Stefan Hartmann

Constructing a schema: Word-class changing morphology in a usage-based perspective

Abstract: This paper discusses how collostructional analysis can be applied to the study of word-formation patterns. Drawing on a diachronic corpus study of German *ung*-nominalization and Infinitival Nominalization using the GerManC Corpus, it is shown that a cross-tabulation analysis comparing the frequencies of word-formation products to those of their respective bases can give valuable clues to the input language users rely on when abstracting a schema (i.e., a word-formation construction) from a quite heterogeneous array of instantiations.

1 Introduction

In usage-based approaches to morphology, numerous quantitative methods have been developed to assess the synchronic productivity as well as the diachronic development of individual word-formation constructions (cf. e.g. Baayen 2009). However, given the key assumption of Cognitive Linguistics and Construction Grammar that language is situated in context and that speakers learn constructions from the statistical features of their input (cf. e.g. Goldberg 2006; Taylor 2012), investigating the *interaction* between constructions is as important to our understanding of linguistic knowledge as assessing individual constructions in isolation. Hilpert (2013) has already proposed a variety of empirical approaches to diachronic morphology in a constructionist framework. This paper adds to this strand of research by investigating German nominalization patterns in the 17th and 18th centuries in comparison to their respective base verbs.

Stefan Hartmann: Johannes Gutenberg-Universität Mainz.
E-mail: stefan.hartmann@uni-mainz.de

The aims of the present paper are twofold: On the one hand, I will discuss methodological approaches to the study of word-class changing morphological constructions. More specifically, I will argue that the basic idea of collostructional analysis as proposed by Stefanowitsch and Gries (2003) can be adapted to the analysis of word-formation patterns. This method provides a valuable means for studying the relationship between word-formation products and their respective base words in quantitative terms. On the other hand, I will discuss how empirical findings on a) the frequency and productivity of word-formation patterns and b) the ratio of word-formation products to their respective base words can feed back into our understanding of morphological constructions. Most importantly, the correlation of constructionalization and frequency patterns will be discussed, and the traditional notion of schema salience employed in Cognitive Grammar and much of usage-based linguistics in general will be critically examined. I will argue that the association strength obtained by means of collostructional analysis can give a clue to the degree to which a word-formation product relies on its schema and thus contributes to strengthening this schema. First, however, I will give a brief overview of the word-formation patterns in question and their diachronic development.

2 German nominalization patterns: A case study

As a case study, I will present a diachronic corpus study of the two most productive German deverbal nominalization patterns, namely nominalization by means of the suffix *-ung* and Infinitival Nominalization. Both patterns have been investigated in a variety of both synchronic and diachronic studies over the past 15 years. Most important for our purposes are the findings of Demske (2000) about the morphological productivity of *ung*-nominalization, complemented by Barz' (1998) considerations on Infinitival Nominalization. Demske (2000) shows that *ung*-nominalization is subject to an increasing number of semantic word-formation constraints: Durative verbs (*glauben* 'believe'), inchoative verbs (*erblühen* 'blossom$_{INCH}$'), iterative verbs (*hüsteln* 'cough (slightly/repeatedly)'), and verbs of transfer (*geben* 'give') cannot function as base verbs of *ung*-nominals any more. Therefore, some derivations in *-ung* that used to be felicitous in the Early New High German period (ENHG, 1350–1650) are nowadays ungrammatical (e.g. *murmelung* 'muttering'). Due to the increase in word-formation constraints, the productivity of the pattern, understood as its ability to be extended to new cases, decreases. This decrease in productivity can be observed both in Demske's newspaper corpus (for which she does not, however, provide a quantitative analysis) and in the GerManC corpus (Durrell et al. 2007), on which the present study relies.
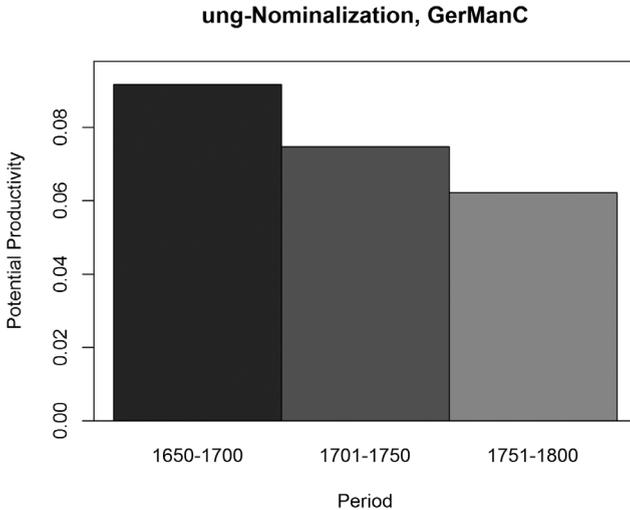
Fig. 1: Diachronic development of the potential productivity of *ung*-nominalization in the GerManC corpus.

