

CLIPP

Christiani Lehmanni inedita, publicanda, publicata

| | |
|---------------------------------------|---|
| titulus | Univerbation |
| huius textus situs retis mundialis | http://www.christianlehmann.eu/publ/lehmann_univerbation.pdf |
| dies manuscripti postremum modificati | 15.12.2018 |
| ocasio orationis habitae | Kolloquium Allgemeine Sprachwissenschaft, Universität Göttingen, 6.11.2018 |
| volumen publicationem continens | ignotum |
| annus publicationis | ignotus |
| paginae | ignotae |

Univerbation

Christian Lehmann
University of Erfurt

Abstract

Univerbation is the syntagmatic condensation of a sequence of words that recurs in discourse into one word, as when the Spanish combination *a tras* ‘to back’ becomes *atrás* ‘behind’. Univerbation is to be distinguished from a set of similar phenomena:

Univerbation is a spontaneous process. In this it differs from the conversion of a syntactic construction into a word, as in Spanish *nomeolvides* ‘forget-me-not’. It is process of *la parole* and therefore not a process of word-formation, which is a process of *la langue*.

This is, at the same time, the difference between univerbation and compounding, the latter being a kind of word-formation. Nouns like Spanish *lavaplatos* ‘dishwasher’ are compounds, since on the one hand, they instantiate a pattern of the system, and on the other hand, they cannot come about by spontaneous coalescence of a syntactic construction *in situ*.

However, univerbations may follow a common syntactic pattern. Spanish *máquina de escribir* ‘typewriter’ and *mercado negro* ‘black market’ are phrasal compounds which have arisen by chunking of syntactic constructions. They differ from univerbations only by not yet being univerbated.

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 Ibero-Romance *de illo castello* becomes Spanish *del castillo*.

Phrasal univerbation follows a functional principle of designating a coherent concept by one word. transgressive univerbation follows a formal principle of chunking elements of discourse that routinely follow each other.

The relationship between univerbation and compounding is complicated. A set of univerbations may share structural features and may therefore evolve into a pattern of compounding. For instance, the Spanish complex words *camposanto* ‘cemetery’ and *pasodoble* ‘double step’ originated by univerbation of a complex nominal, but may now provide a pattern of compounding. As a consequence, sets of complex words may be formed which follow a common pattern, like the Romance adverbs in *-mente*, each of which could theoretically arise by univerbation. As a consequence, univerbation and compounding are not always easily told apart.

The discussion uses empirical evidence and examples adduced in earlier work. Its aim is not to present novel phenomena but to straighten out the relevant conceptual framework, provide a theoretical background for the phenomenology and improve on available analyses.

1 Introduction¹

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 | VULGAR LATIN | | ITALIAN |
| | simplici | mente | > semplicemente |
| | simple:F.ABL.SG | mind(F):ABL.SG | simple:ADVR |
| | ‘in a simple spirit’ | | ‘simply’ |

While this adverbialization is today a synchronic process of word-formation, 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, separating here two morphemes.

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. In the latter case, univerbation is often not easy to distinguish from compounding. Both are processes that produce complex words. It is true that they are interrelated in complex ways, that they are not always easy to distinguish in specific cases and that earlier linguistic terminology sometimes blurred the distinction, as we shall recapitulate in §3. However, as will also emerge from the discussion, they are entirely different processes that 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 | saliscendi |
| ITALIAN | CROSS:way | get.up:get.down |
| | ‘crossroads’ | ‘latch’ |

There are excellent early treatments of our topic,² among them the “usual suspects” Karl Brugmann (1900) and Hermann Paul (1920, chapters 4, 10, 19). Both take it for granted that compounding comes about by univerbation (s. §3 for their use of terms). Brugmann (§4) explicitly excludes “analogische 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 be the leading factor, since in grammati-

¹ Versions of this paper were presented at the Linguistic Colloquia of the Universities of Göttingen (6/11/2018) and Bamberg (12/12/2018). I thank the participants for helpful discussion.

² Among modern book-length historical studies, Baché 2013 for German and Opfermann 2016 for Latin-Romance univerbations may be mentioned.

calization, compositionality is preserved. Even in compounding, it may almost be preserved, as will be acknowledged in §2.³

2 Lexicalization

Univerbation is a feature of both grammaticalization and lexicalization (§5.1). Since the relevant facts of grammaticalization have been treated very often in the past decades (most recently in Haspelmath 2011[g]), more attention will here be paid to univerbation in lexicalization. Taken literally, lexicalization is the entrance of a linguistic expression into the lexicon, i.e. its conversion into a lexical entry. This presupposes that it is not yet a lexical unit before this, which excludes borrowing. Conversion into a lexical entry means that the expression becomes a fixed unit. If it is used in a text, it does not need to be formed by rules of the system; it can be copied as a whole from the inventory. This has obvious psychological correlates relating to the mental lexicon. These may provide an important way of operationalizing and, thus, empirically testing the concept of lexicalization, but will not be pursued here.

Rules of the linguistic system are applied in speech. If a certain mental construct is to be coded in speech which is not matched by a vocabulary item, the speaker composes an expression. The product is, in principle, volatile and not an element of the system. If the mental construct happens to come up again on a different occasion, with the same or another speaker, it may be coded by a different expression. If, however, an expression is taken up and becomes conventional, it becomes an element of the inventory and, thus, of the *langue*. Lexicalization creates conventional signs, i.e. mappings of a signifiens onto a significatum which are not done by the speaker by applying rules of the system and instead are prefabricated elements of the inventory.⁴ Lexicalized formations may survive into later stages of the language independently of whether it remains possible to generate them by rules (Haspelmath 2002, ch. 3.3.1).

An expression instantiating a *langue* without being formed by rules of its system must be an element of its inventory. Consequently, whenever the mapping of expression onto content of a linguistic sign is not compositional, this is a clear indication that the sign is lexicalized. However, lexicalization of a unit does not exclude its accessibility to rules.⁵ This only stops if the unit is isolated (Paul 1920, ch. X and §229) from its paradigmatic context. Putting it in simple general terms, for a linguistic unit to be **isolated** means that it is no longer covered by regular patterns of mapping sound onto meaning. This happens as a consequence of changes of the following kinds (Brugmann 1900, §3, Paul 1920, §229, Bybee 2010, ch. 3.4.2):

- (1) Phonological or semantic changes that operate selectively in only affecting
 - a) signs which are components of syntactic constructions, thus leaving signs which are components of words in a relatively archaic shape
 - b) components of words, thus subtracting them from the general rules of grammar.
- (2) Changes which eliminate one or both components of the compound from the vocabulary so that they no longer occur outside the compound.

³ Recall that the terms ‘composition = compounding’ and ‘compositionality’ are semantically unrelated. A linguistic expression is compositional iff its meaning results exclusively by the application of rules of the language system to the meanings of its components. In Brugmann 1900, §3, *kompositivell* still means ‘related to compounding’.

