

Stefan Hartmann*

Derivational morphology in flux: a case study of word-formation change in German

<https://doi.org/10.1515/cog-2016-0146>

Received December 12, 2016; revised August 14, 2017; accepted August 20, 2017

Abstract: The diachronic change of word-formation patterns is currently gaining increasing interest in cognitive-linguistic and constructionist approaches. This paper contributes to this line of research with a corpus-based investigation of nominalization with the suffix *-ung* in German. In doing so, it puts forward both theoretical and methodological considerations on morphology and morphological change from a usage-based perspective. Regarding methodology, the long-standing topic of how to measure (changes in) the productivity of a morphological pattern is discussed, and it is shown how statistical association measures can be applied to quantify the relationship between word-formation patterns and their bases. These findings are linked up with theoretical considerations on the interplay between constructional schemas and their respective instances.

Keywords: morphology, historical linguistics, construction grammar

1 Introduction

Word-formation patterns and their diachronic development have received a considerable amount of attention in recent cognitive-linguistic and constructionist approaches (cf. e.g. Booij 2010; Hilpert 2013; Traugott and Trousdale 2013). In a Construction Grammar (CxG) framework, they can be conceived of as partially filled (e.g. in the case of English [V-*ation*]) or unfilled (e.g. conversion: *to ride* > *the ride*) constructional schemas. From a cognitive-linguistic point of view, these schemas are of particular interest for at least two reasons. Firstly, the classic view that (at least some) derivational morphemes are meaningless (cf. e.g. Marchand 1969: 215; Spencer 2001: 227) is in sharp contrast with the view held in most cognitive-linguistic and constructionist approaches that “all constructs validly posited for grammatical description [...] must in some way be

*Corresponding author: Stefan Hartmann, German Department, University of Bamberg, Hornthalstrasse 2, 96045, Bamberg, Germany, E-mail: stefan1.hartmann@uni-bamberg.de

meaningful.”¹ (Langacker 2008: 5) As such, unraveling the highly abstract meaning of word-formation constructions is a key challenge for cognitive-linguistic approaches. Secondly, recent years have seen an increasing awareness towards interaction patterns in language. If the view of language as “a network of interlinked constructions” (Hilpert 2014: 57) is correct, it stands to reason that constructions at different levels of abstraction interact with each other. A variety of methodological tools has therefore been developed to study the interaction between words and constructions (e.g. collostructional analysis, cf. Stefanowitsch and Gries 2003) or between words/constructions and their linguistic context (e.g. the methodological toolkit of distributional semantics, cf. Perek 2016). Morphology, which has been described as “the conceptual centre of linguistics [...] at the interface between phonology, syntax and semantics” (Spencer and Zwicky 1998: 1), lends itself particularly well to the analysis of such interaction patterns.

The present paper discusses one specific example of word-formation change. However, it also aims at contributing to the theoretical and methodological discussion on morphological change more generally. Specifically, I will argue that the study of interaction patterns between constructions (in the CxG sense, i.e. form-meaning pairs) at different levels of abstraction can contribute significantly to our understanding of word-formation patterns and their diachronic development.

As a case study, this paper discusses a German word-formation construction that has been investigated quite extensively across different theories and frameworks, namely nominalization with the suffix *-ung*. It has been shown that this construction experiences a significant decrease in morphological productivity from the Early New High German period (1350–1650) onwards (Demske 2000, Hartmann 2014a, Hartmann 2014b, and Hartmann 2016). However, the development of the word-formation pattern in the more recent past has not been systematically investigated so far. This paper therefore focuses on developments in the early stages of the New High German period (1650–today) up to the end of the nineteenth century. After a brief review of previous research in Section 2, Section 3 introduces the corpus compiled for the present study. Section 4 presents a quantitative analysis of frequency and productivity data and discusses the idea of applying collostructional analysis to investigate the interaction between word-formation patterns and word-formation products. Section 5 discusses the implications of the empirical findings for a cognitive-linguistic theory of word-formation change.

¹ Note that Langacker does not use “constructs” in the CxG sense here. In CxG, the term refers to concrete instantiations of a construction (cf. Traugott and Trousdale 2013: 2).

2 The rise of *ung*-nominals and the decline of *ung*-nominalization: a brief review of previous research

Derivations with *-ung* are ubiquitous in the German language. Some like *Erfahrung* ‘experience’, *Meinung* ‘opinion’ and *Leistung* ‘achievement’ are even among the 500 most frequent lemmas in the lemma frequency list of DECOW14AX (Schäfer and Bildhauer 2012; Schäfer 2015), the largest available corpus of present-day German, which comprises more than 5 million different lemma types. However, Demske’s (2000) corpus-based study shows that the pattern used to be less frequent, but more productive in earlier stages of German. In order to measure the pattern’s diachronic productivity, she operationalizes Baayen’s “potential productivity” (Baayen 1992; Baayen 2009), i.e. the ratio of hapax legomena (instances attested only once in a corpus) to the total number of instances of the construction in question. While Demske (2000) focuses on the Early New High German period (1350–1650), Hartmann (2014a, 2014b) takes the early stages of the New High German period into account, drawing on the GerManC corpus (Durrell et al. 2007), which covers the time from 1650 to 1800. Quantitative analyses of the GerManC data show that *ung*-nominalization, during that time, experienced an increase in type and token frequency, while it saw a significant decrease in potential productivity (Hartmann 2016).

Various factors have been proposed to explain these diachronic changes in productivity. Demske (2000) emphasizes the key role of lexicalization: While *-ung* originally derives nouns whose meaning is very close to that of their base verbs, highly frequent *ung*-nominals become lexicalized, e.g. *Lesung* ‘public reading’ (which cannot refer to the process of reading any more) or *Heizung* ‘heating device’ (which cannot refer to the process of heating in present-day German). Thus, the pattern becomes semantically less transparent (cf. Demske 2000: 396). In addition, infinitival nominalization (e.g. *das Lesen* ‘reading’) emerges as a new competitor (cf. Demske 2000; Barz 1998). Although nominalized infinitives are already attested in the Old High German period (cf. e.g. Demske 2001), they seem to become applicable to ever more bases in the Early New High German period (cf. Werner 2012: 168f.).

The diachronic change of *ung*-nominalization qualifies as constructional change. According to Traugott and Trousdale (2013: 26), “[a] constructional change is a change affecting one internal dimension of a construction”. According to Hilpert (2013: 16),

[c]onstructional change selectively seizes a conventionalized form-meaning pair of a language, altering it in its form, its function, any aspect of its frequency, its distribution in the linguistic community, or any combination of these.

These complementary notions of constructional change are in turn compatible with Scherer's (2006, 2007) definition of word-formation change. According to her, "word-formation change in the narrow sense" can be seen as change in word-formation constraints, which in turn is reflected by changes in morphological productivity. Word-formation constraints can in turn be subdivided into input and output constraints. Input constraints affect the constituents of the pattern, e.g. the word classes whose members are eligible as bases for a particular word-formation pattern. Output constraints pertain to the words formed by the pattern, more specifically, to the reading variants afforded by the word-formation products that instantiate the pattern. For instance, English *-er*-nominals like *sleeper* allow for object and person readings, among others (cf. Panther and Thornburg 2001), while *-ation* does not permit person readings (*nominalization* *'someone who nominalizes').² Most research on *-ung* deals with the question of how the input constraints governing *ung*-nominalization in present-day German can be explained in more formally-oriented frameworks (e.g. Shin 2001; Ehrich and Rapp 2000; Demske 2002; Roßdeutscher and Kamp 2010).

From a constructionist perspective, the constraints that govern the application of a word-formation pattern can be seen as part of the language user's linguistic knowledge about the construction in question. For instance, a present-day native speaker of German implicitly "knows" that fairly new formations like *Aufbeulung* 'bulging' or *Versionierung* 'version control'³ are grammatical while others such as **Googelung* 'googling' or **Twitterung* 'tweeting' are not. In fact, some *ung*-nominals attested in earlier stages of German have fallen out of use. Paradigm examples include *Murmelung* 'muttering' (occasionally still attested in the ENHG period) and *Jammerung* 'moaning' (attested in the Middle High German period, 1050–1350), cf. example (1).

- (1) a. kein Fluch/**Murmelung** noch Ungedult würde bey ihnen gefpürt
'no curse, **muttering**, nor impatience was sensed in them'
(Grimmelshausen, *Simplicissimus*, 1699, DTA)

² However, output constraints are seldom, if ever, absolute, and they are subject to change, which is why I prefer to speak of the "output potential" of a word-formation pattern (see Hartmann 2018 for an in-depth discussion).

³ Examples from *wortwarte* (www.wortwarte.de), a collection of neologisms (last checked 26/01/2016).

- b. daz herze ir in dem lîbe spielt/von sender **jâmerunge**.
 the heart her in the body plays from yearning moaning
 ‘Her heart played around in her body due to [its] yearnful **moaning**.’
 (Konrad von Würzburg, Herzmaere, thirteenth century, MHG Dictionary)

These examples indicate – to use Scherer’s (2006) terminology introduced above – that the input constraints governing new coinages in *-ung* have changed significantly. Importantly, these constraints go in tandem with the *function* of the pattern. In Early New High German, for instance, *ung*-nominals are often used to encode progressivity, especially when used as complements of prepositions, e.g. *in Lesung* ‘in reading’ (cf. Demske 2000; Hartmann 2014b). In present-day German, by contrast, they tend to denote what is maybe best described with Langacker’s (e.g. 2008) notion of “bounded region”. Hartmann (2016) shows that many present-day *ung*-nominals refer to bounded regions in space and/or time. For example, *Ausgrabung* ‘excavation’ can refer to an excavation site, and *Lesung* ‘reading’ usually cannot refer to the process of reading any more but only to a public reading event. Taking up Langacker’s (e.g. 2008) distinction between “sequential scanning” and “summary scanning”, it can be argued that ENHG uses of *ung*-nominals like *Lesung* tend to construe the semantic content of the nominalization in a more “verb-like” fashion, viz. as a process unfolding over time, while the predominant present-day usage apprehends the event holistically. Systematic corpus studies have lent support to the hypothesis that the diachronic development of *ung*-nominalization can be understood in terms of increasing “nouniness” (cf. Demske 2000; Demske 2002; Hartmann 2016): *ung*-nominals tend to assume more and more prototypically nominal features – both regarding their semantics and their syntactic behavior – while more “verby” uses of *ung*-nominalization as exemplified in (1) fall out of use.

The subsequent sections of this paper present a corpus study that investigates this development in detail.

3 The corpus: DTA and DTA *baby*

The German Text Archive (*Deutsches Textarchiv*, DTA for short⁴) is a collection of German texts from about 1600 to 1900. As such, it complements other corpora like GerManC for the time from the seventeenth to the eighteenth century, and it

4 www.deutschestextarchiv.de (last checked 26/01/2016).

fills an important gap as it covers the nineteenth century, for which only sparse corpus resources were available until recently. While the 2014 version of the DTA which was used for the present study comprises texts from different areas and different text types, it is a rather opportunistic corpus in that it is very large, but quite unbalanced (note that the current version of the DTA has been considerably extended and is much more balanced). In particular, the amount of text sampled for the different time periods that are covered by the corpus is heavily skewed towards the later decades. In addition, the individual texts differ considerably in length. Therefore, a subset of the corpus was compiled, which will be referred to as *DTAbaby*.⁵ The corpus was designed to be balanced for time periods of 50 years and three different text types: fictional texts, scientific texts, and so-called functional literature (*Gebrauchsliteratur*), which, in the DTA classification scheme, subsumes all kinds of non-fiction literature, not only guides and companions, as the label would suggest, but also (auto)biographies and even sermons. The fourth text type of the DTA classification scheme, viz. newspaper texts, is so heavily underrepresented that it was discarded entirely in compiling the balanced subcorpus. In order to assemble *DTAbaby*, an algorithm was used which, for each of the six fifty-year periods from 1600 to 1900, randomly selects 15 texts with a minimum length of 4000 tokens. From each text, a sample of 4000 consecutive words is extracted; again, the starting point is determined randomly. Thus, we arrive at a balanced corpus comprising 916,786 tokens in 270 texts. The overlap with the GerManC corpus regarding the covered time span allows for comparisons between the results from previous research summarized above and the results of the corpus study reported on here. Figure 1 gives an overview of the composition of DTA and *DTAbaby*.

The corpus was searched with CQP.⁶ In the corpus, 10,946 *ung*-nominals were retrieved (after manual deletion of all false hits). Compounds were only taken into account if the *ung*-nominal was the head of the compound, i.e. a compound such as *Bergerfahrung* ‘mountain-experience’ was included in the present analysis, while compounds like *Erfahrungssache* ‘matter of experience’ were disregarded. This decision was made both for practical and for conceptual reasons. On the practical side, taking all compounds with *ung*-nominals as non-head constituents into account would have entailed a much more cumbersome process of manually checking the results and deleting false hits. Conceptually, it

⁵ This name is of course an allusion to the smaller and balanced version of the British National Corpus, BNC-baby.