Figure 1 plots the potential productivity of *ung*-nominalization, i.e. the number of hapax legomena (words occurring only once in the entire corpus) belonging to a specific construction (here: *ung*-nominalization) in relation to the total number of tokens instantiating this construction (cf. Baayen 2009). As the GerManC corpus is only balanced for 50-year periods, Figure 1 adopts this fairly coarse-grained three-stage periodization. A one-tailed Fisher's exact test confirms our hypothesis that there are significantly less hapax legomena in -*ung* (as compared with the total number of *ung*-nominals) in the second than in the first corpus period ($p < .05$, odds ratio = .81). In addition, we find less hapaxes in the third than in the second period, but the difference is not significant ($p = .054$, odds ratio = .83).

According to Demske (2000), this development is largely due to the lexicalization of highly frequent word-formation products in -*ung*. From a usage-based point of view, this seems highly plausible: Over time, frequent *ung*-nominals assume a range of different meanings, e.g. object (*Heizung* 'heating installation') or even person readings (*Bedienung* 'waiter/waitress'). As the word-formation products are used more frequently in these emerging new meaning variants, they become detached from their respective base verbs. This has ramifications for the word-formation pattern, i.e. for the schematic construction that language users abstract away from actual instances of language use. As the noun *Heizung* is by now almost exclusively used to refer to an object, using it in the processual sense

seems odd, if not ungrammatical (??*Die Heizung des Raums nahm vier Stunden in Anspruch* 'heating the room took four hours'). Instead, Infinitival Nominalization (*das Heizen* 'heating') would be used, which comes in as a 'replacement process' (Barz 1998) for *ung*-nominalization. In the following sections, we will primarily be concerned with the question of how this process of 'detachment' from the base verb can be investigated in quantitative terms.

3 Adapting collostructional analysis to the study of word-formation patterns

3.1 Complex words and their bases

Originally proposed by Stefanowitsch and Gries (2003), collostructional analysis has by now become a cover term for a variety of methodological approaches to study the interaction between constructions (cf. Stefanowitsch 2013 for a recent overview). In this section, we will be concerned with the original approach (simple collexeme analysis). This version of collostructional analysis straightforwardly adapts the logic of classic methods of cross-tabulation analysis like the chi-squared test. Given a construction *c* such as the ditransitive construction (*give sb. sth., tell sb. sth.*), the association of a particular lexical item l_i (e.g. *give*) to this construction is determined by taking into account a) the frequency of l_i in *c*, b) the frequency of l_i in all other constructions ($\neg c$), c) the frequency of other lexemes of the same word class *L* (here: verbs) occurring in *c* ($\neg l_i$), and d) the frequency of all other verbs ($\neg l_i$) in all other constructions ($\neg c$). The *p*-value of a Fisher's exact test then serves as indicator of collostruction strength. As an example highly relevant to our case study introduced in Section 2, consider the [PREP N] construction, in which *ung*-nominals used to occur quite frequently at the beginning of the period covered by the GerManC corpus (cf. Hartmann forthc.). Table 1 lists the

Table 1: *ung*-nominals most strongly attracted to the [PREP N] construction

Collexeme	Frequency	Expected Frequency	*p*-value (collostruction strength)
Ansehung	46	11.60	3.02E−28
Vermeidung	20	5.04	1.09E−12
Befindung	10	2.52	1.05E−06
Ermangelung	9	2.27	4.14E−06
Verfließung	7	1.77	6.51E−05

five *ung*-nominals attracted most strongly to this construction. In the GerManC corpus, *Ansehung* 'view' occurs 46 times in the construction *in Ansehung* 'in view of'. Not once does it occur independently. Thus, $c(l_i) = 46$, and $\neg c(l_i) = 0$. According to the tagged version of the GerManC corpus (LING-COL), there are 31,258 preposition constructions in the corpus which are followed by a bare noun without out a determiner,[1] hence $c(\neg l_i) = 31258 - 46 = 31212$. Finally, 123,870 tokens in the corpus are tagged as nouns. These items can occur in the N slot of the preposition construction. Hence, $\neg c(\neg l_i) = 123870 - 31258 - 0 = 92612$.