⁴ Bybee 2010:35: “A multi-word expression is conventionalized if it has been established (tacitly, through repetition) as the appropriate way to say something in a particular community”.

⁵ Cf. Langacker 1987, ch. I.A.2 on the rule-list fallacy.

Examples:⁶

(1a): In Classical Latin, the original meaning of the preposition *ob* ‘facing, opposite’ is only preserved in compounds like *opponere* ‘confront’, while the word itself acquires the abstract sense of ‘because of’, as in *ob merita sua* ‘because of his merits’. The original phonological form of the preposition *prod* ‘for’ is preserved in *prodesse* ‘be useful’, while the preposition itself is *pro*. The univerbated noun *paterfamilias* ‘family father’ features a form of the genitive on the component *familia* which the noun itself no longer has from Old Latin onwards.

(1b) The Latin numerals *sex* ‘six’ and *decem* ‘ten’ undergo phonological reduction in the compound *sedecim* ‘16’.

(2) The root of PIE **potis* ‘potent’ survives in Latin *potest* (potent:be:3.SG) ‘he/she/it can’, but otherwise does not make it into Latin. The original negator *ne* only survives in Latin complex forms like *nescio* (not:know(PRS):1.SG) ‘I don’t know’ and *neque* (not:and) ‘nor’ (§5.2.3), but disappears as an independent negator. Engl. *mermaid* contains *mer*, the Germanic word replaced by *sea* in English.

These kinds of change contribute to what Paul (1920) and Brugmann (1900) call the “isolation” and, as a consequence, the “unity” of the construction which is or becomes a word.⁷ They concern the isolation of the individual word from the patterns of forming complex words of its kind. This process is to be distinguished from the wholesale isolation of the products of word-formation from syntactic constructions. Structural changes may isolate an entire strategy of compounding from the syntax of the language. The inherited prenominal possessive attribute gave rise to the left-branching determinative compounding that the Germanic languages inherited from Proto-Indo-European, which is illustrated by E3.

| | | |
|--------|----|-------------------|
| E3 | a. | thiudangardi |
| GOTHIC | | king:court:SBVR |
| | | ‘kingdom’ |
| | b. | bātweard |
| O.ENGL | | boat:ward |
| | | ‘ship’s watchman’ |

In the Germanic languages themselves, syntax became predominantly right-branching, including the possessive attribute. So the most important compounding strategy got isolated from syntax (§6.1).

Once the unit is isolated in one or another way, its survival presupposes its lexicalization. The consequence of this for methodology is that the isolation, or relative idiosyncrasy, of a linguistic expression is a safe criterion for ascertaining its lexicalization. The complementary methodological consequence, however, is that if a complex expression is formed in complete consonance with the grammatical, phonological and semantic rules of the system, it is hard to ascertain whether it is lexicalized. If the linguist finds in a text an expression formed according to valid rules: how can he recognize whether it is an element of the inventory?⁸ Linguists

⁶ 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.

⁷ Bybee l.c. speaks of (increase of) autonomy instead of isolation, tacitly rededicating the former term and ignoring the latter.

⁸ This is a vexing and in many instances unsolvable question in the philology of dead languages. The chronologically first occurrence of a certain word-formation in an author may or may not be his innovation; a certain sentence may be a bright *bon mot* or a traditional proverb. Often such questions are

who are pessimistic about this possibility enrich the concept of lexicalization by the additional condition of loss of compositionality⁹ so that this becomes a necessary and sufficient condition of lexicalization. This, however, would seem to imply that compounds formed in a completely regular way (E13a – E15a below) cannot be recognized as lexemes of the language, which is an undesirable result. We will come back to this problem in §4.2.2.

What then is the force behind the lexicalization of an expression? It is to be presupposed that interpersonal recognition of a recurrent semantic configuration is facilitated if it is designated by the same expression. Therefore a linguistic expression will be lexicalized if its designatum is perceived as something recurrent. For an entity to be **recurrent** may imply a variety of things:

- In the simplest case, the designatum is an individual that recurs identically and therefore deserves a name. This gives rise to proper names, including designations like *Rocky Mountains*, *Rio Grande*, NHG *Muttergottes* ‘mother of god’ and *Heiligabend* (holy:evening) ‘Christmas eve’, which could be common nouns, but were lexicalized as proper names from start.
- In a similar case, entities are perceived as instantiating a species which deserves to be individuated. Thus, what is perceived as recurrent is not the individual but the species instantiated by it. This gives rise to common nominals, be they syntactically complex like NHG *dicke Milch* (thick milk) ‘curd’ and *saure Gurke* (sour cucumber) ‘gherkin’, be they compound nouns like *fruit juice*, *apple tree*, *headache* (instead of *pain in the head*), *riverboat* (instead of *boat used on rivers*) and NHG *Waldrand* ‘edge of the woods’.
- Recognizing that a species is just a concept whose extension are reified objects, we extend the idea of “something recurrent” to any phenomenon consistently associated with a unified meaning. It then covers those innumerable mental objects designated by fixed phrases or compound words, among them properties designated by complex adjectives like *man-eating* and *blood-stained* and actions designated by incorporative verbs like *breastfeed*, *babysit* or E4 (s. §3.3 for more precision on this formation).

E4 staubsaugen
 NHG dust:suck:INF
 ‘to vacuum’

Finally, complex signs may get lexicalized which are not generally assumed to designate a concept. This is most clearly the case for (complex) adpositions and conjunctions like *in spite (of)* and *whenever*. Such grammatical formatives are operators in the construction of complex expressions. In order to avoid their subsumption under ‘concept’, we will use ‘mental construct’ instead of ‘concept’ as a cover term.

A recurrent mental construct acquires a unitary quality. Consider the notion of red rose. On a first plane, it is formed in a completely analytical fashion by restricting the notion of a thing by one of its possible colors, entirely in parallel with ‘yellow rose’. And yet, my wife might say to me “You are not going to present a bunch of red roses to the missus, are you?”,

only decidable if relevant metalinguistic information is available.

⁹ Thus, e.g., Lipka 1981. In line with this, authors have repeatedly turned this into the empirical claim that lexicalization is standardly accompanied by loss of compositionality. Some are mentioned in Alonso Calvo 2009: 35f. 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.

with an undertone that she would not use in the case of yellow roses. This shows that, on a second plane, ‘red rose’ has become a recurrent and consequently unitary concept, which moreover has acquired additional semantic or pragmatic features.

The acquisition of a unitary character by a mental construct is a gradual process because it depends on the extent of its recurrence and entrenchment in the speech community. Consequently, lexicalization is a gradual process. Expressions may be more or less lexicalized. The expression *red rose* may not yet be lexicalized in English; but *fruit juice* probably, and *headache* certainly is.