⁶ Cf. <http://cwb.sourceforge.net/index.php> (last checked 26/01/2016)

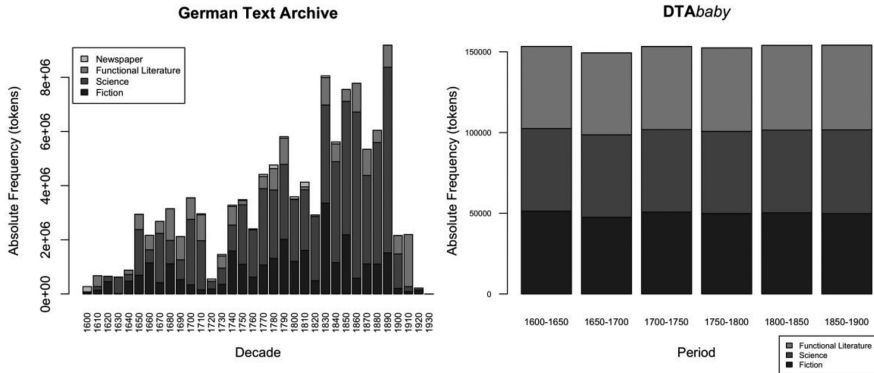


Figure 1: Overview of the absolute token frequencies of DTA (left panel) and DTA *baby* (right panel).

can be expected that productively coined *ung*-nominals will be more prone to appear on their own or as heads of compounds rather than in non-head position.

The lemmatization of the individual types was checked manually. In particular, if *ung*-nominals occurred as last constituents of compounds, the last constituent was taken as the lemma. For instance, a compound like the aforementioned *Bergerfahrung* ‘mountain-experience’ was lemmatized as *Erfahrung* ‘experience’.⁷ This is an important prerequisite for the calculation of potential productivity and the cross-tabulation analysis reported on in the next section to be reliable.

4 Quantitative analysis

This section presents the results of quantitative analyses on the basis of the DTA and DTA_{baby} data. Following a “mixed-methods” approach as advocated in e.g. Fonteyn and Hartmann (2016), several different quantitative approaches are combined. Such a mixed-methods approach follows straightforwardly from the key insight of cognitive linguistics and Construction Grammar that language

⁷ Taking compounds into account could be seen as potentially problematic: As Krott et al. (1999) have shown, the word-formation products of a specific pattern tend to be used more often as input for another word-formation process as the pattern becomes less productive. This suggests that the dynamics of productivity could differ between “simple” *ung*-nouns on the one hand and compounds on the other. (Thanks to Harald Baayen for pointing this out to me.) However, the number of compounds in the data is relatively small (809, amounting to 7% of the data), and all results reported on in the present paper hold even if all compounds are omitted.

users have rich and multi-faceted knowledge about linguistic signs (words and constructions), and that these signs in turn form part of a complex network (cf. e.g. Taylor 2012). As such, different methods can reveal different aspects of what speakers know about a construction. In the case of *ung*-nominalization, frequency and productivity measures, the analysis of interaction patterns, and explorative association measures can reveal formal and semantic preferences and constraints which characterize the construction under investigation.

The remainder of this chapter is structured as follows. Section 4.1 presents frequency and productivity analyses, starting with basic measures of type and token frequency as well as potential productivity based on the *DTAbaby* data, then comparing different productivity models obtained using finite Zipf-Mandelbrot models. Section 4.2 analyzes the interaction between *ung*-nominals and syntactic constructions, while Section 4.3 discusses whether collostructional analysis (Stefanowitsch and Gries 2003) can be fruitfully applied for investigating the association between word-formation constructions and their instantiations.

4.1 Frequency and productivity

The question of how productivity, i.e. the extensibility of a pattern to new cases (cf. Booij 2012: 70), can be measured has been subject to considerable debate in linguistic morphology (cf. e.g. Plag 1999; Bauer 2001). The main upshot of these debates has been the insight that rather than relying on one measure of productivity exclusively, multiple variables should be taken into account and compared. Baayen (2009) mentions a variety of measures for quantitatively assessing morphological productivity. Realized productivity measures the type frequency of a pattern (cf. Hilpert 2013: 132). The measure of expanding productivity relates the number of hapaxes belonging to the category in question to the total number of hapaxes in the corpus. As such, it is “an estimate of the contribution of morphological category C to the growth rate of the total vocabulary.” (Baayen 2009: 902) Finally, potential productivity relates the number of hapax legomena in a specific category to the total number of tokens belonging to that category (cf. Baayen 2009: 906). As Baayen (1993: 189) points out, the number of hapaxes is seen as an index for the likelihood of encountering neologisms – “the probability of encountering neologisms is measured indirectly by means of the probability of encountering hapaxes.” As a thoroughly balanced corpus, *DTAbaby* provides an ideal situation for applying these measures. Table 1 provides a general overview of the token and type frequencies as well as the number of hapax legomena belonging to the construction in question in the corpus.

Table 1: Overview of the corpus data.

	Tokens	Types	Hapax Legomena
Period 1 (1600–1649)	922	335	84
Period 2 (1650–1699)	924	384	93
Period 3 (1700–1749)	1273	395	69
Period 4 (1750–1799)	2106	501	64
Period 5 (1800–1849)	2720	614	102
Period 6 (1850–1899)	3001	663	159
Sum	10946	2892	571

While not being a productivity measure itself, the token frequency of a pattern should be taken into account as well. As, among others, Gaeta and Ricca (2006) point out, the correct interpretation of productivity values partly depends on the token frequencies of the individual patterns, which makes comparisons between different word-formation constructions problematic if they differ considerably in frequency: Given a highly frequent pattern *H* with many established instantiations and a low-frequency pattern *L*, both established and ad-hoc formations coined according to *L* will enter a corpus of size *n* with a much smaller likelihood than instances of *H*. By contrast, in the case of *H*, the highly frequent established instantiations will always outweigh any ad-hoc formations in measuring potential productivity. Taken together, these factors lead to “an overestimation of the values of *P* for the less frequent suffixes” (Gaeta and Ricca 2006: 63).

Like many other criticisms of or reservations against Baayen’s productivity measures, these considerations have been brought forward with regard to comparisons between two different word-formation patterns in a synchronic perspective. However, they also apply for comparisons between productivity values of one pattern at different points in time. For *ung*-nominalization, a significant increase in token frequency can be detected in the *DTAbaby* corpus. Using Kendall’s Tau, a rank-correlation coefficient which Hilpert and Gries (2009) have suggested to be an appropriate measure for assessing frequency changes in diachronic corpora, the diachronic change is identified as highly significant (Kendall’s $\tau=1$, $T=15$, $p<0.01$). The type frequency (realized productivity) of *ung*-nominalization experiences a significant increase as well ($\tau=1$, $T=15$, $p<0.01$). Figure 2 plots these diachronic frequency developments.

Due to the significant differences in token frequency between the six fifty-year periods, the caveats outlined above apply in assessing the potential productivity as well as the expanding productivity of *ung*-nominalization in the

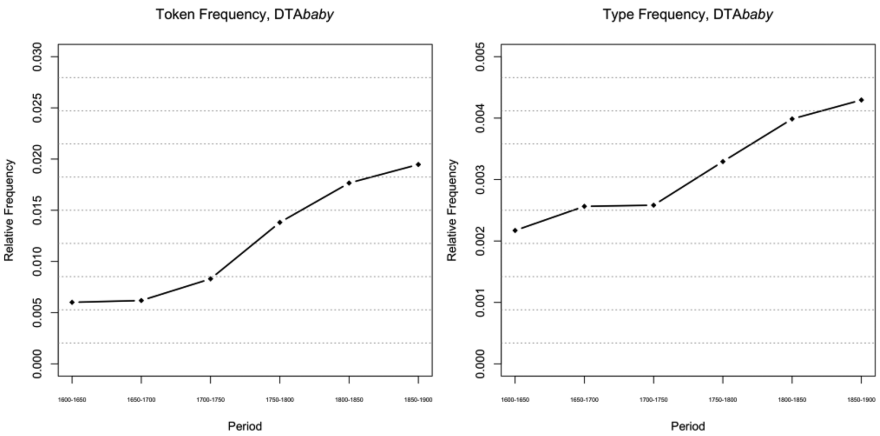


Figure 2: Token frequency (left panel) and type frequency (right panel) of *ung*-nominalization in the DTAaby corpus.

DTAaby corpus. Still, these productivity measures suggest an interesting development (see Figure 3 below). The steep decrease in potential productivity up to about 1800 is in line with Hartmann’s (2016) findings based on the GerManC corpus (see Section 2 above). However, the nineteenth century sees an – albeit very slight – increase in the potential productivity of the pattern. This is quite surprising given that usually, when the number of tokens increases, potential

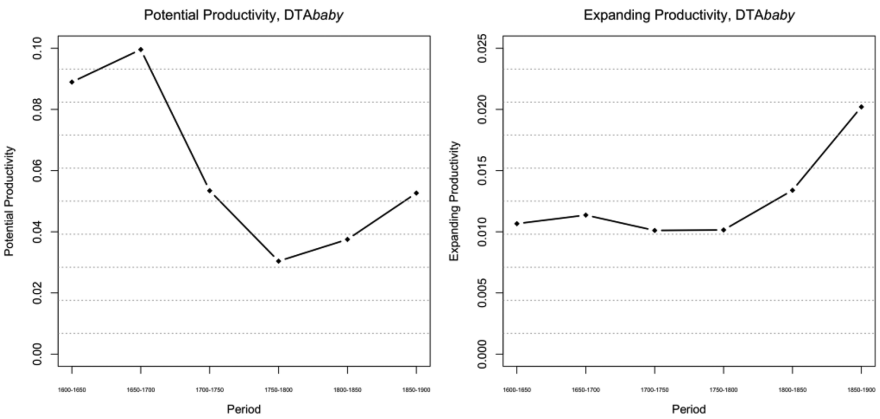


Figure 3: Potential productivity (left panel) and expanding productivity (right panel) of *ung*-nominalization in the DTAaby corpus.

productivity tends to decrease. This might suggest that there are indeed some new *ung*-formations that enter the language after the initial decline in productivity. We will return to this point below. First, however, let us address the question of whether the apparent decline in productivity in the first two centuries reflects an actual trend or if it can just be attributed to the growing number of tokens in the dataset.

In order to compensate for the overestimations mentioned above in measuring potential productivity, a Zipf-Mandelbrot model can be used to determine the expected number of hapaxes for arbitrary sample sizes (cf. e.g. Schneider-Wiejowski 2011).

Drawing on the DTA data, Baroni and Evert's *zipfR* package (cf. Baroni and Evert 2007) was used to obtain the type probability distributions. More precisely, a finite Zipf-Mandelbrot (fZM) model was used, which makes use of the fact that word frequency distributions obtained from random text have been found to be strikingly similar to the distribution described by Zipf's law (cf. Li 1992; Evert 2004), i.e. the observation that the frequency of words is inversely proportional to their rank order (cf. Baayen 2001: 13–16). As Evert (2004) points out, Zipf's law is of a purely descriptive nature; the Zipf-Mandelbrot model (cf. e.g. Mandelbrot 1962), by contrast, is a population model. The mathematical formulation of the Zipf-Mandelbrot law is given in (2).

$$\pi_z = \frac{C}{(z+b)^a} \quad (2)$$

In (2), π_z denotes the population probability of a word with the Zipf rank z , while C is a normalizing constant that makes sure that the probabilities sum up to 1 (cf. Baayen 2001: 14f.) a and b are the distribution parameters, whereby a modifies the slope of the rank-frequency graph, while b introduces a downward curvature which is especially useful for appropriately describing high-frequency ranks⁸ (cf. Baayen 2001: 101f.). Evert (2004) formulates the Zipf-Mandelbrot law in terms of a type density function, which allows for calculating expected values of V (the total number of different types) and V_m (types occurring m times) for a specific sample size N .

fZM models have been used in different ways in the morphological literature. In their demonstration of the *zipfR* package, for instance, Baroni and Evert (2007) compare two Italian suffixes by extrapolating their frequencies to a particular value of N (in this case, the N obtained for the more frequent suffix) and comparing their potential productivities with the help of the estimated numbers of V_1 (i.e., hapax legomena). Schneider-Wiejowski (2011) takes a similar approach, extrapolating, however, to an arbitrary value of 5000. Kempf (2016:

8 Note that Zipf's law is a special case of (2) with $b = 0$.

122) takes a different approach in her diachronic study of German adjectival suffixation, calculating, for each suffix, the N value where V_i reaches a value < 1 , i.e. where virtually no hapaxes belonging to the word-formation pattern in question are expected to be found in the corpus any more.

