germanic languages. Its aim is not to present novel phenomena but to provide a theoretical background for the phenomenology and improve on available analyses.

Keywords: univerbation, chunking, compounding, lexicalization, collocation, word

1 Introduction¹

The primary goal of this contribution is to characterize the concept of univerbation and to give a systematic account of what is and what is not covered by it. The problem with this concept is not so much that it would have been understood in different ways or that there was some terminological confusion. Both the concept and its term have remained rather constant in the discipline for more than a century. What does require some attention are two issues: the concept, its term and the set of phenomena covered by it deserve to be more widely known and to be more firmly established in linguistics. Second, the place of univerbation and its relation to similar concepts in a theory of language need to be clarified. This is what will be attempted here. Examples are variously taken from languages with sufficient historical documentation, chiefly Germanic and Romance languages.²

The section following this introduction will be devoted to the concept and the term of univerbation and its relation to synchrony and diachrony. Next, a systematic survey of relevant

¹ Versions of this paper were presented at the Linguistic Colloquia of Universities of Göttingen (6/11/2018), Bamberg (12/12/2018), Berlin (22/05/2019) and Mainz (21/10/2019). I thank their participants, Beate Hampe and Bridget Drinka for helpful discussion and three anonymous reviewers of this journal for thorough criticism.

² Extensive collections of examples from diverse Indo-European languages are in Brugmann 1900, § 3 and Paul 1920, ch. XIX, many of which are used here without separate acknowledgement.

phenomena will be presented, and two main types, viz. phrasal and transgressive univerbation, will be distinguished and shown to be relevant in empirical generalizations. With a more concrete idea of the kinds of phenomena involved, §4 then proceeds to delimiting univerbation against similar phenomena like clisis, compounding, conversion to word and parasynthesis. §5 compiles the symptoms by which the analyst may recognize univerbation where it has occurred. Among the processes that have been excluded in §4, compounding is particularly problematic because compounding and univerbation are both intimately interrelated in concrete historical cases and are often so similar that they are not easily distinguished. This similarity is the subject matter of §6. §7 finally attempts to identify the causes of univerbation in language activity.

Univerbation is the union of two syntagmatically adjacent word forms into one. It may be formalized as the downgrading of a syntactic to a morphological boundary. A clear example of univerbation to begin with is the original formation of deadjectival adverbs in *-mente* in the Romance languages as illustrated by E1:

E1	LATIN (Cat. <i>Carm.</i> 8, 11)	ITALIAN
	obstinata mente	> ostinatamente
	obstinate:ABL.SG mind:ABL.SG	obstinate:ADVR
	‘in an obstinate spirit’	‘obstinately’

While this adverbialization is today a synchronic process of derivation, its origin in Vulgar Latin is the morphological union of an adverbial phrase in the ablative, consisting of a noun and its adjective attribute. The syntactic boundary separating the two word forms becomes a morphological boundary inside one word.

Univerbation happens both to grammatical formatives and to lexemes. In the former case, no problem of distinguishing it from morphological processes, especially from inflection, arises.³ In the latter case, univerbation contrasts with word formation. Both are processes that produce complex words. It is true that they are interrelated in complex ways, that they are not always easily distinguished in specific cases and that earlier linguistic terminology has sometimes blurred the distinction, as we shall recapitulate in §2.⁴ However, their empirical and historical manifestations differ, so they have to be distinguished on a theoretical basis. The Italian formations of E2 are clearly compounds not traceable to any syntactic constructions in the known history of Italian.

E2	crocevia	nullatenente
ITALIAN	CROSS:way	nothing:having
	‘crossroads’	‘destitute’

There are excellent early treatments of our topic,⁵ among them Karl Brugmann (1900) and Hermann Paul (1920, chapters 4, 10, 19). Both take it for granted that compounding comes about by univerbation (s. §2 for their use of terms). Brugmann (§4) explicitly excludes “ana-

³ To be sure, the gradient character of univerbation (agglutination, in this case) renders it sometimes difficult to tell when a formative has become one of inflection. This problem, however, has never been one of distinguishing univerbation from inflection.

⁴ Apparently, there are entire traditions which ignore the distinction. Practically all of the examples adduced s.v. *mot composé* in the French wikipedia (https://fr.wikipedia.org/wiki/Mot_composé, consulted 26/05/2019) are ones of univerbation. A scientific work representing this tradition is Silberstein 1990.

⁵ Among modern book-length historical studies, Baché 2013 for German and Opfermann 2016 for Latin-Romance univerbations may be mentioned.

logische Nachahmung schon fertiger Komposita” (‘analogical imitation of pre-existent compounds’) from univerbation and from his treatment. In other words, what is compounding in modern terms is, for Brugmann, not a rule-governed process of the language system, but just application of analogy on a pattern in whose origin he is interested. With respect to this origin, Brugmann argues for a strict ‘function leads form’ position: *Komposition* is, in the first place, a semantic process of forming a concept on the basis of two concepts hitherto coded in two words or stems, and of expressing the new concept by combining these pre-existent components. Their structural union is something secondary which may happen with some delay and to different degrees.

In his brief reply, Hermann Paul (1903) reminds Brugmann that univerbation happens both in lexicalization and in grammaticalization (without, of course, using these more recent terms). Only in the former could semantic isolation (§5.3) be the leading factor, since in grammaticalization, compositionality is preserved.

2 Univerbation: concept and term

2.1 The term

A few terminological remarks are necessary.⁶ First, it should be stressed that univerbation is in no way confined to verbs; the general meaning of Latin *verbum* is ‘word’. Second, the concept of ‘univerbation’ was formed by German-speaking Indo-Europeanists. From 1850 on, the term *Zusammenrückung* (action noun from *zusammenrücken* ‘move [sth.] closer together’) was used to designate it⁷ and was regularly contrasted with *Komposition* ‘compounding’. The term *Univerbierung* appears first in Brugmann 1905 and has since been treated as synonymous with *Zusammenrückung*.⁸

2.2 The concept

The term ‘univerbation’ is not rarely (e.g. in Haspelmath 2002, ch. 3.3.2) restricted to the formation of lexemes, excluding the process from the realm of grammaticalization.⁹ Here the term will be used in the sense established in research on grammaticalization and lexicalization to designate the union of two (or, rarely, more) syntagmatically adjacent word forms into one. While there is no doubt that univerbation in this sense is intimately connected with word formation, §6 will conclude that it is not a kind of word formation.

A first methodological consequence may already be derived from the definition: For a complex word to be analyzable as formed by univerbation at a given time presupposes that

⁶ Although the term is defined in all major editions of the wikipedia and other dictionaries, it is not commonly known in linguistics and is, for example, absent from current treatments of chunking in ‘usage-based linguistics’.

⁷ It appears to have been wrongly translated into other languages. It does not mean ‘juxtaposition’ and much less ‘amalgamation’.

⁸ French *univerbation* is a lemma in Marouzeau 1951. First occurrences of English *univerbation* appear in publications of the 1960s.

⁹ Occasionally, the term is even restricted further to patterns of word-formation which create a denomination on the model of a syntactic construction (Martincová 2015: 742, illustrating with Czech *plenární schůze* ‘plenary meeting’ > *plenárka* id.). There, univerbation is a paradigmatic relationship between a more explicit and a more condensed construction.

there be, at that time, a syntactic construction containing the components as words in juxtaposition (cf. Diagram 1). The concept of syntactic construction must here include both fully analytic syntactic and phraseological constructions. Phraseology differs from analytic syntax, among other things, with regard to how extensively constraints may apply. An example of this kind which will be taken up in §5.1.3 consists in nominal groups lacking an article in phraseologisms which they would require in analytic syntax. For instance, English *instead (of)* = German *anstelle (von)* = Spanish *en lugar (de)* comprise a noun without an article which would have to be there if rules of analytic syntax obtained.¹⁰ Such phraseological deviations from analytic syntax must be allowed for by the said condition that univerbation presupposes a syntactic construction. Given this condition, French *pomme-de-terre* (apple-of-earth) ‘potato’ is on its way to univerbation, but compounds like Engl. *pickpocket*, Spanish *girasol* (turn:sun) ‘sunflower’ or those of E2, which resemble certain syntactic constructions, cannot have originated by univerbation, since there is no syntactic construction of English or Spanish which could contain the sequence *pick pocket* or *gira sol*. We will see that this is a rather severe methodological requirement which throws doubt on some extant univerbation analyses.

Univerbation is a phase in a process of coalescence (Haspelmath 2011a). This has the following implications: First, univerbation is itself a gradient process which displays phases of weaker and stronger univerbation. In principle, univerbation takes place at the moment that a construction is converted into a word. Since, however, the concept of word itself does not have neat boundaries, this process is not, in fact, an instantaneous conversion, but rather a transition. Stronger univerbation evolves as the symptoms of univerbation to be reviewed in §5 accumulate.

The second implication of the subsumption of univerbation under coalescence is that, both in grammaticalization and in lexicalization, some phases of coalescence precede and others follow univerbation, as visualized in Diagram 1. Preceding univerbation is the genesis of a collocation and its syntactic fixation. Following it is the loss of the morphological boundary and phonological fusion. The examples of Table 1 illustrate these phases of coalescence; the first and third examples instantiate univerbation of lexemes, while the second is a univerbation of grammatical forms.

Diagram 1 Phases of coalescence

coalescence	incipient			advanced
phase	juxtaposition	> univerbation	> welding	> fusion

¹⁰ Fillmore et al. 1988 offer the concepts of “familiar pieces unfamiliarly arranged” (§1.2.2) and of “extragrammatical constructions” (§1.1.2) for a relevant subset of idioms. However, lack of the article is frequent in the phraseology of the three languages named. Some phraseology is not exactly extragrammatical, but rather exhibits subregularities which deviate from the rules of analytic syntax. The lack of the article in such expressions might signal lack of referentiality.

Table 1 Latin examples of coalescence

<i>terrae mōtus</i> earth(F):GEN.SG movement(M):NOM.SG	>	<i>terraemōtus</i> earthquake(M):NOM.SG	
		<i>*ne-ullus</i> not-any:M.NOM.SG	> <i>nūllus</i> no(body):M.NOM.SG
			<i>*po-s(i)nō</i> > <i>pōnō</i> TERM-let:1.SG put:1.SG

Needless to say, none of this is obligatory in any particular process. However, grammaticalization and lexicalization do not bypass univerbation; the ensuing processes presuppose the unity of the word.

Univerbation is an aspect of linguistic variation. As such, it has two theoretically fundamental properties:

- (1) It happens in speech (*parole*) rather than in the linguistic system (*langue*).
- (2) It has synchronic and diachronic aspects.

As for property #1, the decision of taking an analytic or a holistic approach to a construction happens in speech. If the holistic approach is taken, the construction is treated as unitary, as an element of the inventory. If this approach is taken routinely, the construction is treated like a word. If its components are adjacent, this may well lead to its univerbation. The point here is that univerbation is not an operation of the linguistic system. It does not apply within the system; it changes it. Property #2 will be discussed more fully in the next subsection.

2.3 Synchrony vs. diachrony

Concepts like grammaticalization, lexicalization, coalescence, univerbation, etc., are frequently defined as diachronic processes.¹¹ However, this is insufficient, since all linguistic variation has a synchronic and a diachronic side; the other side of diachronic change is necessarily synchronic variation.¹² All of these processes are kinds of directed linguistic variation; but the direction appears only in the diachronic perspective. This is easily demonstrated for univerbation. The synchronic manifestation of univerbation is variation in the treatment of a construction as one word or as a combination of words. A simple criterion of univerbation is the writing of the expression concerned as one word (s. §5.4.3). Since orthography is conservative, a norm of writing a combination in one word – in an orthography that provides for word separation – is usually a good indicator that univerbation has taken place at some earlier point in time. Now the last German orthography reform was hesitant as to the writing of complex prepositions: During the 20th century, complex denominal prepositions such as *aufgrund* (on:basis; Table 8) had been written together for decades. The first launch of the spelling

¹¹ Not seldom, they are even defined as historical processes. However, a historical phenomenon is by definition unique. A concept of change that is meant to be generally applicable is diachronic, not historical.

¹² The widespread reluctance to acknowledge the synchronic side of such processes is probably due to two features characterizing the history of the linguistic discipline: First, because of the lack of contact between historical and descriptive linguistics, concepts which developed in historical linguistics are not easily adopted into theories of general linguistics. Second, despite the “variationist” trend of the 1970s, general and descriptive linguistics is to this day averse to variation because it is hard to accommodate in theories of the linguistic system.

reform in 1996 required us to write them separately again, as *auf Grund*. The final launch, in 2006, reverted to the joint writing *aufgrund*. Unless we want to assume two diachronic changes back and forth within ten years, it seems more adequate to regard the observable variation as the synchronic side of an ongoing diachronic change.¹³ Univerbation is a process of directed linguistic variation.