The crucial difference between studying the association between words and syntactic constructions on the one hand and the relation between word-formation products and their respective bases on the other is the fact that we are not dealing with lexemes belonging to the same word class within and outside of the construction. Hence, rather than comparing the frequency of $L(l_i)$ vs. $L(\neg l_i)$, we have to compare one word to another word, i.e. a complex word to its base word. A potential objection against such an approach might be that word-formation products – especially highly frequent ones – are often lexicalized. Hence, it seems plausible to assume that an ad-hoc-formation like *Ergoogelung* 'googling' makes both its base verb *googeln* and the word-formation schema [V-*ung*] more salient than does a highly lexicalized nominal such as *Bildung* 'education'. However, this holds for many syntactic constructions, as well. The [PREP NOM] construction discussed above is a case in point: *in Ansehung, bei Vermeidung* 'in order to avoid [a punishment]', and *nach Befindung* 'according to the evidence' are *constructionalized* (cf. Traugott and Trousdale 2013) to a considerable degree. They have non-compositional meaning, and they tend to be "fossilized": There is a highly significant inverse correlation between the collostruction strength of a [PREP N] type ($\pm|\log_{10}(p)|^2$) and the proportion of [PREP N] tokens of this type used with lexical material (e.g. adjectives, prenominal genitives) inserted between PREP and N (Kendall's $\tau = -.1$, $z = -2.7$, $p < .01$).

It is a key hypothesis of usage-based Construction Grammar that language users take note of the frequencies of the linguistic constructions they encounter, thus compiling a structured network of form-meaning pairings, the so-called constructicon (cf. e.g. Taylor 2012; Hilpert 2014). Numerous empirical studies have

---

1 I am grateful to an anonymous reviewer for pointing out that the number of prepositions followed by a bare noun is actually a better reference value than the corresponding "macroconstruction", i.e. all preposition constructions.

2 In order to arrive at a continuous scale ranging from "repelled" to "attracted" values, the collostruction strength (i.e. the *p*-value) is $\log_{10}$-transformed and the sign is set to reflect the direction of association. This is a standard procedure in collostructional analysis (cf. Hilpert 2006: 247).

lent support to this hypothesis (cf. Goldberg 2006; Taylor 2012). Thus it seems reasonable to assume that language users also take note of the relative frequencies of word-formation products – such as, say, *approval* – and their respective bases (*approve*) and that these frequencies determine the association strength between derivations and base words in the constructicon. If the base word is highly frequent, whereas the derivation is not, we can assume the word-formation product to be very strongly associated with its base. In fact, the hearer might even have never encountered that derivation, but she is still able to understand it by virtue of being familiar both with the base word and the morphological construction (the word-formation pattern). If, by contrast, the derivation is highly frequent, while the base word hardly ever occurs, the word-formation product can be regarded as detached from its base. These considerations fit well with the results of psycholinguistic priming studies (e.g. Clahsen et al. 2003) without, however, invoking a dual-route model of morphological processing, as Clahsen and his colleagues do: Instead, we can assume a continuum between schema-reliant and schema-independent processing. While the former refers to understanding a derivation by virtue of recognizing the base word and the morphological construction in which it is used, the latter implies that a (lexicalized) word-formation product is stored holistically independently of its base word.