While the *DTAbaby* corpus is ideally suited for tasks that require equally-sized subcorpora for each time period, the finite Zipf-Mandelbrot model provides an estimate for larger corpus sizes. Therefore, the entire DTA can be used to obtain productivity estimates using large number of rare event (LNRE) models. In order to compare the hapax growth curves diachronically, the DTA, in which altogether 1,713,147 *ung-nouns* (8400 types) were found,⁹ was divided into three subcorpora, one for each century.

For the present fZM model, a more coarse-grained periodization of the *DTAbaby* data as compared to the frequency and productivity calculations reported on above was used. As mentioned above, the corpus was divided into three time slices, one for each century. The extrapolated values were calculated using the *lnre()* function from the *zipfR* package. Figure 4 shows the results.

As Figure 4 shows, the extrapolated hapax curves for the three centuries suggest a somewhat different picture of the pattern's productivity. While the potential productivity calculation based on the *DTAbaby* data suggests a slight increase in productivity throughout the nineteenth century, there is virtually no difference between the extrapolated hapax growth curves for the eighteenth and nineteenth centuries in Figure 4. The curve for the seventeenth century, by contrast, suggests a much higher degree of productivity, which is very much in line with our expectations from previous research as reviewed above.

However, it should be noted that the goodness of fit of these models, as measured by a multivariate chi-squared test, is rather suboptimal (seventeenth century: $\chi^2(13) = 208.06$, $p = 3e-37$; eighteenth century: $\chi^2(13) = 250.08$, $p = 6.17e-46$; nineteenth century: $\chi^2(13) = 183.80$, $p = 2.85e-32$).¹⁰ Baroni and Evert (2005) point out that the goodness of fit is not necessarily a measure of extrapolation quality.¹¹ Nevertheless, the poor goodness of fit gives rise to the suspicion that there will be a great deal of variance between different models fit to random samples of the same data. In order to explicitly take this variance into account, a

⁹ As in the case of *DTAbaby*, the automatic lemmatization of the types was corrected manually. Due to the large amount of types, however, it is possible that a negligible amount of false hits has been overlooked.

¹⁰ In the evaluation of fZM models, low chi-squared statistics and high p-values indicate a high goodness of fit (cf. Baroni and Evert 2014: 10).

¹¹ Changing the model parameters or using a Generalized Inverse Gauss-Poisson model instead of an fZM model yields similarly unsatisfactory results.

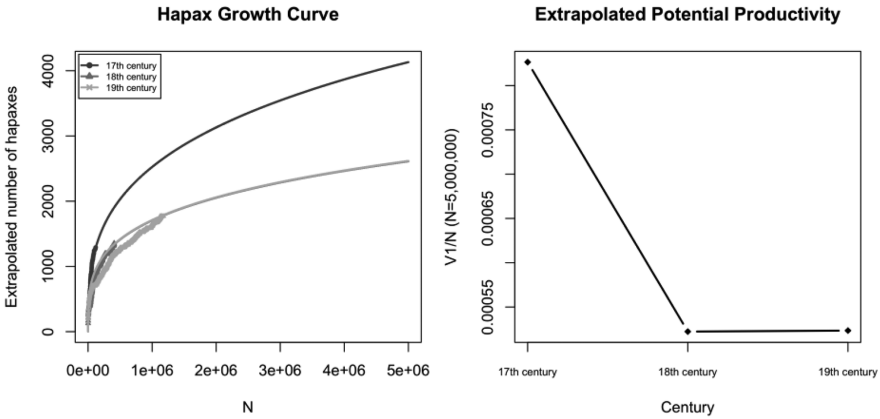


Figure 4: Left: Number of hapaxes, extrapolated using a finite Zipf-Mandelbrot model. Right: Extrapolated potential productivity for $N = 5,000,000$. The points/triangles/crosses show the observed values.

bootstrapping approach was used. For each of the three periods, 100,000 attestations were randomly sampled, and a Zipf-Mandelbrot model was fit to the data. This procedure was repeated 100 times.¹² The left panel of Figure 5

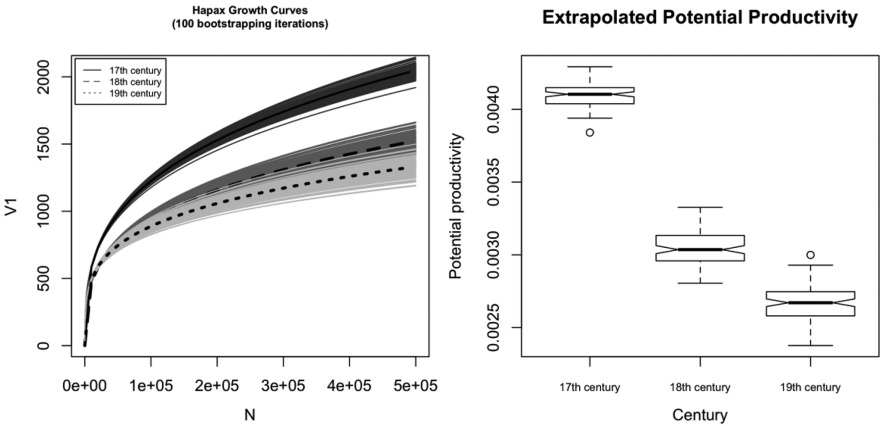


Figure 5: Left panel: Hapax growth curves for random samples of 100,000 tokens per century. Right panel: Extrapolated productivity for $N = 500,000$.

¹² In order to keep the operation computationally simpler, the sample models were fit for a maximum N of 500,000 (as opposed to 5,000,000 in the case of the fZM model fit to the full dataset).

shows the results. What appears, in the plot, as a thick dark-grey line consists of 100 individual lines that represent the fZM models fit to the seventeenth century data. The same goes for the (partly overlapping) areas that appear in somewhat lighter shades of grey which represent the 18th and nineteenth century data, respectively. The black lines represent the average growth curve, obtained by calculating the mean V_I for each N . The right panel shows the extrapolated potential productivity for an arbitrary value of $N = 500,000$.

Note that the majority of models fit to the random samples suggest a steady decline in potential productivity, which would indicate that the trend towards a decline in productivity observed in earlier studies continues in the time span covered by the DTA corpus. Despite the different productivity estimates across different models, a strong decrease in productivity from the seventeenth to the eighteenth century proves a highly robust finding that is also supported by earlier studies on the basis of the GerManC corpus (Hartmann 2016). As for the development from the eighteenth to the nineteenth century, different models arrive at different estimates. While in some models, the decrease continues, others estimate a higher potential productivity for the nineteenth than for the eighteenth century, which is also what we find in the *DTAbaby* data (see Figure 3 above). This might suggest that some of the hapaxes in the later periods might actually be genuine neologisms, while others may be words that are falling out of use (and some are of course perfectly common and only identified as hapaxes by chance). The increase in potential productivity in the *DTAbaby* data might also reflect a well-documented boost in loanwords (so-called *humanistischer Entlehnungsschub*, see e.g. Schmidt 2007 and Hartmann 2018), many of which could be used as bases of *ung*-nominals.

In order to investigate the status of the hapaxes in more detail, a more qualitatively-oriented method proposed by Kempf (2016) and Kempf and Hartmann (forthc.) can be used. Given the availability of fairly large amounts of language data in resources like the German Text Archive, but also GoogleBooks and the Google n-gram database (the latter being a subset of the former, see Michel et al. 2011), we might be able to arrive at a more accurate estimate of a derivative's age by looking up its first attestation in different sources (including e.g. etymological dictionaries). This “comparative dating” method – so called because different resources are taken into account and different dates of first attestation are compared – still entails various problems, as an aggregation of different available resources will necessarily be skewed towards later periods. Applied to hapaxes, however, it might still offer a helpful way of estimating how many of them are actually new coinages.

For the present study, a small sample of 50 hapaxes per century was taken from the *DTAbaby* data, and their first attestation was looked up both in the larger DTA (as of July 2017, i.e. including the newly added texts that are not yet

incorporated in the DTA data used for the analyses reported on above) and in GoogleBooks. Each hapax was annotated with the earlier of the two dates, yielding the distribution shown in Figure 6.

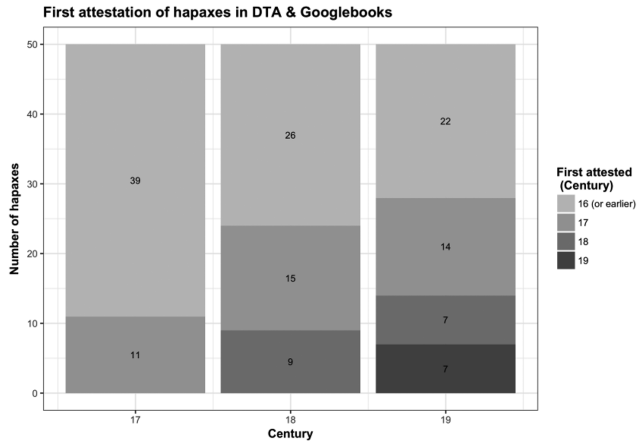


Figure 6: First attestation of DTA *baby* hapaxes in two larger databases.

At first glance, the results of the “comparative dating” analysis are quite unsurprising: While most of the hapaxes, throughout all periods, are actually attested in periods that predate the first century covered by the DTA data, later coinages successively “crawl” into the data, often with some delay. A closer look at the hapaxes in which the century of attestation in the DTA*baby* corpus and the century of first attestation according to the “comparative dating” method coincide (Table 2) suggests that the earlier data contain more genuine “one-offs”, while the later periods boast a number of new formations that prove highly successful and continue to be used to this day, e.g. *Dosierung* ‘dosage’ or *Periodisierung* ‘periodization’, both coined from loan verbs. Also, *Abgrenzung* ‘demarcation, delimitation’ is among the most frequent present-day *ung*-nominals and might have originated in the context of (partly successful) attempts to replace loan words (like Latin *definitio*, which is equivalent to German *Abgrenzung*) by native German formations.

Also note that many of the new formations are based on (phrasal) compound verbs, which is in line with the observation that diachronically, *ung*-nominalization shows a strong tendency towards complex bases (cf. e.g. Demske 2000). From a construction grammar perspective, we can assume that the preference for complex bases has become part of language users’ knowledge about the [V-*ung*] construction (cf. Kempf and Hartmann forthc.).

Table 2: Hapaxes whose first occurrence in the DTA baby corpus coincides with their first occurrence in GoogleBooks (GB) and/or the complete German Text Archive (DTA). The rightmost column shows their last attestation in the aggregated reference and newspaper corpora available at www.dwds.de. (GoogleBooks was not used for querying the last attestations because the database might contain reprints that suggest a much later dating than the actual date when the text in question was produced.).

Lemma	Century	First attestation	Source	Last attestation
Absendigung	17	1641	DTA	1641
Abstrahierung	17	1659	GB	2010
Abtragung	17	1628	DTA	2016
Aufbehaltung	17	1629	DTA	1846
Auszahlung	17	1601	GB	2016
Durchhechelung	17	1663	DTA	1919
Gemeinmachung	17	1616	GB	2012
Gutwartung	17	1669	both	1669
Jahresfristung	17	1679	DTA	1679
Profitierung	17	1653	DTA	2001
Schränkung	17	1639	DTA	1981
Abhelfung	18	1719	GB	1895
Ausstopfung	18	1750	DTA	2013
Befehlshabung	18	1786	DTA	2003
Dosierung	18	1706	GB	2016
Hinaustreibung	18	1705	GB	1705
Hinlenkung	18	1773	GB	1998
Solemnisierung	18	1798	both	1798
Übersuchung	18	1775	DTA	1828
Überwallung	18	1773	DTA	2010
Abgrenzung	19	1884	GB	2016
Buchführung	19	1809	DTA	2016
Neuschöpfung	19	1855	DTA	2016
Periodisierung	19	1843	GB	2016
Traversierung	19	1882	GB	1994
Überkünstelung	19	1810	DTA	1882
Überwächung	19	1894	DTA	1894

In sum, while the pattern still shows some activity in later periods, the analyses reported on above suggest that the seventeenth century was the decisive period for the gradual decline of the word-formation pattern under investigation. While the loss in productivity can consequently be considered a well-established fact, its explanation requires a closer look at the corpus data. Perhaps most importantly, the interaction between *ung*-nominals and the constructions in which they appear has to be taken into account.

4.2 Interaction patterns

It is a key assumption of CxG that constructions at various levels of abstraction interact with each other. Consider Taylor's (Taylor 2012: 158f.) example of the English adjective *unmitigated*, which appears to be strongly associated to a "bad event" reading, even though it can, in principle, also occur with nouns like *joy* or *success*. Its most frequent collocate by far, however, is *disaster*. Similarly, Stefanowitsch and Gries (2003: 218f.) explain the negative connotation of what they term the [N *waiting to happen*] construction with the fact that the noun *accident* and similar words with negative connotation occur significantly more often in the noun slot than would be expected by chance. In the domain of morphology, it has been shown that affixes that used to be semantically neutral can become pejorative as they are frequently combined with negatively-connotated bases (Dammel 2011; Kempf 2016). In these cases, the interaction between words and more abstract constructions, viz. morphological or syntactic constructions, has a strong impact on the diachronic development and synchronic interpretation of the more abstract patterns. In the case of the [N *waiting to happen*] construction, for instance, an utterance like *There's a wedding waiting to happen*, if not used in an ironic sense, will most likely be interpreted as expressing a negative stance towards the wedding in question.