The entire field of word formation is a prime *locus* of the presence of diachrony in synchrony. Early researchers were confused by this fact. As a visible consequence of this confusion, the (synchronic) grammars of most languages lack an account of word formation, on the false assumption that word formation only happens in diachrony. In reality and to put it simply, there are patterns and rules of word formation in every (synchronic) language system. The difference from rules of other parts of the language system is that the products of the grammar are volatile, while the products of word formation have a better chance of getting lexicalized and may thus survive into later stages of the language. Words belonging to present-day English like *enrich* and *enlarge* are clearly products of rules of word formation, but not of rules of present-day English.

Change of an expression from some stage 1 to a later stage 2 may involve a reanalysis, i.e. the result that the expression belongs to different construction types at the two stages. E3 instantiates incorporation synchronically, but at its origin is a backformation from *Staubsauger* ‘vacuum cleaner’, a nominal compound. §5.6 discusses the relation between univerbation and reanalysis.

E3	staubsaugen
NHG	dust:suck:INF
	‘to vacuum’

A typical outcome of lexical change is the following: An item which may have been fully regular at stage 1 becomes only partly analyzable at stage 2. The reasons lie in its isolation, i.e. the process by which a lexicalized item becomes isolated from the regular patterns of the system (Brugmann 1900 and Paul 1920, ch. X and §229).¹⁴ The mapping of meaning on expression may cease to be compositional, as in NHG *versuchen* ‘try’, etymologically ‘forward:seek:INF’. While a sufficient number of products of a certain process that was productive at stage 1 survives into stage 2 so that they remain morphologically analyzable, the process itself is unproductive at stage 2. Examples include English *enrich* and *enlarge* and NHG *wundersam* (wonder:ADJVR) ‘miraculous’ (suffix now unproductive) and *fruchtbar* (fruit:ADJVR) ‘fruitful’ (suffix no longer productive with nominal bases).

In §6, the relationship between univerbation and compounding will be seen as a particularly complex interplay of different factors in synchrony and diachrony.

3 Types of univerbation

Univerbation is by definition spontaneous and happens *in situ*. We can abduce this defining feature from a simple example like E4 (cf. Table 1).

¹³ Fiehler 2011 shows the same for a set of German collocations that are currently in the process of univerbation.

¹⁴ Bybee 2010, ch. 3.4.2 tacitly rededicates the term ‘autonomy’, which from Meillet 1912 to Brinton & Traugott 2005, ch. 2.4 has been used to designate that property of a linguistic sign which gets lost in grammaticalization, to the meaning of ‘isolation’.

E4	CLASSICAL LATIN		LATE LATIN	ITALIAN/SPANISH
	terrae	motus	> terraemotus	> terremoto
	earth(F):GEN.SG	movement(M):NOM.SG		earthquake:M
	‘movement of the earth’			‘earthquake’

Two words, each of their grammatical category, are adjacent in a syntactic construction. They are condensed into a single word (see §5.1.2 for details); but this does not require a specific linguistic operation nor does it presuppose or provoke a change in the context. After univerbation, everything else remains as it was before. Univerbation happens spontaneously *in situ*.

Two theoretically independent binary classifications can be applied to cases of univerbation:

- univerbation joins either lexical forms or grammatical formatives (§3.1)
- univerbation joins words that either belong or do not belong to one syntagma (§3.2).

3.1 Univerbation in lexicon and grammar

Application of the term ‘univerbation’ has been variously restricted to either lexicalization or grammaticalization. However, coalescence is a reductive process. Reduction is common to both lexicalization and grammaticalization. Consequently, univerbation is a phase of both these general processes.¹⁵ And indeed, in many treatments the term has been neutral to the distinction between word and word-form. In keeping with this tradition, it applies just as well to coalescence in lexicalization as it does in grammaticalization.¹⁶ E4 and E5 illustrate univerbation in lexicon and grammar, respectively.

E5	PROTO-ROMANCE		MODERN ITALIAN
	de ab Roma	>	da Roma
	of ABL Rome		from Rome

Univerbation may affect two particular word forms A and B where neither A nor B univerbates with anything else or A and B univerbate at most sporadically with other forms of the adjacent class. This is the case of E4 and, commonly, in univerbation as a phase of lexicalization.

Or, again, item A may univerbate with a whole set of items B belonging to one syntactic class. Item A then usually ends up as a grammatical or derivational formative attached to items B (Himmelman 2004). E1 is an example of this (as are Table 2 and Table 3 below). Finally, formative A may be a member of a paradigm whose members all univerbate with items of class B. Paradigm A then becomes an inflection category of class B. This is a well-known scenario and may briefly be recalled by E6, illustrating the agglutination of personal pronouns or auxiliaries to verbs, thus becoming personal conjugations.

E6	VULGAR LATIN		PROTO IBERIAN ROMANCE	SPANISH
	cantare habeo	>	cantar he	> cantaré

¹⁵ The relation of univerbation to grammaticalization has often been treated in the past decades, most recently in Haspelmath 2011a. Brinton & Traugott 2005 is an exhaustive treatment of lexicalization. It highlights the common reductive character of grammaticalization and lexicalization, but does not mention univerbation.

¹⁶ To be sure, the tradition would also justify other terminological options: Given that ‘coalescence’ is the most general concept in this area, coalescence in lexicalization might be called univerbation, while coalescence in grammaticalization would be called agglutination.

3.2.1.2 Categories of input phrases

Univerbation of a **noun phrase** consisting of an article and a noun occurs in many languages. The Latin demonstrative *ille* was postnominal in Romanian and yielded a suffixal definite article, as in *frate-le* ‘the brother’, plural *frați-i* ‘the brothers’. While this is a case of grammaticalization producing inflection, the massive univerbation of definite noun phrases borrowed from Arabic into Spanish is an aspect of their lexicalization, as when Arabic *al manāḥ* ‘the calendar’ yields Spanish *almanaco* ‘almanac’.

While the univerbated phrases of E4 and E7 are **bare nominals**, things are slightly different if a **cased noun phrase** is univerbated, as already illustrated by E1. In E8, an NP in the ablative functions as an adverbial and is univerbated to an adverb.

E8	LATIN		SPANISH
	<i>hāc</i>	<i>hōrā</i>	> <i>ahora</i>
	D.	hour(F):ABL.SG	now
	‘at this moment’		‘now’

The Germanic instrumental was a polyfunctional case which also served as a locative. This has a temporal function in the Pre-German cased noun phrase of E9, which becomes an adverb in Modern German.

E9	PRE-GERMAN		OHG	NHG
	* <i>hiu</i>	<i>d[ag]u</i>	> <i>hiutu</i>	> <i>heute</i>
	D.PROX:M.INSTR.SG	day(M):INSTR.SG		
	‘on this day’		‘today’	‘today’

The univerbation of a Pre-German instrumental phrase meaning ‘in this year’ to yield NHG *heuer* (*id.*) works analogously (Hackstein 2014). In all such cases, the grammatical category of the product is ‘adverb’. This is still in consonance with the principle of univerbation *in situ*, since the adverb has the distribution of the cased noun phrase.

If the dependent component of the cased noun phrase is an adjective, the univerbation may result in the latter’s **adverbialization**. In E10, a manner adverbial consisting of an ablative noun phrase becomes an adverb.

E10	<i>magnō</i>	<i>ōpere</i>	>	<i>magnōpere</i>
LATIN	great:N.ABL.SG	effort(N):ABL.SG		very
	‘with great effort’			‘very’

This is the same construction type that the Romance adverbialization in *-mente* (E1) is based on, but several centuries older than the latter. In contemporary German, there are at least two patterns of adverbialization of adjectives which are based on the univerbation of a cased nominal construction. Both combine the adjective as an attribute with a head noun of general meaning and put the complex nominal in the genitive. The role of the generic head noun can be taken by either *Weise* ‘manner’ or by *Maßen*, the genitive of a noun meaning ‘measure, extent’ which in Middle High German was feminine. Table 2 and Table 3 present some examples; s. §5.1.3 for further analysis.

Table 2 German adverbs in -weise

form	meaning
dummerweise	stupidly
möglicherweise	possibly
glücklicherweise	fortunately

Table 3 German adverbs in -maßen

form	meaning
solchermaßen	in such a way
dermaßen	to that extent
gleichermaßen	to the same extent
einigermaßen	to some extent

A functionally similar construction which is universally prone to univerbation is the **prepositional phrase**. This differs from the cased noun phrase only in the relative structural autonomy of the case relator. The result of the univerbation is, again, an adverb. Thus, the adverb *overnight* is a univerbation of the prepositional phrase *over night*. Many such products feature a pronoun in the position of the governed NP. German examples include prepositional phrases containing the reciprocal pronoun *einander* ‘each other’, like *nebeneinander* (beside:each.other), and ones containing a demonstrative pronoun, like those of Table 4.

Table 4 German adverbs containing a demonstrative

form	gloss	meaning
trotzdem	despite:that:N.DAT.SG	nevertheless
seitdem	since:that:N.DAT.SG	since (that time)
unterdessen	among:that:N.GEN.SG	meanwhile
stattdessen	stead:that:N.GEN.SG	instead (of that)

Similar evidence could be adduced from numerous languages. As the examples suggest, the model produces coordinating conjunctions, too.

A verb forms a **verbal complex** together with an adverb. The first phase of the coalescence (the “chunking”) of the complex is the formation of what in English grammar is called a particle verb like *put on* (cf. §3.2.1.3). In ancient Indo-European languages, this kind of collocation univerbates to what is traditionally called a *verbum compositum*, i.e. a verb modified by a preverb. This topic is taken up in §5.1.1;¹⁷ examples are in E31 below. On further coalescence, the preverb may become a prefix. PIE **anti* ‘facing’¹⁸ yields the OHG preverb *ant-/ent-*, which appears as a prefix, e.g., in NHG *entsprechen* (facing:speak:INF) ‘correspond’.

The traditional term ‘**incorporation**’ implies that the process is one of univerbation. Deferring the question of the relation between univerbation and compounding to §6, we may observe here that in a language where the incorporation of nouns into verbs is not a productive process, an analysis by univerbation is not only plausible in principle, but is also historically demonstrable in certain cases. Latin *manu tenere* (hand(F):ABL.SG hold:INF) yields French *maintenir* ‘maintain’, Latin *male tractare* (bad:ly treat:INF) yields French *maltraiter* ‘mistreat’. German examples include *haushalten* (house:hold:INF) ‘housekeep’ and *hohnlachen* (derision:laugh:INF) ‘flee’. These French and German products are transitive and intransitive verbs, respectively; so they belong to the same distribution class as the underlying construction. Such examples could be multiplied.¹⁹ These German complex verbs, however, separate unless in final clause position, so the univerbation is either only partially completed

¹⁷ It is treated exhaustively in Brugmann 1900, §5.

¹⁸ The complex question concerning the relationship between adverbial and adpositional function of these formatives in general, and in preverbation in particular, is left alone here.

or is a purely orthographic phenomenon. The participle *feuerspeiend* (fire:spitting), however, is a perfect example, as it is inseparable and intransitive like an incorporative verb.

The univerbation of a **periphrastic verb form**, i.e. of an auxiliary with its (non-finite) full verb, was illustrated above by the Romance synthetic future (E6). Examples sharing this structure are instances of phrasal univerbation. Things are more complicated if the auxiliary and the full verb do not constitute a complete phrase. This may presuppose a reanalysis; see § 5.6.

An adverb modified by an adverb univerbates *in situ*, like NHG *genauso ~ ebenso* (exactly:so). Even particles may univerbate, like German *naja* (hey:yes) ‘well’.

Secondarily, the univerbated product may change its grammatical category. The Spanish noun *aguardiente* changes its gender to masculine. The nominal subconstituent of MHG *ze den wîhhen nahten* (to DEF:DAT.PL sacred:DAT.PL night(F):PL) ‘at Christmas’ got univerbated to NHG *zu Weihnachten*. The form *Weihnachten* was then reinterpreted as a proper noun, possibly even of singular number, which may occupy any nominal syntactic function. The adjective *zufrieden* ‘content’ results from univerbation in contexts like E11.

