

## Compound worlds and metaphor landscapes: Affixoids, allostructions, and higher-order generalizations

- final manuscript, published in *Word Structure* 12.3 (2019): 297–33, DOI:  
10.3366/word.2019.0151 -

(Page numbers from the published paper have been added in parantheses below.)

### [p. 297]

This paper investigates the use of German *-landschaft* ‘landscape’ and *-welt* ‘world’ as compound constituents. Both occur frequently in metaphorical uses such as *Korpuslandschaft* ‘corpus landscape’ or *Arbeitswelt* ‘labor world’. The high productivity of both compound types raises the question of whether [N-*landschaft*] and [N-*welt*] form constructions in their own right, both with a collectivizing meaning, and if so, how they relate to their respective higher-order schemas. More specifically, the question arises how the metaphorical-collectivizing uses relate to non-metaphorical ones. As both metaphorical and non-metaphorical reading variants coexist, these patterns show interesting similarities to those semantically bleached compound constituents that have been discussed under the label of “affixoids” in the morphological literature and in recent constructionist approaches. Drawing on synchronic data from the DWDS reference corpus of the 20<sup>th</sup> century and from the webcorpus DECOW16B as well as diachronic data from the German Text Archive, I argue that they form part of a family of collectivizing constructions. On a more theoretical note, I discuss how the higher-order generalizations connecting this family of constructions can be conceptualized.

### 1. Introduction

This paper presents a corpus analysis of German compounds with the head constituents *-welt* and *-landschaft*, which are often used in a collectivizing sense in coinages such as *Medienlandschaft* ‘media landscape’ and *Bankenwelt* ‘world of banks’. In particular, the present study addresses the question of how the metaphorical uses of the [p. 298] two compound patterns relate to the non-metaphorical ones as well as to potential higher-level constructions. While periphrastic variants such as *Welt der Banken* ‘world of banks’ will be discussed as well, the main focus of the empirical study presented in this paper will be on the compound constructions.

Compounding is arguably the domain of word-formation where the description of morphology as “the syntax of morphemes” (Booij 2012: 8) seems most appropriate. Especially in German, which is well-known for making extensive use of this word-formation pattern (Schlücker 2012), it offers a highly flexible way of forming complex expressions. From a descriptive point of view, an account of compounding seems straightforward: According to Bauer (2003: 40), compounding is simply “the formation of a new lexeme by adjoining two or more lexemes.” Moreover, quite unlike most derivational word-formation patterns, there are hardly any restrictions on compounding, at least in the case of N+N compounds.

Simple as it may seem, there are at least two challenges that the description of compounding faces from a Construction Grammar perspective. Both are interconnected, and they relate to the question of how schemas at different levels of abstraction interact with each other (also see Audring 2019, in this issue, on this topic). The first relates to the question of whether compounding patterns can be treated as constructions, viz. form-meaning pairings. If we assume that all constructions have meaning, the question arises whether we can legitimately posit a meaning side for a simple concatenative pattern. If we do, the meaning has to be sufficiently abstract: Booij (2010: 17) therefore simply describes the semantics of the nominal compound construction  $[[a]_{xk}[b]_{Ni}]_{Nj}$  as ‘SEM<sub>i</sub> with relation R to SEM<sub>j</sub>’. This leads to the second question: Given the semantic heterogeneity of compounds, it seems

reasonable to posit a number of subconstructions. But how many subconstructions do we need, and how can we conceptualize the division of labor between higher-level and lower-level schemas?

Approaching these and related questions requires an in-depth investigation of lower-level compounding patterns. The present paper contributes to this line of research by investigating two particular German compound patterns. These compound patterns are interesting for a number of reasons. Firstly, they have seen a steep increase in frequency since the 18<sup>th</sup> century, as will be shown in § 2 below. Secondly, they tend to have quite abstract semantics. Thus, they show interesting similarities to those elements that have been described as “semi-affixes” or “affixoids” especially in the literature on German word-formation (Tellenbach 1985, Schmidt 1987, Motsch 1996, Stevens 2005, Ascoop 2005, Elsen 2009, Leuschner 2010, among others).

To investigate these compound patterns in more detail, I combine an explorative approach based on synchronic data with a more deductive approach investigating the diachronic development of one of the two patterns, namely [*X-landschaft*]. While the synchronic analysis helps characterize the present-day usage preferences of the constructions, a diachronic approach can reveal patterns that may indicate how a subschema emerges and how it gradually emancipates itself from its parent schema(s).

The remainder of this paper is structured as follows. In § 2, I give a brief overview of the corpus data used for the present study. § 3 discusses the word-formation patterns in question in more detail. Also, the concept of affixoids is introduced and linked up to [p. 299] recent theoretical discussions on hierarchical relations in the construction. § 4 analyzes the corpus data against the backdrop of these theoretical conceptions, investigating the hypothesis that metaphorical *-welt* and *-landschaft* are subconstructions with specific collocational preferences that have diachronically diverged from other compounds where the same constituents are used in a literal sense. The main results are discussed in more detail in § 5. § 6 sums up the empirical findings as well as the theoretical approach: Empirically, it can be shown that the collectivizing use is a more recent development in the case of *-landschaft*. From a theoretical point of view, I will argue that *-welt* and *-landschaft* can be seen as part of a family of collectivizing constructions and that it makes sense to assume higher-order generalizations connecting the different members of the construction family.

## 2. Corpora used for the present study

The main analyses reported on in § 3 draw on two reference corpora – the German Text Archive (*Deutsches Textarchiv*, DTA) as well as the DWDS Core Corpus of the 20<sup>th</sup> century (DWDS20) – and on the webcorpus DECOW16B (Schäfer & Bildhauer 2012). Additional explorative analyses also use the aggregated DWDS reference and newspaper corpora. This resource comprises all data from the DTA, the DWDS core corpora of the 20<sup>th</sup> and 21<sup>st</sup> centuries as well as two newspaper corpora.

DTA and DWDS20 aim to capture the “standard” language use of the time they cover (see e.g. Geyken 2009, Geyken & Gloning 2015). Their design follows similar principles – for instance, they are balanced for the same coarse-grained text type categories. It therefore seems legitimate to aggregate the results of these corpora (which I will do in the diachronic analyses), even though cross-corpus analyses always require some caution (see Kopf in press).

As of October 2018, the DTA comprises around 214 million words and the DWDS20 around 121 million<sup>1</sup>. Data collection for the analysis of reading variants reported on in § 3.3, however, already took place in February 2015, using the now-legacy search interface still

---

<sup>1</sup> See <https://www.dwds.de/r> (last accessed 10/10/2018).

available via <http://eins.dwds.de/>. As the DTA has grown considerably since then<sup>2</sup> and as the new search interface yields slightly different results for the DWDS as well, query results obtained via the current interface differ from the results that were used for annotating the reading variants of *-landschaft*. For the frequency analyses and collocation analyses reported on in § 3.1, which do not require extensive manual annotation, the current versions of DTA and DWDS were used. To ensure reproducibility, all data and annotations have been made available on the Open Science Framework (OSF)<sup>3</sup>.

**[p. 300]**

The DTA and DWDS corpora were queried for instances of *-welt(en)* and *-landschaft(en)*<sup>4</sup> preceded by at least one other character in the case of *-landschaft* and at least three other characters in the case of *-welt* (to exclude cases like *Umwelt* ‘environment’ with a particle rather than a noun as first constituent). The search was conducted on the token level rather than the lemma annotation level to ensure that all instances were found, as some compounds may not be correctly lemmatized. The data were then cleaned manually.

The other main resource used for the present study, DECOW16B, is a webcorpus of German comprising c. 20 billion words (Schäfer & Bildhauer 2012; Schäfer 2015). Using DECOW16B can offer important additional insights on the patterns under investigation for two reasons: Firstly, it is much bigger than DWDS20. Thus, it is well-suited for investigating low-frequency phenomena. Secondly, it largely consists of present-day data, and it contains a considerable amount of non-standard language data. This is ideal for investigating very recent developments. Table 1 shows the number of hits for [X-*landschaft*] across the three corpora and [X-*welt*] in the two corpora of present-day German (as the diachronic analyses presented in § 4 are limited to *-landschaft*, the numbers in the DTA and in the 2015 version of DWDS20 are omitted here).

Table 1: Overview of the frequencies of [X-*landschaft*] and [X-*welt*] in the corpora under investigation.

	<i>-landschaft</i>		<i>-welt</i>	
	Tokens	Types	Tokens	Types
DTA (2015)	301	82	n.a.	n.a.
DWDS20 (2015)	1,039	305	n.a.	n.a.
DWDS20 (2018)	1,126	341	7,969	697
DECOW16B	25,917	1,901	1,492,946	812

### 3. Compounds in *-welt* and *-landschaft* and their theoretical status

This section first outlines the different reading variants of the patterns under discussion (3.1) before we turn to a more detailed discussion of their theoretical status. As mentioned above (§ 1), they show certain similarities to so-called affixoids, which in turn are often discussed in the context of grammaticalization scenarios as a “bridge” between free words and affixes.

<sup>2</sup> The 2014 version of the DTA is still available for download via <http://www.deutschestextarchiv.de/download> (last accessed 10/10/2018).

<sup>3</sup> <https://osf.io/d4umx/>. A commented Rmd notebook can be found there in the folder R\_scripts/welten\_landschaften\_wordStructure.

<sup>4</sup> *-en* is the plural suffix. For both lexemes, all case forms in the singular and in the plural are identical, i.e. *Welt* and *Landschaft* across all singular cases and *Welten* and *Landschaften* across all cases in the plural.

This is why § 3.2 discusses the concept of affixoids and its relation to grammaticalization in more detail.

### [p. 301]

#### 3.1 Reading variants of *-welt* and *-landschaft*

Compounds with the head constituents *-welt* and *-landschaft* can exhibit a fairly broad array of different meaning variants in present-day German, as examples (1) through (6) show.

- (1) karge **Hochgebirgslandschaft** im westl. Himalaya ‘barren **high mountains landscape** in Western Himalaya’ (Everding, Lexikon des Buddhismus, 1998, DWDS).
- (2) Die **Verlagslandschaft** des Dritten Reiches ‘The **publishing landscape** of the Third Reich’ (Wittmann, Geschichte des deutschen Buchhandels, 1992, DWDS)
- (3) Weiter oben in der dünnen **Gesichtslandschaft** stehen zwei leuchtende Augen ‘Further up in the lean **face landscape** there are two glowing eyes’ (Strittmatter, Der Laden, 1983, DWDS)
- (4) eine bunte **Bilderwelt** ‘a colorful **picture world**’ (Aktuelles Lexikon, 1995, DWDS)
- (5) ein der **Fachwelt** bisher unbekanntes Wesen ‘a being so far unknown to the **world of the discipline**’ (ZEIT, 1998, DWDS)
- (6) Die Anpassung der **Arbeitswelt** an die Lebenssituation der Arbeitnehmer [...] ‘Adapting the **labor world** to the life situation of the employees and their families [...]’ (ZEIT, 1986, DWDS)

Example (1) demonstrates the literal use of *-landschaft* as it refers to an actual landscape in the geographic sense, i.e., in the definition of the *Wörterbuch der deutschen Gegenwartssprache* (Dictionary of Contemporary German), a “part of the Earth’s surface that has gained a special character through ground conditions, vegetation, or settlement”<sup>5</sup>. (2) exemplifies a more abstract use. This reading variant is the most interesting one for the purposes of the present study. Here, *-landschaft* has undergone semantic bleaching. The use of *-landschaft* here can be interpreted as collectivizing – *Verlagslandschaft* can roughly be paraphrased as ‘the entirety of publishing houses’. By contrast, (3) exemplifies cases in which *-landschaft* in its literal, physical/geographical sense is metaphorically mapped to another concrete entity, namely a face. These innovative metaphorical mappings are, however, relatively rare in the data analyzed for the present study.