In the case of derivational morphology, these observations suggest that thoroughly investigating the constructions in which word-formation products appear can be highly informative in understanding the respective word-formation pattern. For *ung*-nominalization, several interaction patterns have been discussed in the previous literature (especially Demske 2000) which, taken together, suggest that the construction becomes more "nouny" over time. For instance, Demske (2000) observes that in the ENHG period, nominals in *-ung* are often used as prepositional complements without a determiner, yielding very "verb-like" structures, many of which would be ungrammatical in present-day German. Hartmann (2016) shows that the frequency of nominals in *-ung* used as prepositional complements in this way decreases drastically from the ENHG to the NHG period.

This [P NOM] construction, which also occurs with other nominalizations such as nominalized infinitives as in *nach Verlangen* 'upon demand' or implicit derivations as in *aus Trieb* 'from (the) drive' (cf. Hartmann 2016), is remarkable in that it usually evokes a highly processual construal that is fairly untypical for *ung*-nominals in present-day German.¹³ Consider the following examples:

¹³ By "processual / progressive construal", I mean that the context suggests an interpretation of the event unfolding over time, rather than referring to the event as a whole, with its

- (2) allermaffen der gũnftige Lefer **in durchlefung** gedachter Reyfe [...] zu erfehen haben wird.
 ‘as the interested reader will see **in reading** the aforementioned [lit. thought-of] journey.’
 (Hulsius, Schifffahrt, 1649, DTababy)
- (3) wie etwa David **bei heimführung** der bundeslade mit aller macht tantzete
 ‘as David was dancing vigorously [lit. with all might] **when taking home** the Ark of the Covenant.’
 (Geier, Heinrich Schütz, 1672, DTababy)

The [P NOM] construction – whose instances in the corpus were identified following the formal criteria defined by Hartmann (2016: 196–198)¹⁴ – can encode simultaneity (as in the case of *in* and *bei* in the examples above), anteriority (if a preposition like *vor* ‘before’ occurs in the P slot), posteriority (e.g. in the case of *nach* ‘after’), or causality (in the case of *wegen* ‘because of/ due to’). All these readings require a fairly “verb-like” construal of the nominal. As Vogel (2000: 269) points out, “deverbal nouns constitute a continuum with regard to ‘nouniness’, which is why some are more ‘nouny’ and others are more ‘verby’.” Adopting Croft’s (2001: 87) universal-typological parts of speech model, German nouns can be seen as prototypically denoting objects, which are defined as non-relational, stative, and non-transitory, whereas German verbs prototypically denote processes, defined as relational, processual, and transitory. The *ung*-nominals occurring in the [P NOM] construction tend to refer to dynamic events which are processual (rather than stative) and transitory (rather than permanent). In addition, they tend to make the relationality exhibited by the base verb transparent by encoding what would be the (obligatory) direct object of the verb in a genitival complement, e.g. *in durchlefung gedachter Reyfe* in (3) above. This has already been shown for data from the GerManC corpus (cf. Hartmann 2016). In the DTababy data, *ung*-nominals in a [P NOM]

boundaries “in view”, as is arguably the case for present-day German *Lesung* ‘(public) reading’ or *Vorlesung* ‘lecture’. For a more in-depth discussion of the aspectual construal of *ung*-nominals, cf. e.g. Ehrich and Rapp (2000), Hartmann (2014, 2016).

14 Specifically, these criteria are: a) The PP is an independent element that is, in principle, omissible. For instance, the sentence in (3) would still be grammatical if *in durchlefung gedachter Reyfe* would be left out. b) The *ung*-nominal must be in the singular and must not be accompanied by a determiner. c) If it is clear from the context that the *ung*-nominal is lexicalized and shows an object reading, the PP is not considered an instance of the [P NOM] construction, e.g. *in königlicher Kleidung* ‘in royal clothing’ (Opitz, Barclayens Argenis, 1626, DTababy).

construction also occur with a genitival complement significantly more often than other *ung*-nominals ($\chi^2(1) = 642.8$, $p < 0.001$), even though the effect size is rather small ($\phi = 0.24$).

The observation that this construction seems to have fallen out of use, apart from a few relics that preferentially appear in more formal text types (cf. Demske 2000: 397), ties in with the hypothesis that *ung*-nominals tend to become “nounier” over time. The DT*Ababy* data also bear witness to the decline of this construction, which becomes significantly less frequent (Kendall’s $\tau = -1$, $T = 0$, $p < 0.01$), as shown in Figure 6.

Another tendency that can be interpreted in terms of increasing “nouniness” is the observation that *ung*-nominals tend to occur ever more frequently with a determiner. Langacker (e.g. 1991, 2004) describes determiners as grounding elements, relating a discourse referent to the speech event, its participants, and the immediate context. More specifically, grounding pertains to whether an instance of a type, or a set of instances, is uniquely apparent to both the speaker and the hearer (cf. Langacker 1991: 53). Note that nominal grounding – at least to a certain degree – always entails a count-noun construal (cf. Vogel 1996: 131; Smirnova and Mortelmans 2010: 113). Vogel (1996: 115) mentions the example of *beauty*, which refers to an abstract property, while *a beauty* refers to one specific entity characterized by that property.¹⁵ While the term is more mass-noun-like if referring to a property, its meaning shifts towards a much more concrete reading if a determiner is used. Likewise, in the case of *ung*-nominals, a great variety of different meaning variants can be distinguished, some of which are much more nuanced than the distinction between *beauty* and *a beauty*. For instance, *Dämmerung* ‘dusk/dawn/twilight’ in (5) describes an abstract state. Taking up Croft’s aforementioned properties characterizing prototypical objects and events, it describes something that is stative and non-transitory. The state described here can be interpreted as inherently relational in that the phrase *tiefe Dämmerung* ‘deep twilight’ ascribes a property to a specific setting/location (*Draußen auf der Straße* ‘Outside on the street’). It is used for *predication*, i.e. specifying “what the speaker intends to *say about* what he is talking about (the referent)” (Croft 1991: 52; emphasis original). This function is traditionally associated with verbs (cf. *ibid.*), and tellingly, (5) could easily be paraphrased using a verb: *Draußen auf der Straße dämmerte es*. In (6), by contrast, *Dämmerung* appears in a more reified reading variant, referring to one specific instance of

¹⁵ With a definite determiner, *beauty* of course also refers to an abstract property. However, *the beauty (of sth.)* of course singles out one specific instance of this property and can therefore also be considered more concrete and more count-noun-like than its counterpart without any determiner.

a type (cf. Langacker 2004). Even though the term is used metaphorically here, referring to a pre-enlightenment way of thinking that allegedly mixes up fact and fiction, it can thus be considered more count-noun-like and hence more concrete than its counterpart in (5).

- (4) Draußen auf der Straße war es tiefe **Dämmerung**.
 ‘Outside on the street, there was deep **twilight**.’
 (Storm, Immensee, 1852, DTA**aby**)
- (5) Wir find aus diefer **Dämmerung** hinaus, und wollen durchaus Mährchen als Mährchen, Gefchichte als Gefchichte lesen.
 ‘We are out of this **twilight**, and we are willing to read fairy tales as fairly tales, history as history.’
 (Herder, Humanität, 1796, DTA**aby**)

Similar considerations apply to other *ung*-nominals as well. It has already been shown that the relative frequency of *ung*-nominals with a determiner increases significantly from the sixteenth to the eighteenth century (Hartmann 2016). The DTA**aby** data confirm this observation and show that the use of *ung*-nominals with determiners – which, following Thielmann’s (2007) definition, include definite and indefinite articles, demonstratives, prenominal genitives, and personal pronouns – still experiences a slight increase throughout the nineteenth century. Overall, the increase in the DTA**aby** data, shown in the left panel of Figure 7, proves to be slight, but steady (Kendall’s $\tau = 1$, $T = 15$, $p < 0.01$). The development is a bit less clearly unidirectional in the case of pluralization, which can be considered another diagnostic for the shift towards a more “nouny” construal of *ung*-nominals (cf. e.g. Vogel 1996: 115; Demske 2000; Hartmann 2016). As Langacker (2008: 130) points out, “[o]nly a count noun can be pluralized.” Thus, pluralization by necessity entails a reading that conceptualizes the referent as a group of individuated entities rather than a mass. This goes for concrete mass nouns like *wine* (*I tasted two wines, one from France and one from South Africa*) but also for abstract nouns like the *ung*-nominal *Hoffnung* ‘hope’, which, in (7), refers to several distinct things the country has been hoping for.

- (6) die schönsten **Hoffnungen** des Landes waren erfüllt
 ‘the most beautiful **hopes** of the country were fulfilled’
 (Pahl, Geheimnisse, 1799, DTA**aby**)

As the right panel of Figure 8 shows, the use of pluralized *ung*-nominals experiences a steep increase from the early seventeenth to the mid-eighteenth

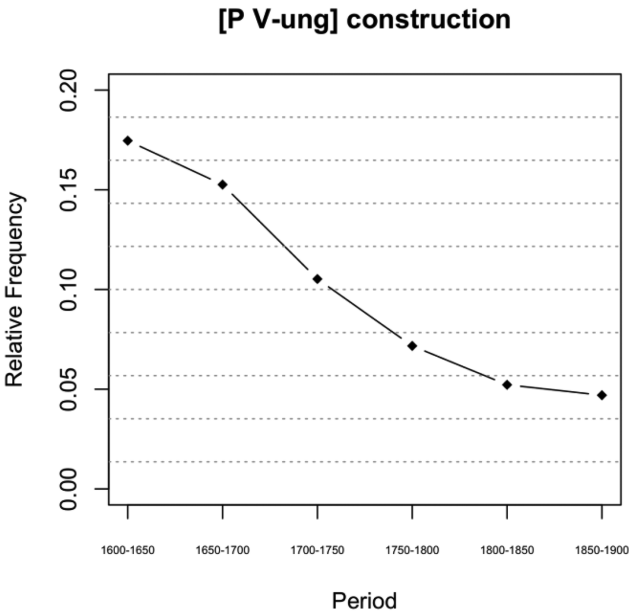


Figure 7: Relative frequency of *ung*-nominals in [PREP NOM] constructions in relation to the total number of *ung*-nominals in the respective corpus period.

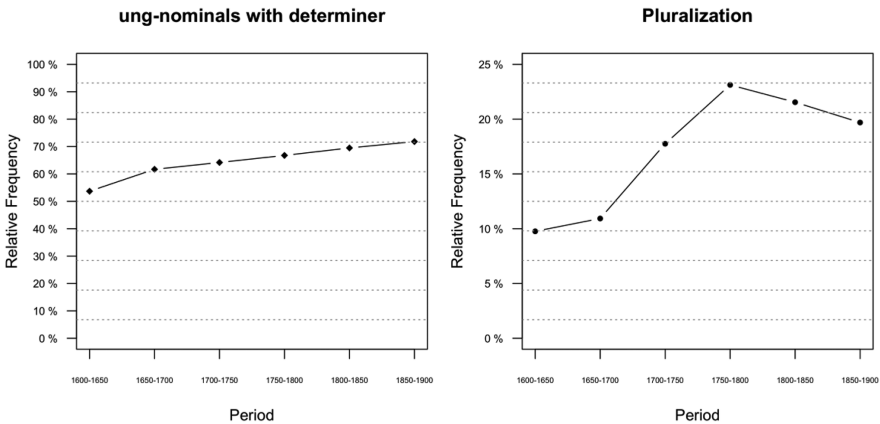


Figure 8: Relative frequency of *ung*-nominals with a determiner (left panel) and of pluralized *ung*-nominals (right panel) in relation to the total number of *ung*-nominals in the respective corpus period.

century. During the nineteenth century, the relative frequency of pluralized *ung*-nominals drops slightly. All in all, however, the development of pluralization is in line with previous observations as well, and like the increase in the use of determiners, it lends further support to the idea of an increasing “nouniness” of *ung*-nominals.

Another aspect that has been discussed in connection with the “nouniness” or “verbiness” of *ung*-nominals and other nominalizations is the use of genitival complements. According to Vogel (1996: 130), genitival complements are characteristic of “nominalizations proper”, as they can incorporate the subject or – perhaps more importantly – the object of the verb phrase, e.g. *x zerstört die Stadt* ‘x destroys the city’ > *die Zerstörung der Stadt* ‘the destruction of the city’. Taking up Croft’s aforementioned properties of prototypical nouns and verbs once again, one could say that in such cases, the relationality of the verb carries over to the nominal (cf. Knobloch 2002: 353). Of course, not all genitives encode a subject or object argument. Especially in the case of possessive pronouns (which are functionally equivalent to prenominal genitives), the genitive frequently has to be interpreted in a possessive sense. However, in the DTA*baby* data, even possessive pronouns often encode the subject argument of the corresponding verb phrase, as in (8), or its direct object, as in (9).