The lexicalization of an expression is the more probable the simpler it is. Nevertheless, as the examples have shown, beside the lexicalization of words we find lexicalization of phrases and even of sentences, in the form of phraseologisms, idioms and proverbs.

| | | | |
|-----|----|----------------|-----------------------------|
| E5 | 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 E5 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’).

3 Univerbation: concept and term

3.1 The term

A few terminological remarks are necessary.¹⁰ First, it should be made sure that univerbation has nothing to do with verbs; the general meaning of Latin *verbum* is ‘word’. Second, the first phase of the fate of the term reveals some uneasiness. The concept of ‘univerbation’ was formed by 19th century German-speaking Indo-Europeanists; but they did not call it *Univerbierung* from the outset. From 1850 on, the term *Zusammenrückung* (action noun from *zusammenrücken* ‘move sth. closer together’) became established in linguistics¹¹ and was regularly contrasted with *Komposition* ‘compounding’. Brugmann (1900), however, does not use *Zusammenrückung* and instead the terms *Komposition* and *Wortzusammensetzung* (word:composition) to designate ‘univerbation’; and likewise Paul (1920) calls it *Zusammensetzung*. The term *Univerbierung* appears first in Brugmann 1905 and has since been treated as synonymous with *Zusammenrückung*.¹² At the same time, compounding was called *Zusammensetzung*.

¹⁰ Although the term is defined in all major editions of the wikipedia and other dictionaries, it is not commonly known in linguistics and, e.g., absent from current treatments of chunking in ‘usage-based linguistics’.

¹¹ It appears to have been wrongly translated into other languages. It does not mean ‘juxtaposition’ and much less ‘amalgamation’.

¹² French *univerbation* is a lemma in Marouzeau 1951. First occurrences of English *univerbation* appear in publications of the 1960s.

3.2 The concept

The term ‘univerbation’ is not rarely (e.g. in Haspelmath 2002, ch. 3.3.2) restricted to word-formation, excluding the process from the realm of grammaticalization. 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. 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, §7 below 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 there be, at that time, a syntactic construction containing the components in juxtaposition (cf. Diagram 1). 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 renders some extant univerbation analyses problematic.

Univerbation is a phase in a process of coalescence (Haspelmath 2011[g]). This has the following implications: First, it 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 is not actually an instantaneous conversion, but rather a transition. Stronger univerbation evolves as the symptoms of univerbation to be reviewed in §6 accumulate.

The second implication of the subsumption of univerbation under coalescence is that, both in grammaticalization and in lexicalization, there are phases of coalescence preceding, and others following, univerbation. Preceding univerbation is the genesis of a collocation and its syntactic fixation. Following it is the loss of the morphological boundary and phonological fusion as visualized in Diagram 1. The examples of Table 1 illustrate these phases of coalescence; moreover, the first and third 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 |

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 all this is obligatory in any particular process. However, grammaticalization and lexicalization do not skip 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 access to a construction happens in speech. If the holistic access 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 leads to its univerbation. The point here is that univerbation is not an operation of the linguistic system. It does not apply the system; it changes it. Aspect #2 will be discussed more fully in the next section.

3.3 Synchrony vs. diachrony

Concepts like grammaticalization, lexicalization, coalescence, univerbation etc. are mostly defined as diachronic (or even worse, historical) processes.¹³ However, this is not possible since all linguistic variation has a synchronic and a diachronic side; the other side of diachronic change is necessarily synchronic variation. All the processes named are kinds of directed linguistic variation; but the direction appears only in the diachronic perspective. This is easily shown for univerbation. The synchronic manifestation of univerbation as a diachronic process 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. §6.4). Since orthography is conservative, a norm of writing a combination in one word – in an orthography that comprises 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:base; Table 8) had been written together for decades. The first launch of the spelling reform in 1996 required us to write them separately, like *auf Grund*. The final launch, in 2006, reverted to the earlier joint writing *aufgrund*. Unless we want to assume two diachronic changes back and forth within ten years, it seems

¹³ 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 lack of contact between historical and descriptive linguistics, concepts 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.

more adequate to recognize the observable variation as the synchronic side of an ongoing diachronic change. Fiehler 2011 shows the same for a set of German collocations that are currently in the process of univerbation. Univerbation is a directed process of linguistic variation.

The entire field of word-formation is a prime *locus* of the presence of diachrony in synchrony. Early researchers got much 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 stage 1 to stage 2 may bring it about that the expression belongs to different construction types at the two stages. E4 instantiates incorporation synchronically, but at its origin is a backformation from *Staubsauger* ‘vacuum cleaner’, a nominal compound.

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 are those enumerated as responsible for isolation in §2: The mapping of meaning on expression may cease to be compositional, as in German *versuchen* ‘try’, originally ‘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 German *wundersam* (wonder:ADJVR) ‘miraculous’ (suffix now unproductive) and *fruchtbar* (fruit:ADJVR) ‘fruitful’ (suffix no longer productive with nominal bases).

In §7, the relationship between univerbation and compounding will be seen as a particularly complex interplay of different factors in synchrony and diachrony.

4 Phenomena akin to univerbation

Univerbation is to be delimited against a few similar phenomena. In §3, 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, it is restricted to the agglutination of grammatical formatives, which univerbation is not. Moreover, agglutination has come to designate a technique of morphological structure in which every morph is the exponent of a value of one morphological 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* probably¹⁴ results from the univerbation of the infinitive of the full verb with a form of an auxiliary meaning ‘do’ (E6), which latter was enclitic, witness the unstressability of this past desinence.

¹⁴ 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.

| | | | |
|----|-----------------|-------------|-----------------|
| E6 | 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. E7 is a simple example demonstrating this: *que* ‘and’ is enclitic.

| | |
|-------|---------------------|
| E7 | servusque |
| LATIN | slave(M):NOM.SG:and |
| | ‘and the slave’ |

E8 illustrates the integration of enclitic grammatical formatives with their syntactic and phonological host into a phonological word (Haspelmath 2011[g]: 343).

| | | | | | |
|----|-------------------|----------|-----------------|---|--------------------|
| E8 | 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’ |

| | |
|---------|----------------------------|
| E9 | 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 provided with clitics which latter would precede other conjugated forms of the same word, as in E9.

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. 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 (E10) and conditional (E26), the desinence is stressed.

| | | | |
|-----|------------------|---------------|---------------|
| E10 | PROTO ROMANCE | | SPANISH |
| | cantar | he | > cantaré |
| | sing:INF | have.PRS.1.SG | sing:FUT:1.SG |
| | ‘I have to sing’ | | ‘I will sing’ |

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 has been 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 a lexical stem, 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 com-

pounding.¹⁵ Instead, it may characterize a certain strategy of compounding of a specific language system. NHG *Tellerwäscher* ‘dishwasher’ is left-branching, while its Spanish counterpart *lavaplatos* (wash:plates) is right-branching. 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.