[p. 302] Turning to *-welt*, its primary literal meaning is harder to pin down. According to the DWDS definition, the word *Welt* denotes ‘the Earth, globe’ or ‘the entirety of all people’<sup>6</sup> (note that the latter meaning variant is already inherently collectivizing). These more concrete meaning variants, however, hardly ever occur – in compounds, the only instances approaching the first meaning variant ‘globe’ are cases like (7) where *-welt* refers to a planet (Mars).

- (7) Heute ist unser Nachbarplanet eine lebensfeindliche **Wüstenwelt**. ‘Today our neighbor planet is an inhospitable **desert world**.’ (ZEIT, 02.08.2012, DWDS)

However, we do find some instances where the reading of *-welt* can be considered fairly concrete, as in (4), which refers to a concrete visual impression. By contrast, (5) and (6) show

---

<sup>5</sup> “Teil der Erdoberfläche, der durch Bodengestalt, Bewachsung, Besiedlung sein besonderes Gepräge erhalten hat” (<https://www.dwds.de/wb/Landschaft>; my translation).

<sup>6</sup> “die Erde, Erdkugel (...) Gesamtheit der Menschen” (<https://www.dwds.de/wb/Welt>; my translation)

a collective meaning, referring to research(ers) in a specific discipline in the case of *Fachwelt* and to the organization of labor in the case of *Arbeitswelt*.

These examples already point to some important commonalities and differences between the two constructions under discussion. While *-landschaft* can be used in a very literal sense, this is only rarely the case for *-welt*.<sup>7</sup> However, the productive reading variants of *-welt* are arguably more nuanced, including person collectives like *Damenwelt* ‘world of ladies’ and imaginary or abstract/immaterial “worlds” like *Geisteswelt* ‘intellectual world’ or *Lebenswelt* ‘living environment’. What both *-welt* and *-landschaft* have in common is their use in a semantically bleached, collectivizing sense. This semantic bleaching of compound constituents has often been put forward as a defining criterion for affixoids. This is the topic to which we now turn.

### 3.2 Affixoids and grammaticalization: Some theoretical and terminological clarifications

As mentioned above (§ 1), both *-welt(en)* and *-landschaft(en)* have seen an increase in frequency over time. This is shown in Figure 1 based on the current data obtained from the DTA and DWDS corpora.<sup>8</sup>

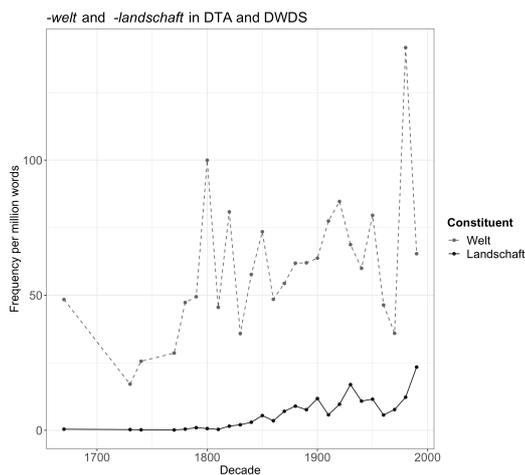


Figure 1: Frequencies of *-welt* and *-landschaft* as compound constituents in the German Text Archive (DTA) and the DWDS Core Corpus of the 20<sup>th</sup> century.

Such increases in frequency are also fairly typical of grammaticalization phenomena, as are processes of semantic bleaching. The development of so-called affixoids – and the subsequent development of some of those affixoids into full-fledged affixes – has frequently been discussed in the context of grammaticalization. From a Construction Morphology point of view (see e.g. Booij 2010), compound constituents like *-welt* and *-landschaft* might potentially be considered affixoids. According to Battefeld et al. (2018: 236), “affixoids can insightfully be modelled as the lexically specified parts of ‘constructional idioms’ at the word level”. Constructional schemas at the word level are in turn defined as word-formation patterns with one filled slot.

[p. 303] The concept of affixoids, or semi-affixes, has experienced some ups and downs in recent decades: An early use of the term “semi-affix” can be found in Marchand

<sup>7</sup> See e.g. the frequency list for *-welt* for the DTA and DWDS corpora, which is available in the OSF repository that contains the supplementary material for this paper.

<sup>8</sup> See the OSF repository for the exact queries and some caveats.

(1969: 326), who defines semi-suffixes as “such elements as stand midway between full words and suffixes.” As Schmidt (1987) shows, the concept of semi-affixes or affixoids has been widely adopted in textbooks and reference guides on German morphology. However, Schmidt himself makes a strong case that no intermediate category between full word and affix is needed (also see Olsen 1986). Stevens (2005), though, argues that the concept of affixoids can still be useful especially from a diachronic point of view, and he suggests a number of tests for affixoids, quoted here from Ascoop & Leuschner (2006: 243f.), who add illustrative examples to Stevens’ (2005: 74) original diagnostics:

- (i) Affixoids exist alongside a formally identical, and usually free, “parent” morph. Examples are the prefixoid *Riese(n)-* as in *Riesenlärm* ‘huge noise’ vis-à-vis the noun *Riese* ‘giant’, and the suffixoid *-arm* as in *schadstoffarm* ‘low in pollutants’ vis-à-vis the adjective *arm* ‘poor’.  
[p. 304]
- (ii) The affixoid is semantically more generalized and abstract and may have special metaphorical uses (as when *Lärm* is described metaphorically in terms of ‘giant’).
- (iii) The other member, not the affixoid, determines the basic meaning. For instance, *Riesenkraich* (with the augmentative prefixoid *Riesen-*) designates a giant noise, not a giant’s noise, and *Schulwesen* (with the collectivizing suffixoid *-wesen*) designates everything to do with schools and schooling rather than, say, a creature (*Wesen*) at school.
- (iv) Affixoids (in particular suffixoids) are in competition or complementary distribution with affixes. For example, *-werk* as in *Buschwerk* alternates with *ge-* as in *Gebüsch* (both meaning ‘bushes, undergrowth’), *-los* as in *gefahrlos* with *un-* as in *ungefährlich* (‘not dangerous’), *-voll* as in *sehnsuchtsvoll* with *-ig* as in *sehnsüchtig* (‘nostalgic’).
- (v) Affixoids are serial and usually highly productive, especially in spoken registers. Thus, the prefixoid *Riesen-* may combine not only with *Lärm* but also with *Ärger* (*Riesenärger* ‘huge trouble’), *Gewinn* (*Riesengewinn* ‘huge profit’) and many other roots, the suffixoid *-frei* may combine with *Atomwaffen* (*atomwaffenfrei* ‘free of nuclear weapons’), *Kosten* (*kostenfrei* ‘free of charge’) etc., and the prefixoid *Scheiß-* is routinely being combined in substandard German with any number of roots from *Abend* to *Zustand* (*Scheißabend* ‘unpleasant evening’, *Scheißzustand* ‘unbearable situation’, etc. etc.).

Criteria (i), (iii), and (v) obviously apply to [*X-welt*] and [*X-landschaft*] in the collective sense: Both exist alongside the free words *Welt* and *Landschaft*, and in both cases, the left constituent arguably carries the main meaning (e.g. *Verlagslandschaft* ‘publishing landscape’ refers to publishing houses, *Bankenwelt* ‘world of banks’ to banks, etc.). Also, both are productive, as they are used to coin nonce words like *Soaplandschaft* ‘soap opera landscape’. As for criterion (iv), one could argue that they fulfil a similar function as collectivizing affixes, including the prefix *Ge-* (*Gebüsch* ‘bushes’), the circumfix *Ge-X-e* (*Gerede* ‘talk, gossip’), or the suffix *-schaft* (*Kundschaft* ‘customers’), which, incidentally, also occurs in *Landschaft*.<sup>9</sup> However, criterion (ii) does not apply as straightforwardly: The abstract meaning of the constituents *Welt* and *Landschaft* does not only occur in the compound constructions but also in periphrastic variants like *Welt der Politik* ‘world of politics’ or *Landschaft der Parteien* ‘landscape of political parties’. We will return to this issue in §§ 4 and 5 below.

Despite Stevens’ (2005) attempt to “rescue” the concept of affixoids, the question of whether it is actually necessary remains a matter of debate. According to Bauer et al. (2013: 440), it seems questionable that positing an intermediate category can lead to a better understanding of the processes involved in the development of an affix out of a free [p. 305] word, and Booij (2005: 114) argues that the notion of affixoid should “not be seen as a theoretical notion, only as a provisional classificatory term”.

<sup>9</sup> Thanks to an anonymous reviewer for pointing out that *Landschaft* actually contains a collectivizing suffix. However, it can be assumed that the internal structure of the word is not transparent to present-day language users any more: As Fleischer & Barz (2012: 221) point out, the suffix is not productive any more when it comes to deriving collective nouns from anything else but lexemes referring to persons.

From the perspective of Construction Grammar, which assumes a lexicon-syntax continuum instead of clear-cut boundaries between different levels of linguistic organization (see e.g. Goldberg 1995, 2006; Croft 2001), such cases in which the boundaries between compounding and derivation are blurred are of particular interest. This is why the concept of affixoids, despite Booij’s above-cited caveat, enjoys quite some popularity in constructionist approaches to morphology (see e.g. Booij 2010: 57–66; Battfeld 2014; Hüning & Booij 2014; Norde & Van Goethem 2015; Battfeld et al. 2018). As Battfeld (2014: 20f.) points out, constructionist approaches use the term “affixoid” in a descriptive way, i.e. without necessarily assigning categorial status to them, in order to describe elements in-between free lexemes and affixes. As an example, he mentions German *Haupt-*, which has a free counterpart in the (now rare) lexeme *Haupt* ‘head’ and can be paraphrased as ‘very important, significant, relevant’ (*Hauptbedeutung* ‘main meaning’) or ‘person on top of a hierarchy’ (*Hauptmann* ‘captain’). In line with the idea of the hierarchical lexicon that is central to Booij’s construction morphology (see e.g. Booij 2010: 25–31), he assumes different degrees of schematicity for compounds with *Haupt-*, with the constructions (8) a. and b. sharing a common, more abstract parent construction  $[[Haupt]_N [y]_{N_i} N_j] \leftrightarrow$  ‘ $N_i$  of high importance’ (see Booij 2010: 62 for an equivalent analysis of Dutch *hoofd-*).

- (8)  $[[Haupt]_N [y]_{N_i} N_j] \leftrightarrow$  ‘ $N_i$  of high importance’  
 a.  $[[Haupt]_N [y]_{N_i} N_j] \leftrightarrow$  ‘very important, significant, relevant  $N_i$ ’  
 b.  $[[Haupt]_N [y]_{N_i} N_j] \leftrightarrow$  ‘main  $N_i$ ’

Importantly, the meanings in (8) are only available when *Haupt-* occurs in a compound, which is why Booij (2010: 62) speaks of a “bound meaning”. The compound-conditioned interpretation of *Haupt-* can be assumed to form part of the linguistic knowledge of language users. Construction Morphology accommodates this assumption by positing subschemas for the different bound meaning variants (see Booij 2010: 62).

This line of argument can easily be applied to the word-formation patterns under investigation in this paper. However, the necessity to posit constructional subschemas has been called into question by Meibauer (2013: 23). Discussing examples of pejorative compound constituents such as Dutch *kanker-* ‘cancer’ in *kanker-school* ‘bloody school’, he argues that “these putative prefixoids are ordinary lexemes used metaphorically”. A similar position is adopted by Fleischer & Barz (2012: 142–145), who treat e.g. augmentative patterns like *Riesen-N* ‘giant N’ as ordinary compounds in which one constituent is used metaphorically.