E11	jemanden	zu	Frieden	stellen
NHG	someone:Acc	to	peace(M):Dat.Sg	put:Inf
	‘to satisfy somebody’ (Brugmann 1900, §3.4)			

The immediate product is *zufrieden* as an adverb. This, however, is secondarily recategorized as an adjective, witness its declinability.

Another such change is the retransitivization of incorporative verbs (Hackstein 2012: 89). Older NHG *wahrnehmen* (attention:take:INF) ‘perceive’ and more recent *gewährleisten* (warranty:provide:INF) ‘warrant’ are complex verbs univerbated from a verb phrase, i.e. a transitive verb governing a noun in direct object function. Both verbs are separable to this day. Although incorporation of the direct object originally produced an intransitive verb, this has meanwhile become transitive again: both complex verbs take a direct object. The older one takes it unconditionally, the more recent one only when not separated.²⁰

3.2.1.3 Phrasal compounding

A **phrasal compound** or multi-word lexeme is a lexicalized collocation at the syntactic level of the phrase. Its lexicalization implies loss of most of the variability available for a collocation at that level. Apart from this, a phrasal compound has the same structure as a syntactic construction. It is a hermaphrodite between the levels of the word and the phrase. Three subtypes of phrasal compounds may be specified:

a) Phrasal compounds consisting of a nominal head and an adjectival attribute are a frequent input to univerbation, as already indicated by E7. Spanish examples include *aire acondicionado* ‘air conditioning’, *corto circuito* ‘short circuit’ and *mercado negro* ‘black market’; their English translation equivalents are phrasal compounds, too. German examples include those of E12.

¹⁹ The specialized literature adduces examples like *blankwischen* (shining:wipe:INF), *schönfärben* (pretty:dye:INF), *hängenbleiben* (hang:INF:stay:INF), *spazierengehen* (stroll:INF:go:INF) etc., all of which are separable.

²⁰ A rather idiosyncratic pattern of phrasal univerbation in German, illustrated by nouns like *Zeitlang* (time:long) ‘while’ and *Handbreit* (hand:wide) ‘a few inches’, is analyzed in Donalies 2016.

- E12 a. grüne Welle
 NHG green:F.SG.NOM wave(F)
 ‘progressive signal system’
- b. toter Punkt
 dead:M.SG.NOM point(M)
 ‘deadlock’

Lexical items like those of E12 are lexicalized nominals, but not nouns since they are not words: their components remain accessible to rules of grammar (genitive *des toten Punktes*, modification *ein ziemlich toter Punkt* ‘a rather dead lock’).

b) A phrasal compound whose head is modified by a prepositional phrase has been called a *synapsis*²¹ in Romance linguistics. Spanish examples include *olla a presión* (pot at pressure) ‘pressure cooker’ and *máquina de escribir* (machine of write:INF) ‘typewriter’.

c) As anticipated in §3.2.1.2, verbs forming a phrasal compound with an adverb, viz. particle verbs like German *zurückschicken* ‘send back’, may likewise be univerbated.

Phrasal compounding is a frequent, though not necessary, preliminary stage of phrasal univerbation. A phrasal compound is an input candidate to univerbation. In none of the examples adduced is univerbation completed. However, there are also examples of orthographic variation testifying to the ongoing univerbation of a phrasal compound, like Spanish *medio ambiente* or *medioambiente* (middle ambient) ‘environment’.

As will be discussed more fully in §6, compounding and univerbation are distinct processes. The hermaphrodite status of phrasal compounding concerns not only its intermediate syntactic level between the phrase and the word, but also its ambivalent origin: On the one hand, it is compatible with rules of syntax, but on the other, like word formation, it creates potential lexical entries. In diachronic perspective, the first instances of a pattern of phrasal compounding may result from the consolidation of a syntactic collocation. These may then establish a pattern for the formation of complex lexical items. This will be taken up in §6.1.2.

3.2.2 Transgressive univerbation

As we have seen, any univerbation by definition occurs across a syntactic boundary, viz. the word boundary. There are, however, univerbations across stronger boundaries, viz. a phrase or even clause boundary. The general schema of transgressive univerbation²² takes the form of Diagram 3, illustrated by E13.

²¹ Benveniste 1966[1974: 174] characterizes the ‘synapsie’ as follows: « ...tous les éléments sont en principe idiomatiques et de forme libre et ... peuvent être eux-mêmes des synapsies, ils sont reliés par des joncteurs, principalement *de* et *à*, et leur ordre est toujours déterminé + déterminant. » French examples include *machine-à-calculer* (machine-to-calculate) ‘mechanical calculator’ and *pomme-de-terre* (apple-of-earth) ‘potato’. The literal English equivalent of French *synapsie* would be *synapsy*. However, the closely related *synapsis* is already existent, with the desired meaning (although outside the field of linguistics).

²² The generic meaning of *transgressive* is ‘crossing a border’. It has been used with more specific senses in other disciplines, but these are of no relevance to linguistics. *Cross-border* (or *trans-border*) *univerbation* may be more idiomatic in English, but is inapt as an international term.

Diagram 3 Transgressive univerbation

input	[A [B C] _{XP}] _Y
output	[AB C] _Y

E13 SPANISH

input	[a [el [inicio] _N] _{NP}] _{PrepP} to DEF.M.SG start
output	[al [inicio] _N] _{PrepP}

E13 represents a common scenario, viz. the univerbation of a primary preposition with the following definite article. This happens in the Romance languages (Bybee 2007: 330), but also with German primary prepositions like *an* ‘at’, *zu* ‘to’ and *von* ‘from’, which univerbate (among others) with the masculine or neutral singular definite article *dem* to yield *am*, *zum* and *vom*. Diagram 3 shows that, before undergoing univerbation, one of the items, B, forms a syntagma of category XP – NP in E13 – with other material (C). Univerbation reduces the syntactic boundary separating this syntagma from the other univerbated component (A) to a morphological boundary, whereby the construction XP becomes unrecognizable. Unless some syntactic reanalysis (§5.6) precedes the process, transgressive univerbation will ignore and destroy syntactic structure, viz. the syntactic integrity of XP.

Prepositions also univerbate with each other if a preposition governs a prepositional phrase, as in Italian *da* (E5), English *onto* and Spanish *atrás* ~ *detrás* ‘behind’ (Lehmann 2019, §6.1). A preposition coalesces with the governed relational noun which is the head of its complement in Spanish *encima* (Loc:top) ‘on’ and numerous similar examples including those of Table 7 and Table 8 below. The univerbation of a preposition (and analogously, a postposition in languages which have clause-final subordinating conjunctions) with a subordinator is a common sight in European languages (Kortmann 1997, §8.3), with examples like Spanish *porque* (for:that) ‘because’. The place of the preposition may also be taken by an adverb, as in Spanish *aunque* (still:that) ‘although’, French *puisque* (then:that) ‘because’ and German *sodass* (so:that).

Table 5 German univerbated subordinating conjunctions

form	gloss	meaning
indem	in:that:N.DAT.SG	while
seitdem	since:that:N.DAT.SG	since
nachdem	after:that:N.DAT.SG	after

A few German subordinating conjunctions, enumerated in Table 5, are univerbations of a preposition with the demonstrative form *dem*. In contrast with those of Table 4, these go back to a construction of the form [preposition [*dem* [*dass* ...]_S]_{NP}]_{PrepP}, where *dem* is cataphoric to the following subordinate clause (cf. French *parce que* (for:DEM SR) ‘because’). The univerbation produces a subordinating conjunction which renders the subordinator *dass* redundant (Lehmann 1993, §4.2).

Periphrastic verb forms may comprise more than the two words provided by the auxiliary and the full verb. English *want to* and *be going to* are such cases. They are particularly prone to univerbation, since once the erstwhile main verb has become an auxiliary, it no longer com-

bines with the dependent full verb via the particle *to*. In general, since it now signals only a conjugation category, more than one morpheme for it is excessive. These sequences are consequently univerbated (*to wanna* and *gonna*, respectively), again across a syntactic boundary, since the *to* forms a constituent with the following full verb.

In certain cases, the auxiliary may form a phrasal unit with the full verb and then univerbate with it (cf. §5.6.1). In other cases, it instead univerbates with pronouns or pronominal indexes referencing the actants of the full verb. The English contractions *I'll*, *I've*, *he's* etc. are well-known examples (Bybee 2007: 327; Haspelmath 2011a, §4). A less widely known case comes from Yucatec Maya. An independent verbal clause is headed and introduced by an impersonal auxiliary. It is followed by the verbal group, starting with the pronominal index representing the subject. These indexes are, however, enclitic. They therefore univerbate with the preceding auxiliary. For instance, although in E14, *in* forms a constituent with *meyah*, univerbation first produces *ts'o'kin meyah* and then, by welding, the final result shown at the right-hand side of E14 (Lehmann 2017, §4.7.4).

E14	ts'o'k	in	meyah	>	ts'in	meyah
YUCATEC	TERM	[SBJ.1.SG	work]			
	'my work(ing) is finished'				'I have worked'	

Finally, transgressive univerbation may even act across a finite-clause boundary. The syntactic construction of E15 has the compositional reading indicated.

E15	nescio	quis	p
LATIN	[not:know(PRS):1.SG	[who:NOM.SG	p]s]s
	'I do not know who p'		

In E15, the position of *nescio* (itself a univerbated form [§3.2.3]) may be taken by any conjugated form of *nescire* 'to not know', and the position of *quis* may be taken over by any interrogative proform. However, this particular conjugated form of *nescire* univerbates with the interrogative pronoun to the indefinite pronoun *nescioquis* 'some(one)' (Haspelmath 1997: 132).

Transgressive univerbation is a case of "chunking in violation of constituent structure" which demonstrates "the dominance of sequentiality over hierarchy" (Bybee 2007, ch. 15.12).

3.2.3 Phrasal and transgressive univerbation

The distinction between phrasal and transgressive univerbation depends, of course, on the constituent structure analysis, in particular, on the structure of the sequence forming the input in Diagram 3. The negator often univerbates with neighboring words. The constituency of a negator is sometimes not easy to ascertain. If we assume that the expression constituting the scope of negation is a co-constituent of the negator, then the following examples are cases of phrasal univerbation. The negator and an indefinite proform are often univerbated to a negative proform.²³ They provide particularly clear evidence for the principle that adjacency is the *conditio sine qua non* for univerbation. Univerbation occurs in English *no one*, *nobody*, *nothing*, *never*. It does not occur in their French equivalents *ne ... aucun*, *ne ... personne*, *ne ... rien* and *ne ... jamais*. The two components here may be regarded as a discontinuous negative proform. They are never adjacent in syntactic constructions and consequently never univer-

²³ In the morphological and etymological analysis of any negative proform, it is generally a fruitful heuristic to check whether it results from the univerbation of a negator with an indefinite proform.

bated. What happens to them, instead, is the loss of the initial particle of the discontinuous expression.

The negator also coalesces with the verb, as in Latin *nescio* (E15). From a logical point of view, one may think that the entire verb phrase headed by *scio* forms the scope of the negator. However, *nescio* is a close synonym of *ignoro* ‘I ignore’, which is a product of word formation based on a pattern which combines a negative prefix (*in-*) with a stem. In many languages, the existential verb or verboid has a negative counterpart.²⁴ In such cases, it is possible that univerbation is preceded by a reanalysis which converts the sequence A B of Diagram 3 into a syntagma (§5.6.1). Finally, the complex postposition *notwithstanding* arose by univerbation of the sequence *not withstanding*.

Summarizing the two classifications of §3 in one cross-classification, we get the picture of Table 6. One member of each example subset filling the table cells is repeated below for convenience. Inclusion of additional examples will not change the distribution shown in Table 6.

Table 6 Types of univerbation

domain level	lexicalization	grammaticalization
phrasal	E4, E7, E16	E17, E19
transgressive	X	E5, E14, E15, E18

E16 LATIN
(= E10) *magnō* *ōpere* > *magnōpere*
great:N.ABL.SG effort(N):ABL.SG very
‘with great effort’ ‘very’

E17 LATIN (Cat. *Carm.* 8, 11) ITALIAN
(= E1) *obstinata* *mente* > *ostinatamente*
obstinate:ABL.SG mind:ABL.SG obstinate:ADV
‘in an obstinate spirit’ ‘obstinately’

E18 SPANISH
(= E13)

input	[a [el [inicio] _N] _{NP}] _{PrepP} to DEF.M.SG start
output	[al [inicio] _N] _{PrepP}

As may be seen, transgressive univerbation is limited to univerbation as an ingredient of grammaticalization. That is, A and/or B of Diagram 3 are already grammatical formatives. In this, transgressive univerbation differs from phrasal univerbation, which applies to lexical and grammatical words alike. It is evident that new concepts are formed only by phrasal univerbation. Transgressive univerbation is restricted to grammaticalization and may therefore generate new inflection paradigms. This finding also sheds new light on the continuum of sequential closeness of the exponents of morphological categories to the lexical root. According to Bybee 1985, its functional basis is semantic relevance. However, to the extent that

²⁴ Yucatec *yaan* EXIST vs. *mina'n* NEG.EXIST is an example. Cf. Hackstein 2014: 41 on Old Russ. *ŋe[s-tŷ]=tu > nētū > nēt > Russ. net* (NEG.EXIST).

morphological structure results from univerbation, the sequential distance of morphological material also correlates with the strength of the underlying syntactic boundary.