- E11 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 E11a is left-branching, but the corresponding phrase of #b is right-branching. Word-formation has its own rules, and these may correspond to rules of syntax of an earlier stage of the language.

4.2.1 Types of compounds

In the 19th century, the concept of compounding was formed on the basis of compounding techniques of archaic Indo-European languages. These involved using a mere stem as the *determinans*, as in E3 and E12.

- E12 dēmokratía
 A.GREEK people:rule(F):NOM.SG
 ‘democracy’

This very narrow concept of compounding excluded formations such as German *Tageslicht* because the *determinans* is in the genitive. Grimm 1826, ch. III, called the former type “proper compounds” and the latter “improper compounds”. In the middle of the 19th century, the latter were dubbed *Zusammenrückung* ‘univerbation’ (s. §3.1). The point here is that the “proper compound” cannot, in the language system generating it, be analyzed by univerbation since the naked 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 is neutralized in Modern German. NHG compounds like *Bücherkiste* (book:PL:box) ‘book box’ and *Sonnenwende* (sun:CAT:turn) ‘solstice’ do not use the naked stem of the *determinans* and nevertheless cannot be explained by univerbation, either. Most current compounds contain the juncture element (CAT), which may or may not be identical with a declension desinence (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 (§5.2.1.3), while items like *moonlike* and *error-free* (cp. the derivate *flawless*) mark the transition into derivation.

¹⁵ 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.

4.2.2 Compounding and compositionality

All of word-formation creates linguistic signs which are destined for holistic access. Word-formation does not code the semantic relation between the components of a binary construction. This relation is partly based on the meaning of these components and otherwise is inferred by world knowledge (Motsch 1981: 99f). Examples of this are well-known: Minimal pairs like English *alligator shoe* vs. *horseshoe*, NHG *Brillenträger* (glass:CAT:bearer) ‘person wearing glasses’ vs. *Hosenträger* (trousers:CAT:bearer) ‘suspenders’ or *Holzofenpizza* ‘pizza baked in an oven burning wood’ vs. *Steinofenpizza* ‘pizza baked in an oven of stone’ show clearly that the linguistic system does not reliably determine the semantic relation between members of a compound. The words in E13a – E15a are sometimes adduced as examples of fully compositional compounds with a regular mapping of meaning onto form. But even these have a conventional meaning involving a specific semantic relation between their components which is based on world knowledge, as is revealed by a comparison with the #b examples.

- E13 a. fruit juice
 b. jungle juice (mix of liquors)
- E14 a. apple tree
 b. princess tree (*Paulownia tomentosa*)
- E15 a. Diskussionsbeitrag ‘contribution to a discussion’
NHG b. Mitgliedsbeitrag ‘membership fee’

Thus, lexicalization leads to the fixation of a specific semantic relation between the components of a complex expression.

On the other hand, compounds like *white-collar*, *sky-scraper*, *cutthroat*, *dreadnought* ‘heavily armed battleship’ and NHG *Geisterfahrer* (ghost:CAT:driver) ‘person driving in a highway lane in the wrong direction’ have been predominantly non-compositional from the moment of their formation (Paul 1920: 90). The latter set includes many designations of plants and animals like *Vampyroteuthis infernalis* ‘Vampire squid’, *harvestman* = German *Weberknecht* (kind of spider), *Vergissmeinnicht* ‘forget-me-not’, *Maßliebchen* ‘daisy’ etc.

Compounding involves some kind of semantic union between the components of the compound. This includes a fixation of the semantic relation between these. Once the compound is lexicalized, it is available for holistic access. Once it is holistic, there is no need for compositionality. Compounding therefore favors loss of compositionality (“isolation”).

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 2 and illustrated by E16.

Diagram 2 Conversion to word

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

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

As E16 suggests, the verbal group (VG) 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 E16, 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 present case.

In conversion to word, the enforcing context Z2 is by definition a syntactic context. It is a noun phrase introduced by an article in E16. 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*.¹⁶ French *rendez-vous* (move.to:2.PL-you(PL)) is based on a 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. At any rate, conversion to word is a process of word-formation, not one of univerbation.

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* (together:formation) in German linguistics). Here a derivational process applies to a combination of two words or stems which looks like a compound except that such a compound has no independent existence. Spanish examples include *sietemesino* (seven:month:SBVR:M) ‘baby born two months premature’ and *quincañero* (fifteen:year:SBVR:M) ‘teenybopper’. A German example is *Grablegung* (grave:lay:SBVR) ‘burial’. Parasynthesis is like conversion to word in that two stems are forced into one word by a process of word-formation, which in this case is some derivation instead of conversion.

The concept of parasynthesis has been expanded, and may be further expanded, to include other processes of word-formation which force a combination of lexical or derivational items into a unit by combining them with an operator which, in turn, may be a lexical or grammatical item. The enforcing context is the head of a determinative compound in NHG *Sauregurkenzeit* (sour:cucumber:time) ‘silly season’. It is the conjugation morphology in French *embarquer* (in:ship:INF) ‘embark’. 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 enforces this. This is, again, unlike univerbation, which happens *in situ*. Unlike univerbation, compounding, conversion to word and parasynthesis are processes of word-formation.

¹⁶ The verb form in these formations needs further analysis. It is obviously an imperative in *forget-me-not* and its German counterpart *Vergissmeinnicht*. However, an imperative makes little sense in *Tunichtgut*. In Romance formations like Spanish *lavaplatos* ‘dishwasher’, Italian *grattaciello* ‘skyscraper’ etc., the verb form looks like the imperative, too, but actually is either the third singular or the verb stem.

5 Types of univerbation

Once the related phenomena of §4 have been excluded from the concept of univerbation, univerbation is by definition spontaneous and happens *in situ*. We can abduce this defining feature from a simple example like E17 (cf. Table 1).

| | | | |
|-----|-------------------------|--------------------|-----------------|
| E17 | LATIN | | ITALIAN/SPANISH |
| | terrae | motus | > 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, 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 classification can be applied to cases of univerbation:

- univerbation joins either lexical forms or grammatical formatives (§5.1)
- univerbation reduces either a word boundary or a higher-level boundary (§5.2).

5.1 Univerbation in lexicon and grammar

Application of the term ‘univerbation’ has variously been 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. 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 equally well to coalescence into one word in either lexicalization or grammaticalization. E17 and E18 illustrate univerbation in lexicon and grammar, resp.

| | | | |
|-----|---------------|---|----------------|
| E18 | PROTO-ROMANCE | | MODERN ITALIAN |
| | de ab Roma | > | da Roma |
| | of ABL Julia | | from Rome |

Univerbation may affect two particular word forms A and B where neither A nor B univerbates with anything else or univerbates at most sporadically with other forms of the adjacent class. This is the case of E17 and, more generally, in univerbation as a phase of lexicalization.