While it is true that “[m]etaphors are quite routinely parts of compounds” (Meibauer 2013: 23), this line of reasoning neglects an important aspect that speaks in favor of an account positing subconstructions for patterns like Dutch [*kanker-N*], German [*Haupt-N*], or [*N-welt*] and [*N-landschaft*]: While conceptual metaphor can indeed be assumed to play a key role in the emergence of compounds of this type, it is striking to observe that in nearly all of the examples discussed here, different compounds sharing [p. 306] the same first or head constituent exhibit a very similar, abstract semantics in which the original metaphor is bleached. This indicates that a new, productive subpattern has emerged. Importantly, the patterns exhibiting these bleached meaning variants can easily be transferred to new coinages. If the lexemes were used metaphorically on a case-by-case basis, we would expect that different compounds highlight different aspects of the source. For example, the KEY metaphor is used differently in the two German compounds *Schlüsselerlebnis* ‘key experience’ and *Schlüsselroman* ‘roman à clef’. They select different elements of the source concept (Lakoff & Johnson 2003: 84; “selective projection” in the terminology of Fauconnier & Turner’s [2002] conceptual integration framework; also see Kövecses 2015: 24–26). In the case of *Schlüsselerlebnis*, the source concept is mapped to a fairly abstract target domain that can

roughly be described as ‘importance’ – a key experience is an experience that bears a special importance for somebody’s life. In *Schlüsselroman*, by contrast, the metaphor highlights the “encrypted” nature of the novel. Here, the reader needs additional knowledge as a “key” to relate the novel to the real-world events it is based on. Note that the first, more abstract usage can easily be transferred to new compounds (other coinages include e.g. *Schlüsselwort* ‘key word’, *Schlüsselfigur* ‘key figure’), whereas it would seem odd to transfer the meaning of *Schlüssel-* in *Schlüsselroman* to new compounds such as <sup>?</sup>*Schlüsselfilm* ‘key film’.<sup>10</sup> This is why *Schlüssel-* in the first sense can be seen as serial, i.e. instantiated in numerous different coinages, and productive, according to Stevens’ (2005) criteria cited above. A very similar case can be made for *-welt* and *-landschaft*. While some rare examples like *Gesichtslandschaft* in (3) above are innovative ad-hoc metaphors, most productively coined compounds in *-landschaft* seem to exhibit the same abstract semantics as e.g. *Verlagslandschaft* in (2). In this case, then, one particular metaphorical mapping has become conventionalized.

Importantly, this usage variant has important commonalities with many uses of *-landschaft* in the literal sense. In many cases, the first constituent of literal *-landschaft* compounds refers to specific characteristics of an area (*Alpenlandschaft* ‘Alps landscape’, *Seenlandschaft* ‘landscape of lakes’). The collectivizing reading is already inherent in compounds like *Hügellandschaft* ‘hill landscape’ or *Seenlandschaft* ‘landscape of lakes’, which invoke a “summary view” of multiple hills or lakes. These compounds already afford the shift of semantic weight from the second to the first constituent that is characteristic of the use of *-landschaft* in the abstract-collectivizing sense and that is also an important feature of suffixoids according to Stevens (2005).

Hierarchical models of the type proposed by Booij (2010: 62) for Dutch *hoofd-* and by Battefeld (2014: 21) for its German counterpart *Haupt-* lend themselves well to modeling the different meanings of *-welt* and *-landschaft*, respectively. Figure 2 shows a simplified hierarchy for *-landschaft*, which could, of course, be extended with a more fine-grained array of subconstructions. Note that Figure 2 assumes that the abstract [p. 307] reading variant of *-landschaft* is a subconstruction in its own right; in §§ 4 and 5, I will discuss to what extent this assumption is backed up by the corpus data.

Just as in the case of *Haupt-*, formations in *-welt* and *-landschaft* differ in their degree of schematicity. Following Trousdale (2014: 565), schematicity pertains to the semantic generality of a construction. Drawing on the example of English quantifiers like *a lot*, Trousdale (2014: 565f.) shows how the schematicity of micro-constructions tends to change in the development of new grammatical constructions. While, for instance, Old English *hlot* refers to a concrete ‘portion’, as in *hlot landes* ‘a lot of land’, the new micro-construction that develops out of the older construction encodes abstract, more grammatical meaning (see e.g. Traugott & Trousdale 2013: 23–29). Similarly, the meanings of *-welt* and *-landschaft* can be assumed to become more abstract and schematic over time.

---

<sup>10</sup> This is of course not entirely impossible, but it would be an example of creative/innovative language use. In Barðdal’s (2008) terms, it would therefore be an example of analogical extension rather than productive use (see § 4.4 below).

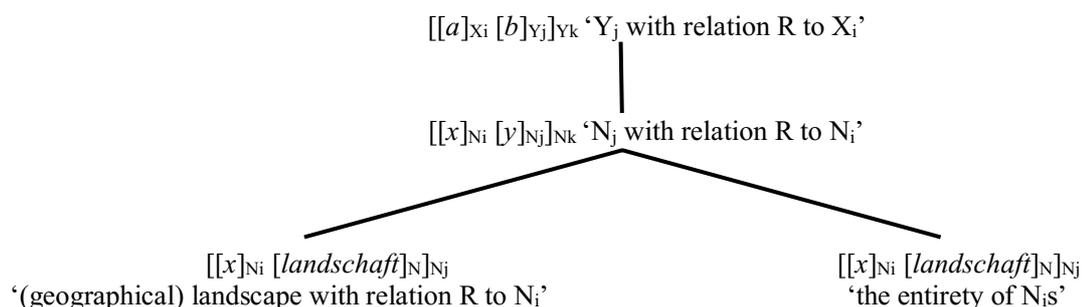


Figure 1: Hierarchy of compound constructions, based on Booij (2010: 62).

Booij (2010: 58), discussing the development of Dutch prefixoids such as *poep-* in e.g. *poep-lekker* ‘very pleasant’ (< *poep* ‘shit’), considers the changes in meaning of such “bound lexemes” a case of grammaticalization. In particular, he mentions host-class expansion as one “symptom” of grammaticalization (see e.g. Himmelmann 2004), as the affixoids under discussion combine with new compound constituents. Van Goethem & Hiligsmann (2014: 51f.) argue that the development of Dutch *reuze-* ‘giant’ and its German cognate *Riesen-* shows many classic “symptoms” of grammaticalization, e.g. divergence (from its root), persistence (of the original meaning in the free form), and decategorialization (some properties of the original form are lost, e.g. in the case of Dutch *reuze-*, the prefixoid cannot be pluralized: \**reuzen-leuk* ‘giants fun’).

Traugott & Trousdale (2013) make a similar case, discussing the emergence of English derivational suffixes like *-dom* from original free words via an intermediate stage of affixoids. However, they see this development as a case of lexicalization, or more precisely lexical constructionalization, arguing that lexical items, like grammatical ones, can be complex or atomic to various degrees (Traugott & Trousdale 2013: 163).

According to Smirnova (forthcoming), the development of affixoids does not necessarily go in tandem with grammaticalization: Investigating German noun-participle compounds, she argues that some highly frequent participial heads like *-bezogen* ‘related to’ and *-orientiert* ‘oriented’ can be seen as affixoids as they have undergone semantic bleaching. However, she argues that while they show some potential for [p. 308] grammaticalization, it would be too early to actually speak of a grammaticalization process here as long as no other changes in these head constituents can be observed (such as e.g. phonological reduction that leads to a differentiation between the emerging affix and the original free lexeme). The same is true for many other elements that qualify as affixoids according to the definitions cited above: They do share some features with affixes, but there is – at least at present – no indication that they are on their way to actually becoming affixes.

Summing up, then, there is a broad consensus in usage-based approaches to language that there is a continuum between highly frequent compound constituents and (emergent) affixes, while it is a matter of debate whether or not a term like “affixoid” is needed for this interim category. Also, there is some discussion about whether the development of affixes out of compound constituents can be accounted for as grammaticalization and/or (lexical) constructionalization. A major argument in favor of the notion of affixoid is that affixoid-formation might be seen as a word-formation pattern in its own right, as similar processes of semantic bleaching can be observed for many different compound constituents (Elsen 2009). On the other hand, the existence of periphrastic equivalents (like [*Landschaft des/der N*], to be discussed in more detail in § 4.2) indicates that this recurrent pattern of semantic bleaching is not limited to word-formation but also occurs in syntactic constructions. It might therefore

be preferable to find an overarching term for lexemes that have developed semantically bleached meaning variants that are tied to specific constructions, regardless of whether they are morphological or syntactic. The question of how the relationship between such similar patterns can be modeled will be taken up again in § 5.

In sum, however, it has become clear that much of the debate is terminological rather than conceptual. Irrespective of the terminology, the main question that will be discussed in the remainder of this paper is whether we can assume a more or less independent subconstruction for the semantically bleached compounds. The empirical study in § 4 first offers a more detailed descriptive account of compounds in *-welt* and *-landschaft*. This is complemented by a diachronic analysis of *-landschaft*, which emerges as the more “affix(oid)-like” of the two constituents. It is shown that the metaphorical readings become more prominent over time as compared to the literal ones, which can be interpreted as evidence in favor of the emergence of a new subschema.

#### 4. Quantitative analysis

This section presents four quantitative analyses of the attestations for compounds in *-welt* and *-landschaft* in the DTA and DWDS corpora as well as in the webcorpus DECOW16B. In § 4.1, I will present an explorative analysis using collocation analysis (Stefanowitsch & Gries 2003, Gries & Stefanowitsch 2004) based on the attestations of the pattern in synchronic corpora. The compound constructions will be compared with their syntactic equivalents [*Welt des/der X*] and [*Landschaft des/der X*], respectively, and it will be shown that *-landschaft* as a compound head behaves quite differently from *Landschaft* as a free word. *-welt*, by contrast, behaves very similarly to its free counterpart. Thus, the description as an affixoid seems altogether more [p. 309] fitting for *-landschaft* than for *-welt*. This is why §§ 4.2 through 4.4 focus on *-landschaft*, analyzing the diachronic distribution of literal and metaphorical reading variants in the data.

##### 4.1 Explorative collocation analysis I: Compound constituents vs. free words

The family of collocation analyses, which aims at quantifying the association between words and constructions, can be used for an explorative analysis of the present data in two ways: On the one hand, to detect association patterns between lexemes and the word-formation patterns they occur in (see e.g. Hartmann 2014; Smirnova forthcoming), on the other hand, to compare the compound patterns with their syntactic counterparts, viz. [*Welt des/der X*] and [*Landschaft des/der X*]. This can help answer the question to what degree the ‘bound meanings’ that are characteristic of affixoids are actually tied to the compound patterns.

All collocation analyses reported below were performed using R (R Core Team 2018) and Flach’s (2018) R package *collocations*. Following e.g. Wulff (2006), Ebensgaard-Jensen (2013), and Flach (2015), the log-likelihood ratio score  $G^2$  (Dunning 1993), which is particularly well-suited for large samples, is used as an association measure.

First, let us take a look at the nouns that occur with above-chance frequency as first constituents in the two compound patterns. This analysis was performed with both the DWDS20 data and the DECOW16B data. Distinctive collexeme analysis (Gries & Stefanowitsch 2004) was applied to compare the frequency of the first constituents in the two word-formation patterns to the frequency of their free counterparts. For the DWDS corpus, the frequency of the free counterparts was retrieved with the help of a frequency list of all

types tagged as nouns (NN or NE) in the DWDS corpus.<sup>11</sup> For the DECOW16B corpus, the pre-compiled frequency lists were used.<sup>12</sup>

The log-likelihood ratio  $G^2$  was calculated on the basis of  $2 \times 2$  tables as exemplified in Table 2: *Kultur* occurs 87 times as first constituent of compounds with the head constituent *-landschaft* in the DWDS20 corpus, while it occurs 14,129 times as a free word. In total, 1,126 [*X-landschaft*] compounds are attested in the data. Thus, the data contain 1,039 [*X-landschaft*] tokens with first constituents other than *Kultur-*. The fourth cell was filled with the sum count of tokens of all free words that occur as first constituents in the [*X-landschaft*] construction, subtracting the number of free attestations of the current collexeme, i.e. *Kultur* in the example in Table 2.

Table 2: Example for the application of distinctive collexeme analysis to compare the frequency of first constituents in the [*X-landschaft*] construction to their use as free words ([*X*]).