4 Phenomena akin to univerbation

Univerbation is to be distinguished from several similar phenomena. In §2, it was subsumed under coalescence. Taken as a term, **agglutination** is similar to univerbation, as it designates the gluing of something to its host. Traditionally, the concept is restricted to the agglutination of grammatical formatives, while univerbation is not so restricted. Moreover, agglutination has come to designate a technique of morphological structure in which every morph is the exponent of a value of one category which is morphologically optional.

4.1 Clisis

The relationship of univerbation and clisis is more complex. Clisis is the prosodic combination of a stressless word with an adjacent stressed word into one phonological word. It leads potentially to univerbation. For instance, the Germanic weak past in *-d* probably²⁵ results from the univerbation of the infinitive of the full verb with a form of an auxiliary meaning ‘do’ (E19); this auxiliary must have been enclitic, as witnessed by the unstressability of this past desinence.

E19	PROTO-GERMANIC		GOthic
	*salbo-n	ded-um	> salbodedum
	anoint-INF	do:PST-1.PL	anoint:PST:1.PL
	‘we did anoint’		‘we anointed’

Clisis does not necessarily lead to univerbation, though, since the resulting phonological word is not necessarily a word-form by grammatical criteria. E20 is a simple example demonstrating this: *que* ‘and’ is enclitic.

E20	dominus	servusque
LATIN	master(M):NOM.SG	slave(M):NOM.SG:and
	‘master and slave’	

E21 illustrates the integration of enclitic grammatical formatives with their syntactic and phonological host into a phonological word (Haspelmath 2011a: 343).

E21	LATIN		SPANISH		
	dare	mihi	illud	>	dar=me=lo
	give:INF	1.SG:DAT	D.DIST:N.ACC.SG		give:INF=me=it
	‘to give me that’				‘to give it to me’

E22	me	lo	da
SPANISH	me	it	gives
	‘(he/she/it) gives it to me’		

Here again, clisis does not lead to univerbation because the result is not a word-form (of the conjugation system), but a word-form followed by clitics; these clitics would appear before other conjugated forms of the same word, as in E22.

²⁵ The past of the verb ‘do’ is not attested separately in Gothic, nor is there a regular change by which the infinitive ending would have been lost.

Sometimes the phonological rules acting across a clitic boundary differ from the rules acting inside the word. In the Cabecar language (Chibchan, Costa Rica), focus particles like /nẽ/
CONTRASTIVE IDENTIFIER are enclitic to the focal constituent, yielding, e.g., /hé=nẽ/ (D.MED=IDNT) ‘exactly that’. The language has a phonological rule of nasality propagation which copies the feature [nasal] of a vowel on vowels to its left inside the word, but not across a word boundary, not even a clitic boundary. Now sequences of the medial demonstrative with a following enclitic are often univerbated. In the present case, this yields /ẽ’nẽ/ (with additional suppression of initial /h/). The univerbated form thus differs phonologically from the clitic combination.

Several treatments of grammaticalization (Hopper & Traugott 1993: 7; Harris & Campbell 1995: 337) consider clisis a phase in the gradual coalescence of a grammatical formative with its host, thus, a phase in Diagram 1. However, clisis is neither a precondition for grammaticalization nor for univerbation. In German separable compounds like *aufhören* ‘stop’, the preverb is stressed. In the Romance synthetic future (E6) and conditional (E30), the desinence is stressed. These univerbated formatives have never been clitic. To conclude, clisis is neither a necessary nor a sufficient condition for univerbation, but is frequently involved in it.

4.2 Compounding

The term ‘composition’ has had a complicated fate in the history of the discipline and was narrowed down to its present meaning, now mostly termed ‘compounding’ in English, at most a hundred years ago. It now designates a technique of word formation whose prototypical instantiation is the combination of two lexical stems into one, as in *playboy* and *search machine*. What must be recorded, at any rate, is that compounding is part of a language system viewed synchronically.

Compounding differs from other techniques of word formation by combining more than one lexical stem. The branching direction makes no difference for the general concept of compounding.²⁶ Instead, it may characterize a certain strategy of compounding of a specific language system. *Fountain pen* is left-branching, while its Spanish counterpart *pluma fuente* is right-branching, as are the Italian examples in §5.2. It is, however, to be noted that the branching direction in a compounding pattern may be opposite to the branching direction in the corresponding syntactic pattern.

- E23 a. Tageslicht
NHG day(M):GEN.SG:light
‘daylight’
- b. Licht des Tages
light DEF:M.GEN.SG day(M):GEN.SG
‘light of the day’

Thus, the compound of E23a is left-branching, but the corresponding phrase of #b is right-branching. Word formation has its own rules, which may or may not correspond to the rules of syntax of the same language system. This will be taken up in §§4.2.2 and 6.2.

²⁶ Some Germanist and Anglicist definitions of nominal compounding require a left-branching structure, i.e. that the head be the right-hand member of the construction. This, however, is a germanocentric requirement.

4.2.1 Types of compounds

In the 19th century, the concept of compounding was formed on the basis of compounding techniques of ancient Indo-European languages. These involved using a mere stem as the *determinans*, as in E24 – E26.

E24	dēmokratía
A.GREEK	people:rule(F):NOM.SG 'democracy'
E25	thiudangardi
GOTHIC	king:court:SBVR 'kingdom'
E26	bātweard
O.ENGL	boat:ward 'ship's watchman'

This very narrow concept of compounding excluded formations such as German E23a because the *determinans* is in the genitive. Jacob Grimm 1826, ch. III, called the former type (as exemplified in E24 – E26) “proper compounds” and the latter (as in E23a) “improper compounds”. In the middle of the 19th century, the latter were dubbed *Zusammenrückung* ‘univerbation’ (s. §2.1). The point here is that the “proper compound” cannot, in the language system generating it, be analyzed by univerbation since the bare stem does not occur in syntactic constructions. While it is possible that both types of compounding coexist at a synchronic state, the difference between them has been neutralized in Modern German. NHG compounds like *Bücherkiste* (book:PL:box) ‘book box’ and *Sonnenwende* (sun:CAT:turn) ‘solstice’ do not use the bare stem of the *determinans*, but cannot be explained by univerbation, either. Most current compounds contain the catalytic element (CAT), which may or may not be identical with a declension morph, as in *Bücherkiste* (Schlücker 2012, §1).

As a structural technique of forming complex expressions, compounding is placed between the formation of syntactic phrases, on the one hand, and derivation, on the other. Phrasal compounds like *peanut butter* and *Stockholm based* mark the transition into syntactic phrases (§3.2.1.3), while items like *moonlike* and *error-free* (cp. the near-synonymous derivate *flawless*) mark the transition into derivation. A compounding process which productively employs a particular lexical component may develop into a derivational process. Examples are in E1, E10, Table 2 and Table 3. While the underlying compounding process may originate in univerbation, univerbation does not seem to directly establish a pattern of derivation. This is why the present delimitation of univerbation against word formation focuses primarily on compounding.

4.2.2 Compounding and phrasal univerbation

Univerbation and word formation are distinct in principle. In many cases, the methodological problem of attributing a given complex expression to either process is easily solved. First of all, transgressive univerbation is unmistakable, as it always involves grammatical formatives and therefore does not produce a complex lexical concept. Forms produced in this way, like Eng. *I'll* or French *au* DAT:DEF.M.SG, run no risk of being mistaken for compounds. It is phrasal univerbation which shares with compounding the production of complex concepts and whose products can be mistaken for compounds. Again, this is not true of all of phrasal univerbation. Whenever a strongly grammaticalized formative is involved, as in Romanian

fratele ‘the brother’ or Eng. *nevertheless*, the defining criteria of compounding clearly do not apply. It is with cases like Eng. *blackbird* and Spanish *aguardiente* that doubts can arise. The principle of distinction is clear enough: given a complex word *W* created at a certain stage of the language, then:

- If there are no contemporary rules of compounding to generate *W*, it cannot be a compound. *W* can then be due to univerbation. For instance, German *Muttergottes* follows no pattern of compounding, but does follow a syntactic pattern and is therefore not a compound, but due to univerbation.
- If there are no contemporary rules of syntax to generate the phrase formed by replacing a morphological boundary in *W* by a word boundary, *W* cannot come about by univerbation. It can then be a compound. For instance, Eng. *apple tree*, Greek *dēmokratía* and Italian *crocevia* ‘crossroads’ cannot be univerbations (and must consequently be compounds), as there are, in the historical periods in which they originated, no rules of syntax which could have generated nominal groups like *apple tree*, *dēmo kratía* and *croce via*, respectively.

A methodological problem arises if compounding rules of one stage of a language correspond to rules of syntax of an earlier stage. For instance, NHG *Hahnenfuß* ‘buttercup’ corresponds to contemporary rules of compounding, but nevertheless originated as E27 by univerbation at some stage antedating Old High German.

E27	haninfuoʒ
OHG	rooster(M):GEN.SG:foot ‘buttercup’

Moreover, at that stage, the compounding rules would not have allowed the formation of such a complex stem. We will come back to diachronic relations between univerbation and compounding in §6.1.1.

4.3 Conversion to word

Conversion to word is the recategorization of a syntactic construction of category *XP* (fitting a context *Z1*) as a word of category *Y* ≠ *X* by using it in the distribution class of *Y*, i.e. in a context *Z2* ≠ *Z1* (cf. Eisenberg 1998: 224ff and Schlücker 2012, §4). The process is formalized in Diagram 4 and illustrated by E28.

Diagram 4 Conversion to word

input	[... [A B] _{XP} ...] _{Z1}
output	[... [AB] _Y ...] _{Z2}

E28 NHG

input	[Sie [möchte gern] _{VC} ...] _S she would.like:3.SG with.pleasure
output	[ein [Möchtegern] _N] _{NP} a would-be

As E28 suggests, the verbal complex (VC) that forms the input is converted into a noun by combining it with an article. The target category *Y* could, in principle, be a phrasal category. In E28, combination with the article only requires a nominal expression, not necessarily a

noun. However, a phrasal category brings with it constraints on its internal subconstituency, which the categories of the component items A and B do not meet. The only way of analyzing such a complex expression is therefore to treat AB as a word of the target category, N in the case of E28.

In conversion to word, the enforcing context Z2 is by definition a syntactic context. In E28, it is a noun phrase introduced by an article. The same applies to English *good-for-nothing*, a substantivized adjectival, and its German counterpart *Tunichtgut* (do:not:good), a substantivized clause, and likewise to *forget-me-not*.²⁷ French *rendez-vous* (move.to:2.PL-you(PL)) is based on a finite reflexive verb form, *pourboire* (for:drink:INF) ‘tip’ on a prepositional phrase and *tête-à-tête* (head-to-head) on an adverbial, all of them used in a syntactic context requiring a noun. Thus, conversion to word is a recategorization forced by the context (Vogel 1996). Therefore, it is not a process of univerbation; univerbation happens *in situ*.²⁸

A comparison of Diagram 2 with Diagram 4 shows two things: First, phrasal univerbation and conversion to word share the presupposition that the input structure form a syntagma. Second, they differ in that the grammatical context remains the same (Y in Diagram 2) for the input and the output of univerbation, while they differ for the input and the output of conversion to word (Z1 and Z2 in Diagram 4). With phrasal univerbation, the category of the output structure results automatically by dropping the properties of the phrase enumerated in § 3.2.1.1, thus lowering the level of structural complexity, while in conversion to word, the output category is forced by using the expression in a context which requires a syntagma of that category.²⁹ It is this contrast which is meant to be covered by the terms ‘spontaneous’ vs. ‘coercive’.