These considerations also tie in neatly with the notion of schema salience (or schema strength) proposed in usage-based linguistics (e.g. Bybee 1995). Taylor (2002: 291) describes schema salience as a function of type frequency, which is in line with Bybee's (e.g. 2010) view that it is only or at least predominantly the type frequency of a construction that contributes to its schema strength. However, I will argue that a usage-based conceptualization of linguistic knowledge as a fine-grained, taxonomic network of constructions, "involving massive redundancy and vastly rich detail" (Traugott and Trousdale 2013: 53), entails a more complex conceptualization of schema salience. The strength of a schema depends on the degree to which it is strengthened by its instantiations. To take up the example mentioned above, it can be assumed that a lexicalized formation like *Bildung* 'education' makes the word-formation schema less salient than an ad-hoc-formation like *Ergoogelung*. One determinant of lexicalization is of course high token frequency, which may be accompanied by a decline of the base word's frequency or even by developments leading to a formal "detachment" of the word-formation product from its base (cf. *nähren* 'nourish' – *Nahrung* 'food'; *ábteilen* 'separate' – *Abteílung* 'department'). The association strength between base and derivation can thus be considered the flip side of schema salience, as it were.

To illustrate this, let us return to our case study of German *ung*-nominalization and Infinitival Nominalization. First, we take a look at the absolute frequencies of *ung*-nominals and Nominalized Infinitives (NIs). Figure 2 plots

Fig. 2: Absolute frequencies of *ung*-nominals (upper panel) and NIs (lower panel) and their respective base verbs (y axis) in the three corpus periods of GerManC

the derivations' frequencies in each of the three periods of the GerManC corpus on the x-axis and the frequencies of their respective base verbs on the y-axis.

The frequency distributions in Figure 2 are in line both with the expectations that can be derived from our discussion of the two word-formation patterns in Section 2 and with the theoretical considerations on association strength: Concerning the latter, we find that fairly lexicalized (and highly polysemous) *ung*-nominals such as *Handlung* ('act(ion)', but also 'plot') or *Ordnung* ('order', 'regulation', but also 'tidiness') assume positions on the far right of the x-axis, but rather low on the y-axis, i.e. the *ung*-nominals are very frequent, whereas their base verbs are relatively infrequent. Semantically transparent nominals such as *Haltung* 'keeping'[3] and *Findung* 'finding', by contrast, are derived from highly frequent base verbs, while they themselves are quite infrequent. Diachronically, we can observe that these latter nominals fall out of use.

Comparing the upper panel of Figure 2 with the lower one, we also find that most NIs are not particularly token-frequent. However, they continue to be derived from highly frequent verbs. Given the above considerations, these distributions are hardly surprising. The word-formation pattern of Infinitival Nominalization is characterized by a high degree of schema salience. Conversely, most NIs are strongly associated with their respective base verbs by virtue of being occasional formations. Exceptions include *Leben* 'life', which can be considered as at least in part lexicalized since it refers to the life span as a whole rather than to the process of living, *Ansehen* 'reputation', and *Schreiben* 'letter'.

While the plots in Figure 2 are quite revealing, the methodology of collostructional analysis provides a means to investigate the association strength between the word-formation products and their respective bases in quantitative terms. Since I am not investigating collexemes, but rather the ratio of derivations to base words, I will plainly refer to the method to be outlined below as 'morphological cross-tabulation analysis.'

Table 2 summarizes the main idea: Given a word-formation pattern w deriving word-formation products of the word class W (in our case study: noun), we can determine the association strength between this pattern and a specific base word b_i belonging to word class B (here: verb) by taking into consideration a) the derivation in question, here labeled as $W [w (b_i)]$, indicating that the derivation of the word class W comes about by inserting b_i in the open slot of the constructional schema w; b) the frequency of all other derivations formed according to the

---

3 In Present Day German, *Haltung* 'posture, attitude' is lexicalized, but in the 17th/18th century, it is used in a fairly transparent reading.

Table 2: Contingency table for morphological cross-tabulation analysis

	Base b_i of class B	Other words of class B
Word-formation construction w deriving words of word class W	$W[w(b_i)]$	$W[w(\neg b_i)]$
Word class B of the base	$B(b_i)$	$B(\neg b_i)$

Table 3: An example of the application of morphological cross-tabulation analysis

	Base b_i of class B	Other words of class B
Word-formation construction w deriving words of word class W	*Handlung* 167	all other *ung*-nominals 6872
Word class B of the base	*handeln* 129	all other verbs 80245

pattern w; c) the frequency of the base b_i in its original word class B, and d) the total number of instances of the word class B apart from b_i.

Applied to *ung*-nominalization, this means that we compute a) the frequency of a specific *ung*-nominal, e.g. *Handlung* (167); b) the token frequency of all *ung*-nominals apart from *Handlung* (7039 – 167 = 6872), c) the frequency of the base verb (129), d) the frequency of all other verbs (apart from auxiliaries, which arguably have to be treated separately from other verbs[4]) (80374 – 129 = 80245).