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). Finally, formative A may be a member of a paradigm all of whose members univerbate with items of class B. Paradigm A then becomes an inflection category of class B. The agglutination of personal pronouns to verbs to become their personal conjugation, or the agglutination of postpositions to nouns to become their case suffixes, are relevant examples. This is what happens in univerbation as a phase of grammaticalization.

The distinction between univerbation in the lexicon and in the grammar will occasionally come up again since certain formal types of univerbation are restricted to one or the other.

5.2 Univerbation across different syntactic levels

There are two subtypes of univerbation by the criterion of the syntactic level at which it works, phrasal univerbation and transgressive univerbation.

5.2.1 Phrasal univerbation

5.2.1.1 Fundamentals

The phrase and the word form constitute two neighboring levels of the hierarchy of grammatical complexity. They differ in their relative susceptibility to operations of the domains of the language system. These characterize the phrase as follows:

- Its components may be distantiated and permuted and may take their own dependents which, in turn, may be intercalated.
- Each of its components may be inflected.
- Its semantics is relatively compositional.
- It is relatively immune to phonological adaptations and may get variable stress assignment.

For the word form, the contrary of all of these properties applies. The correlation among these properties, however, is less than complete. For instance, German separable compounds like *kennenlernen* ‘get acquainted’ are phrases by the syntactic criteria, but words by the semantic criterion. If words are by definition non-separable (impossibility of intercalating material), these are phrases, not words (Schlücker 2012, §4).

Phrasal univerbation is the univerbation of a phrase. An expression which, in its context, is of a phrasal syntactic category is condensed into a word. At this point, the word inherits the category of the head. This is represented in Diagram 3 and illustrated by E19.

Diagram 3 Phrasal univerbation

| | |
|--------|--|
| input | [... [A B] _{XP} ...] _Y |
| output | [... [AB] _X ...] _Y |

E19 SPANISH

| | |
|--------|---|
| input | [el [agua ardiente] _{Nom}] _{NP} DEF.SG water:F burning |
| output | [el [aguardiente] _N] _{NP} DEF.M.SG brandy(M) |

As may be seen, the construction of E19 functions as a nominal in its context before univerbation; after univerbation, it is a noun, which can still function as a nominal. The same goes for E17. Phrasal univerbation is commonplace with such nominal constructions as Spanish *nochebuena* (night(F):good:F) ‘Christmas eve’ and English *blackbird*. More examples are in §6.

A comparison of Diagram 3 with Diagram 2 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 3) 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 2). With phrasal univerbation, the category of the output

structure results automatically by lowering the level of structural complexity given with univerbation, while in conversion to word, the output category is forced by using the expression in a context which requires an expression of that category.¹⁷ It is this contrast which is meant to be covered by the terms ‘spontaneous’ vs. ‘coercive’.

5.2.1.2 Categories of input phrases

While E17 and E19 represent a **bare nominal**, things are slightly different if a **cased noun phrase** is univerbated. In E20, an NP in the ablative functions as an adverbial and is univerbated to an adverb.

| | | | |
|-----|------------------|----------------|---------|
| E20 | LATIN | | SPANISH |
| | hāc | hōrā | > ahora |
| | D.PROX:F.ABL.SG | 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 E21, which becomes an adverb in Modern German.

| | | | | |
|-----|-------------------|-----------------|---------|---------|
| E21 | PRE-GERMAN | | OHG | NHG |
| | *hiu | d[ag]u | > hiutu | > heute |
| | 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.¹⁸

If the dependent component of the noun phrase is an adjective, the univerbation may amount to its **adverbialization**. In E22, a manner adverbial consisting of an ablative noun phrase becomes an adverb.

| | | | |
|-------|---------------------|------------------|-------------|
| E22 | magno | ōpere | > magnōpere |
| 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. 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

¹⁷ 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 an amazing amount of these examples, it is not clear whether they instantiate univerbation or conversion to word; i.o.w. 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.

¹⁸ A rather idiosyncratic pattern of univerbation in German, illustrated by nouns like *Zeitlang* (time:long) ‘while’ and *Handbreit* (hand:wide) ‘a few inches’, is analyzed in Donalies 2016.

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. §6.1 for further analysis.

Table 2 German adverbs in *-maßen*

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

Table 3 German adverbs in *-weise*

| form | meaning |
|------------------|-------------|
| dummerweise | stupidly |
| möglicherweise | possibly |
| glücklicherweise | fortunately |

Another 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*. Most 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 coordinative conjunctions, too.

The univerbation of an auxiliary with its (non-finite) full verb is an instance of phrasal univerbation only if these two components exhaust a phrase. This may presuppose a reanalysis; s. §6.6.

A verb forms a **verbal complex** together with an adverb. The first phase of the coalescence of the complex is the formation of what in English grammar is called a particle verb like *put on* (cf. §5.2.1.3). In ancient Indo-European languages, this kind of construction univerbates to what is traditionally called a *verbum compositum*, i.e. a verb modified by a preverb. The topic is taken up in §6.1; examples are in E27.¹⁹ On further coalescence, the preverb may become a prefix. PIE **anti* ‘facing’²⁰ yields the OHG preverb *ant-/ent-*, which appears as a prefix, e.g., in NHG *entsprechen* (facing:speak:INF) ‘correspond’.

¹⁹ The topic is treated exhaustively in Brugmann 1900, §5.

²⁰ 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.

The traditional term ‘incorporation’ implies that the process is one of univerbation. Deferring the question of the relation between univerbation and compounding to §7, 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 also historically demonstrable in certain cases. Latin *manu tenere* (hand(F):ABL.SG hold:INF) yields French *maintenir* ‘maintain’, Latin *male tractare* (badly treat:INF) yields French *maltraiter* ‘mistreat’. German examples include *haushalten* (house:hold:INF) ‘to housekeep’ and *hohnlachen* (derision:laugh:INF) ‘to flear’. These French and German products are transitive and intransitive verbs, resp.; so they belong to the same distribution class as the underlying construction. Such examples could be multiplied.²¹ These German complex verbs, however, separate unless in final clause position, so the univerbation is either only partly completed or a purely orthographic phenomenon. The participle *feuerspeierend* (fire:spitting), however, is a perfect example, as it is inseparable and intransitive like an incorporative verb.

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

Secondarily, the univerbated product may change its grammatical category. 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 syntactic function. The adjective *zufrieden* ‘content’ results from univerbation in contexts like *jemanden zu Frieden stellen* (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. Cf. also the retransitivization of incorporative verbs (§6.1).

5.2.1.3 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 the phrase 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. The following subtypes of phrasal compounds may be singled out:

- Phrasal compounds consisting of a nominal head and an adjectival attribute are a frequent input to univerbation, as already indicated by E19. Spanish examples include *aire acondicionado* ‘air conditioning’, *corto circuito* ‘short circuit’ and *mercado negro* ‘black market’. German examples include those of E5. These are not compounds proper because they are not words (§2).