4.4 Parasynthesis

Another phenomenon which bears some resemblance to conversion to word and likewise needs to be excluded from univerbation is parasynthesis (called *Zusammenbildung*

²⁷ Nouns produced by the conversion of a verbal construction involve an intriguing form-function relation concerning the verb form. In Romance formations like Spanish *lavaplatos* ‘dishwasher’, Italian *grattaciello* ‘skyscraper’ etc., the verb form is ambiguous between imperative, third person singular and bare verb stem. However, in Italian nouns like *rompighiaccio* (break:ice) ‘icebreaker’, only the imperative fits formally. The same goes for *forget-me-not* and its German counterpart *Vergissmeinnicht*. In its Spanish counterpart *nomeolvides*, even the syntactic rule conditioning subjunctive instead of imperative applies. Semantically, however, an imperative makes little sense in *Tunichtgut* or Italian *dormialfuoco* (sleep:at:the:fire) ‘lazybones’ and *batticuore* (beat:heart) ‘palpitations’. Apparently the structural position in which just a verb stem is needed is occupied by the imperative, regardless of its meaning. Thanks are due to Livio Gaeta for relevant facts and analyses.

²⁸ The task here is to contrast univerbation with similar processes, not to classify processes of word formation. Consequently, questions such as whether the concept of conversion to word includes or excludes exocentric (bahuvrihi) compounds like Spanish *ciempiés* ‘centipede’ need not be addressed here.

²⁹ Like most of our forebears, Hermann Paul (1920) adduces almost all of his examples of univerbation (“Komposition”) in isolation without consideration of the context in which they originated. For a remarkably high number of these examples, it is not clear whether they instantiate univerbation or conversion to word; in other words it is not clear what kind of syntactic environment would be presupposed for the syntactic construction from which they stem. Here more historical investigation of texts is required to clarify the mechanisms involved.

(together:formation) in German linguistics). A word produced by it has the general structure schematized by Diagram 5 and illustrated by E29.

Diagram 5 *Parasynthesis*



E29	quince	+añ	-ero
SPANISH	fifteen	+year	-NR:M
	'teenager'		

B of Diagram 5 is generally a lexical stem. A may be one, too, as in E29 and in German *Grablegung* (grave:lay:SBVR) 'burial'; or it may be a formative, as in French *embarquer* (in:ship:INF) 'embark'. C may be a lexeme of the target category Y, as in German *Sauregurkenzeit* (sour:cucumber:time) 'silly season', or a formative determining Y, as in E29, or just the inflection morphology of Y, as again in *embarquer*. The details differ and may require different analyses. What matters here are two traits shared by relevant cases: First, the construct contains the combination AB which looks like a compound or derived stem except that such a combination has no independent existence as a word of the inventory. Second, the category of the complex is determined by C, which thus has semantic scope over AB. In general, parasynthesis is like conversion to word in that it converts a complex construction into a stem by using it in a context which compels this usage – a syntactic context in conversion to word, a morphological context in parasynthesis. This is, again, unlike univerbation, which happens *in situ*.

To summarize, compounding and parasynthesis differ from univerbation in being processes of word formation. Conversion to word and parasynthesis differ from univerbation by forcing the union into one word by means of the context.

5 Symptoms of univerbation

Since grammatical boundaries are imperceptible, the linguist needs other methodological cues to detect univerbation. The reduction of a syntactic boundary separating two items to a morphological boundary renders possible all the changes distinguishing a word from a phrase enumerated in §3.2.1.1. The following symptoms are the methodological reflection of these changes.

5.1 Reduction of structural variability

5.1.1 Enforcement of continuity

Preceding univerbation, a complex unit may still be discontinuous. Thus, the preposition *instead* (§2.2) is univerbated from *in stead*. *Stead* may still be preceded by its dependent, e.g. *in the king's stead*. Enforcement of continuity of *in stead* produces *instead of the king*. The same goes for the German counterpart *anstatt*.

E30 illustrates the coalescence of the erstwhile conditional auxiliary with the full verb in Portuguese and Spanish (the same goes for the future auxiliary illustrated by E6). In Proto-Romance, it takes a preceding infinitive as its complement. This infinitive, in turn, can take

enclitic pronouns. If both the dependent pronoun and the auxiliary attach to the infinitive, the result is as shown in E30.

E30	PORTUGUESE	SPANISH
	ajudar=me-ia	me ayudaría
	help:INF=me-COND	me help:INF:COND
	‘(he/she/it) would help me’	

In Portuguese (as still in Old Castilian), the conditional auxiliary follows any enclitics attached to the infinitive, so that it is not yet univerbated with the latter. In Modern Castilian, on the other hand, univerbation of the auxiliary with the infinitive is completed, nothing can intervene between them. The clitic pronoun therefore precedes what is now a finite verb form.

Preverbation is the result of the univerbation of an adverb with a verb (cf. § 3.2.1.2 on the verb complex). Originally, the construction comprising the adverb and the verb could be discontinuous. In Latin, traces of this are visible in what is traditionally called *tmesis* ‘cutting’ (as if the discontinuous order were a secondary deviation from the norm). It is no longer possible in Classical Latin. But the lexicographer Festus still notes survivals of *tmesis* in the 2nd century AD, viz. in E31.

E31	<i>ob vos sacro</i> in quibusdam precationibus est pro <i>vos obsecro</i> , ut <i>sub vos placo</i> pro <i>supplico</i> (Festus XIII, §175)	
LATIN	‘In certain prayers it says <i>ob vos sacro</i> [facing 2.PL.ACC implore:1.SG] instead of <i>vos obsecro</i> [‘I implore you’], like <i>sub vos placo</i> [beneath 2.PL.ACC supplicate:1.SG] instead of <i>supplico</i> [‘I supplicate (you)’].	

Something similar happened in the development of such adverbs into adpositions. Colloquial Northern German still allows fronting a demonstrative or interrogative pronoun, in the form of a local adverb, while stranding the adposition governing it, as in E32 (Brugmann 1900, § 5.4).

E32	Da kann ich nichts für.
NHG	there can I nothing for
	‘For that I’m not responsible.’

Such a distancing does not occur in standard German, which has put the univerbation of the adposition and its complement (*dafür kann ich nichts*) into effect.

5.1.2 Fixation of order

Many examples of univerbation concern a nominal construction consisting of a head noun and a possessive attribute which univerbation condenses into the equivalent of a determinative compound. The position of the attribute may be variably pre- or postnominal; but the univerbation presupposes its fixation in one position. Examples from Latin include E4: Classical Latin allows both *terrae motus* (Cic. *div.* 1, 101, 12) and *motus terrae* (Plin. *nat.* 2, 191, 1). The univerbation *terraemotus* ‘earthquake’ is first documented in St. Jerome (~385) and presupposed by all the Romance descendants. Likewise, Classical Latin has both *aquae ductus* (Iust. *Dig.* 8, 3, 2, 2, 1) and *ductus aquae* (*op.cit.* 43, 20, 3, 4, 1), which in Late Latin is exclusively *aquaeductus*.

The same can be shown by German examples. The position of the genitive attribute used to be variable in Proto-Germanic times. From the prenominal genitive attribute we get such products as *Waldesruhe* ‘quiet of the woods’ and *Gottesurteil* (god’s judgement) ‘trial by ordeal’. Only in one case, both orders were univerbated: we have both *Gottesmutter* and,

exceptionally, *Muttergottes* ‘mother of God’. While determinative compounds are always left-branching, as in *Waldesruhe*, *Gottesurteil*, and *Gottesmutter*, in possessive attribution itself, the pronominal order is in most cases no longer possible in Modern German.

Fixation of order is also presupposed by univerbation of nominals comprising an adjective attribute. Latin *res publica* (thing(F):NOM.SG public:F.NOM.SG) ‘republic’ is never inverted to *publica res*, although it always shows introflexion, like the accusative *rem publicam*.

German complexes consisting of a primary adposition and its pronominal complement provide an exception to the generalization that univerbation presupposes fixation of order. With lexical complements, these adpositions are prepositions during the entire documented history of German. With certain pronominal complements, they are univerbated in the order ‘complement-adposition’, as in *dafür* ‘therefore’ (cf. E32), *damit* ‘therewith’, *davon* ‘therefrom’ etc. They thus display a postpositional word order which, as their English equivalents demonstrate, must date back at least to West Germanic.³⁰ Despite this archaic order, they remain separable as shown before (Brugmann 1900, §1). This is one of many examples showing that the symptoms of univerbation do not always march in step.

5.1.3 Reduction of syntactic structure

As long as the construction to be condensed is a syntactic one, its components may establish syntactic relations. Once the construction is a word, it sets up syntactic relations as a whole. For nominal compounds resulting from univerbation, this fact has, *inter alia*, the consequence that the *determinans* no longer takes dependents of its own. For instance, the German compound *Freudenfest* ‘feast of joy’ originates in the univerbation of E33. The current compound no longer takes an article fitting the first component.

E33	(der)	vreuden	fest
MHG	DEF:F.SG.GEN	joy(F):SG.GEN	feast(N)
	‘(the) joy’s feast’		

The German adjectives which univerbated with *-maßen* and *-weise* (Table 2 and Table 3) no longer take their own adverbs. Thus we have *freimütig zugegeben* ‘frankly admitted’ and *zugegebenermaßen* ‘admittedly’, but not **freimütig zugegebenermaßen*. Likewise, there is *übertrieben vorsichtig* ‘overly cautious’ and *vorsichtigerweise* ‘cautiously’, but not **übertrieben vorsichtigerweise*.³¹

Reduction of syntactic structure has been studied extensively in the univerbation of denominal complex prepositions (Lehmann & Stolz 1992, §6). Omission of the definite article is an early symptom of imminent univerbation. Spanish *a base de X* ‘on the basis of X’ would be *a la base de X* (at DEF.F.SG base(F) of X) if it were a regular syntactic construction (§2.2). In German, what started out as *im Bezug auf* (in:DEF.M.SG relation on) is now *in bezug auf* ‘with regard to’. Orthographic univerbation (*inbezug auf*, cf. §2.3), although currently not standard, is attested as early as 1913.³² Again, the nouns in such complex prepositions do not

³⁰ The model is the same as in the combinations of the deictic particles *her* VENTIVE and *hin* ANDATIVE with some adverbs and adpositions, like *her-/hinein* ‘into it’, *her-/hinzu* ‘to it’.

³¹ *Sehr dummerweise* ‘very stupidly’ does occur, but instantiates modification of an adverb by an adverb.

³² Medizinalabteilung des Königlich Preußischen Ministeriums des Innern (ed.) 1913, *Krankenpflege-Lehrbuch*. Berlin & Heidelberg: Springer; 317.

take their own attribute; beside *aufgrund von Informationen* ‘on the basis of information’ we do not have **auf sicherem Grund von Informationen* (on safe ground of information).³³

Recall from §§2.2 and 4.2.2 the methodological requirement that for a complex expression to be analyzable as formed by univerbation, there must be a syntactic construction of the same form. Now constructions like *in Bezug auf X*, *an Hand von X*, *auf Grund von X* (cf. Table 8 below) are attested in German. They thus fulfill the methodological condition no matter whether they instantiate rules of analytic syntax or subregularities of phraseology.

5.2 Adaptation of morphological structure

Once univerbation is accomplished, the morphological – chiefly, inflectional – structure of the complex is adapted to the model of the simple word: internal structure is suppressed, external structure is added as needed (Haspelmath 1993). While Late Latin *terraemotus* has an internal genitive, this disappears in Spanish and Italian *terremoto*. Italian has phrasal compounds of the structure $[N_1 N_2]_N$, where N_1 is the head and is inflected for number, which produces internal inflection, as in *pesce farfalla* (fish butterfly) ‘butterflyfish’, plural *pesci farfalla*. Univerbation of these formations leads to the externalization of inflection, as in *pescecane* (fish:dog) ‘shark’, plural *pescecani* (Gaeta 2011). If the first component of a phrasal compound is an agreeing adjectival attribute, inflecting it leads to internal inflection. In 1910, Christian Morgenstern still wrote *mit langer Weile* (with long:F.DAT.SG while(F)) ‘with boredom’; today this is *mit Langeweile*, with internal inflection again expunged. The superlative of such univerbated German adjectives as *schwerwiegend* (heavy:weighing) ‘grave’ and *naheliegend* (near:lying) ‘obvious’ would be *schwerstwiegend* and *nächstliegend* (and analogously for the comparative); but now internal inflection is mostly avoided, producing *schwerwiegendst* and *naheliegendst*. Colloquial Northern German univerbates the complex determiner *so ein* (so a) ‘such’ to *son* [zo:n]. In principle, this product shows the declension of the indefinite article. However, since it is synonymous with the determiner *solch* ‘such’, it also inflects for plural: *sone Sachen* (such:PL thing:PL) features a plural suffix incompatible with the base *ein*. S. also §3.2.1.2 for inflection of *zufrieden*.