- (7) [sie] hielten ihm auch **feine** nachlässige **Regierung** für
 ‘they also held **his** negligent **governance** against him’
 (Lohenstein, Ibrahim Sultan, 1673, DTA *baby*)
- (8) du wirft/durch **ihre Belobung**/dich bey uns belobt machen.
 ‘you will, by **her prasing** [i.e. by praising her], make yourself praised by us.’
 (von Birken, Gesprächspiel-Gesellschaft, 1815, DTA *baby*)

In many contexts, however, the different types of genitive cannot be easily told apart. For example, (9) is ambiguous between a so-called *genitivus subiectivus* and a *genitivus objectivus*, and the larger context has to be taken into account to determine that the possessive pronoun encodes the object, not the subject, of the corresponding verb phrase (as the addressee is about to sing a song praising a beautiful woman). The use of genitives has therefore been annotated in purely formal terms in the DTA*baby* data, i.e. distinguishing between prenominal and postnominal usages (and a few instances in which both a prenominal and a postnominal genitive are used). Figure 9 shows that the use of genitival complements with *ung*-nominals decreases significantly in the DTA *baby* period (Kendall’s $\tau = -0.87$, $T = 1$, $p < 0.05$).

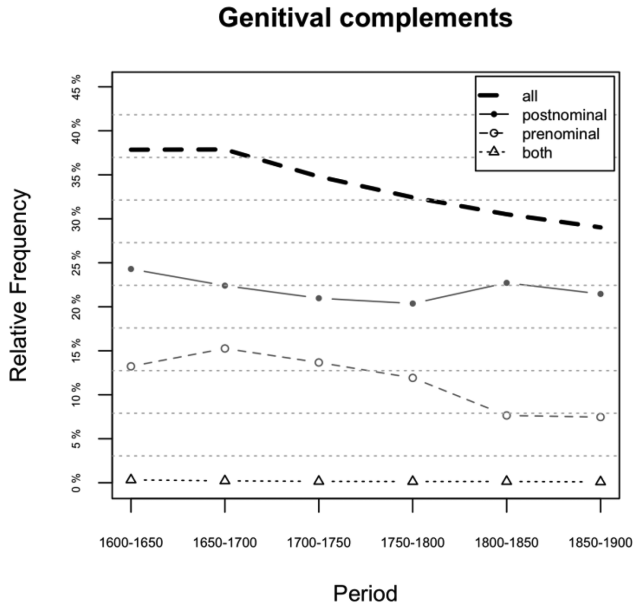


Figure 9: Relative frequency of *ung*-nominals with a genitival complement in relation to the total number of *ung*-nominals in the respective corpus period. “both” represents the few cases in which an *ung*-nominal occurs with both a prenominal and a postnominal genitive. The curve labeled “all” represents the sum of the three other curves.

Unsurprisingly, prenominal genitives suffer the most significant decrease – in fact, the downward curve would be much steeper without the data for possessive pronouns and proper names. Throughout the Early New High German period and the early stages of the New High German period, the default position of the genitive in German has gradually shifted from prenominal to postnominal (cf. Nübling et al. 2013: 107–110). In present-day German, the prenominal genitive is largely restricted to proper names and possessive pronouns. The declining use of *ung*-nominals with a prenominal genitive is not made up for by an increasing use of postnominal genitives, however, even though the relative frequency of postnominal genitives increases slightly in the mid-eighteenth century. All in all, the use of genitives decreases, which points to an increase in “non-relational” uses of *ung*-formations.

However, it should be pointed out that the “relationality” expressed with the help of genitival complements can also be expressed differently: The subject or object of the verb can also be expressed as first constituent of a compound. (10) shows how the patient argument of the nominal *Einpflropfung* ‘grafting’ is

expressed once in a genitival complement and once as first constituent of a compound.

- (9) Auch ist es mir unbegreiflich, wie Haller die sogenannte **Einfropfung der Blutgefäße des Hühnchens** in die Gefäße des Dotters so allgemein, und als ungereimt verwerfen konnte, da doch Er selbst bey der Verbindung des menschlichen Eyes mit der schwangern Gebärmutter, eine vollkommen ähnliche **Gefäßeinpfropfung** vertheidigte.

‘Also, I cannot understand how Haller can reject the so-called **grafting of the chicken’s blood vessels** in the vessels of the yolk so generally as inconsistent, given that he himself defended a perfectly similar **grafting of vessels** regarding the connection of the human egg with the pregnant uterus.’

(Bodmer, Sammlung, 1742)

Arguably, however, the N + N compound entails a less processual construal than the use of the genitival complement, partly as it seems to summarize the event as a whole to a higher degree than adding the patient in the genitive. Also, note that the syntactic option allows for a more detailed description of the process: Although a massive compound like *die Blutgefäße-des-Hühnchens-in-die-Gefäße-des-Dotters-Einpfropfung* ‘the-chicken’s-blood-vessels-into-the-yolk-grafting’ would not be entirely impossible, it is highly unlikely to be actually used, be it in present-day German or in earlier stages of (New High) German.

In sum, the DTAbaby data are in line with the assumption that word-formation products in *-ung* become “nounier” over time, which is reflected in a variety of changing usage patterns. If we assume an interaction between constructions and their instantiations, it seems reasonable to expect that the changing use of *ung*-nominals, i.e. concrete instances of the pattern, also has ramifications for the word-formation pattern itself. Thus, for understanding the word-formation pattern, it seems pivotal to take a closer look at these interaction patterns. The next section deals with the question of how they can be approached empirically.

4.3 Collostructional analysis as a diagnostic for lexical strength

While the frequency and productivity measures reported on above can give insights into the development of the pattern as a whole, they do not take the relationship between the construction and the actual constructs, i.e. its

instantiations, into account. However, in order to grasp the semantics of a word-formation pattern, it is essential to take a look at the bases it takes. To this end, *collostructional analysis* can be used. Collostructional analysis has become a cover term for a family of cross-tabulation methods (cf. Stefanowitsch 2013 for a recent overview). In order to investigate the relationship between word-class changing derivation patterns and their bases, morphological cross-tabulation analysis (Hartmann 2014c) can be used, which adapts the basic logic of simple collexeme analysis (Stefanowitsch and Gries 2003). Four values are needed as input for morphological cross-tabulation analysis: a) the token frequency of a specific instance w of a word-formation pattern W ; b) the frequency of its base b in its original word class B ; c) the frequency of all other instances of word-formation pattern W ; d) all other instances of word class B . Then, a cross-tabulation test is performed over this 2×2 matrix. While the Fisher Exact test is the most widely-used method for this purpose (cf. Flach 2015: 241), I follow e.g. Wulff (2006), Ebensgaard-Jensen (2013), and Flach (2015) in using the log-likelihood ratio score instead (Dunning 1993). This allows for a more fine-grained ranking (cf. Ebensgaard-Jensen 2013: 167) and avoids infinite or 0 values in the case of large samples (cf. Wulff 2006; Flach 2015).¹⁶ In order to allow for a more intuitive interpretation, the sign was set to reflect the direction of association by negativizing the log-likelihood ratio values when the observed frequency is smaller than the expected one (cf. Hilpert 2006: 247; Levshina 2015: 232). The analysis was performed using R (R Core Team 2015) and snippets of code from Gries (2007) and Flach (2016).

Table 3 reports the results of three morphological cross-tabulation analyses on the basis of the *DTA_{baby}* data for *ung*-nominalization (one for each century). For each of the three time slices, the top 20 attracted and repelled items are reported.

The results reported in Table 3 are very much in line with those obtained in a morphological cross-tabulation analysis on the basis of data from the GerManC corpus (cf. Hartmann 2014c). A main finding of that study was that strongly lexicalized *ung*-nominals tend to be identified as significantly attracted to the word-formation pattern. Ad-hoc formations and derivatives that are about to fall out of use at the time covered by the corpus, by contrast, tend to be identified as repelled collexemes. The same goes for the *DTA_{baby}* data, even though the results of morphological cross-tabulation analysis have to be regarded with some caution as the high frequency of individual *ung*-nominals or the low frequency of their corresponding base verbs (or vice versa) might in some cases be corpus artefacts. For example, *Entscheidung* ‘decision’ is significantly

16 For a comparison of different association measures, see Wiechmann (2008).

Table 3: Results of morphological cross-tabulation analysis for *ung*-nominalization in the DTAbaby corpus. For each century, only the top 20 attracted and repelled collexemes are displayed.

Attracted				Repelled			
Lemma	Freq. (obs.)	Freq. (exp.)	LogLik	Lemma	Freq. (obs.)	Freq. (exp.)	log ₁₀ (p)
Seventeenth century							
1 Ordnung ‘order’	107	9.3	437.78	Haltung ‘holding/ attitude/posture’	2	29.88	−47.22
2 Meinung ‘opinion’	88	10.56	272.1	Bringung ‘bringing’	1	20.52	−34.51
3 Hoffnung ‘hope’	55	6.31	175.92	Nennung ‘mentioning’	1	16.6	−26.75
4 Regierung ‘government’	32	3.78	99.33	Schreibung ‘writing’	1	12.28	−18.35
5 Wirkung ‘effect’	35	5.25	87.93	Ziehung ‘drawing’	1	12.15	−18.09
6 Verwaltung ‘administration’	16	1.99	47.37	Führung ‘leading’	1	10.89	−15.69
7 Veränderung ‘change’	15	1.86	44.58	Erkennung ‘recognition’	1	10.69	−15.31
8 Anmerkung ‘remark’	12	1.2	43.07	Legung ‘laying’	1	9.43	−12.94
9 Anfechtung ‘challenge/ appeal’	11	1	42.88	Weisung ‘instruction’	1	7.97	−10.24
10 Erfindung ‘invention’	21	3.92	42.48	Ehrung ‘honor’	1	5.18	−5.31
11 Erklärung ‘explanation’	20	3.85	39.13	Stellung ‘position’	1	5.11	−5.2
12 Anleitung ‘instruction’	8	0.6	37.28	Erweisung ‘bestowing’	1	4.85	−4.76
13 Verwunderung ‘wonderment’	16	2.59	37.25	Annehmung ‘acceptance’	1	4.45	−4.11
14 Wohnung ‘flat’	21	4.78	34.15	Ansehung ‘regard’	2	6.18	−4.05
15 Verfassung ‘constitution’	10	1.06	33.94	Erlangung ‘gaining/ obtaining’	1	4.32	−3.89
16 Rüstung ‘armor’	8	0.66	33.69	Richtung ‘direction’	1	4.18	−3.68
17 Rechnung ‘calculation/bill’	16	2.99	32.3	Lesung ‘reading’	2	5.78	−3.49
18 Kleidung ‘clothing’	11	1.39	32.03	Begebung ‘incident’	1	3.78	−3.06

(continued)

Table 3: (continued)

	Lemma	Attracted			Lemma	Repelled		
		Freq. (obs.)	Freq. (exp.)	LogLik		Freq. (obs.)	Freq. (exp.)	log ₁₀ (p)
19	Beschreibung 'description'	21	5.58	28.23	Befindung 'finding'	2	5.44	-3.04
20	Auferziehung 'education'	10	1.33	27.93	Begreifung 'grasping'	1	3.72	-2.95
Eighteenth century								
1	Handlung 'action/shop'	98	18.88	225.67	Findung 'finding'	1	45.44	-87.71
2	Wirkung 'effect'	93	17.54	219.57	Haltung 'holding/ posture/attitude'	2	37.52	-64.23
3	Ordnung 'order'	67	9.87	212.26	Zeigung 'showing'	1	20.59	-35.74
4	Gleichung 'equation'	59	8.89	181.54	Führung 'guidance'	4	17.79	-17.14
5	Meinung 'opinion'	84	17.67	173.67	Bestehung 'standing (trans.)'	1	11.21	-16.86
6	Gattung 'species'	28	3.53	109.65	Erkennung 'recognition'	1	9.02	-12.61
7	Veränderung 'change'	50	10.36	104.97	Lesung 'reading'	2	10.6	-11.49
8	Bewegung 'movement'	46	8.89	104.96	Unterscheidung 'distinction'	1	7.68	-10.07
9	Sammlung 'collection'	38	6.58	98.85	Entstehung 'evolution'	3	11.21	-9.34
10	Neigung 'inclination'	35	5.97	92.65	Tretung 'kicking'	1	7.07	-8.93
11	Schöpfung 'creation'	25	3.29	91.68	Bekennung 'confession'	1	5.97	-6.93
12	Erfahrung 'experience'	45	10.36	82.72	Verlassung 'leaving'	1	5.73	-6.49
13	Rechnung 'calculation/bill'	38	8.28	74.76	Befindung 'determining'	1	5.6	-6.27
14	Hoffnung 'hope'	44	11.45	68.68	Teilung 'division'	3	8.89	-5.81
15	Verfassung 'constitution'	24	3.9	67.27	Begegnung 'encounter'	1	5	-5.21
16	Anmerkung 'remark'	30	5.97	66	Stellung 'position'	7	13.89	-4.67
17	Empfindung 'sentiment'	46	12.79	65.43	Anführung 'leading'	1	4.39	-4.17
18	Verzweiflung 'despair'	15	1.95	55.99	Erweckung 'awakening'	1	3.78	-3.17