²¹ The specialized literature adduces examples like *blankwischen* (shining:wipe), *schönfärben* (pretty:dye), *hängenbleiben* (hang:stay), *spazierengehen* (stroll:go) etc., all of which are separable.

- A phrasal compound whose head is modified by a noun is called *synapsis*²² in Romance linguistics. Spanish examples include *olla a presión* (pot at pressure) ‘pressure cooker’ and *máquina de escribir* (machine of write:INF) ‘typewriter’.
- As anticipated in §5.2.1.2, verbs forming a phrasal compound with an adverb, viz. particle verbs like NHG *zurückschicken* = English *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, the seeming structural difference between Spanish *aire acondicionado* and *aguardiente* may be a merely orthographic one.

As will be discussed more fully in §7, 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, it creates lexical entries like word-formation. In diachronic perspective, the first instances of a pattern of phrasal compounding may result from the consolidation of a syntactic collocation. These may then found a pattern for the formation of complex lexical items. This will be taken up in §7.1.2.

5.2.2 Transgressive univerbation

As we saw before, any univerbation necessarily happens 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 univerbation²³ takes the form of Diagram 4, illustrated by E23.

Diagram 4 Transgressive univerbation

| | |
|--------|---|
| input | [A [B C] _x] _Y |
| output | [AB C] _Y |

E23 SPANISH

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

E23 represents a common case, 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*.

²² 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* ‘computing machine’ 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).

²³ *Cross-border* (or *trans-border*) *univerbation* may be more idiomatic in English, but is inapt as an international term.

Diagram 4 generalizes over the particulars, indicating that preceding univerbation, one of the items undergoing it (B) forms a syntagma of category X (NP in E23) with other material (C). The univerbation reduces the syntactic boundary separating this syntagma from the other univerbated component (A) to a morphological boundary, whereby the construction X becomes unrecognizable. Unless some syntactic reanalysis (§6.6) precedes the process, transgressive univerbation ignores and destroys syntactic structure, viz. the syntactic integrity of X.

Prepositions also univerbate with each other if a preposition governs a prepositional phrase, as in Italian *da* (E18), English *onto* and Spanish *atrás ~ detrás* ‘behind’. A governing preposition coalesces with the relational noun which is the head of its complement in Spanish *encima* (Loc top) ‘on’ and numerous similar examples including those of Table 8. The univerbation of a preposition (and analogously, a postposition in languages which have clause-final subordinative 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 NHG *sodass* (so:that).

Table 5 German univerbated subordinative 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* (that:N.DAT.SG). 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. 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 signaling a conjugation category, more than one morpheme for it is a luxury. These sequences are consequently univerbated (to *wanna* and *gonna*, resp.), again across a syntactic boundary, since the *to* forms a constituent with the full verb.

In certain cases, the auxiliary may form a phrasal unit with the full verb and then univerbate with it. In other cases, it instead univerbates with pronouns or pronominal indexes representing the actants of the full verb. The English contractions *I’ll*, *I’ve*, *he’s* etc. are well-known examples (Bybee 2007: 327; Haspelmath 2011[g], §4). A less widely known case comes from Yucatec Maya. An independent verbal clause is headed and introduced by an auxiliary. It is followed by the verbal group, starting with the pronominal index representing the subject. These indexes are enclitic. They therefore univerbate with the preceding auxiliary. For instance, although in E24, *in* forms a constituent with *meyah*, univerbation produces *ts’o’kin meyah* and *ts’in meyah* by welding (Lehmann to app., §2.5.2.2).

| | |
|---------|----------------------|
| E24 | ts’o’k in meyah |
| YUCATEC | TERM [SBJ.1.SG work] |
| | ‘I have worked’ |

Finally, transgressive univerbation may even act across a finite-clause boundary. The syntactic construction of E25 has the compositional reading indicated.

E25 *nescio* *quis* *p*
 LATIN [ignore(PRS):1.SG [who:NOM.SG *p*]s]s
 ‘I do not know who *p*’

In E25, the position of *nescio* may be taken by any conjugated form of *nescire* ‘to ignore’, and the position of *quis* may be taken 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).

5.2.3 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 4. 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.²⁴ 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 univerbated. What happens with them, instead, is loss of the initial particle of the discontinuous expression.

The negator also coalesces with the verb, as in Latin *nescio* (not:know:1.SG) ‘I don’t know’. From a logical point of view, one may think that the entire clause 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 with a stem. In many languages, the existential verb or verboid has a negative counterpart.²⁵ In such cases, it is possible that univerbation is preceded by a reanalysis which converts the sequence A B of Diagram 4 into a syntagma; s. §6.6. Finally, the complex postposition *notwithstanding* arose by univerbation of the sequence *notwithstanding*.

Summarizing the two classifications of §5 in one cross-classification, we get the picture of Table 6. Inclusion of the examples following further down would not change the logic of Table 6.

²⁴ 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.

²⁵ Ch. Hackstein 2014: 41 on Old Russ. *ně[s-ti]=tu > neŭ > neŭ > Russ. net* (not:EXIST).

Table 6 Types of univerbation

| domain level | lexicalization | grammaticalization |
|-----------------|----------------|--------------------|
| phrasal | E17, E19, E22 | E1, E6 |
| transgressive | X | E18, E23, E24, E25 |

As may be seen, transgressive univerbation is limited to univerbation as an ingredient of grammaticalization. That is, A or B of Diagram 4 or both 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 give origin to new inflection paradigms. This 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 morphological structure results from univerbation, the sequential distance of morphological material also correlates with the strength of the underlying syntactic boundary.

6 Symptoms of univerbation

Since grammatical boundaries are imperceptible, the linguist needs other methodological cues to detect univerbation. The conversion of a syntactic boundary separating the two items in question into a morphological boundary renders possible all the changes distinguishing a word from a phrase enumerated in §5.2.1.1. The following symptoms are the methodological reflection of these changes.

6.1 Reduction of structural variability

Enforcement of continuity: At the initial stage of univerbation, a complex unit may still be discontinuous. Thus, the preposition *instead* is univerbated from *in stead*. *Stead* may still be preceded by its dependent, e.g. *in her stead*. Enforcement of continuity of *in stead* produces *instead of her*. The same goes for the German counterpart *anstatt*.

E26 illustrates the coalescence of the erstwhile conditional auxiliary with the full verb in Portuguese and Spanish (the same goes for the future auxiliary). In Proto-Romance, it takes a preceding infinitive as its complement. This infinitive, in turn, can bear enclitic pronouns. If both the dependent pronoun and the auxiliary attach to the infinitive, the result is as shown E26.

| | | |
|-----|-----------------------------|------------------|
| E26 | PORTUGUESE | SPANISH |
| | ajudar-me-ia | me ayudará |
| | help:INF-me-COND | me help:INF:COND |
| | '(he/she/it) would help me' | |

In Portuguese, the conditional auxiliary follows any enclitics attached to the infinitive, so it is not yet univerbated with the latter. In Spanish instead, univerbation of the auxiliary with the infinitive is completed, nothing can intervene between them. The clitic pronoun therefore has to precede what is now a finite verb form.