5.3 Semantic change

Lexical semantic change occurs both in phrases, producing idioms, and in words, be they monomorphemic or morphologically complex. If it occurs in complex constructions, it reduces and possibly annihilates compositionality. This is the chief semantic aspect of isolation (§2.3). A regular mapping of content on expression is relatively more normal for a complex construction, while it is all but excluded for a monomorphemic sign. Given this continuum, loss of compositionality is a process accompanying univerbation in lexicalization. One example suffices to show what is meant: The German syntactic combination *auf Grund* has a rather literal sense ‘on ground’. The preposition *aufgrund* means ‘on the basis (of)’, with metaphorical extension of the lexicalized product and subsequent grammaticalization to a causal preposition. While such cases are legion, they play little criterial or diagnostic role in the analysis of univerbation, since loss of compositionality is neither a necessary nor a suffi-

³³ Another process which normally presupposes analytic syntax is anaphora on a nominal constituent: it tends to become impossible once this is turned into the *determinans* of a compound. However, the empirical situation is far from clear, so I leave this criterion alone.

cient condition for univerbation: it does not happen in the agglutination of a grammatical formative, but it may also happen in the idiomaticization of a phrase.³⁴

5.4 Phonological and orthographic adaptation

5.4.1 Prosodic adaptation

Converting an expression into a word also implies subjecting it to rules of word prosody. The product assumes one stress according to the word-stress rules of the language. Whether or not this continues the primary stress of the underlying phrase does not matter. English *blackbird*, *holiday* and NHG *Langeweile*, *Herrgott* (lord:god) ‘Lord’ have word-initial stress in adaptation to the compounding pattern, while the underlying phrases had main stress on the right-hand head. At the same time, the stress that the other component of the univerbated construction may have borne disappears: these examples have no secondary stress on the second component, nor do Late Latin *terraemotus* or *aquaeductus* have a secondary stress on the first syllable. German *Muttergottes* is, again, idiosyncratic in keeping the main stress of the phrase *Mutter Gottes* (on the possessive attribute) instead of obeying the general rule for word stress.

5.4.2 Segmental adaptation

Phonological adaptations following univerbation may be subsumed under welding (Diagram 1). They are of two kinds: some are just consequences of rules of word sandhi replacing phrase sandhi, others are more or less irregular reductive changes. When Pre-Latin *de inde* (from thence) is univerbated to Latin *deinde* ‘thence’, regular synizesis diphthongizes the vowels meeting at the erstwhile word boundary so the product becomes disyllabic. When Vulgar Latin *de ex* (from out) univerbates to Old Castilian *des* ‘since’ (later *desde*), the /ks/ is reduced to /s/, a regular sound change in the language history. On the other hand, the reduction of German *so ein* to [zo:n] (§5.2) is completely irregular.

Diagram 6 Genesis of Spanish *delante*

language	VULGAR LATIN	OLD CASTILIAN			
	attested	384	13 th cent.	(~ 9 th cent.)	950
stage	1	2	3	4	5
form	in ante	enante	*de enante	denante	delante
gloss	in before		from front		
meaning	forward	before		in front	in front

Diagram 6 summarizes the changes leading to the Modern Spanish preposition *delante*, most of which are documented (Lehmann 2019). The transition from stage 1 to stage 2 is a univerbation (which happens to be documented late). At stage 3, the univerbated product is made dependent on a governing preposition. At stage 4, this combination again univerbates, with

³⁴ Sometimes, e.g. in Lipka 1981, loss of compositionality is regarded as a necessary and sufficient condition of lexicalization. In line with this, authors (some are mentioned in Alonso Calvo 2009: 35f) have repeatedly turned this into the empirical claim that lexicalization is normally accompanied by loss of compositionality. Brugmann (1900: 138f), who takes “Modifikation der Bedeutung des syntaktischen Wortverbands” (‘modification of the meaning of the syntactic word association’) to be a characteristic of compounding, is an early advocate of this approach.

fusion of the two /e/s. Finally, the transition to stage 5 consists in an irregular change from /n/ to /l/.

5.4.3 Orthographic adaptation

It goes without saying that orthography is an unreliable witness of changes in the language. Still, variation in writing without consideration of a norm may reflect variation at other levels of the language system. And given that orthographic norms are conservative, one can be relatively sure that an adjustment of the norm is a reaction to a change that occurred earlier in the history of the language. Variation of this kind can be observed extensively in the writing of complex prepositions as it appears, e.g., in documents of Late and Vulgar Latin (Lehmann 2019). Thus, the undated Latin inscription from ancient times in E36 below has *ab ante* ‘from the front’, while the Itala (~300) has *abante*, the word which is now Italian *avanti* ‘forward’ and French *avant* ‘before’. An inscription from AD 188 has *de post* ‘from after’, while the grammarian Pompeius (~500) has *depost*. More recent examples include the Castilian ones adduced in §5.4.2.³⁵

Italian “phonosyntactic doubling” is a less common example of the reflection of univerbation in writing. The Tuscan Italian lengthening of a word-initial consonant occurs if the word is preceded by one of a set of other words – typically, a preposition – in the same phrase, like *a casa* [a'k:aza] ‘at home’. This phonological change is canonized as orthographical gemination only if the sequence is univerbated. This produces lexemes like *davvero* ‘really’ and grammatical forms like *della* (of:DEF.F.SG), *alla* (to:DEF.F.SG) etc. In other words, appearance of phonosyntactic doubling in the orthography presupposes univerbation.

5.5 The gradience of univerbation

Conceived in simple terms, univerbation is the replacement of a syntactic by a morphological boundary. This would seem to be an abrupt shift. However, the above discussion has shown that univerbation is actually a gradient phenomenon, and this in two respects:

First, as visualized by Diagram 1, univerbation is just a phase of a more extensive, but uniform process of coalescence which comprises phases preceding and following it. Univerbation is therefore one or more steps in a gradience.

Second, as noted in §3.2.1.1, a grammatical boundary is a linguistic construct marking levels of grammatical complexity. These form a hierarchy which is, again, uniform. In other words, levels $n+1$ and n differ along the same parameters as levels n and $n-1$. The parameters are situated in the various domains of the linguistic system, i.e. in phonology, semantics and grammar. Shifting a grammatical construction from one to another level involves the application of changes along these parameters. Since not all of these changes happen simultaneously, the shift is gradient.³⁶ From a methodological perspective, these changes are the symptoms of univerbation (cf. Haspelmath 2011a, §3). They allow us to recognize a process which itself is unobservable.

³⁵ Fiehler 2011 produces evidence of variation concerning joint or separate writing of five common Modern German phrases.

³⁶ Haspelmath 2011b investigates a set of properties that are widely assumed to distinguish the levels of the word and the phrase and shows that they do not correlate sufficiently to base a definition of the ‘word’ as a cross-linguistically applicable notion on them.

5.6 Univerbation and reanalysis

Reanalysis is the alternation between two structural analyses of a construction. The construction in question involves a grammatical boundary between its components; its reanalysis changes this boundary, possibly relocating it or reducing its strength (Lehmann 2015, ch. 4.3.2). The change of a grammatical boundary is a feature shared by univerbation and reanalysis. In this respect, these two processes may be related, in any given instance of a grammatical change, in two ways: Either the univerbation is preceded by a reanalysis or it constitutes a reanalysis in itself.

5.6.1 Univerbation preceded by reanalysis

The former possibility concerns transgressive univerbation. Consider again the univerbation of the auxiliary with the full verb shown in E6, E19 and E30. Auxiliary univerbation is widely attested in the languages of the world; the complex of conjugation prefixes in Swahili is another case. E34 (= E6) and E35 are standard examples of the grammaticalization of the synthetic Romance future, forming a pair of an intransitive and a transitive full verb.

E34	VULGAR LATIN		PROTO IBERIAN ROMANCE		SPANISH
	cantare habeo	>	cantar he	>	cantaré
	sing:INF have(PRS):1.SG		sing:INF have.PRS.1.SG		sing:FUT:1.SG
	‘I have to sing’				‘I will sing’

E35	VULGAR LATIN				SPANISH
	librum comparare habeo	>	compraré un libro		
	[book(M):ACC.SG buy:INF] have(PRS):1.SG		buy:FUT:1.SG INDF(M.SG) book(M)		
	‘I have to buy a book’				‘I’ll buy a book’

In the input construction of E35, the entire phrase headed by the full verb depends on the auxiliary as indicated by the brackets. In the output structure, the auxiliary is univerbated only with the full verb. This looks like a transgressive univerbation. However, with increasing grammaticalization, the structural scope of the auxiliary shrinks, and prior to univerbation, it combines directly with the full verb. This produces an intermediate stage in which the complex composed of the full verb and the auxiliary forms a periphrastic verb form: *librum [comparare habeo]* ‘I will buy a book’. This step involves a reanalysis which shifts the syntactic boundary initially separating the full verb and the auxiliary to a position outside their complex. This change is led by intransitive full verbs as in E34, whose dependents, if any, are less intimately related to them. At any rate, what univerbates is the periphrastic verb form [verb:INF aux:FIN]. This, in turn, is a phrasal, not a transgressive univerbation.

The same approach may apply to the univerbation of complex adpositions. One strategy of their formation, common in the history of the Romance languages (cf. Lehmann 2019), is illustrated by E36.

E36	hunc	ab	ante	oculis	parentis
LATIN	D.PROX:M.ACC.SG	[from	[before	eye(M):ABL.PL	parent(M):GEN.SG]

rapuerunt nymphae in gurgite
 rob:PRF:3.PL nymph(F):NOM.PL in whirlpool(M):ABL.SG
 ‘this [child] was robbed before his father’s eyes by nymphs in a whirlpool’
 (Gruter, *Inscriptiones antiquae totius orbis Romani*, 717, 9 [undated])

The initial syntactic constituency of the expression relevant here is as indicated in the gloss. Again, one may assume the existence of periphrastic prepositions in the language. Latin possesses monomorphemic prepositions combining in their meaning a local relation with a spatial region. For instance, *ex* ‘out of’ combines the ablative relation with the interior region. Likewise, there can be a preposition combining the ablative relation with the front region. It would be the periphrastic preposition *ab ante*. In the Romance languages, this univerbates as seen in §5.4. The univerbation would, thus, presuppose a phrase [*ab ante*] and would therefore be phrasal univerbation.

Such an analysis seems appropriate in these cases and may apply in other cases that have been considered examples of transgressive univerbation up to now. However, it does not carry over in a straightforward way to all such cases. Consider again the univerbation of E13 (Span. *a el > al*). For it to be a phrasal univerbation would imply the assumption of periphrastic cased articles, which in turn presuppose synthetic cased articles. These things do not exist in Spanish. The conclusion at this point therefore seems to be: Some transgressive univerbations are only apparent since they involve a prior reanalysis which converts the sequence in question into a periphrastic word form, so that the ensuing univerbation is phrasal univerbation. However, there remain cases of transgressive univerbation not amenable to this analysis.

5.6.2 Univerbation as a reanalysis

We now come to the second way mentioned above in which univerbation relates to reanalysis, viz. that it is itself a kind of reanalysis. The possible theoretical obstacle to this subsumption consists in the fact that univerbation is a gradient phenomenon (§5.5) while reanalysis has always been conceived of as an abrupt switch from one structure to a different structure. However, a grammatical boundary is not either present or absent. It is stronger or weaker; and the difference between two levels of grammatical complexity is constituted by a whole set of applicable or inapplicable operations of the language system (§3.2.1.1). Given this, a reanalysis involving weakening or loss of a grammatical boundary can also be seen as a kind of gradient variation. Under this condition and only then³⁷ can univerbation be seen as a kind of reanalysis, viz. one in which a syntactic boundary separating two words gradually becomes a morphological boundary.

Whether a given construction is treated by speakers as having one or another structure is not something perceptible. Univerbation behaves like any reanalysis in that the linguist cannot identify the moment that it happens. What can be observed are structural adaptations that presuppose the assignment of a different structure. If such changes happen, one knows that the reanalysis has been completed. In the case of univerbation, we have seen evidence of these various symptoms in §§5.1 – 5.4.