Applying morphological cross-tabulation analysis to both *ung*-nominalization and Infinitival Nominalization yields the results given in Tables 4 and 5, respectively.

These results are very much in line with the semantic characterization of *ung*-nominalization and Infinitival Nominalization in previous research (e.g. Demske 2000; Shin 2001). *ung*-nominals in present-day German tend to construe actions and events in a fairly "nouny" manner as bounded regions in time, whereas those occurring in the [PREP N] construction discussed above tend to evoke a progressive and hence quite verb-like construal (cf. Hartmann 2014). It is

---

4 An anonymous reviewer correctly points out that in terms of their availability for *ung*-nominalization, this applies for copulas and modals, as well. However, these verbs are not precluded from nominalization in general since they can function as bases of Infinitival Nominalization. Therefore, it seems appropriate to exclude only verbs used as auxiliaries, which can be regarded as less contentful than their full-verb counterparts.

Table 4: Results of morphological cross-tabulation analysis for *ung*-nominalization

Rank	Lemma	Frequency	Expected Frequency	Association strength (*p*-value)
ATTRACTION				
1	*Ordnung* 'order'	246	22.4	9.75E−231
2	*Meinung* 'opinion'	180	21.1	4.82E−132
3	*Handlung* 'act(ion)'	167	23.8	4.36E−102
4	*Wirkung* 'impact'	103	11.8	5.95E−78
5	*Hoffnung* 'hope'	133	22.0	6.90E−71
6	*Gattung* 'species'	54	4.35	6.90E−60
7	*Regierung* 'government'	75	9.02	1.71E−54
8	*Bewegung* 'movement'	100	16.4	5.61E−54
9	*Versammlung* 'assembly'	57	6.2	1.01E−45
10	*Veränderung* 'change'	70	10.1	2.59E−43
REPULSION				
1	*Machung* 'making'	1	84.4	6.17E−37
2	*Haltung* 'holding'	3	74.7	1.98E−29
3	*Nehmung* 'taking'	1	65.4	1.95E−28
4	*Findung* 'finding'	1	64.1	1.00E−27
5	*Setzung* 'setting/putting'	1	36.4	2.43E−15
6	*Nennung* 'mentioning'	1	28.7	6.21E−12
7	*Ziehung* 'drawing'	1	28.4	9.78E−12
8	*Legung* 'laying'	1	21.3	1.01E−08
9	*Erkennung* 'recognition'	1	18.5	1.53E−07
10	*Pflegung* 'fostering'	1	14.8	7.67E−06

hardly surprising that the ad-hoc formations identified by means of morphological cross-tabulation analysis tend to occur in this construction, e.g. *wegen nehmung einiger Fahrzeuge* 'due to the taking of some vehicles' (NEWS_P2_WMD_1701_ hanau1), *Faullentzen mit Machung so nennend blauer Montaegen* 'lazing about [along] with making so-called blue mondays' (LEGA_P2_WOD_1738_Constantz). *Machung, Nehmung, Findung, Legung,* and *Pflegung* can be considered infelicitous in present-day German. By contrast, the repelled NIs are all grammatical in present-day German; *Wissen* 'knowledge' may even be considered lexicalized. The attracted NIs, like the attracted *ung*-nominals, are partly lexicalized. Again in line with previous semantic characterizations of both patterns, the "attracted" *ung*-nominals denote reified, "count-noun-like" entities, e.g. *Regierung* 'government' (a collective of persons), *Handlung* 'action/act' (a single, individuated action), *Gattung* 'species'. With the exception of *Verbrechen* 'crime', which mostly refers to one specific act of crime, *Schreiben* in the lexicalized reading of 'letter',

Table 5: Results of morphological cross-tabulation analysis for Infinitival Nominalization