	X = <i>Kultur</i>	X = others
[ <i>X-landschaft</i> ]	87	1,039
[ <i>X</i> ]	14,129	1,812,709

The results can be found in the Appendix: Table 3 (DWDS20) and Table 4 (DECOW16B) list the top 20 distinctive collexemes for [*X-landschaft*], while the top 20 distinctive collexemes for [*X-welt*] are given in Table 5 (DWDS20) and Table 6 (DECOW16B). All four tables show the top 20 attracted collexemes. All collocation strengths reported in the four tables are significant at  $p < 0.00001$ .

The first constituents in both patterns can roughly be grouped in three categories: nature or geography (e.g. *Gebirge* ‘mountains’, *Fluss* ‘river’), concrete entities (e.g. *Sauna* ‘sauna’, *Tier* ‘animal’), and abstract concepts (e.g. *Bildung* ‘education’, *Geschäft* [p. 310] ‘business’). In the top 20 collexemes of [*X-landschaft*] in the DWDS 20 corpus (Table 3), we almost exclusively find items from the first domain. An exception is *Kultur* ‘culture’.<sup>13</sup> However, the term “cultural landscape” still refers to a concrete landscape in its default use. Thus, the [*X-landschaft*] compounds containing first constituents that occur in the pattern with above-chance frequency in the DWDS20 corpus all refer to concrete areas of land. It is only in the DECOW16B data in Table 4 that we see compounds referring to abstract concepts “crawling in”, as it were, namely *Medienlandschaft* ‘media landscape’ and *Bildungslandschaft* ‘education landscape’. This might have to do with the composition of the two corpora but it might also reflect a change in usage as DECOW16B represents present-day language use while DWDS20 captures the entire 20<sup>th</sup> century. In other words, it might indicate that the use of *-landschaft* in the collectivizing sense has become more widespread over the recent decades. This hypothesis is supported by the frequency development of some highly frequent compounds, some of which have occurred in the tables above, while others will surface again in the quantitative comparison between [*X-landschaft*] and [*Landschaft des/der X*] below. The frequency developments of the exemplary compounds shown in Figure 3 have been obtained on the basis of the DWDS collection of reference and newspaper corpora<sup>14</sup> using the Time Series tool of DWDS’ Dstar interface<sup>15</sup>. The frequency data show

<sup>11</sup> This frequency list was obtained via DWDS’ Dstar interface using the query *count( \$p=/N./g #sep ) #by[\$I]*.

<sup>12</sup> Available at <https://www.webcorpora.org//opendata/frequencies/german/decow16b/> (last checked 30/10/2018).

<sup>13</sup> A reviewer points out that given its high frequency, *Kulturlandschaft* could also be considered a lexicalized item. This is true, but assuming that lexicalization is a matter of degree, many compounds identified as attracted collexemes can be seen as lexicalized to varying extents. This is why I have decided not to discard any (potentially) lexicalized items.

<sup>14</sup> See <https://www.dwds.de/r> for details on the corpus composition (last accessed 01/11/2018).

<sup>15</sup> See <https://kaskade.dwds.de/dstar/> (last accessed 01/11/2018).

that compounds like *Bildungslandschaft* ‘education landscape’ and *Medienlandschaft* ‘media landscape’, which have emerged as attractor collexemes in the DECOW16B data but not in the DWDS20 data, only really start to gain ground in the second half of the 20<sup>th</sup> century. By way of comparison, the right panel of Figure 3 shows some of the literal uses that have emerged as collexemes on the basis of the DWDS20 data. Here, the frequencies remain approximately unchanged or decrease over the course of the 20<sup>th</sup> century.

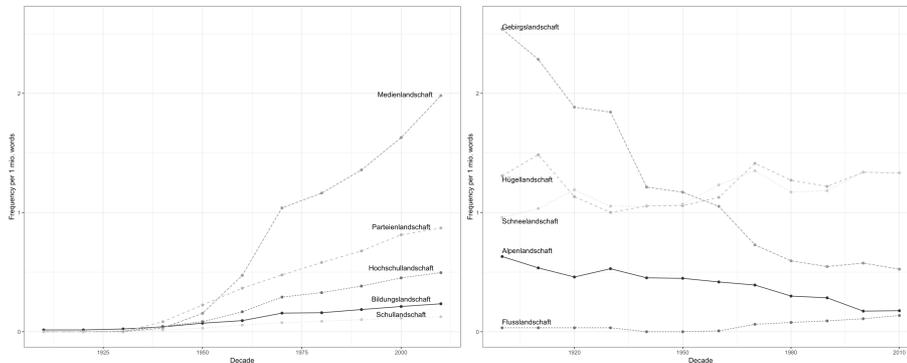


Figure 3: Examples for the frequency development of some collectivizing (left) and literal (right) uses of *-landschaft* in the DWDS collection of reference and newspaper corpora.

In the case of *-welt*, the differences between the results for DWDS20 in Table 5 and those for DECOW16B in Table 6 are probably more symptomatic of shifts in the public discourse than of changes in the use of the pattern: Topics such as the organization of [p. 312] work (*Arbeitswelt* ‘labor world’) or financial issues (*Finanzwelt* ‘financial world’) lead to the highly frequent use of the respective collectivizing compounds in *-welt*. In most cases across both corpora, the compounds with *-welt* can be interpreted in a collectivizing sense. This suggests that the abstract-metaphorical/collectivizing use is predominant in the case of *-welt* throughout the period under investigation, whereas it is a fairly new development in the case of *-landschaft*.

#### 4.2 Explorative collocation analysis II: Compounds vs. syntactic counterparts

According to Stevens’ (2005) tests for affixoids cited in § 2, affixoids exist alongside a free, formally identical parent morph, but their meaning is more generalized and abstract than that of the free word. In the case of *-welt* and *-landschaft*, their meaning in compounds in which they serve a collectivizing function is clearly more abstract compared to the default meaning of the free words. However, the free words can also be used in a collectivizing sense, as the examples in (9) and (10) show.

- (9) Die **Landschaft der Bildungsmedien** wird unübersichtlich ‘The **landscape of educational media** is becoming confusingly vast’  
(251aa7b1acf333ed5901691709869422639b<sup>16</sup>, DECOW16B)
- (10) Diese in der **Welt des Motorsports** noch sehr seltene Technologie ‘This technology, which is still very rare in the **world of motor sports**’  
(9efaaef3851e49ccff24cbe6c77d0bc63b00, DECOW16B)

<sup>16</sup> Attestations from the DECOW corpus are cited with their ID (even though this is, admittedly, not the most reader-friendly identifier).

Therefore, the question arises to what degree the ‘bound meaning’ that e.g. Booij (2010) identifies as an important characteristic of affixoids is actually tied to the compound pattern. This leads us to the second way in which distinctive collexeme analysis is applied in this study. As a rough proxy for comparing the behavior of *-welt* and *-landschaft* as head constituents to that of their free counterparts, a distinctive collexeme analysis was performed to compare it with the syntactic constructions [*Welt des/der X*] ‘world of X’ and [*Landschaft des/der X*] ‘landscape of X’. For this study, only the webcorpus DECOW16B was used because the number of attestations for the syntactic patterns in the DWDS corpus is fairly low (31 for *Welt(en) des/der X* and 153 for *Landschaft(en) des/der X*). In DECOW16B, the number of hits is of course much higher (236,969 for *Welt des/der X* and 21,556 for *Landschaft(en) des/der X*). Tables 7 and 8 show the top 20 items that occur significantly more often in the synthetic constructions (i.e., the compound patterns) than in the analytic ones (their syntactic counterparts). Conversely, Tables 9 and 10 show the top 20 items that occur significantly more often in the analytic than in the synthetic constructions. All collocation strengths reported in the four tables are significant at  $p < 0.00001$ .

[p. 313] Given the very high frequency of [*X-welt*] and [*Welt des/der X*], only items with at least 100 attestations were taken into account. Also note that the distinctive collexemes for [*Landschaft des/der X*] as compared to [*X-landschaft*] and vice versa are based on a subset of the [*Landschaft des/der X*] data where all instances in which the construction is immediately preceded by an adjective were omitted. The reason behind this is that the most frequent instances include many superlatives like *die schönste Landschaft der Welt* ‘the most beautiful landscape in the world’. Unlike e.g. *Landschaft der Alpen* ‘landscape of the Alps’, which is (near-)synonymous to *Alpenlandschaft*, *die schönste Landschaft der Welt* cannot be paraphrased as *\*die schönste Weltlandschaft*. This subset only contains 7,505 tokens.<sup>17</sup>

The comparison between the syntactic and the analytic construction reveals a number of interesting tendencies. First of all, note that no less than seven of the top 20 distinctive collexemes of [*X-landschaft*] compared to [*Landschaft des/der X*] – *Medien, Bildung, Schule, Forschung, Hochschule, Partei, Museum* – refer to abstract concepts or to institutions and trigger a collective reading of the head constituent *-landschaft* – not counting the top collexeme *Kulturlandschaft*, which is used referring to cultural institutions (rather than a concrete geographical landscape) in quite a few cases as well, as exemplified in (11).

- (11) Es wird in Zukunft die subventionierten Theater-, Tanz- und Opernhäuser nach wie vor brauchen, wenn Berlin seine reiche **Kulturlandschaft** nicht preisgeben will ‘The subsidized theaters, dancing and opera houses will still be needed in the future if Berlin doesn’t want to abandon its rich **cultural landscape**.’  
(5732e2b285642a9cbf4dc9c31c79eeb9cf87, DECOW16B)

By contrast, the distinctive collexemes for [*Landschaft des/der X*] largely come from the domain of nature and geography. The most notable exception is *Heimatlosigkeit* (roughly ‘lack of a homeland’), which is only attested in quotes from the Swedish Academy’s announcement of the Nobel Prize for Herta Müller. Other exceptions are *Jahr* ‘year’ and *Zukunft* ‘future’. *Jahr* almost exclusively occurs in attestations referring to the award *Landschaft des Jahres* ‘landscape of the year’, while *Zukunft* is used in contexts like (12). Thus, both *Landschaft des Jahres* and *Landschaft der Zukunft* also refer to geographical spaces. The same is true for *Landschaft(en) der Bibel* ‘landscape(s) of the Bible’, which refers to the places where the Biblical stories are set.

<sup>17</sup> The two datasets as well as R scripts for performing the collexeme analysis on the basis of each of them are available on the OSF repository mentioned above. – What still remains in the data, however, are implicitly superlative uses like *Landschaft des Jahres* ‘landscape of the year’.

- (12) Frage ist, wie die **Landschaft der Zukunft** aussehen soll ‘[The] question is how the **landscape of the future** should look like’ (ae35ce3e79cf223374bbb6a10d88031f93f4, DECOW16B)

[p. 314] This suggests that the more abstract meaning variant is strongly connected with the synthetic pattern, even though it does occur with the analytic one as well, as exemplified in (9) above. By contrast, both [*X-welt*] and [*Welt des/der X*] attract more abstract concepts such as *Arbeit, Beruf, Geschäft* on the one hand and *Wissenschaft* or *Technik* on the other (see Tables 8 and 10). Also, person collectives like *Damen, Männer*<sup>18</sup> or *Lebende, Erwachsene, Reiche, Stars* occur in both Table 8 and Table 10. Thus, the synthetic and analytic variants can potentially be treated as allostructions, i.e. “variant structural realizations of a construction that is left partially underspecified” (Cappelle 2006: 18). This means that the two constructions are linked via a joint supercategory (see also Perek 2015: 152).

The factors that drive the choice between the synthetic and analytic patterns should be investigated in more detail in future research; candidates are e.g. morphological complexity (*Naturlandschaft* vs. *Landschaft des Naturparks*) as well as phonological and prosodic factors. Semantically, however, both constructions seem largely exchangeable. This is supported by a comparison of the two frequency columns in each table: Especially in Table 10, the frequency differences between both constructions are often quite low, whereas in the case of the *Landschaft*-patterns, the top collexemes of one pattern tend to vastly outnumber their counterparts from the competing construction. It therefore seems fair to say that *-landschaft* altogether that the ‘bound meaning’ of *-landschaft* is much more strongly tied to the compound pattern than in the case of *-welt*, where the free word exhibits the more abstract meaning as well on a regular basis.