6 Univerbation and morphological processes

Univerbation has now been characterized as a type of gradient variation which changes a linguistic system. This entails that it is not, in itself, a component of the language system. On the other hand, there are morphological processes, viz. processes of inflection and word forma-

³⁷ Earlier discussions (Lehmann 2004, Haspelmath 2011a, §2) have demonstrated that gradient phenomena cannot be accounted for as kinds of reanalysis if this is regarded as an abrupt switch, which was then generally the case.

tion. These are integral components of a language system.³⁸ It follows that univerbation is not a kind of word formation (Eisenberg 1998: 232-235), just as the agglutination of a grammatical formative to a host is not a process of inflection. To put it in Saussurean terms, morphological processes are processes of *la langue*, while coalescence is a process of *la parole*.

6.1 Patterns of univerbation

In this section, we look at the systematic and diachronic relations between univerbation and processes of word formation. Among these, analogy plays an important role. Analogy is a more general and pervasive notion than coalescence or morphological process. Any item of *langue* or *parole* may serve as a model for an analogical formation. As a consequence, it is both possible that a univerbation targets a result already represented by a type of complex words of the system (§6.1.1) and that a compounding pattern develops on the basis of a set of lexicalized univerbations (§6.1.2).

6.1.1 Univerbation oriented by compounds

It is possible that a particular univerbation is analogically oriented by existing words of the target category, including existent compounds of the target structure (as already observed by Brugmann 1900). However, one must be careful not to abuse the concept of analogy. Analogy needs a model. Where a specific model is missing, it is always possible to posit a model at a more abstract and general level so that apparently any linguistic formation can be guided by an analogical model. For instance, there is in German no specific model for the univerbation of nominals comprising a postnominal possessive attribute, like *Muttergottes*. At this level of specificity, this univerbation is unique. There is, however, the more general model of the compound noun. And, of course, any univerbation targets the word, which might be said to provide the analogical model for it. The more abstract the model, the weaker its force in linguistic argumentation. In the following, we will restrict our attention to models whose structure is closely matched by univerbations.

Consider again compounding in the Germanic languages as an example. The left-branching type of nominal compounding can be safely reconstructed for Proto-Indo-European. It consists in combining a nominal stem as the *determinans* with a noun as the *determinatum*, as in E24. This compounding pattern was inherited by the Germanic languages, witness such old compounds as those of E25 and E26. And it is still alive in these languages to this day, witness such recent German formations as *Waschmaschine* (wash:machine) ‘washing machine’, although the details have been renewed. At the same time, Proto-Germanic had inherited from Proto-Indo-European the prenominal position of the possessive attribute (the option of its postnominal position notwithstanding). Moreover, since univerbation is omnipresent, this has been available during the entire history of the Germanic languages, too. Thus, nominals consisting of a prenominal possessive attribute and its head often got univerbated; E27 (OHG *haninfoz*) is an early example. This process had the model of the inherited compound in front, which only differed by the absence of a case suffix on the *determinans*. Proof of this analogical effect is the stress adaptation noted in §5.4.1.

³⁸ Word-formation has been viewed as the introduction of regularity into the lexicon, as a “grammaticalization of the primary lexicon” (Coseriu 1982: 8).

6.1.2 Univerbation creating new patterns

Particular instances of univerbation may or may not constitute a pattern. The formation of *Muttergottes* did not. Other instances of univerbation form sets which do establish a pattern. If there is no preexistent analogical model for some process of univerbation, then any pattern created by it will be novel in the language. One historical case of univerbation which took place in the absence of any structural model is provided by Spanish phrasal compounds like *mercado negro* (§3.2.1.3). Several of these have been univerbated, including *camposanto* (field:holy) ‘cemetery’ and *pasodoble* ‘double step’, and may now provide a pattern of compounding.

Classical Latin did not possess many more complex prepositions resulting from the combination of two prepositions than the items listed in Table 7.

Table 7 Latin complex prepositions

form	gloss	meaning
incircum	in:around	around
insuper	in:above	on top
desuper	down.from:above	from above

The Roman grammarians condemned this formation because, as they said, a preposition does not combine with a preposition. And indeed, the syntactic construction illustrated by E36 was narrowly restricted in the language, so it did not, at first, serve as a productive source of such univerbations. Nevertheless, in the further course of the language history, this construction did become current, and the univerbation of the two adjacent prepositions became a productive pattern. Spanish *delante* (Diagram 6), *despues* ‘after’ < Latin *de ex post* (of out.of after) and many more are instances of it (Lehmann 2019).

Table 8 German denominal prepositions

form	gloss	meaning
aufgrund	on:ground	on the basis (of)
anhand	at:hand	using
infolge	in:consequence	owing to
zugunsten	to:favor	in favor (of)

Another structural type of complex adposition is illustrated in Table 8. These (and many more; Lehmann & Stolz 1992, §4.2) are prepositions univerbated from a combination of a simple preposition and a relational noun. This type is more easily fed by the univerbation of a syntactic construction than the previous one since phrasal prepositions of the kind *in Folge von X* (in consequence of X) are omnipresent in German as in many other languages. This pattern has been productive in the language for several centuries. Some of these prepositions are old and not easily traced back to univerbations because the presupposed syntactic constructions are not attested. An appropriate assumption appears to be that once some early instances had been produced by the univerbation of phraseological constructions (§2.2), their set established a pattern on which further complex prepositions could be formed by compounding.

Such phrasal univerbations like E27 were sufficiently uniform and numerous to establish a pattern of nominal compounding.³⁹ The pattern of the German left-branching determinative compound containing a juncture element which underlies formations from E27 to NHG *Blumenladen* ‘flower shop’ is highly productive although it is no longer fed by univerbation. Modern products like these cannot be derived by univerbation, as there is in present-day German no rule of syntax which would generate a syntactic construction *Blumen Laden*. As a consequence, German now has compounding both with and without a juncture element (§ 4.2.1).⁴⁰

The Proto-Romance univerbation of the adverbials of the type *obstinata mente* to complex words like Italian *ostinatamente* (E1) had a rather weak analogical model in seven extant univerbations with *opere* like E10. It nevertheless became immensely productive, as *-mente* is now a formative for the adverbialization of adjectives. In cases like this, univerbation affects a set of phrases sharing a specific item as one of their components. Afterwards, the latter becomes a derivational or grammatical formative. NHG *-maßen* (Table 3) and *-weise* Table 2 constitute other cases of this kind.

6.2 The diachronic relation between univerbation and compounding

Brugmann (1900, §2) assumes that any process of compounding originates in the univerbation of syntactic constructions. In order to appreciate this hypothesis, one has to keep in mind that different stages of the language may be involved. A certain pattern may be one of compounding at stage 2 and may be inexplicable, at this stage, by univerbation because there is no syntactic construction corresponding to it. The original formations corresponding to the pattern may, however, go back to some stage 1 where the syntax was yet different and did generate constructions which could univerbate in the way hypothesized. The German left-branching nominal compounds like E23a and E27 are a case in point. Another example is provided by German complex adjectives like *mannstoll* (man:GEN:crazy) ‘nymphomaniac’ and *geisteskrank* (mind:GEN:ill) ‘mentally ill’. Preceding the univerbation of these adjectivals in Middle High German, the adjectives governed a genitive, which they no longer do in Modern German. Thus, at the present stage, such adjectives may or may not instantiate a compounding pattern; but they cannot be derived from contemporary syntactic constructions.

Similarly, German has compounds like *Junggeselle* (young:companion) ‘bachelor’ and *Bösewicht* (wicked:wimp) ‘villain’. They cannot be formed by univerbation of a Modern German syntactic construction since the adjective attribute has to agree with its head. This was not, however, obligatory in Middle High German, when they did originate in this way (Paul 1920, §135).

While such examples are amenable to Brugmann’s analysis once diachrony is taken into account, it is not clear how Brugmann would explain the “proper compounds” of Proto-Indo-

³⁹ Brugmann (1900, §2) speaks of “Nachahmung” (‘imitation’).

⁴⁰ Compounds having the *determinans* in the genitive like *Waldesruhe* can, in principle, originate by univerbation of a possessive attribute construction. In a language like German, the condition of being insensitive to the context would, however, be violated by such genitive attributes which require an article, like *des Waldes Ruhe*, because the article would have to change after univerbation (*die Waldesruhe*; cf. §5.1.3 on *Freudenfest*). Univerbation would therefore only apply to such attributive constructions whose genitive attribute requires no article, like *Gottesurteil*; and these would provide the model for compounding. It remains to be checked whether the historical documentation bears this out.

European and its descendants. This particular compounding pattern cannot be directly based on univerbation because univerbation presupposes a syntactic construction, and a bare stem is not a possible component of a syntactic construction in Proto-Indo-European. Given the diachronic staggering just conceded, one might postulate the required syntax for a Pre-Proto-Indo-European stage of the language. Since this is not, of course, documented, such a postulate would amount to a *petitio principii*. In order to uphold it, one would have to show that no origin other than univerbation can be demonstrated for any known pattern of compounding.

A methodological consequence of this state of affairs for the analysis of neologisms is the following: At a stage of the language where compounding patterns differ from the patterns of phrasal syntax and where, in particular, the order of constituents differs at these two levels, it is easy to distinguish a compound from a product of univerbation. At a stage where the branching direction is the same at the two levels, it may be possible to tell, for a given neologism, whether it originated by compounding or by univerbation only if one has historical evidence of its origin; and even then it may be impossible. In the absence of historical evidence, undecidable cases arise. In Cabecar, both possessive attribution and nominal compounding are left-branching and lack any morphological marking. For instance, *kɔnó yéria* (paca hunter) could be the complex nominal ‘hunter of pacas’ or the compound ‘paca hunter’ [own data]. In such a situation, only prosody could distinguish the two processes. Likewise, under such circumstances, a univerbation pattern may be impossible to distinguish from a compounding pattern.

7 Causes of univerbation

The set of processes promoting the unity of the word enable us to form a dynamic concept of the word, in the sense of ‘lexeme’: a significative unit emerging as the product of a set of processes which exempt it from regular operations of the system, thereby converting it into an element of the inventory. The coactive force of this set of processes is unmistakable; their target is the word. This is so despite the impossibility of delimiting the word consistently against neighboring concepts in a static view (Haspelmath 2011a, §2 and 2011b). The fruitful question is not: “Is this item a word or not?” but rather: “What is the position of the item on the various parameters characterizing its autonomy?”

Concerning the forces promoting univerbation, it should first be recalled that this question is not identical to the question of the forces leading to lexicalization, for two reasons: First, univerbation occurs both in lexicalization and in grammaticalization. Second, it is but one phase of both of these processes.

In inquiring into the underlying mechanisms, it is instructive to return to the discussion between Brugmann (1900) and Paul (1903) mentioned in §1. Concerning examples like Latin **ne-ullus* (NEG-any) > *nullus* ‘none’ (cf. §3.2.3), Brugmann takes it for granted that the component meanings ‘not’ + ‘any’ have been united into one concept, so the univerbation follows the semantic union. Paul argues that no semantic modification takes place in *nullus*; the combination is just as compositional as in *non alius* ‘not another’.

Champions of function-leads-form conceptions have adduced semantic unity as a condition and trigger of univerbation; Brugmann is an early proponent of this theorem. According to Nerius et al. 2000: 163, univerbation affects “Wortgruppen, die eine relativ abgegrenzte, einheitliche Gegebenheit der objektiven Realität (Gegenstand, Eigenschaft, Vorgang, Beziehung) benennen und damit eine den Einwortlexemen ähnliche Nominationsfunktion

übernehmen”⁴¹. This, however, is only true for noun-creating phrasal univerbation as, e.g., in E27. It does not hold for univerbation feeding other word classes, and much less for transgressive univerbation. In negative proforms like Latin *nullus* < *ne ullus*, the univerbation appears to be a consequence of the regular combination of a scope particle with the focused constituent.

This leads us to the opposite position, taken by advocates of usage-based approaches.⁴² For them, the routinization of collocations is the triggering factor of univerbation. “Chunking is the process by which sequences of units that are used together cohere to form more complex units.” (Bybee 2010: 7) The main responsible factors are automation and predictability (*op.cit.* 316). Chunking is a very general cognitive and neuromotor process within which the formation of linguistic units is but a variety (Haiman 1994: 8, Bybee 2007, 2010, ch. 3.2). The main factor is frequent cooccurrence (Taylor 2010). The units thus formed may be continuous or discontinuous. They can, however, be univerbated only if they are continuous. Moreover, frequency of cooccurrence can hardly be held responsible for univerbations like E27. Consequently, chunking may be a decisive factor in transgressive univerbation, but is insufficient to account for phrasal univerbation.