Preverbation is the result of the univerbation of an adverb with a verb. 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 E27.

E27 *ob vos sacro* in quibusdam precationibus est pro *vos obsecro*, ut *sub vos placo* pro *supplico* (Festus XIII, §175)

LATIN ‘In certain prayers it says *ob vos sacro* [facing 2.PL.ACC implore:1.SG] instead of *vos obsecro* [‘I implore you’], like *sub vos placo* [beneath 2.PL.ACC supplicate:1.SG] instead of *supplico* [‘I supplicate (you)’].

Something similar happened in the development of such adverbs into adpositions. Colloquial German still allows fronting a demonstrative or interrogative pronoun, in the form of a local adverb, while stranding the adposition governing it, as in E28 (Brugmann 1900, §5.4).

E28 Da kann ich nichts für.
GERMAN there can I nothing for
‘For that I’m not responsible.’

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

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 E17: Classical Latin allows both *terrae motus* (Cic. *div.* 1, 101, 12) and *motus terrae* (Plin. *nat.* 2, 191, 1); but all the Romance descendants presuppose *terraemotus* ‘earthquake’. Likewise, Classical Latin has both *aquae ductus* (Iust. *Dig.* 8, 3, 2, 2, 1) and *ductus aquae* (o.c. 43, 20, 3, 4, 1), which in Late Latin is only *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 the first three forms, in possessive attribution itself, the prenominal 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* (matter(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 an adposition and its pronominal complement provide an exception to the generalization that univerbation presupposes fixation of order. They are univerbated in the order ‘complement-adposition’, as in *dafür* ‘therefore’ (cf. E28), *damit* ‘therewith’, *davon* ‘therefrom’ etc. They thus display a postpositional word order which these adpositions have otherwise never had during the documented history.²⁶ 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.

²⁶ The model is possibly provided by the yet older combinations of the deictic particles *her* VENITIVE and *hin* ANDATIVE with some adverbs and adpositions like *her-/hinein* ‘into it’, *her-/hinzu* ‘to it’.

Reduction of syntactic structure: As long as the construction to be condensed is a syntactic one, its components may contract syntactic relations. Once the construction is a word, it contracts syntactic relations as a whole. For nominal compounds resulting from univerbation, this 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 (*der*) *Freuden Fest* (DEF:F.SG.GEN joy(F):SG.GEN feast(N)) ‘(the) joy’s feast’. The current compound no longer takes an article fitting the first component.

The German adjectives 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*. A search machine does turn up a few examples of *sehr dummerweise* (very stupidly) on the web, which again testify to an incomplete univerbation of this formation.

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 an analytic syntactic construction. 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*), although currently not standard, is attested as early as 1913.²⁷ Again, the nouns in such complex prepositions do not 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).²⁸

Recall that the methodology requires that for a complex expression to be analyzable as formed by univerbation, there must be a syntactic construction of the same form. Now constructions of like *in Bezug auf X*, *an Hand von X*, *auf Grund von X* are attested in German. The problem is that the lack of the definite Artikel violates a rule of syntax. While we are not at present concerned with the theoretical question of what the status of such phraseologisms between syntax and morphology is, we must modify the methodological requirement to admit not only regular syntactic constructions, but also phraseologically impoverished syntactic constructions as sources of univerbation.

Sometimes a univerbated form undergoes further syntactic changes. For instance, the noun created in E19 changes its gender to masculine. Another such change is the retransitivization of incorporative verbs (Hackstein 2012: 89). Older German *wahrnehmen* (attention:take:INF) ‘to perceive’ and more recent *gewährleisten* (warranty:provide:INF) ‘to 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.

²⁷ Medizinalabteilung des Königlich Preußischen Ministeriums des Innern (ed.) 1913, *Krankenpflege-Lehrbuch*. Berlin & Heidelberg: Springer; 317.

²⁸ 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.

6.2 Adaptation of morphological structure

Once univerbation is accomplished, the morphological structure – 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. While Late Latin *terraemotus* has an internal genitive, this disappears in Spanish and Italian *terremoto*. Internal inflection is expunged. In 1910, Christian Morgenstern still wrote *mit langer Weile* (with long:F.DAT.SG while(F)) ‘with boredom’; today this is *mit Langeweile*. If the first component of a univerbation is an adjectival attribute, inflecting it leads to internal inflection. Thus, the superlative of NHG *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* (Haspelmath 1993). Colloquial Northern German univerbates the complex determiner *so ein* (so a) ‘such’ to *son* (with long *o*). 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 §5.2.1.2 for inflection of *zufrieden*.

6.3 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. 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.²⁹ 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, because the same may happen in the idiomatization of phrases.

6.4 Phonological and orthographic adaptation

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 head. At the same time, the stress that the other component of the univer-

²⁹ The discussion between Brugmann and Paul mentioned in §1 raises some serious theoretical problems. For instance, concerning examples like Latin **ne-ullus* (not-any) > *nullus* ‘none’ (cf. §5.2.3), 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’. While Paul seems to be right in principle, the question arises why so many languages besides Latin have negative pronouns. This may be a consequence of the regular combination of a scope particle with the focused constituent.

bated construction may have born disappears: these examples have no secondary stress on the second component, nor do Late Latin *terrae motus* or *aquaeductus* have a secondary stress on the first syllable. NHG *Muttergottes* is, again, idiosyncratic in keeping the main stress of the phrase *Mutter Gottes* instead of obeying the general rule for word stress.

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 synzesis diphthongizes the vowels meeting at the erstwhile word boundary so the product becomes bisyllabic. When Vulgar Latin *de ex* (from out) univerbates to Castilian *des* ‘since’ (later *desde*), the /ks/ is reduced to /s/, a regular sound change in the language history.

Diagram 5 Genesis of Spanish *delante*

| language attested stage form gloss meaning | VULGAR LATIN | OLD CASTILIAN | | | |
|---|--------------|------------------------|---------------------------|----------|----------|
| | 384 | 13 th cent. | (~ 9 th cent.) | 950 | 1042 |
| | 1 | 2 | 3 | 4 | 5 |
| | in ante | enante | *de enante | denante | delante |
| | in before | | from front | | |
| | forward | before | | in front | in front |

Diagram 5 summarizes the changes leading to the Modern Spanish preposition *delante*, most of which are documented. 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 fusion of the two /e/. Finally, the transition to stage 5 consists in an irregular change from /n/ to /l/.