(continued)

Table 3: (continued)

Attracted				Repelled			
Lemma	Freq. (obs.)	Freq. (exp.)	LogLik	Lemma	Freq. (obs.)	Freq. (exp.)	log ₁₀ (p)
19 Ausschweifung 'digression'	14	1.83	51.91	Verderbung 'tainting'	1	3.66	-2.98
20 Übersetzung 'translation'	25	5.24	51.60	Bezahlung 'payment'	2	4.87	-2.42
Nineteenth century							
1 Bewegung 'movement'	125	34.44	228.48	Findung 'finding'	1	76.15	-160.84
2 Bedingung 'condition'	63	13.49	181.77	Sonderung 'separation'	1	64.32	-134.3
3 Beziehung 'relationship'	74	19.71	142.81	Führung 'guidance'	3	37.35	-61.51
4 Untersuchung 'examination'	79	22.41	137.02	Haltung 'holding/ posture/attitude'	12	52.7	-53.99
5 Ordnung 'order'	63	16.18	129.25	Erhaltung 'preservation'	6	36.1	-45.02
6 Beobachtung 'observation'	79	24.28	119.3	Erkennung 'recognition'	1	18.67	-33.62
7 Verhandlung 'negotiation'	38	8.09	110.88	Sitzung 'session'	3	21.37	-28.89
8 Wirkung 'effect'	81	27.18	104.47	Gewinnung 'winning/ extraction'	1	14.52	-24.8
9 Dichtung 'poetry'	35	7.47	101.58	Annehmung 'acceptance'	2	17.01	-24.75
10 Regierung 'government'	38	8.51	99.65	Ergebung 'submission'	1	13.9	-23.49
11 Entwicklung 'development'	66	21.16	92.34	Schaffung 'creation'	1	12.86	-21.31
12 Bedeutung 'meaning'	62	19.71	88.05	Hebung 'holding/ lifting'	2	13.69	-18.16
13 Vorstellung 'imagination'	46	12.45	86.25	Beweisung 'proof'	1	10.17	-15.73
14 Anwendung 'application'	57	18.26	79.84	Bewahrung 'preservation'	1	9.13	-13.61
15 Verbindung 'connection'	79	30.29	78.76	Erreichung 'reaching'	5	14.73	-10.27
16 Entscheidung 'decision'	53	16.6	77.26	Aufnehmung 'admitting'	1	6.43	-8.27

(continued)

Table 3: (continued)

Attracted				Repelled			
Lemma	Freq. (obs.)	Freq. (exp.)	LogLik	Lemma	Freq. (obs.)	Freq. (exp.)	log ₁₀ (p)
17 Richtung 'direction'	69	25.73	72.71	Verschließung 'locking'	1	6.02	-7.47
18 Einbiegung 'bend'	64	23.45	69.8	Empfehlung 'recommendation'	3	9.13	-6.61
19 Lösung 'solution'	47	14.73	68.38	Versuchung 'temptation'	2	7.05	-5.97
20 Empfindung 'sensation'	50	60.81	64.14	Geltung 'validity'	11	19.5	-5.36

more frequent than its base verb in the nineteenth century. However, this is mainly due to its high frequency in one legal text, the “Civilprozessordnung” (‘civil procedure rules’) of the city of Berlin (1877). Conversely, the observation that *gewinnen* ‘win’ (also: ‘extract’) is significantly more frequent than the corresponding *ung*-nominal in the nineteenth century can partly be attributed to the fact that the DTAbaby corpus contains a chess manual (“Lehrbuch des Schachspiels”, Halle 1856), in which *gewinnen*, for obvious reasons, occurs more frequently than in any other text in the corpus. This entails that if we can interpret the results of morphological cross-tabulation analysis in terms of the lexical strength of the items under investigation, we have to keep some potential confounding factors in mind.

Despite such skews in the distribution of the individual lexemes that call for caution in the interpretation of the cross-tabulation results, this method can provide interesting clues regarding the lexical strength of the individual nominals. Lexical strength refers to the degree of independence between a word-formation product and the schema from which it is derived. According to Bybee (1995: 428),

[w]ords with high lexical strength are easy to access, serve as the bases of morphological relations and exhibit an autonomy that makes them resistant to change and prone to semantic independence.

Bybee (*ibid.*) sees token frequency as the main determinant of lexical strength. Conversely, she sees the strength of a schema as “based directly on its type frequency” (Bybee 1995: 430). As Taylor (2002: 277) puts it,

- a. High token frequency of an instance entrenches the instance and weakens (or at least, does not strengthen) the schema.
- b. High type frequency of the instances entrenches the schema and weakens (or at least does not strengthen) the instances.

These ideas are highly compatible with the conceptualization of grammar as “a dynamic system of emergent categories and flexible constraints” (Diessel 2015: 296), which is constitutive of most strands of construction grammar as well as other usage-based approaches. In CxG terms, *ung*-nominalization can be conceived of as a construction which is connected to its instances by means of so-called inheritance links (cf. Goldberg 1995; Hilpert 2014). As soon as a new instance is coined, however, it is subject to *constructionalization* (Traugott and Trousdale 2013). This assumption follows straightforwardly from a usage-based view which holds that “[e]very usage event has some impact (even if very minor) on the structure of the categories it invokes.” (Langacker 1987: 376) As soon as a word-formation product is used, it has the potential to be stored as a sign in its own right – hence: an independent form-meaning pair, a construction.¹⁷ Extending the network metaphor that is key to usage-based CxG (cf. Diessel 2015), one could say that in such cases, the link between the instance and its superordinate construction becomes weaker. Morphological cross-tabulation analysis, then, provides a tool for quantitatively assessing important aspects of the association strength between an individual instance and the corresponding derivational pattern. However, it should be kept in mind that there are almost certainly many more factors that have an impact on the development of a word-formation product and on the degree to which it is associated with its superordinate construction (hence, strengthens the corresponding schema). Before discussing the limitations of the method in more detail, however, let us take a closer look at the results in Table 3 and discuss what they *can* tell us about the development both of the individual instances and of *ung*-nominalization in general.

First of all, note that many of the most strongly attracted or repelled collexemes recur across all three time periods. However, an interesting pattern of change can be detected on closer inspection of the top repelled collexemes. The seventeenth century data contain many derivatives based on simplex verbs,

¹⁷ Note that not all construction grammarians agree that words can be considered constructions. For instance, Stefanowitsch (2009: 569) reserves the term “construction” for complex grammatical patterns with non-compositional properties, while Dąbrowska (2009: 217) argues that only relational words such as verbs are to be considered constructions. I follow Croft (2001: 17) in regarding constructions as a uniform representation of all kinds of structures from lexical items to highly abstract syntactic patterns.

e.g. *Bringung* ‘bringing’, *Legung* ‘laying’, *Lesung* ‘reading’. (10) exemplifies the use of *Legung* in a context where a nominalized infinitive would be preferred in present-day German, even though *Legung* is occasionally still used in similar contexts.¹⁸ *Lesung*, by contrast, is fairly frequent in present-day German but can only be used in the lexicalized meaning of ‘public reading, reading event’. In the DTA*baby* data, however, the “verby” processual reading exemplified in (11) is the default use of the nominal. Thus, the fact that these items are identified as repelled might bear witness to their falling out of use (even though *Lesung* is “rescued” later on in the lexicalized reading).

- (10) Man hat aber in **legung** deß Oftertags mit keinem andern Voll Mond zu thun/als mit dem jenigen/der in den Gleichtag fellet ‘In **determining** the day of Easter, one has to do with no other full moon than the one that coincides with the equinox’ (Crüger, Cupediæ Astrosophicæ, 1631, DTA*baby*)
- (11) Jch muß bekennen/ich habe nach **Lefung** dieses feines Buchs am erften rechtſchaffene Gelegenheit bekommen/die Sache ein wenig reiffer zu überlegen ‘I have to confess that I only had the opportunity to think a bit more carefully about the issue after **reading** [of] his book.’ (Thomasius, Außübung der Vernunft, 1691, DTA*baby*)

In the second and third time period, many more complex *ung*-nominals can be found, derived from prefixed or particle verbs (e.g. *Verlassung* ‘leaving/departure’, *Annehmung* ‘acceptance’). This might indicate that in the later periods, ad-hoc formations are more likely to be derived from complex verbs, which is in line with the observation that in present-day German, prefixed and particle verbs are more eligible as bases for *ung*-formations than simplex verbs (cf. Demske 2000: 368). Regarding the meaning of the individual word-formation products, it is striking that many of the repelled collexemes actually belong to the semantic classes that Demske (2000: 369), based on a host of previous literature, has identified as ineligible for *ung*-nominalization in present-day German, namely verbs with durative *aktionsart* (cf. *Ansehung* < *ansehen* ‘to look at’) and inchoative verbs (cf. *Erweckung* < *erwecken* ‘awaken’),

¹⁸ For instance, there is one attestation for *Legung der Brände* ‘setting the fires’ from a 2006 issue of the newspaper “Die Zeit” in the German Reference Corpus (DeReKo; Z06/MAR.00092). The overall frequency of *Legung*, however, is marginal, with only 26 hits in the DeReKo and 27 hits in the DWDS Core Corpus of the twentieth century, most of which date to the first three decades of the century.

Erreichung < *erreichen* ‘to reach’). In addition, she mentions change-of-possession verbs which do not occur among the repelled colllexemes of the DTA_{baby} data but are attested in the larger DTA corpus, e.g. *Nehmung* ‘taking’, *Verkaufung* ‘selling’. This observation corroborates the hypothesis that these constraints are just emerging during the period covered by the DTA corpus. It has to be added that these constraints are not absolute in present-day German. For example, *Erweckung* is still perfectly grammatical. All in all, however, the constraints posited by Demske seem to capture the general tendencies in present-day German quite well. In addition, it should be pointed out that the “repelled” items are not in all cases word-formation products that are falling out of use. Instead, ad-hoc formations can also be identified as repelled items. Also, the confounding factors discussed above of course apply to the repelled items as well. As for *Ansehung* and *Befindung*, for instance, which are among the repelled colllexemes in the earlier periods but do not occur in the nineteenth century data, it should be noted that they almost exclusively occur as parts of constructional idioms that are falling out of use in the period covered by the corpus. Both in the GerManC data reported on in Hartmann (2014c) and in the DTA_{baby} data, *Befindung* exclusively occurs in the constructional idiom *nach Befindung* (roughly: ‘after judging’), as in (13). Likewise, *Ansehung* is strongly tied to *in Ansehung* ‘with regard to’, which is gradually replaced by its equivalent *in Betracht*.

- (12) fo pflegt der Richter nach **Befindung** der Umftänden die Schwerdt-Straff oder nur ein Leibs-Straff aufzulegen
 ‘so the judge, after **assessing** the circumstances, inflicts punishment by sword or only corporal punishment.’
 (Mutach, Substantzlicher Vnderricht, 1709, DTA_{baby})
- (13) und ein so grosser Herr, wie unser gnädigster Herzog ist, soll sich in **Ansehung** des Hofglanzes von diesen kleinen übertreffen lassen, und sich in einem ewigen Kraise von Pöbel herumdrehen?
 ‘and such a great lord as our gracious duke is, is supposed to be surpassed by these little ones with **regard** to the courtly glamour, and to spin around in an eternal circle of the mob?’
 (Käsebohrer, Libell, 1797, DTA *baby*)

These observations point to the multidimensionality and complexity of the dynamics involved in constructionalization and constructional change. Constructionalization does not only seize individual words but also larger units. This is of course an aspect that morphological cross-tabulation analysis – at least

in the way it is applied here – cannot capture. If we take the idea seriously that constructional change can only be fully understood by taking a schema-based perspective, this raises the question, in each individual case, at what level of abstraction a process of change is taking place. For example, the observation that *Ansehung* is identified as a repelled collexeme in the earlier periods might be misinterpreted as an indicator that the constraint against durative bases is not yet in effect in the seventeenth century. If we know about the existence of the constructional idiom *in Ansehung*, which used to be relatively frequent in the Early New High German period, we can realize that *Ansehung* is not an ad-hoc formation, as its identification as repelled collexeme might wrongly suggest, but rather part of a lexicalized structure that is falling out of use.