Since a word is a linguistic sign, combining content with expression, it is to be expected that forces targeting the word are both of the functional and of the structural variety. The former are decisive in phrasal univerbation, the latter are decisive in transgressive univerbation. The mental construct fixed by univerbation is a concept in the former, but an operation in the latter case. Given that univerbation is an aspect of both grammaticalization and lexicalization and that both of these are characterized by the interaction of functional and formal parameters, this association applies likewise to univerbation: univerbation is driven by functional and formal factors.

8 Conclusion

The main theses of the present article are the following: univerbation is a spontaneous process taking place at the level of *parole*. It affects expressions *in situ*; in other words, it does not presuppose or involve any transposition into a different context. It takes place either inside a syntactic phrase and then may form new concepts, or it works across a higher-level syntactic boundary and then produces inflected forms and complex grammatical formatives. Word formation and, in particular, compounding differ from univerbation in being processes of the language system, thus, at the level of *langue*. Several processes of word formation whose results look like univerbations actually involve some transposition of an expression into a different context, thus being coercive rather than spontaneous. Nevertheless, univerbation and compounding are intimately related since a pattern of compounding may be based on prior cases of univerbation.

⁴¹ ‘word groups designating a relatively delimited, unitary phenomenon of objective reality (object, property, process, relation) and thereby taking on a function of nomination similar to one-word lexemes’.– Baché (2013) speaks of “Begriffskonsolidierung” (‘concept consolidation’), though he does not claim this to be the only factor at work.

⁴² It should be noted that most of what 21st century advocates of a “linguistics of usage” have to say about the formation of complex units in speech (dubbed “chunking”) is a reinvention of what can be read in Paul 1920, with the difference that Paul discusses the problems with much more circumspection and ten times as many examples of different types and languages.

Abbreviations in glosses

ABL	ablative	INDF	indefinite
ACC	accusative	INF	infinitive
ADJVR	adjectivizer	INSTR	instrumental
ADVR	adverbializer	M	masculine
CAT	catalytic element	N	neuter
COND	conditional	NOM	nominative
D.DIST	distal demonstrative	NR	nominalizer
D.MED	medial demonstrative	PL	plural
D.PROX	proximal demonstrative	PRF	perfect
DAT	dative	PRS	present
DEF	definite	PST	past
F	feminine	SBJ	subject
FIN	finite	SBVR	substantivizer
FUT	future	SG	singular
GEN	genitive	TERM	terminative
IDNT	identifier		

Other abbreviations

MHG	Middle High German	OHG	Old High German
NHG	New High German	PIE	Proto-Indo-European

References

- Alonso Calvo, Raquel 2009, "Lexicalización y colocaciones: una introducción a su estudio diacrónico." *Onomázein* (Santiago de Chile) 19: 33-56.
- Baché, Nathanael 2013, *Der Weg von syntaktischer Fügung zum Wort. Eine Analyse deutscher substantivischer Univerbierungen unter synchroner, diachroner und sprachvergleichender Perspektive*. Frankfurt/M etc.: P. Lang (Europäische Hochschulschriften, 2030).
- Benveniste, Émile 1966, "Formes nouvelles de la composition nominale". *BSL* 61: 82-95. Reimpr.: Benveniste, Emile 1974, *Problèmes de linguistique générale II*. Paris: Gallimard (Bibliothèque des Sciences Humaines – NRF); 163-176.
- Brinton, Laurel J. & Traugott, Elizabeth C. 2005, *Lexicalization and language change*. Cambridge etc.: Cambridge University Press (Research Surveys in Linguistics).
- Brugmann, Karl 1900, "Über das Wesen der sogenannten Wortzusammensetzung." *Berichte der Königlich-Sächsischen Gesellschaft der Wissenschaften, philologisch-historische Classe* 52: 359-401. (Reimpr.: Lipka & Günther (eds.) 1981:135-178)
- Brugmann, Karl 1905, "Ἐκὼν und seine griechischen Verwandten". *Indogermanische Forschungen* 17/1: 1-11.
- Bybee, Joan L. 1985, *Morphology. A study of the relation between meaning and form*. Amsterdam: J. Benjamins.
- Bybee, Joan L. 2007, *Frequency of use and the organization of language*. Oxford etc.: Oxford University Press (Online ed.: 2011).
- Bybee, Joan L. 2010, *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Coseriu, Eugenio 1982, "Les procédés sémantiques dans la formation des mots." *Cahiers Ferdinand de Saussure* 35 [1981]: 3-16.
- Donalies, Elke 2016, "Eine Zeitlang – Über die ärgerliche Univerbierung". *Sprachreport* (Mannheim) 2/2016: 34-39.
- Eisenberg, Peter 1998, *Grundriß der deutschen Grammatik*. Band 1: *Das Wort*. Stuttgart & Weimar: J.B. Metzler.

- Fiehler, Reinhard 2011, "Korpusbasierte Analyse von Univerbierungsprozessen." Konopka, Marek & Kubczak, Jacqueline & Mair, Christian & Šticha, František & Waßner, Ulrich Hermann (eds.), *Grammatik und Korpora 2009. Dritte internationale Konferenz. Mannheim, 22.-24.9.2009*. Tübingen: G. Narr (Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache, 1); 141-155.
- Fillmore, Charles & Kay, Paul & O' Connor, Mary 1988, "Regularity and idiomaticity in grammatical constructions: The case of *let alone*." *Language* 64: 501-538.
- Gaeta, Livio 2011, "Univerbazione". Simone, Raffaele & Berruto, Gaetano & D'Achille, Paolo (eds.), *Enciclopedia dell'italiano*. Vol. 2. Roma: Istituto dell'Enciclopedia Italiana G. Treccani; 1533-1534.
- Grimm, Jacob 1826, *Deutsche Grammatik*. Zeiter Teil. Göttingen: Dieterich.
- Hackstein, Olaf 2012, "When words coalesce: chunking and morphophonemic extension." Melchert, H. Craig (ed.), *The Indo-European verb. Proceedings of the Conference of the Society for Indo-European Studies, Los Angeles 13-15 September 2010*. Wiesbaden: L. Reichert; 87-104.
- Hackstein, Olaf 2014, "Univerbierung und irreguläre Reduktion in temporalen Adverbien: uridg. *gestern* von Bopp bis heute." Melchert, H. Craig & Rieken, Elisabeth & Steer, Thomas (eds.), *Munus amicitiae Norbert Oettinger a collegis et amicis dicatum*. Ann Arbor & New York: Beech Stave; 32-45.
- Haiman, John 1994, "Ritualization and the development of language." Pagliuca, William (ed.), *Perspectives on grammaticalization*. Amsterdam & Philadelphia: J. Benjamins (Current Issues in Linguistic Theory, 109); 3-28.
- Harris, Alice C. & Campbell, Lyle 1995, *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
- Haspelmath, Martin 1993, "The diachronic externalization of inflection." *Linguistics* 31(2): 279-309.
- Haspelmath, Martin 1997, *Indefinite pronouns*. Oxford etc.: Oxford University Press (Oxford Studies in Typology and Linguistic Theory) (2nd ed.: 2004).
- Haspelmath, Martin 2002, *Understanding morphology*. London: Arnold (Understanding Language Series).
- Haspelmath, Martin 2011a, "The gradual coalescence into "words" in grammaticalization." Narrog, Heiko & Heine, Bernd (eds.), *The Oxford handbook of grammaticalization*. Oxford etc.: Oxford University Press (Oxford Handbooks in Linguistics); 324-355.
- Haspelmath, Martin 2011b, "The indeterminacy of word segmentation and the nature of morphology and syntax." *Folia Linguistica* 45: 31–80.
- Himmelmann, Nikolaus P. 2004, "Lexicalization and grammaticization: Opposite or orthogonal?" Bisang, Walter & Himmelmann, Nikolaus & Wiemer, Björn (eds.), *What makes grammaticalization? A look from its components and its fringes*. Berlin & New York: Mouton de Gruyter (Trends in Linguistics, 158); 21-42.
- Hopper, Paul J. & Traugott, Elizabeth C. 1993, *Grammaticalization*. Cambridge: Cambridge University Press (Cambridge Textbooks in Linguistics) (2nd ed. 2003).
- Kortmann, Bernd 1997, *Adverbial subordination. A typology and history of adverbial subordinators based on European languages*. Berlin & New York: Mouton de Gruyter (Empirical Approaches to Language Typology, 18).
- Lehmann, Christian 1993, "Theoretical implications of processes of grammaticalization". Foley, William A. (ed.), *The role of theory in language description*. Berlin & New York: Mouton de Gruyter (Trends in Linguistics, 69); 315-340.
- Lehmann, Christian 2004, "Theory and method in grammaticalization". *Zeitschrift für Germanistische Linguistik* 32,2 [publ. 2005]; 152-187.
- Lehmann, Christian 2015, *Thoughts on grammaticalization*. 3rd edition. Berlin: Language Science Press (Classics in Linguistics, 1).
- Lehmann, Christian 2017, "Grammaticalization of tense/aspect/mood marking in Yucatec Maya". Bisang, Walter & Malchukov, Andrej (eds.), *Unity and diversity in grammaticalization scenarios*. Berlin: Language Science Press (Studies in Diversity Linguistics, 16); 173-237.

- Lehmann, Christian 2019, "Complex spatial prepositions from Latin to Castilian". Fagard, Benjamin & Pinto de Lima, José & Stosic, Dejan (eds.), *Les prépositions complexes dans les langues romanes*. Numéro spécial de *Revue Romane* 54/1: 93-125.
- Lehmann, Christian & Stolz, Christel 1992, *Bildung von Adpositionen im Deutschen*. Erfurt: Seminar für Sprachwissenschaft der Universität (ASSidUE, 6).
- Lipka, Leonhard 1981, "Zur Lexikalisierung im Deutschen und Englischen." Lipka & Günther (eds.) 1981: 119-132.
- Lipka, Leonhard & Günther, Hartmut (eds.), *Wortbildung*. Darmstadt: Wissenschaftliche Buchgesellschaft (Wege der Forschung, 564).
- Marouzeau, Jules 1951, *Lexique de la terminologie linguistique. Français - allemand - anglais - italien*. Paris: Librairie Orientaliste Paul Geuthner (3e éd. augm. et mise à jour).
- Martincová, Olga 2015, "Multi-word expressions and univerbation in Slavic." Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-formation. An international handbook of the languages of Europe*. 5 vols. Berlin & Boston, Mass.: de Gruyter Mouton (Handbücher zur Sprach- und Kommunikationswissenschaft, 40); 1: 742-756.
- Meillet, Antoine 1912, "L'évolution des formes grammaticales." *Scientia (Rivista di Scienza)* 12: 384-400.
- Nerius, Dieter et al. 2000, *Deutsche Orthographie*. Mannheim etc.: Dudenverlag (3. Aufl).
- Opfermann, Andreas 2016, *Univerbierung. Der passive Wortbildungsmechanismus*. Hamburg: Baar (Studien zur historisch-vergleichenden Sprachwissenschaft, 8).
- Paul, Hermann 1903, "Das Wesen der Wortzusammensetzung." *Indogermanische Forschungen* 14: 251-258 (Reimpr.: Lipka & Günther (eds.) 1981: 179-186).
- Paul, Hermann 1920, *Prinzipien der Sprachgeschichte*. Tübingen: M. Niemeyer (5th ed.). (English ed.: *Principles of the history of language*. Transl. from the 2nd ed. of the orig. by Herbert A. Strong. London: Sonnenschein, 1888. Reimpr.: College Park : McGrath, 1970.)
- Schlücker, Barbara 2012, "Die deutsche Kompositionsfreudigkeit. Übersicht und Einführung". Gaeta, Livio & Schlücker, Barbara (eds.), *Das Deutsche als kompositionsfreudige Sprache. Strukturelle Eigenschaften und systembezogene Aspekte*. Berlin, New York: De Gruyter (= Linguistik - Impulse & Tendenzen 46); 1-25.
- Silberztein, Max 1990, "Le dictionnaire électronique des mots composés". *Langue française* 87: 71-83.
- Taylor, John 2010, "Language in the mind". De Knop, Sabine & Boers, Frank & De Rycker, Antoon (eds.), *Fostering language teaching efficiency through cognitive linguistics*. Berlin etc.: de Gruyter Mouton; 29-58.
- Vogel, Petra M. 1996, *Wortarten und Wortartenwechsel. Zu Konversion und verwandten Erscheinungen im Deutschen und in anderen Sprachen*. Berlin & New York: W. de Gruyter (Studia Linguistica Germanica, 39).