Rank	Lemma	Frequency	Expected Frequency	Association strength (*p*-value)
ATTRACTION				
1	*Leben* 'life'	455	20.90	0
2	*Vergnuegen* 'joy'	87	2.97	1.97E–114
3	*Ansehen* 'reputation'	72	5.53	3.87E–59
4	*Mitleiden* 'sympathy'	33	0.88	9.84E–53
5	*Vermoegen* 'ability / wealth'	54	3.24	1.04E–51
6	*Verbrechen* 'crime'	31	0.88	7.10E–47
7	*Vorhaben* 'plan'	32	1.07	2.33E–43
8	*Verlangen* 'desire, urge'	48	3.56	8.99E–41
9	*Schreiben* 'writing / letter'	56	7.78	4.34E–31
REPULSION				
1	*Gehen* 'going'	3	27.70	4.89E–09
2	*Stehen* 'standing'	3	16.90	9.02E–05
3	*Ziehen* 'pulling'	1	9.44	0.0014
4	*Sprechen* 'speaking'	1	7.78	0.0058
5	*Wissen* 'knowing/knowledge'	11	21.50	0.0201
6	*Denken* 'thinking'	1	5.91	0.0342
7	*Erkennen* 'recognizing'	1	6.15	0.0361

and *Vorhaben* 'plan', the attracted NIs denote more "mass-noun-like" concepts, e.g. states of mind like *Vergnügen* 'joy' and *Mitleiden* 'sympathy'.

3.2 Competing patterns

Since *ung*-nominalization and Infinitival Nominalization fulfill a similar function in that they both derive nouns from verbs, adapting another variant of collostructional analysis, namely distinctive collexeme analysis (Gries and Stefanowitsch 2004), suggests itself. Originally developed to capture the association strength of words with regard to two competing syntactic constructions (e.g. ditransitive vs. *to*-dative construction, *give sb. sth.* vs. *give sth. to sb.*), its logic can be directly adapted to derivational morphology: Given two word-formation constructions w_1 and w_2, we can assess how strongly a given base verb b_i is attracted to one of the two competing patterns by using the contingency table given in Table 6.

Again, the *p*-value computed with the help of Fisher's exact test is taken as an indicator of collostruction strength. Altogether, 113 verbs are used as bases for

Table 6: Contingency table for distinctive cross-tabulation analysis

	Base b_i of class B	Other words of class B
Word-formation construction w_1	$W[w_1(b_i)]$	$W[w_1(\neg b_i)]$
Word-formation construction w_2	$W[w_2(\neg b_i)]$	$W[w_2(\neg b_i)]$

Table 7: Results of distinctive cross-tabulation analysis

ung-Nominalization		Infinitival Nominalization	
Base Verb	**Distinctiveness (*p*-value)**	**Base Verb**	**Distinctiveness (*p*-value)**
wirken 'work/act' (103:1)	1.27E–11	*schreiben* 'write' (4:56)	3.41E–30
hoffen 'hope' (133:6)	3.39E–10	*vergnügen* 'enjoy oneself' (41:87)	8.23E–26
erscheinen 'appear' (48:1)	3.25E–05	*ansehen* 'look at' (46:72)	1.38E–17
untersuchen 'investigate' (39:1)	0.0003	*bedenken* 'think (about)' (2:30)	8.70E–17
rechnen 'calculate' (54:5)	0.0033	*verbrechen* 'break / commit a crime' (6:31)	3.40E–14
beobachten 'observe' (29:1)	0.0045	*verderben* 'spoil/degenerate' (3:27)	3.96E–14
fordern 'demand' (26:1)	0.0111	*belieben* 'like' (6:26)	1.67E–11
versichern 'assure' (23:1)	0.0165	*versprechen* 'promise' (7:20)	6.64E–08
ausführen 'execute' (21:1)	0.0255	*absterben* 'die (off)' (1:9)	2.24E–05
auflösen 'dissolve' (22:1)	0.0257	*vernehmen* 'hear/perceive/ interrogate' (2:7)	0.0011

both *ung*-nominalization and Infinitival Nominalization in the GerManC corpus (out of 1,157 verbs functioning as base verbs for *ung*-nominalization and 430 verbs used as bases for Infinitival Nominalization).[5] Since most of the verbs are highly polysemeous and some of them do not occur outside of specific idiomatic constructions (e.g. *belieben*, which is tied to *etw. zu tun belieben* 'deign to do sth.' and *nach Belieben* 'at (someone's) convenience'), the translations given in Table 7 are necessarily very rough.