More generally speaking, one important limitation of morphological cross-tabulation analysis is that it cannot take context into account, which, apart from mere frequency, can be considered an important factor in lexicalization and constructionalization. Every usage event is tied to specific contexts, both linguistic and non-linguistic. For example, repeated use in a specific set of semantically similar contexts can trigger the emergence of specific non-compositional meaning variants, while repeated use in the same context can lead to the fixation of multi-word patterns through chunking (cf. e.g. Bybee 2010: 7; Traugott and Trousdale 2013: 123). Taking the interaction patterns discussed in Section 4.2 above into account can partly make up for these limitations and help us find the right level of abstraction at which a constructionalization pattern or a pattern of constructional change can be detected. In the case of the [P NOM] construction, for instance, the development of the syntactic pattern can be seen as interacting with the development of the word-formation pattern. Many processual uses of *ung*-nominals are tied to this syntactic pattern. Thus, the question emerges if it is actually the word-formation pattern of *ung*-nominalization that is productively used to coin highly processual nominals or if it is rather the [P V-*ung*] construction that productively generates linguistic units with a highly progressive reading. From a constructionist point of view, however, these alternatives do not exclude each other. Instead, language users' knowledge about the [P NOM] construction is shaped by the nominals occurring in the noun slot just as their knowledge about *ung*-nominalization is partly shaped by the constructions in which word-formation products in *-ung* appear.

Summing up, then, the different methods discussed in this section have shed light on various aspects of *ung*-nominalization. The productivity measures have corroborated the finding that the potential of *ung*-nominalization to coin new words decreases over time, while the analysis of interaction patterns has revealed that the semantics of *ung*-nominals shifts towards a more “nouny”

construal. Finally, the morphological cross-tabulation analysis could corroborate some of the observations put forward in earlier research with regard to the changing selectional preferences of *ung*-nominalization (inclination to take complex base verbs; *ung*-nominals derived from certain verb classes falling out of use). Taken together, these findings allow for a coherent interpretation of the pattern's development which can also feed back into a general cognitive-linguistic theory of morphology and morphological change, as will be discussed in the next section.

5 Discussion and conclusion

The previous sections have offered a fairly in-depth analysis of one specific word-formation pattern and its diachronic development. This section sums up the results and connects them to current lines of discussion in (diachronic) construction grammar. I will argue that the network metaphor that is key to cognitive and usage-based linguistics can offer a unified account of word-formation change and other types of constructional changes. More specifically, Bybee's (e.g. 1995) notions of lexical strength and schema strength can be operationalized to model the cyclic relationship between instance and generalization that drives constructional change.

As outlined in the course of this paper, *ung*-nominalization in present-day German is subject to a variety of constraints which different researchers have struggled to explain from more formally-oriented perspectives (e.g. lexicalist and transformationalist approaches in the generative tradition, cf. Römer 1987; decompositional semantics, cf. Ehrich and Rapp 2000; discourse representation theory, cf. Roßdeutscher and Kamp 2010). A usage-based approach to word-formation and word-formation change can offer a fairly straightforward explanation of the present-day situation and the diachronic developments that gave rise to it.

The data from DTA and DTAbaby show that *ung*-nominals tend to occur less frequently in contexts that require a “verby”, i.e. relational and processual, construal of the word-formation product and more frequently in contexts that entail a “nouny” construal. In these latter cases, the nominal denotes a fairly individuated, albeit often abstract, entity: for example, *Lesung* ‘reading’ or *Veranstaltung* ‘event’ refer to single events with clear boundaries in space and time, rather than processes. Thus, the meaning of these nominals is shifted towards the prototypical meaning of nouns, which tend to denote objects, which in turn are characterized by non-relationality and stativity. Cases like *Heizung*

‘heating installation’ or *Kupplung* ‘gearshift’ even denote what Langacker (2004) calls the “conceptual archetype” of nouns, namely physical objects.

On the instantiation level, i.e. on the level of the individual word-formation product, *ung*-nominals tend to emancipate themselves from their superordinate construction. Using Bybee’s (1995) terms of lexical strength and schema strength, they exhibit a high degree of lexical strength – also because of their high frequency –, but they do not strengthen the schema, i.e. *ung*-nominalization. The lexicalization of many different *ung*-nominals and the subsequent lack of strengthening of the schema might be one reason for the declining productivity of the pattern, which is also mirrored by the fact that some less frequent formations fall out of use and must be considered marked or ungrammatical in present-day German.

While the usage-based approach to word-formation change outlined in this paper is highly compatible with recent work in diachronic construction grammar (e.g. Traugott and Trousdale 2013), the notions of lexical strength and schema strength require a much more gradualistic view of constructions than is usually adopted in constructionist approaches. As Hilpert (forthc.) points out, it is hard to decide when exactly the line between constructional change, i.e. the change of some internal dimension(s) of a construction, and constructionalization, i.e. the emergence of a new construction, is crossed. Schmid (2017: 25) therefore proposes a model that “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”. The association-based view has important advantages over the node-and-relations view: As Hilpert (forthc.) points out, it lends itself more naturally to be mapped onto neurophysiological processes as well as to computational implementation. In addition, it may allow for taking more complex sets of associations into account – not only between forms and meanings, but also, for instance, between contexts and forms/meanings. However, it can still be argued that “nodes”, i.e. constructions, constitute a valuable heuristic tool for the analyst to detect patterns in the associative network. These nodes, however, are highly dynamic, and they are contingent upon the associations they form part of, not least with regard to their strength. For instance, the word-formation pattern of *ung*-nominalization can be regarded as a “node” in the constructional network. This way to describe the pattern captures the insight that *ung*-nominalization is a generalization over its instances, i.e. over concrete constructs in *-ung* and the associations that hold between them. Importantly, the associations between *ung*-nominals and the syntactic constructions they appear in (see Section 4.3 above) form part of language users’ implicit knowledge about the pattern as well. Furthermore, associations between *ung*-nominals and other constructions

with a similar function may be seen as part of this knowledge. For example, in many cases in present-day German, the Nominalized Infinitive is the more natural choice where *ung*-nominals were used in earlier stages of German (cf. Demske 2000; Hartmann 2016), which can be explained as a case of statistical preemption (cf. Goldberg 2001): in a context where, in principle, an *ung*-nominal could be expected, another word-formation product occurs. Thus, the *ung*-nominal is “negatively entrenched”, as it were. This pertains especially to highly lexicalized nominals. Based on a corpus analysis using the DWDS Core Corpus of the twentieth century, Hartmann (2016: 264f.) shows that *Heizung* (from *heizen* ‘to heat’) is increasingly used referring to an object, while the processual reading variant falls out of use. In more processual contexts, alternative formations like the nominalized infinitive *das Heizen* are used. This in turn contributes to weakening the link (i.e., the association) between *ung*-nominalization and processual meaning.

At the same time, the network metaphor can prove helpful in describing and explaining the heterogeneity of different reading variants that *ung*-nominals have developed via lexicalization (see Hartmann 2014a). In line with the radial category approach proposed by e.g. Panther and Thornburg (2002) for English *-er*-nominals, the different senses of present-day *ung*-nominals can be conceived of as a conceptual network with more and less prototypical reading variants that has grown over time. Importantly, however, such polysemy networks emerge from the interaction between words and constructions. Diachronically, they can be considered a “by-product”, as it were, of semantic change and constructional change as well as the interaction between those two.

Summing up, then, the usage-based network metaphor provides a helpful framework for understanding constructional change in general and word-formation change in particular. In constructionist terms, linguistic knowledge can be conceptualized as a network of form-meaning pairs, the so-called constructicon (e.g. Goldberg 2006; Taylor 2012; Hilpert 2014). Every usage event can potentially lead to a reconfiguration of this network. This also applies to word-formation constructions. Language users’ knowledge about a word-formation pattern encompasses knowledge about the bases it can take, about the semantics of its instances, about the co-occurrence patterns of the instances, and about the registers in which these instances preferentially occur. Importantly, this knowledge comes about through abstractions and generalizations over the actual instances of a pattern. If an instance changes, e.g. via semantic change/lexicalization, or if even many instances undergo such changes, these developments will most likely have ramifications on the word-formation pattern as a whole.

In such a framework, then, word-formation change can be conceived of in terms of the strengthening and weakening of associations between a) word-

formation constructions and their instantiations (i.e., word-formation products); b) word-formation patterns and the contexts in which their instantiations occur; c) word-formation patterns and competing constructions. The first two of these dimensions have been explored empirically in the present paper. Following up on previous research on the diachrony of German *ung*-nominalization, it has been shown that the word-formation pattern is subject to a growing number of constraints and that the contexts in which its instantiations preferentially occur are subject to change as well. Interestingly, cross-linguistic research suggests that the development of English *ing*-nominals follows a fairly similar path (cf. Fonteyn and Hartmann 2016). It would be interesting to test the model of word-formation change presented in this paper on further languages. In addition, patterns of competition could be explored in more detail. For instance, it has been suggested that infinitival nominalization serves as a “replacement”, as it were, for *ung*-nominalization. But even though Hartmann’s (2016) findings on the basis of the GerManC corpus seem to support the idea that infinitival nominalization incorporates parts of *ung*-nominalization’s functional domain, this hypothesis has not been systematically tested yet. Focusing on patterns of competition between constructions seems all the more promising as recent research adopting this approach has shown that competition indeed seems to play a pivotal role in the diachronic change of word-formation patterns (cf. e.g. Kempf 2016).

Further issues that could be addressed in more detail in future research are of a more methodological nature. For example, the question of how accurately the productivity measures applied in this paper capture the actual productive use of the construction(s) in question needs to be addressed. Another open question relates to the level of abstraction at which constructions should be posited in the first place: in the case of *ung*-nominalization, for instance, multiple subconstructions might be legitimately posited, some of which are still quite productive in present-day German (e.g. [(Prefix)-X-*ierung*] as in *Gentrifizierung* ‘gentrification’, [*Be*-X-*ung*] as in kolloq. *Bespaßung* ‘entertainment’, lit. ‘befunning’). Given a sufficient amount of data, explorative methods from distributional semantics (Perek 2016) could be a helpful tool for detecting word-formation products that behave similarly, thus identifying potential subconstructions.

6 Datasets and scripts

The datasets and scripts that were used for the present study are available at github.com/hartmast/ung_DTA.

Acknowledgements: I am grateful to three reviewers and particularly Harald Baayen for helpful comments and suggestions. In addition, the project reported on here as well the paper itself have benefitted from the input of many different people, including Damaris Nübling, Renata Szczepaniak, Michael Pleyer, Luise Kempf, Susanne Flach, Daniela Schröder, Lisa Dücker, Eleonore Schmitt, Melitta Gillmann, and Annika Vieregge. Remaining errors are of course mine.

References

Corpora

DECOW14AX = Corpora from the Web. <http://www.corporafromtheweb.org/>
 DTA = Deutsches Textarchiv (German Text Archive), deutschestextarchiv.de
 DeReKo = Deutsches Referenzkorpus (German Reference Corpus), <https://cosmas2.ids-manheim.de/cosmas2-web/>
 DWDS = Digitales Wörterbuch der deutschen Sprache – Kernkorpus 20. Jahrhunderts. <http://www.dwds.de/>
 FnhdC = Bonner Frühneuhochdeutschkorpus (Bonn Early New High German Corpus). <https://korpora.zim.uni-due.de/Fnhd/>

Software

R Core Team. 2015. *R. A software and environment for statistical computing*. Vienna: R Foundation.
 Scripts and packages:
 Baroni, Marco & Stefan Evert. 2007. *zipfR. Lexical statistics in R*. <http://zipfr.r-forge.r-project.org/> (accessed 26 January 2015).
 Flach, Susanne. 2016. Collostructions. An R implementation for the family of collostructional methods. www.bit.ly/sflach
 Gries, Stefan Th. 2007. Collostructional analysis: Computing the degree of association between words and words/constructions.

Research literature

Harald. 1992. Quantitative Aspects of Morphological Productivity. In Geert E. Booij & Jaap Van Marle (eds.), *Yearbook of morphology 1991*, 109–149. Dordrecht: Kluwer.
 Baayen, R. Harald. 1993. On Frequency, Transparency, and Productivity. In Geert E. Booij & Jaap Van Marle (eds.), *Yearbook of morphology 1992*, 181–208. Dordrecht: Kluwer.
 Baayen, R. Harald. 2001. *Word Frequency Distributions*. (Text, Speech and Language Technology, 18). Dordrecht: Kluwer.