While both patterns are interesting and merit further investigation, we will now focus on *-landschaft* and the diachronic development of the bound meaning variant. The results of the collexeme analysis constitute one piece of evidence for the hypothesis that [*X-landschaft*] can be treated as a construction in its own right, i.e. that compounds in *-landschaft* are not just “ordinary” compounds in which the head constituent is used metaphorically on a case-by-case basis (see discussion in § 2 above). A diachronic perspective can lend further support to this idea: If we are dealing with ad-hoc metaphorical mappings, then we would not necessarily expect the frequency of metaphorical uses to increase. If, by contrast, we are dealing with a fairly new constructional schema that has been spreading over, say, the last century (as the comparison between the DWDS and the DECOW data suggests), then we would expect the proportion of abstract/metaphorical uses to increase. Also, the availability of the new constructional schema might be reflected in the morphological productivity of the compound pattern as a whole. These two aspects – the diachronic frequency development of different reading variants of [*X-landschaft*] and the productivity of the pattern – will be addressed in §§ 3.3 and 3.4, respectively.

#### [p. 315]

##### 4.3 Literal and metaphorical uses of *-landschaft*

All 1,340 compounds in *-landschaft* from the 2015 DTA and DWDS corpora (see § 2 above) were coded in context for the reading variant in which they are used: apart from literal (geographical) uses, two different collectivizing usage variants were distinguished. On the one hand, we find collectivizing uses that refer to a concrete area, but not to a landscape in the usual sense (see DWDS definition cited in § 3.1). This category also includes innovative metaphorical mappings such as *Gesichtslandschaft* in (3). In addition, a fourth category

---

<sup>18</sup> As the tables above use lemmas rather than word forms, these occur as *Dame* and *Mann* in **Error! Reference source not found.**

captures the fairly frequent uses of *-landschaft* referring to a painting of a landscape (especially *Winterlandschaft* ‘winter landscape’).

In some cases, the same compound varies between the different reading variants, as the example of *Kulturlandschaft* ‘cultural landscape’ shows. On the one hand, *Kulturlandschaft* can refer to an actual physical space, on the other hand, it can be used in a collective sense to refer to the entirety of cultural institutions or to cultural life in general, as in (11) above. This is why all instances were annotated in their respective contexts.

As Figure 4 shows, abstract-collectivizing uses have seen a steady increase in the second half of the 20<sup>th</sup> century. By the end of the time span covered by the DWDS20 corpus, they account for approximately a quarter of all attestations.

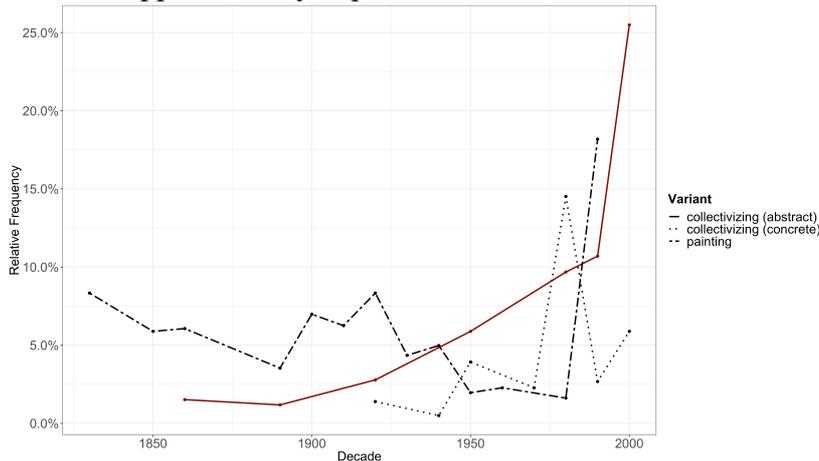


Figure 4: Proportion of non-literal uses of *-landschaft* in the DTA and DWDS data.

Also note that the increase in collectivizing/metaphorical uses over the 20<sup>th</sup> century can be detected across different text types.<sup>19</sup> Thus, the increase in metaphorical uses is probably not just a side effect of *-landschaft* being used more frequently in an abstract sense in one specific variety (e.g. newspaper language).

The surge in the proportion of metaphorical uses is partly due to some highly token-frequent instances such as *Medienlandschaft* and *Parteienlandschaft* [p. 316] (see Figure 3 above), which in turn may act as “attractors” (see e.g. Traugott 2008, Claidière et al. 2014) for new coinages. This entails a diversification in the first constituents that *-landschaft* combines with, i.e. the host-class expansion that Booij (2010) sees as symptomatic of affixoid formation (see § 2 above). This will be discussed in more detail in the next section.

In sum, the results of the diachronic analysis show that collectivizing uses of *-landschaft* have become more prominent over the course of the 20<sup>th</sup> century, which fits in well with the hypothesis that abstract [N-*landschaft*] constitutes a (sub)construction in its own right.

#### 4.4 Productivity development of [X-*landschaft*]

The host-class expansion that goes along with the increase in metaphorical uses is reflected in the type-token ratio and the productivity of the pattern. It is often assumed that an increasing number of types can be indicative of increased schematicity (see e.g. Smirnova forthcoming). As can be seen in Figure 5, both the type-token ratio and the so-called potential productivity of [X-*landschaft*] show a fairly consistent increase over the time span covered by the two corpora. Following Baayen (2009), potential productivity is calculated by dividing the hapax legomena belonging to a specific construction (in this case [X-*landschaft*]) by the total

<sup>19</sup> See the analysis in the OSF repository for more details.

number of tokens that instantiate the construction.<sup>20</sup> Given that the measure is very sensitive for corpus size (see e.g. Gaeta & Ricca 2006, Hartmann 2018, Berg forthcoming), the results have to be taken with some caution, but they can still give interesting hints to the development of the construction. Note that the graphs in Figure 5 only display the decades from 1880 onwards – the decades from 1780 to 1880 were omitted because the number of attestations is rather consistently below 20, in many cases even below 10.

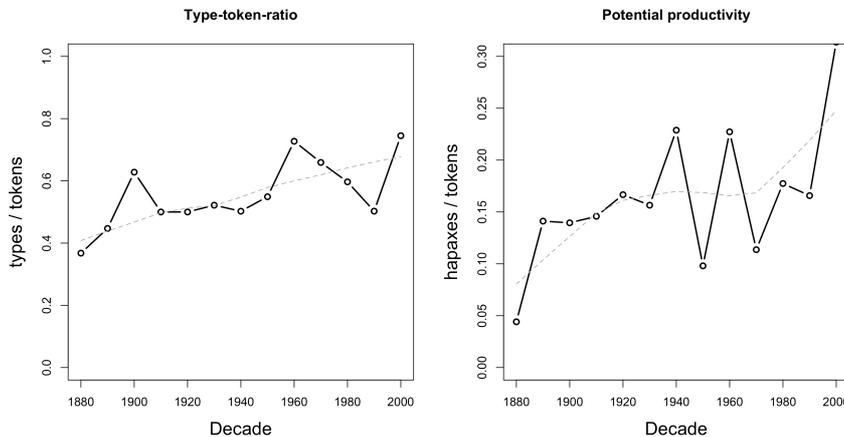


Figure 5: Type-token-ratio and potential productivity of the [X-landschaft] construction in the DTA/DWDS20 data. The decades before 1860 were dropped because the total frequencies of attestations were too low.

Following Guardamagna’s (2018: 192f.) account of the host-class expansion of a syntactic construction in Latin ([*secundum* NP]), the development of [X-landschaft] can be explained with the help of Barðdal’s (2008) model of productivity. Barðdal assumes an inverse relationship between type frequency and coherence: A construction with high type frequency is fully productive, it forms an open schema whose instantiations are often heterogeneous, i.e. characterized by low coherence. A construction with low frequency, on the other hand, is extensible via *analogy* if its instantiations are coherent by virtue of sharing important features, e.g. by belonging to the same semantic class (semantic coherence). In the case of [X-landschaft], we can assume that one or more coinages in which *-landschaft* was used metaphorically experienced an increase in token frequency and acted as attractor(s) for new, analogical formations. As the type frequency of metaphorical [X-landschaft] increased, the productivity of the construction shifted towards the “open-schema” pole of Barðdal’s productivity gradient. In other words, we can assume that the pattern started out with innovative formations that were extended analogically to new cases. This then gave rise to a productive subschema in which *-landschaft* is systematically used in a collectivizing sense.

## [p. 317]

### 5. Discussion

The empirical findings discussed above support the idea that in the case of *-welt* and especially *-landschaft*, we are not dealing with ad-hoc innovative metaphorical uses of the constituents but rather with productive subschemas. Recapitulating the four empirical studies reported on above, we can summarize the main results as follows: (i) The collexeme analysis showed that in the case of *-landschaft*, more abstract usage variants emerge as attracted collexemes for the data from the webcorpus DECOW, but not for the 20<sup>th</sup>-century DWDS core corpus. This suggests that the abstract reading variant is a fairly recent development in

<sup>20</sup> For in-depth discussions about the application of productivity measures to compounds, see e.g. Roth (2014) and Hein & Engelberg (2017).

this case. For *-welt*, by contrast, no difference between the two databases can be detected. Instead, the collexemes for both DWDS and DECOW data reflect the broad semantic spectrum covered by these compounds. This is quite consistent with the behavior of *Landschaft* and *Welt* in other constructions: (ii) Comparing the compound constructions with their periphrastic counterparts [*Landschaft des/der N*] and [*Welt des/der N*] on the basis of DECOW data, it could be shown that items that are clearly tied to the abstract reading variant are strongly drawn to the compound pattern in the case of *Landschaft*, which is not the case for *Welt*. This indicates that *Welt* is used with abstract semantics more often regardless of the specific construction in which it occurs. (iii) Zooming in on the diachronic semantic development of [*N-landschaft*], we have seen a steep rise in the proportion of collectivizing-metaphoric uses, which is consistent with the assumption that the abstract use of *-landschaft* is a recent phenomenon. (iv) Finally, regarding the productivity development of [*X-landschaft*], we have seen an increase in type-token ratio and potential productivity, which may be indicative of host-class expansion.

Cases like these (and many others that have been discussed as “affixoids”) raise the question of how they can be modeled in a Construction Grammar framework. While the [p. 318] Construction Morphology approach of treating them as affixoid subschemas seems straightforward, this account entails some problems: The characteristic feature of an affixoid is that it exhibits what Booij (2010) calls a “bound meaning” when it occurs in compounds, but not when it occurs in free words (also see Stevens’ criterion ii in § 2 above). Both in the case of *-landschaft* and especially in the case of *-welt*, this is not true: Their abstract, collectivizing meaning variants also occur in specific syntactic constructions such as [*Landschaft des/der X*] and [*Welt des/der X*]. In addition, they also exhibit this meaning variant in combination with adjectives, e.g. *politische Landschaft* ‘political landscape’, a construction not discussed in the present paper.

Even though in the case of *-landschaft* the more abstract, collectivizing meaning seems to be strongly connected to the compound pattern, as the distinctive collexeme analysis in § 3 has shown, it is clear that in both cases, the semantically bleached meaning variant is by no means exclusive to the compound construction. The same is true for other items that have been discussed as affixoids, e.g. *-papst* ‘pope’ (see Schmidt 1987, Elsen 2009, among others), which can mean ‘authority’ in compounds like *Literaturpapst* ‘literature pope’ or *Modepapst* ‘fashion pope’ but also in syntactic constructions like *Papst der Geldpolitik* ‘pope of financial politics’ (ZEIT, 10.01.1997, DWDS).