Orthographic adaptation: It goes without saying that orthography is an unreliable witness of changes in the language system. 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 univerbated prepositions in one word as it appears, e.g., in documents of Late and Vulgar Latin (Lehmann to app.). Thus, the undated Latin inscription from ancient times in E30 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 the preceding paragraph.³⁰

Italian “phonsyntactic doubling” is a less common example of the reflection of univerbation in writing. The Tuscan Italian gemination 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 in the orthography 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. I.o.w., appearance of phonsyntactic doubling in the orthography presupposes univerbation.

³⁰ Fiehler 2011 produces evidence of variation concerning joint or separate writing of five common Modern German phrases.

6.5 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 §5.2.1.1, 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.³¹ In a methodological perspective, these changes are the symptoms of univerbation (cf. Haspelmath 2011[g], §3). They allow us to recognize a process which itself is unobservable.

6.6 Univerbation and reanalysis

Reanalysis is the alternation between two analyses of a construction. The construction in question involves a grammatical boundary between its components, and its reanalysis changes this boundary. It may relocate it or reduce 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, they may be related, in any given instance of a grammatical change, in two ways: Either the univerbation is preceded by a reanalysis or it consists in a reanalysis.

6.6.1 Univerbation preceded by reanalysis

The former possibility concerns transgressive univerbation. Consider again the univerbation of the auxiliary with the full verb shown in E6, E10 and E26. It is widely attested in the world's languages; the complex of conjugation prefixes in Swahili is another case. E29 is a standard example of the grammaticalization of the synthetic Romance future.

| | | |
|-----|--|---|
| E29 | 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, 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*

³¹ Haspelmath 2011[i] 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 define the 'word' as a cross-linguistically applicable notion.

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 (standard example: E10), 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 to app.), is illustrated by E30.

E30 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 §6.4. 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 ones 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 E23. 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.

6.6.2 Univerbation as a reanalysis

We now come to the second way mentioned above in which univerbation is related to reanalysis, viz. that it is a kind of reanalysis. The possible theoretical obstacle to this subsumption consists in the fact that univerbation is a gradient phenomenon (§6.5) while reanalysis has always been conceived as an abrupt switch from one structure to a different structure. However, a grammatical boundary is not either there 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 (§5.2.1.1). 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 then³² is univerbation a kind of reanalysis,

³² Earlier discussions (Lehmann 2004, Haspelmath 2011[g], §2) have demonstrated that gradient phenomena cannot be accounted for as kinds of reanalysis if this is regarded as an abrupt switch, which

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, this concerns its various symptoms as analyzed in §§6.1 – 6.4.

7 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-formation. These are integral components of a language system.³³ 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*.

7.1 Patterns of univerbation

7.1.1 Univerbation oriented by compounds

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 compounding pattern develops on the basis of a set of lexicalized univerbations and that a univerbation targets a result already represented by a type of complex words of the system. I.o.w., it is possible that a particular univerbation is analogically oriented by existing words of the target category, including existent compounds of the targeted 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 just any linguistic formation is 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 may 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 E12. This compounding pattern was inherited by the Germanic languages, witness such old

was then generally the case.

³³ Word-formation has been viewed as the introduction of regularity into the lexicon, as a “grammaticalization of the primary lexicon” (Coseriu 1982: 8).

compounds as those of E3. 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 (regardless of the option of its postnominal position). 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. E31 is an early example.

| | |
|-----|-------------------------------------|
| E31 | haninfuoꝝ |
| OHG | rooster(M):GEN.SG:foot |
| | ‘buttercup’ (NHG <i>Hahnenfuß</i>) |

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 §6.4.

7.1.2 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 found 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 and did found a pattern is provided by the Latin complex prepositions resulting from the combination of two prepositions. Classical Latin did not possess many more 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 the formation because, as they said, a preposition does not combine with a preposition. And indeed, the syntactic construction illustrated by E30 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 5), *despues* ‘after’ < Latin *de ex post* (of out.of after) and many more are instances of it (Lehmann to app.).

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 syntactically impoverished constructions (§6.1), their set founded a pattern on which further complex prepositions could be formed by compounding.

Such phrasal univerbations like E31 were sufficiently uniform and numerous to found a pattern of nominal compounding.³⁴ The pattern of the German left-branching determinative compound containing a juncture element which underlies formations from E31 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).

The Proto-Romance univerbation of the adverbials of the type *simplici mente* to compounds like Italian *semplicemente* (E1) had a rather thin analogical model in E22. It nevertheless became immensely productive, as *-mente* is now a grammatical suffix for the adverbialization of adjectives. In cases like this, univerbation affects a set of phrases sharing a specific item as one of their components. In the sequel, the latter becomes a derivational or grammatical formative. NHG *-maßen* (Table 2) and *-weise* Table 3 constitute other cases of this kind.

7.2 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 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

³⁴ Brugmann (1900, §2) speaks of “Nachahmung” (‘imitation’).

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-European and its descendants. This particular compounding pattern cannot directly be 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 syntax presupposed 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. More in general, under such circumstances, a univerbation pattern may be hard to distinguish from a compounding pattern.

8 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 2011[i] and 2011[g], §2). 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 (§2), for two reasons: First, univerbation occurs both in lexicalization and in grammaticalization. Second, it is but one phase of both of these processes.

Champions of function-leads-form conceptions have adduced semantic unity as a condition and trigger of univerbation. As noted in §1, Brugmann 1900 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".³⁵ This, however, is only true for noun-creating phrasal univerbation like, e.g., in E31. It does not hold for univerbation feeding other word classes,

³⁵ 'word groups designating a relatively delimited, unitary phenomenon of objective reality (thing, 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.

and much less for transgressive univerbation. To recall just one example: the elements do not obviously form a semantic unit in negative proforms like Latin *nullus < ne ullus*.

An almost opposite position has been taken by advocates of usage-based approaches.³⁶ 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 (o.c. 316). Chunking is a very general cognitive and neuromotor process, of 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. 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 E31. 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, thus, content with expression, it is to be expected that forces targeting it are both of the functional and 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 both of grammaticalization and lexicalization and that both of these are characterized by the interaction of functional and formal parameters, this carries over to univerbation: Univerbation is driven by functional and formal factors.

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 | 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 |

Other abbreviations

| | |
|-----|---------------------|
| NHG | New High German |
| OHG | Old High German |
| PIE | Proto-Indo-European |

References

³⁶ 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.

- Alonso Calvo, Raquel 2009, "Lexicalización y colocaciones: una introducción a su estudio diacrónico." *Onomázein* 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.
- 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.
- 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 2011, "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 2011, "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).
- Langacker, Ronald W. 1987, *Foundations of cognitive grammar. I: Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- 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 to app. "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*.
- 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.
- Motsch, Wolfgang 1981, "Der kreative Aspekt in der Wortbildung". Lipka & Günther (eds.) 1981: 94-118.
- 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.; 1st ed. 1880).
- 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.