- Baayen, R. Harald. 2009. Corpus Linguistics in Morphology: Morphological Productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics* (HSK 29.2), 899–919. Berlin, New York: De Gruyter.
- Baroni, Marco & Stefan Evert. 2005. Testing the extrapolation quality of word frequency models. *Proceedings of Corpus Linguistics 2005*. <http://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2005-journal/Lexiconodf/EvertBaroni2005.pdf>. (accessed 15 October 2016).
- Baroni, Marco & Stefan Evert. 2014. The zipfR package for lexical statistics: A tutorial introduction. <http://zipfr.r-forge.r-project.org/materials/zipfr-tutorial.pdf>.
- Barz, Irmhild. 1998. Zur Lexikalisierungspotenz nominalisierter Infinitive. In Irmhild Barz & Günther Öhlschläger (eds.), *Zwischen Grammatik und Lexikon*, 57–68. Tübingen: Niemeyer.
- Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.
- Booij, Geert E. 2010. *Construction Morphology*. Oxford: Oxford University Press.
- Booij, Geert E. 2012. *The grammar of words: An introduction to linguistic morphology*. 3rd edn. Oxford: Oxford University Press.
- Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5). 425–455.
- Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Croft, William. 1991. *Syntactic categories and grammatical relations*. Chicago: University of Chicago Press.
- Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Dąbrowska, Ewa. 2009. Words as constructions. In Vyvyan Evans & Stéphanie Pourcel (eds.), *New directions in cognitive linguistics* (Human Cognitive Processing 24), 201–223. Amsterdam and Philadelphia: John Benjamins.
- Dammel, Antje. 2011. Wie kommt es zu rumstudierenden Hinterbänkclern und anderen Sonderlingen? Pfade zu pejorativen Wortbildungsbedeutungen im Deutschen. In Jörg Riecke (ed.), *Historische Semantik*. (Jahrbuch für germanistische Sprachgeschichte 2). Berlin, New York: De Gruyter.
- Demske, Ulrike. 2000. Zur Geschichte der ung-Nominalisierung im Deutschen: Ein Wandel morphologischer Produktivität. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 122. 365–411. doi: 10.1515/bgsl.2000.122.3.365.
- Demske, Ulrike. 2001. Zur Distribution von Infinitivkomplementen im Althochdeutschen. *Linguistische Berichte* 9. 61–86.
- Demske, Ulrike. 2002. Nominalization and argument structure in early new high German. In Ewald Lang & Ilse Zimmermann (eds.), *Nominalisations* (ZAS Papers in Linguistics), 67–90. Berlin ZAS.
- Diessel, Holger. 2015. Usage-based construction grammar. In Ewa Dąbrowska & Dagmar Divjak (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.
- Durrell, Martin, Astrid Ensslin & Paul Bennett. 2007. The GerManC project. *Sprache und Datenverarbeitung* 31. 71–80.
- Ebengaard-Jensen, Kim. 2013. Semantic coherence in English accusative-with-bare-infinitive constructions. *Rask* 38. 161–176.

- Ehrich, Veronika & Irene Rapp. 2000. Sortale Bedeutung und Argumentstruktur: -ung-Nominalisierungen im Deutschen. *Zeitschrift für Sprachwissenschaft* 19. 245–303.
- Evert, Stefan. 2004. A simple LNRE model for random character sequences. In Gérald Purnelle, Cédric Faron, Anne Dister (eds.), *Proceedings of JADT 2004*. 411–422. Louvain: Presses universitaires de Louvain.
- Evert, Stefan & Marco Baroni. 2007. zipfR: Word frequency distributions in R. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Posters and Demonstrations Sessions*, Prague, 29–32.
- Flach, Susanne. 2015. Let's go look at usage. (Ed.) Thomas Herbst & Peter Uhrig. *Yearbook of the German Cognitive Linguistics Association* 3(1). 231–251. doi:10.1515/gcla-2015-0013.
- Fonteyn, Lauren & Stefan Hartmann. 2016. Usage-based perspectives on diachronic morphology: A mixed-methods approach towards English ing-nominals. *Linguistics Vanguard* 2(1). doi:10.1515/lingvan-2016-0057.
- Gaeta, Livio & Davide Ricca. 2006. Productivity in Italian word-formation. *Linguistics* 44(1). 57–89.
- Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago and London: The University of Chicago Press.
- Goldberg, Adele E. 2001. Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics* 22(1). 131–153.
- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Hartmann, Stefan. 2014a. The Diachronic Change of German Nominalization Patterns: An Increase in Prototypicality. In Gabriella Rundblad, Aga Tytus, Olivia Knapton & Chris Tang (eds.), *Selected Papers from the 4th UK Cognitive Linguistics Conference*, 152–171. London: UK Cognitive Linguistics Association.
- Hartmann, Stefan. 2014b. “Nominalization” Taken Literally: A Diachronic Corpus Study of German Word-Formation Patterns. *Italian Journal of Linguistics* 26(2). 123–155.
- Hartmann, Stefan. 2014c. Constructing a Schema: Word-Class Changing Morphology in a Usage-Based Perspective. In Martin Hilpert & Susanne Flach (eds.), *Yearbook of the German Cognitive Linguistics Association*, Vol. 2, 235–252. Berlin, New York: De Gruyter.
- Hartmann, Stefan. 2016. *Wortbildungswandel. Eine diachrone Studie zu deutschen Nominalisierungsmustern*. Berlin, Boston: De Gruyter.
- Hartmann, Stefan. 2018. *Deutsche Sprachgeschichte. Grundzüge und Methoden*. Tübingen: Francke.
- Hilpert, Martin. 2006. Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory* 2(2). 243–256.
- Hilpert, Martin. 2013. *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press.
- Hilpert, Martin. 2014. *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.
- Hilpert, Martin. forthc. Three open questions in diachronic construction grammar. In Evie Coussé, Peter Andersson & Joel Olofsson (eds.), *Grammaticalization meets construction grammar*. Amsterdam, Philadelphia: John Benjamins.
- Hilpert, Martin & Stefan Th Gries. 2009. Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing* 24(4). 385–401.

- Kempf, Luise. 2016. *Adjektivsuffixe in Konkurrenz. Wortbildungswandel vom Frühneuhochdeutschen zum Neuhochdeutschen*. Berlin, Boston: de Gruyter.
- Kempf, Luise & Stefan Hartmann. forthc. Schema unification and morphological productivity: A diachronic perspective. To appear in Geert E. Booij (ed.): *The construction of words. Advances in construction morphology*. Berlin: Springer.
- Knobloch, Clemens. 2002. Zwischen Satz-Nominalisierung und Nennerivation: -ung-Nomina im Deutschen. *Sprachwissenschaft* 27. 333–362.
- Krott, Andrea, Robert Schreuder & R. Harald Baayen. 1999. Complex words in complex words. *Linguistics* 37(5). 905–926.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar. Vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald W. 1991. *Foundations of cognitive grammar. Vol. 2: Descriptive application*. Stanford: Stanford University Press.
- Langacker, Ronald W. 2004. Remarks on Nominal Grounding. *Functions of Language* 11(1). 77–113.
- Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.
- Levshina, Natalia. 2015. *How to do linguistics with R. Data exploration and statistical analysis*. Amsterdam, Philadelphia: John Benjamins.
- Li, Wentian. 1992. Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory* 38(6). 1842–1845.
- Mandelbrot, Benoît. 1962. On the theory of word frequencies and on related Markovian models of discourse. In Roman Jakobson (ed.), *Structure of language and its mathematical aspects*, 190–219. Providence: American Mathematical Society.
- Marchand, Hans. 1969. *The categories and types of present-day English word formation: A synchronic-diachronic approach*. München: Beck.
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak & Erez Lieberman Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014). 176–182.
- Nübling, Damaris, Antje Dammel, Janet Duke & Renata Szczepaniak. 2013. *Historische Sprachwissenschaft des Deutschen: Eine Einführung in die Prinzipien des Sprachwandels*. 4th edn. Tübingen: Narr.
- Panther, Klaus-Uwe & Linda Thornburg. 2001. A conceptual analysis of English -er nominals. In Martin Pütz, Susanne Niemeier & René Dirven (eds.), *Applied cognitive linguistics* (Cognitive Linguistics Research 19.2), 149–200. Berlin, New York: De Gruyter.
- Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony. A case study. *Linguistics* 54(1). 149–188.
- Panther, Klaus-Uwe & Linda Thornburg. 2002. The roles of metaphor and metonymy in English -er Nominals. In René Dirven & Ralf Pörings (eds.), *Metaphor and metonymy in comparison and contrast*, 279–319. Berlin, New York: De Gruyter.
- Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation*. (Topics in English Linguistics, 28). Berlin, New York: De Gruyter.
- Römer, Christine. 1987. Transformationalistische und lexikalistische Erklärung von Wortbildungen - dargestellt am Beispiel deverbaler -ung-Substantive. *Deutsch als Fremdsprache* 24. 217–221.

- Roßdeutscher, Antje & Hans Kamp. 2010. Syntactic and Semantic Constraints on the Formation and Interpretation of -ung-Nouns. In Monika Rathert & Artemis Alexiadou (eds.), *The semantics of nominalizations across languages and frameworks* (Interface Explorations 22), 169–214. Berlin, New York: De Gruyter.
- Schäfer, Roland. 2015. Processing and querying large corpora with the COW14 architecture. *Challenges in the Management of Large Corpora (CMLC-3)*. <http://corpora.ids-mannheim.de/cmlc.html>.
- Schäfer, Roland & Felix Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. In Cicoletta Calzolari, Khalid Choukri, Terry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of LREC 2012*, 486–493. Istanbul: European Language Resources Association.
- Scherer, Carmen. 2006. Was ist Wortbildungswandel? *Linguistische Berichte* 205. 3–28.
- Scherer, Carmen. 2007. The Role of Productivity in Word-Formation Change. In Joseph C. Salmons & Shannon Dubenion-Smith (eds.), *Historical linguistics 2005* (Current Issues in Linguistic Theory, 284), 257–271. Amsterdam and Philadelphia: John Benjamins.
- Schneider-Wiejowski, Karina. 2011. *Produktivität in der deutschen Derivationsmorphologie*. University of Bielefeld PhD thesis.
- Schmid, Hans-Jörg. 2017. A framework for understanding entrenchment and its psychological foundations. In Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning. How we reorganize and adapt linguistic knowledge*, 9–39. Berlin, Boston: De Gruyter.
- Schmidt, Wilhelm. 2007. *Geschichte der deutschen Sprache: Ein Lehrbuch für das germanistische Studium*. 10th ed. Stuttgart: Hirzel.
- Shin, Soo-Song. 2001. On the event structure of -ung-nominals in German. *Linguistics* 39. 297–319.
- Smirnova, Elena & Tanja Mortelmans. 2010. *Funktionale Grammatik: Konzepte und Theorien*. Berlin, New York: De Gruyter.
- Spencer, Andrew. 2001. Morphology. In Mark Aronoff & Janie Rees-Miller (eds.), *The handbook of linguistics*, 213–237. Oxford: Blackwell.
- Spencer, Andrew & Arnold M. Zwicky. 1998. Introduction. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology*, 1–10. Oxford: Blackwell.
- Stefanowitsch, Anatol. 2009. Bedeutung und Gebrauch in der Konstruktionsgrammatik: Wie kompositional sind modale Infinitive im Deutschen? *Zeitschrift für Germanistische Linguistik* 37. 562–592.
- Stefanowitsch, Anatol. 2013. *Collostructional analysis*. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of construction grammar*, 290–306. Oxford: Oxford University Press.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Taylor, John R. 2002. *Cognitive grammar*. Oxford: Oxford University Press.
- Taylor, John R. 2012. *The mental corpus: How language is represented in the mind*. Oxford: Oxford University Press.
- Thielmann, Winfried. 2007. Substantiv. In Ludger Hoffmann (ed.), *Handbuch der deutschen Wortarten*, 791–822. Berlin, New York: De Gruyter.

- Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. (Oxford Studies in Diachronic and Historical Linguistics 6). Oxford: Oxford University Press.
- Vogel, Petra M. 1996. *Wortarten und Wortartenwechsel: Zur Konversion und verwandten Erscheinungen im Deutschen und in anderen Sprachen*. Vol. 39. (Studia Linguistica Germanica). Berlin, New York: De Gruyter.
- Vogel, Petra M. 2000. Grammaticalization and part-of-speech systems. In Petra Maria Vogel & Bernard Comrie (eds.), *Approaches to the typology of word classes* (Empirical Approaches to Language Typology 23), 259–284. Berlin, New York: De Gruyter.
- Werner, Martina. 2012. *Genus, Derivation und Quantifikation: Zur Funktion der Suffigierung und verwandter Phänomene im Deutschen*. (Studia Linguistica Germanica, 114). Berlin, New York: De Gruyter.
- Wiechmann, Daniel. 2008. On the computation of collocation strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory* 4(2). 253–290.
- Wulff, Stefanie. 2006. Go-V vs. go-and-V in English: A case of constructional synonymy? In Stefan Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in cognitive linguistics. Corpus-based approaches to syntax and lexis*, 101–126. Berlin, New York: Mouton de Gruyter.