Given these observations, two different accounts are possible: On the one hand, we can explain them in terms of a general semantic development of the items in question, i.e. *Landschaft*, *Welt*, and also *Papst* and others have developed a semantically bleached meaning variant that occurs both within and outside of compounds. On the other hand, however, there may be good reasons to stick to the idea that we are dealing with “bound” meanings, even though they are not exclusively bound to the compound patterns. Instead, they are bound to various partially schematic constructions that form constructional families, such as [*Welt des/der X*], [*X-Welt*], and [*ADJ Welt*], instantiated in e.g. *Welt der Politik*, *Polit(ik)welt*, and *politische Welt*. The two accounts do not necessarily exclude each other: If we take seriously the core ideas of Construction Grammar, then we might argue that all meanings are actually bound to the contexts in which they occur – and hence, to specific constructions. In this view, all meanings are bound meanings, and to the extent that context-free meanings exist, they are generalizations over context-dependent meanings. As Langacker (1999: 108) puts it, “the claim that linguistic meanings are clearly distinguishable from contextual understandings is vacuous, rendered true just as a matter of definition.” Consequently, it stands to reason that the “bound” meanings sooner or later have an impact on the meaning of the free words, as can also be seen in cases like the debonding of affixoid patterns discussed in Battefeld et al. (2018; see also Norde & Van Goethem 2015).

I have argued above (§ 3) that [*Welt des/der X*] and [*X-welt*] can potentially be described as allostructions. For [*Landschaft des/der X*] and [*X-landschaft*], this is less clear because the abstract meaning seems to be more specific to the compound pattern, i.e. the two patterns are much less exchangeable than in the case of the *Welt* patterns: Does it make sense to speak of allostructions here as well, and can we legitimately posit a joint higher-order schema? There are at least two reasons why it seems plausible to treat them as allostructions. Firstly, there are few if any allostruction pairs where the two constructions are fully exchangeable. Secondly, we can conceive of higher-order [p. 319] generalizations as emergent relationships between constructions (Langacker 2008: 249). They link up different constructions that share common properties, and of course the association between constructions can be stronger or weaker. We could therefore say that the allostructional relationship is stronger between [*Welt des/der X*] and [*X-welt*] than between [*Landschaft des/der X*] and [*X-landschaft*]. But in both cases, we can assume that generalizations over the patterns – in other words: higher-order schemas – are part of speakers’ linguistic knowledge.

Given the semantic similarity between *Welt* and *Landschaft* in those constructions – and others that have not been discussed here, e.g. *-szene* ‘scene’ or *-bereich* ‘area’ – we can assume yet another generalization and conceive of these different patterns as a constructional family of collectivizing constructions. They are not completely interchangeable as they combine with different nouns and arguably convey different nuances of meaning. Yet they can be considered near-synonymous, and they partly tend to combine with the same nouns – for instance, *Bank* ‘bank’, *Jugend* ‘youth’, *Kultur* ‘culture’, *Kunst* ‘art’, *Musik* ‘music’, *Schule* ‘school’, and *Verlag* ‘publishing house’ combine with all of the four head constituents *-welt*, *-landschaft*, *-szene*, and *-bereich* in the DWDS20 corpus.<sup>21</sup> Also, many compound constituents that have been argued to be affixoids are used in a collectivizing sense, e.g. *-gut*, *-zeug*, *-werk*, and *-wesen*. This family of collectivizing constructions arguably fills a gap in the constructional network of German as there are no productive collectivizing affixes that could be used with the first constituents they combine with.

## 6. Conclusion

As Goldberg (2006: 18) famously stated, “the network of constructions captures our grammatical knowledge of language *in toto*, i.e. it’s constructions all the way down.” But how far down – and how far up – can we go when we posit constructions? The “right” level of abstraction for positing constructions is a recurring open question in Construction Grammar, and it relates to the equally challenging question of when a construction can actually be considered a construction (see Hilpert 2013, 2018). The domain of compounding is particularly well-suited to address these challenges. There are various competing accounts for the empirical observations discussed for the compound patterns [*X-welt*] and [*X-landschaft*] in the present paper, and these in turn partly tie in with the debate about the concept of affixoids reviewed in § 3.2. In this last section, I will try to tease these different perspectives apart, and I will argue – as already hinted at in § 3.2 – that much of the potential disagreement might actually be more terminological than conceptual.

In § 1, I have argued that it seems reasonable to posit subconstructions for serial compound patterns, and I have asked how many subconstructions we can legitimately posit. An approach that keeps the number of constructions at a minimum would not treat [p. 320] [*X-welt*] and [*X-landschaft*] as constructions but rather see compounds with *-welt* and *-landschaft* as head constituents as instantiations of the more general N+N compounding construction, or even a more abstract compounding construction without any part-of-speech specification. A “maximalist” approach, on the other hand, might go so far as to posit a

<sup>21</sup> However, not all uses are necessarily collective: As pointed out multiple times, *Kulturlandschaft* usually refers to a landscape in the literal sense; also, *Musikszene* can refer to a music scene in a film.

construction for each individual instantiation. It seems quite uncontroversial that an approach that offers additional insights beyond the trivial statement that we are dealing with compounds (or that we are dealing with words) should strike a good balance between both extremes. Given the evidence from an explorative analysis of synchronic data, I have suggested that we can assume subconstructions for the collectivizing use of [*X-welt*] and [*X-landschaft*]. While there is not necessarily a need for positing e.g. a subconstruction for [*X-landschaft*] in its literal reading as it perfectly compositional, the metaphorical reading of *Landschaft* seems to be bound to a limited number of constructions, including the compound pattern. A diachronic analysis has shown that the metaphorical use of *-landschaft* has gained prominence over time, which can also be seen as evidence in favor of the gradual emergence of a new subschema.

I have suggested in § 5 that higher-order generalizations can be assumed between a number of different collectivizing constructions such as the *Welt*- and *Landschaft*-constructions on the one hand and e.g. compounds in *-szene* or *-bereich* on the other. However, this hypothesis is currently based solely on introspective evidence and an impressionistic look at shared first constituents, as mentioned in § 5. A more detailed study would be necessary to investigate this hypothesis in more detail and to carve out the commonalities and differences between these constructions. From a more theoretical point of view, however, the question remains what kind of entities higher-order generalizations are: Are they constructions in their own right, or are they merely links between constructions or between different meanings of one construction, like Goldberg's (1995) polysemy links? In the end, this might just be a terminological issue and depend on one's definition of constructions – after all, constructions are simply generalizations as well, but generalizations that meet a specific definition. But as is well-known, definitions of constructions in Construction Grammar differ considerably. Goldberg & Jackendoff (2004: 532f.), for instance, offer two different definitions in their text and in a footnote (“pieces of *syntax* connected to meaning in a conventionalized and *partially idiosyncratic* way” vs. “any stored pairings of form and function”, emphases added). Such challenges do not pose themselves in frameworks like Schmid's (2017) entrenchment-and-conventionalization model, which “rejects the distinction between constructions serving as nodes in the network and relations between nodes and instead assumes that linguistic knowledge is available in one format only, namely, associations.” (Schmid 2017: 25)

Given these considerations, one could ask whether constructions are the new affixoids: While the conceptualization of linguistic knowledge as a network is fairly consensual in usage-based linguistics (see e.g. Diessel 2015), it is a matter of debate when exactly a form-meaning connection can be seen as a construction. In a way, this situation is similar to the one in the affixoid debate: There is a broad consensus that free morphemes and affixes do not belong to two clearly separate categories but that there is a continuum [p. 321] between both poles. The question, however, is whether we should assume an in-between category of affixoids and if so, which linguistic units can be legitimately termed affixoids. Most researchers in Construction Morphology take the compromise position of treating the notion of affixoid as a heuristic and descriptive term rather than a category in its own right. Analogously, one could conceive of constructions as a largely heuristic concept. And arguably this is exactly what constructions are: Although one goal of Construction Grammar is to be cognitively realistic, nobody would claim that constructions have one-to-one neural correlates in any meaningful sense. The question of whether higher-order generalizations constitute constructions in their own right is therefore largely terminological. A more interesting question that should be addressed in future research is to what degree speakers actually make these generalizations. For the constructional family discussed here, this could, for instance, be investigated from a psycholinguistic point of view using a priming paradigm.

Follow-up corpus-based studies, on the other hand, could add to the results of the present paper by taking a cross-linguistic perspective. Importantly, the use of words meaning ‘world’ and ‘landscape’ in a collectivizing sense can be found in various languages. The use of [N-*landschap*] in Dutch in compounds like *kunstlandschap*, for instance, shows many similarities to its German counterpart. Also, the TIME Magazine Corpus as well as the Corpus of Historical American English (COHA) show a steep increase in the use of the perhaps most prominent example in English, namely *political landscape*, from the mid-1950s onwards. It seems not too far-fetched to assume that different languages have influenced each other in the development of these patterns. The collectivizing constructions discussed here might therefore be a good candidate for a cross-linguistic construction family (also see Höder 2019, in this issue, on multilingual constructions).

In sum, the case of *-welt* and *-landschaft* gives interesting clues to the complex interaction between higher-order and lower-level schemas in morphology and beyond. The studies presented in this paper have only covered a small part of the constructional family the compound patterns are part of, but they have already revealed interesting patterns regarding the division of labor between different collectivizing constructions. What makes them particularly interesting is that they are in-between categories in many different ways: Not only are they in-between compound constituents and affixes, but they also share many features with affixoids without fulfilling all of the widely-accepted criteria. Their development shows many commonalities with grammaticalization processes without necessarily qualifying as grammaticalization (see § 2 above). And last but not least, their relationship to each other and to similar collectivizing constructions raises many interesting questions that call for a closer investigation of the landscape of collectivizing constructions.

### Acknowledgments

I am grateful to Martin Hilpert and two anonymous reviewers for helpful comments and suggestions on a previous draft of this paper. Also, the refined annotation criteria used in this final version of the paper have benefited from discussions with Tom Bossuyt. Remaining shortcomings are of course my own.

### References

- Ascoop, Kristin. 2005. Affixoidhungrig? Skitbra! Status und Gebrauch von Affixoiden im Deutschen und Schwedischen. *Germanistische Mitteilungen* 62. 17–28.
- Baayen, R. Harald. 2009. Corpus Linguistics in Morphology: Morphological Productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics*, 899–919. (HSK 29.2). Berlin, New York: De Gruyter.
- Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. (Constructional Approaches to Language 8). Amsterdam, Philadelphia: John Benjamins.
- Battefeld, Malte. 2014. Konstruktionsmorphologie sprachübergreifend: Perspektiven eines Vergleichs von “Affixoiden” im Deutschen, Niederländischen und Schwedischen. *Germanistische Mitteilungen* 40(1). 15–29.
- Battefeld, Malte, Leuschner, Torsten & Rawoens, Gudrun. 2018. Evaluative morphology in German, Dutch and Swedish: Constructional networks and the loci of change. In Van Goethem, Kristel & Norde, Muriel & Coussé, Evie & Vanderbauwhede, Gudrun (eds.), *Category change from a constructional perspective*, 229–262. Amsterdam, Philadelphia: John Benjamins.
- Bauer, Laurie. 2003. *Introducing linguistic morphology*. 2nd ed. Washington, DC: Georgetown University Press.

- Bauer, Laurie & Lieber, Rochelle & Plag, Ingo. 2013. *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.
- Berg, Kristian. forthcoming. Changes in the productivity of German word-formation patterns: Some methodological remarks. *Linguistics*.
- Booij, Geert E. 2005. Compounding and derivation. Evidence from construction morphology. In Dressler, Wolfgang U. & Kastovsky, Dieter & Pfeiffer, Oskar E. & Rainer, Franz (eds.), *Morphology and its demarcations*, 109–132. Amsterdam, Philadelphia: John Benjamins.
- Booij, Geert E. 2010. *Construction morphology*. Oxford: Oxford University Press.
- Booij, Geert E. 2012. *The grammar of words: An introduction to linguistic morphology*. 3rd ed. Oxford: Oxford University Press.
- Börjars, Kersti & Vincent, Nigel & Walkden, George. 2015. On Constructing a Theory of Grammatical Change. *Transactions of the Philological Society* 113(3). 363–382.
- Cappelle, Bert. 2006. Particle placement and the case for “allostructions.” *Constructions* 7(S11). urn:nbn:de:0009-4-6839.
- Claidière, Nicolas & Scott-Phillips, Thomas C. & Sperber, Dan. 2014. How Darwinian is cultural evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1642). doi:[10.1098/rstb.2013.0368](https://doi.org/10.1098/rstb.2013.0368).
- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Davies, Mark. 2010. The Corpus of Historical American English (COHA): 400+ million words, 1810–2009. <http://corpus.byu.edu/coha> (last accessed 01/11/2018).
- Diessel, Holger. 2015. Usage-Based Construction Grammar. In Dąbrowska, Ewa & Divjak, Dagmar (eds.), *Handbook of Cognitive Linguistics*, 296–322. Berlin, New York: De Gruyter.
- Dunning, Ted. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19(1). 69–74.
- Ebengaard-Jensen, Kim. 2013. Semantic coherence in English accusative-with-bare-infinitive constructions. *Rask* 38. 161–176.
- Elsen, Hilke. 2009. Affixoide: Nur was benannt wird, kann auch verstanden werden. *Deutsche Sprache* 37. 316–333.
- Fauconnier, Gilles & Turner, Mark. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.
- Flach, Susanne. 2015. Let's go look at usage. In Thomas Herbst & Uhrig, Peter (eds.). *Yearbook of the German Cognitive Linguistics Association* 3(1). 231–251.
- Flach, Susanne. 2018. collostructions: An R Implementation for the Family of Collostructional Methods. [bit.ly/sflach](http://bit.ly/sflach) (last accessed 24/10/2018)
- Fleischer, Wolfgang & Barz, Irmhild. 2012. *Wortbildung der deutschen Gegenwartssprache*. 4th ed. Berlin, New York: De Gruyter.
- Gaeta, Livio & Ricca, Davide. 2006. Productivity in Italian word-formation. *Linguistics* 44(1). 57–89.
- Geyken, Alexander. 2009. The DWDS corpus: A reference corpus for the German language of the 20th century. In Fellbaum, Christiane (ed.), *Idioms and collocations: corpus-based linguistic and lexicographic studies*, 23–40. London: Continuum.
- Geyken, Alexander & Gloning, Thomas. 2015. A living text archive of 15th-19th century German: Corpus strategies, technology, organization. In Gippert, Jost & Gehrke, Ralf (eds.), *Historical Corpora*, 165–180. Tübingen: Narr.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, London: The University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

- Goldberg, Adele E. & Jackendoff, Ray. 2004. The English resultative as a family of constructions. *Language* 80(3). 532–568.
- Gries, Stefan Th. & Stefanowitsch, Anatol. 2004. Extending collocation analysis: A corpus-based perspective on “alternations.” *International Journal of Corpus Linguistics* 9(1). 97–129.
- Guardamagna, Caterina. 2018. Type frequency, productivity and schematicity in the evolution of the Latin secundum NP construction. In Coussé, Evie & Andersson, Peter & Olofsson, Joel (eds.), *Grammaticalization meets construction grammar*, 169–201. Amsterdam: John Benjamins.
- Hartmann, Stefan. 2014. Constructing a Schema: Word-Class Changing Morphology in a Usage-Based Perspective. In Hilpert, Martin & Flach, Susanne (eds.), *Yearbook of the German Cognitive Linguistics Association 2*, 235–252. Berlin, New York: De Gruyter.
- Hartmann, Stefan. 2018. Derivational morphology in flux: a case study of word-formation change in German. *Cognitive Linguistics* 29(1). 77–119.
- Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043–1068.
- Hein, Katrin & Engelberg, Stefan. 2017. Morphological variation: the case of productivity in German compound formation. In Koutsoukos, Nikos & Audring, Jenny & Masini, Francesca (eds.), *Morphological variation: synchrony and diachrony. Proceedings of the 11th Mediterranean Morphology Meeting*, 36–50. <http://hdl.handle.net/2078.1/203214> (last accessed 28/10/2018).
- Hilpert, Martin. 2013. *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press.
- Hilpert, Martin. 2018. Three open questions in diachronic construction grammar. In Coussé, Evie & Andersson, Peter & Olofsson, Joel (eds.), *Grammaticalization meets construction grammar*, 21–39. Amsterdam, Philadelphia: John Benjamins.
- Himmelman, Nikolaus P. 2004. Lexicalization and Grammaticization: Opposite or Orthogonal? In Bisang, Walter & Himmelman, Nikolaus P. & Wiemer, Björn (eds.), *What makes grammaticalization?*, 21–42. Berlin, New York: De Gruyter.
- Hüning, Matthias & Booij, Geert. 2014. From compounding to derivation: The emergence of derivational affixes through “constructionalization.” *Folia Linguistica* 48(2). 579–604.
- Koch, Peter. 1999. Frame and contiguity: On the cognitive bases of metonymy and certain types of word formation. In Panther, Klaus-Uwe & Radden, Günter (eds.), *Metonymy in Language and Thought*, 139–167. Amsterdam, Philadelphia: John Benjamins.
- Kopf, Kristin. in press. Von Korpus zu Korpus: Chancen und Risiken korpusübergreifenden Arbeitens. In Szczepaniak, Renata & Hartmann, Stefan & Dücker, Lisa (eds.), *Historische Korpuslinguistik*. Berlin, Boston: De Gruyter.
- Kövecses, Zoltán. 2015. *Where metaphors come from: Reconsidering metaphor in context*. Oxford: Oxford University Press.
- Langacker, Ronald W. 1999. *Grammar and conceptualization*. Berlin, New York: De Gruyter.
- Langacker, Ronald W. 2008. *Cognitive Grammar: A basic introduction*. Oxford: Oxford University Press.
- Leuschner, Torsten. 2010. Ausnahmepianist fettgeschreckt - inbleich! Deutsche, niederländische und schwedische Präfixoide im Spannungsfeld von Genealogie, Kreativität und Norm. In Dammel, Antje & Kürschner, Sebastian & Nübling, Damaris (eds.), *Kontrastive germanistische Linguistik*, 863–892. Hildesheim: Olms.
- Lakoff, George & Johnson, Mark. 2003. *Metaphors we live by*. Updated Edition. Chicago: University of Chicago Press.
- Marchand, Hans. 1969. *The categories and types of present-day English word formation: A synchronic-diachronic approach*. 2nd ed. München: Beck.

- Meibauer, Jörg. 2013. Expressive compounds in German. *Word Structure* 6(1). 21–42.
- Motsch, Wolfgang. 1996. Affixoide. Sammelbezeichnung für Wortbildungsphänomene oder linguistische Kategorie? *Deutsch als Fremdsprache* 3. 168–168.
- Munske, Horst Haider. 2002. Wortbildungswandel. In Habermann, Mechthild & Müller, Peter O. & Munske, Horst Haider (eds.), *Historische Wortbildung des Deutschen*, 23–40. Tübingen: Niemeyer.
- Norde, Muriel & Van Goethem, Kristel. 2015. Bleaching, productivity and debonding of prefixoids: A corpus-based analysis of ‘giant’ in German and Swedish. *Linguisticae Investigationes* 37(2). 256–274.
- Norde, Muriel & Sippach, Sarah. 2019. Nerdalicious scientainment: A network analysis of English libfixes. *Word Structure*. (in this issue)
- Olsen, Susan. 1986. “Argument-Linking” und unproduktive Reihen bei deutschen Adjektivkomposita. *Zeitschrift für Sprachwissenschaft* 5(1). 5–24.
- Perek, Florent. 2015. *Argument structure in usage-based construction grammar: experimental and corpus-based perspectives*. Amsterdam, Philadelphia: John Benjamins.
- Roth, Tobias. 2014. *Wortverbindungen und Verbindungen von Wörtern. Lexikografische und distributionelle Aspekte kombinatorischer Begriffsbildung zwischen Syntax und Morphologie*. Tübingen: Francke.
- Schäfer, Roland. 2015. Processing and querying large corpora with the COW14 architecture. In Bański, Piotr & Biber, Hanno & Breiteneder, Evelyn & Kupietz, Marc & Lungen, Harald & Witt, Andreas (eds.), *Challenges in the Management of Large Corpora (CMC-3)*, 28–34. <http://ids-pub.bsz-bw.de/frontdoor/index/index/docId/3826> (last accessed 28/10/2018)
- Schäfer, Roland & Bildhauer, Felix. 2012. Building large corpora from the web using a new efficient tool chain. In Calzolari, Cicoletta & Choukri, Khalid & Declerck, Terry & Doğan, Mehmet Uğur & Maegaard, Bente & Mariani, Joseph & Moreno, Asuncion & Odijk, Jan & Piperidis, Stelios (eds.), *Proceedings of LREC 2012*, 486–493.
- Schlücker, Barbara. 2012. Die deutsche Kompositionsfreudigkeit. Übersicht und Einführung. In Gaeta, Livio & Schlücker, Barbara (eds.), *Das Deutsche als kompositionsfreudige Sprache: Strukturelle Eigenschaften und systembezogene Aspekt*, 1–25. Berlin, New York: De Gruyter.
- Schmid, Hans-Jörg. 2017. A framework for understanding entrenchment and its psychological foundations. In Schmid, Hans-Jörg (ed.), *Entrenchment and the psychology of language learning. How we reorganize and adapt linguistic knowledge*, 9–39. Berlin, Boston: De Gruyter.
- Schmidt, Günter Dietrich. 1987. Das Affixoid: Zur Notwendigkeit und Brauchbarkeit eines beliebten Zwischenbegriffs der Wortbildung. In Hoppe, Gabriele & Kirkness, Alan & Link, Elisabeth & Nortmeyer, Isolde & Rettig, Wolfgang & Schmidt, Günter Dietrich (eds.), *Deutsche Lehnwortbildung*, 53–101. Tübingen: Narr.
- Smirnova, Elena. forthcoming. Deutsche Partizipialkomposita aus diachroner Perspektive. In Schwarz, Christian & Ganslmayer, Christine (eds.), *Historische Wortbildung: Theorie - Methode - Perspektiven*. Hildesheim: Olms.
- Stefanowitsch, Anatol & Gries, Stefan Th. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Stevens, Christopher M. 2005. Revisiting the affixoid debate: On the grammaticalization of the word. In Leuschner, Torsten & Mortelmans, Tanja & Groodt, Sarah (eds.), *Grammatikalisierung im Deutschen*, 71–84. Berlin, New York: De Gruyter.
- Tellenbach, Elke. 1985. Wortbildungsmittel im Wörterbuch. Zum Status der Affixoide. In Bahner, Werner (ed.), *Beiträge zu theoretischen und praktischen Problemen in der Lexikographie der deutschen Gegenwartssprache*, 266–315. Berlin: Akademie der Wissenschaften der DDR.

- Traugott, Elizabeth Closs. 2008. Grammaticalization, constructions and the incremental development of language: Suggestions from the development of degree modifiers in English. In Eckhart, Regine & Jäger, Gerhard & Veenstrs, Tonjes (eds.), *Variation, selection, development*, 219–250. Berlin, New York: De Gruyter.
- Traugott, Elizabeth Closs & Trousdale, Graeme. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
- Trousdale, Graeme. 2014. On the relationship between grammaticalization and constructionalization. *Folia Linguistica* 48(2). 557–577.
- Van Goethem, Kristel & Hiligsmann, Philippe. 2014. When two paths converge: Debonding and clipping of Dutch *reuze-*. *Journal of Germanic Linguistics* 26(1). 31–64.
- Wulff, Stefanie. 2006. Go-V vs. go-and-V in English: A case of constructional synonymy? In Gries, Stefan Th. & Stefanowitsch, Anatol (eds.), *Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis*, 101–126. Berlin, New York: De Gruyter.

## Appendix

### A.1 Distinctive collexeme analysis: [X-*landschaft*] and [X-*welt*] vs. free constituents

Table 3: Top 20 distinctive collexemes of [X-*landschaft*] compared to the frequency of the first constituents as free words in the DWDS20 corpus.

Collexeme	Freq [X- <i>landschaft</i> ]	G <sup>2</sup>
Kultur ‘culture’	87	252.98
Gebirge ‘mountains’	36	191.59
Fluss ‘river’	13	165.07
Winter ‘winter’	46	147.49
Mond ‘moon’	27	99.49
Park ‘park’	24	90.68
Alpen ‘alps’	11	88.03
Schnee ‘snow’	27	81.69
Hügel ‘hill’	21	80.93
Heide ‘heath’	18	77.18
Küste ‘coast’	20	54.81
Wunsch ‘wish’	38	51.52
Berg ‘mountain’	26	51.15
Moor ‘moor’	9	37.89
Wald ‘wood’	22	30.79
Düne ‘dune’	6	30.06
Löss ‘loess’	2	29.67
Geest ‘geest’	4	29.24
Ruine ‘ruin’	9	27.43
Hochgebirge ‘high mountains’	5	26.99

Table 4: Top 20 distinctive collexemes of [X-landschaft] compared to the frequency of the first constituents as free words in the DECOW16B corpus.

Collexeme	Freq [X-landschaft]	G <sup>2</sup>
Kultur ‘culture’	2894	16229.65
Sauna ‘sauna’	704	4954.46
Hügel ‘hill’	567	3638.7
Medium ‘medium’ <sup>22</sup>	1121	3433.54
Natur ‘nature’	958	3144.16
Fluss ‘river’	593	2661.4
Berg ‘mountain’	599	2164.77
Agrar ‘agriculture’	231	2079.95
Bildung ‘education’	609	1959.06
Düne ‘dune’	262	1896.63
Mittelgebirge ‘low mountains’	212	1844.29
Gebirge ‘mountains’	266	1647.64
Heide ‘heath’	309	1623.44
Winter ‘winter’	516	1504.97
Bodden ‘bodden’	167	1498.37
Park ‘park’	426	1482.09
See ‘lake’	463	1383.75
Alpen ‘alps’	65	1209.56
Wüste ‘desert’	246	1125.82
Mond ‘moon’	285	1106.55

Table 5: Top 20 distinctive collexemes of [X-welt] compared to the frequency of the first constituents as free words in the DWDS20 corpus. First constituents that are not nouns (e.g. prepositions like mit ‘mit’, nach ‘after’) were omitted.

Collexeme	Freq [X-welt]	G <sup>2</sup>
Leben ‘life’	888	2045.77
Tier ‘animal’	361	1079.6
Pflanze ‘plant’	168	585.33
Geist ‘spirit’	213	355.92
Gedanke ‘thought’	216	329.69
Wahrnehmung ‘perception’	91	307.54
Idee ‘idea’	168	305.57
Erscheinung ‘apparition’	143	296.1

<sup>22</sup> A reviewer points out that it is problematic to conflate *Medien* ‘media’ with *Medium* ‘medium’, as the former can be considered lexicalized. However, the collexeme analysis presented here relies on the lemmatization of the DECOW corpus, where *Medien* is tagged with the singular form *Medium*. But as *Medien* makes up for the vast majority of all instances tagged as *Medium* in DECOWB (1,133,564 out of 1,758,353 tokens), other uses can be considered marginal.

Fach ‘discipline’	64	290.3
Vorstellung ‘imagination’	153	282.76
Kultur ‘culture’	151	268.11
Wirklichkeit ‘reality’	139	260.76
Geschäft ‘business’	104	178.89
Alpen ‘alps’	22	153.12
Körper ‘body’	112	119.24
Sinn ‘sense’	181	115.02
Märchen ‘fairy tale’	41	113.23
Gelehrte <sup>23</sup> ‘scholar’	44	107.74
Erwachsene ‘adult’	44	106.5
Alltag ‘everyday’	33	96.2

Table 6: Top 20 distinctive collexemes of [X-welt] compared to the frequency of the first constituents as free words in the DECOW16B corpus. First constituents that are not nouns (e.g. prepositions like mit ‘with’, nach ‘after’) were omitted. Also, immo- and kugelblitzdemoheimat- were omitted as the respective compounds are proper names (of the journal Immowelt and a blogger with the nickname Kugelblitzdemoheimatwelt).

Collexeme	Freq [X-welt]	G <sup>2</sup>
Arbeit ‘work’	111022	251481.26
Pflanze ‘plant’	31363	165336.89
Tier ‘animal’	60694	138114.82
Finanz ‘finance’	27190	133804.18
Netz ‘web’	43117	127642.36
Berg ‘mountain’	29602	87965.02
Erlebnis ‘experience’	26369	84338.29
Fach ‘discipline’	20812	79999.69
Leben ‘life’	66933	56688.45
Traum ‘dream’	18064	34358.36
Kino ‘cinema’	14362	33096.65
Beruf ‘occupation’	17963	32911.62
Klang ‘sound’	11195	30166.21
Geschäft ‘business’	17010	28860.7
Lyrik ‘poetry’	6323	27798.7
Mode ‘fashion’	10735	27185.1
Schein ‘appearance’	8017	27087.11
Vogel ‘bird’	12837	25640.31
Ware ‘product’	2621	23677.62
Alpen ‘alps’	2022	22360.72

<sup>23</sup> Nouns with the suffix -er such as *Erwachsener* are lemmatized in their female/plural form *Erwachsene* in the DWDS20 corpus.

## A.2 Distinctive collexeme analysis: [X-landschaft/welt] vs. [Landschaft/Welt des/der X]

Table 7: Top 20 distinctive collexemes for [X-landschaft] compared to [Landschaft des/der X] in DECOW16B

Collexeme	Freq. [X-landschaft]	Freq. [Landschaft des/der X]	G <sup>2</sup>
Kultur 'culture'	2894	9	3550.18
Medium 'medium'	1121	6	1313.91
Natur 'nature'	958	10	1080.11
Sauna 'sauna'	704	0	861.02
Bildung 'education'	609	1	730.51
Hügel 'hill'	567	4	650.62
Fluss 'river'	593	12	624.88
Winter 'winter'	516	3	597.11
Berg 'mountain'	599	47	467.71
Schule 'school'	381	4	425.53
Forschung 'research'	331	0	402.62
Hochschule 'University'	332	1	391.79
See 'lake'	463	29	389.04
Partei '(political) party'	305	2	349.86
Museum 'museum'	300	5	321.55
Heide 'heath'	309	9	307.92
Düne 'dune'	262	2	297.99
Agrar 'agriculture' <sup>24</sup>	231	0	280.57
Mond 'moon'	285	14	255.32
Gebirge 'mountains'	266	10	252.94

Table 8: Top 20 distinctive collexemes for [X-welt] compared to [Welt des/der X] in DECOW16B. Only items with at least 100 attestations were taken into account (hence, "0" in the second frequency column just means "less than 100")

<sup>24</sup> It can be debated whether *Agrar* can actually occur as a free word or if it only occurs in bound form (maybe qualifying as a so-called prefix, i.e. non-native morphemes that only occur in compounds and derivations, see e.g. Fleischer & Barz 2012: 63f.). *Agrar* does occur as a free word in the DWDS corpora, but usually it is used as shorthand for compounds like *Agrarministerium* 'ministry of agriculture'.

attestations”). First constituents that are not nouns (e.g. prepositions like *mit* ‘mit’, *nach* ‘after’) were omitted. Also, *immo* was omitted as *Immo*welt is a proper name (of a journal).

Collexeme	Freq. [X-welt]	Freq. [Welt des/der X]	G <sup>2</sup>
Arbeit ‘work’	111022	789	14031.26
Leben ‘life’	66933	153	10212.3
Netz ‘web’	43117	142	6226.63
Pflanze ‘plant’	31363	0 (<100)	5383.53
Tier ‘animal’	60694	1183	4666.66
Erlebnis ‘experience’	26369	0 (<100)	4518.93
Thema ‘topic’	21298	0 (<100)	3643.88
Fach ‘discipline’	20812	0 (<100)	3560.17
Berg ‘mountain’	29602	297	3227.16
Beruf ‘occupation’	17963	0 (<100)	3069.97
Geschäft ‘business’	17010	0 (<100)	2906.2
Finanz ‘finance’	27190	401	2470.06
Insel ‘island’	9681	0 (<100)	1650.11
Gedanke ‘thought’	18437	287	1617.15
Spiel ‘game’	33416	1151	1378.78
Nacht ‘night’	8085	0 (<100)	1377.36
Vogel ‘bird’	12837	173	1216.95
Dame ‘lady’	6847	0 (<100)	1165.99
Mann ‘man’	16541	357	1149.31
Lyrik ‘poetry’	6323	0 (<100)	1076.58

Table 9: Top 20 distinctive collexemes for [Landschaft des/der X] compared to [X-landschaft] in DECOW16B. One item – “Leibniz” – was omitted because it only occurs in “Fakultät für Architektur und Landschaft der Leibniz Universität Hannover” ‘faculty of architecture and landscape of the Leibniz University Hannover’.

Collexeme	Freq. [Landschaft des/der X]	Freq. [X-landschaft]	G <sup>2</sup>
Region ‘region’	171	1	502.09
Insel ‘island’	217	43	441.28
Jahr ‘year’	130	0	390.11
Naturpark ‘nature park’	92	0	275.71
Toskana ‘Tuscany’	92	1	265.16
Erde ‘Earth’	74	0	221.63
Nationalpark ‘national park’	73	1	208.53
Land ‘land’	62	0	185.61

Heimatlosigkeit 'lack of a homeland'	58	0	173.61
Schweiz 'Switzerland'	57	0	170.61
Umgebung 'surroundings'	66	4	168.97
Provence 'Provence'	45	0	134.64
Bibel 'Bible'	41	0	122.65
Norden 'North'	41	0	122.65
Staat 'state'	40	0	119.66
Seele 'soul'	96	69	98.18
Provinz 'province'	32	0	95.7
Zukunft 'future'	41	6	89.79
Halbinsel 'peninsula'	29	0	86.72
Kreis 'county'	28	0	83.73

Table 10: Top 20 distinctive collexemes for [X-welt] compared to [Welt des/der X]. Only items with at least 100 attestations in were taken into account (hence, "0" in the second frequency column just means 'less than 100 attestations'). "Welt der Amélie" was omitted because "Die fabelhafte Welt der Amélie" is a proper name (a movie title). "Landschaften der Bibel" often occurs in a proper name (a book title) as well, but not exclusively, which is why it was not omitted.

Collexeme	Freq. [Welt des/der X]	Freq. [X- welt]	G <sup>2</sup>
PC 'PC'	1950	0 (<100)	9806.81
Physik 'physics'	1438	0 (<100)	7226.74
Internet 'internet'	2600	2038	7066.25
Sims 'Sims'	1287	0 (<100)	6466.52
Wissenschaft 'science'	1812	694	6270.18
Papst 'pope'	1240	0 (<100)	6229.97
Lebende 'living'	1128	0 (<100)	5666.38
Frieden 'peace'	956	0 (<100)	4801.21
Mensch 'human'	2977	5704	4780.18
Mittelalter 'Middle Ages'	854	0 (<100)	4288.34
Buch 'book'	3230	8236	4009.2
Islam 'Islam'	660	0 (<100)	3313.28
Videospiel 'video game'	1172	831	3309.4
Zahl 'number'	1038	531	3294.89
Star 'celebrity'	656	0 (<100)	3293.18

Reiche 'rich'	653	0 (<100)	3278.11
Magie 'magic'	633	0 (<100)	3177.62
Erwachsene 'adult'	1979	4132	2949.43
Technik 'technology'	1207	1221	2903.78
Ritter 'knight'	765	228	2809.42