Although many of the nominals derived from the verbs mentioned in Table 7 are lexicalized, the list can give a clue to the division of labor between *ung*-

---

5 In principle, all base verbs can be used as input for distinctive collexeme analysis. If this approach is used, however, the results – at least for *ung*-nominalization – hardly differ from those obtained by means of simple collexeme analysis as presented in section 3.1 above.

nominalization and Infinitival Nominalization. With the exception of *schreiben* 'write' and *versprechen* 'promise', the NIs tend to denote states of mind or activities in which the agent remains fairly passive. The *ung*-nominals, by contrast, are derived from verbs in which the agent is construed as active; however, the corresponding nominals do not refer to those actions directly. Instead, they refer to their outcome (*Rechnung* 'calculation/bill' – the result of calculating), and/or they reify the semantic frame evoked by the verb (or a part of it) in a count noun fashion. For example, *Erscheinung* 'apparition' refers to the entity which appears, and *Forderung* 'demand' refers to what is demanded.

4 Productivity and schema salience

The results of both analyses lend support to the theoretical considerations discussed above. First of all, they quantify what has already become obvious by superficially eyeballing Figure 2. While there are a number of *ung*-nominals that massively exceed their base verbs in frequency, this hardly ever occurs in the case of Infinitival Nominalization.

For reasons of space, only the items most strongly attracted to the respective construction as well as those hardly attracted at all were listed in the above tables. However, even the snippets provided by these tables demonstrate clearly the range of association strength values: Some base verbs of *ung*-nominals and NIs appear hardly at all in their original word class; in other cases, the frequency of the nominals and their respective base verbs is fairly balanced; finally, there are some cases in which the base verb is frequent, whereas the derivation is not. As Stefanowitsch and Gries (2003: 236) point out with regard to syntax,

> If syntactic structures served as meaningless templates waiting for the insertion of lexical material, no significant associations between these templates and specific verbs would be expected in the first place.

The same goes for morphological patterns: If their function amounted to nothing more but syntactic transposition (in our case study, from verb to noun), we would expect all bases and derivations to be distributed evenly. In addition, we would not expect any verb to be significantly attracted to either *ung*-nominalization or Infinitival Nominalization (notwithstanding the fact that in such a model, the coexistence of two patterns deriving nouns from verbs wouldn't make much sense in the first place). To be sure, no one would seriously entertain the hypotheses just mentioned. Nevertheless, word-formation patterns such as *ung*-nominalization and Infinitival Nominalization are still too often reduced to the

function of syntactic transposition (e.g. Donalies 2005: 104). The analyses presented here thus lend further support to the hypothesis that the word-formation patterns under discussion are (meaningful) constructions in their own right. As discussed in detail in Hartmann (forthc.), they impose specific construal patterns on the verbs inserted into the open slot of the respective constructional schema.

But how and to which degree do the attracted and the repelled cases contribute to the language user's construction of a morphological schema from actual usage events? As argued above, it seems reasonable to assume that ("repelled") ad-hoc formations, which require schema-reliant processing, significantly strengthen the schema. But assuming that "attracted" formations do not strengthen the schema at all would be what Langacker (1987: 28) calls an "exclusionary fallacy". As discussed in Sections 3.1 and 3.2, the results of morphological cross-tabulation analysis are compatible with the semantic characterization of *ung*-nominalization as denoting more reified, count-noun-like concepts, one the one hand and Infinitival Nominalization as denoting less "nouny", mass-noun-like or even verb-like concepts on the other. As the attracted items are consistent with this characterization, they can be assumed to play a significant role in constituting the morphological schema.

As for the repelled items, we have seen that the repelled *ung*-nominals fall out of use. The repelled NIs, by contrast, can be considered prototypical examples of Infinitival Nominalization in present-day German. This makes sense if we reconsider the concepts of a) morphological productivity, b) schema salience, and c) word-formation constraints in the light of Construction Grammar. More specifically, the concept of a continuum ranging from lexical/contentful to grammatical/processual constructions (cf. Traugott and Trousdale 2013) can be fruitfully applied to word-formation patterns: Productive patterns such as Infinitival Nominalization can be allocated at the "grammatical" end of this scale. The semantics of their word-formation products can be fairly straightforwardly predicted by the conceptual content of the base and the abstract semantics of the derivation pattern. By contrast, unproductive word-formation constructions (e.g. English -*th* in *length, width*) are constrained to a fix set of instantiations, which can, in principle, be extended by means of analogy (cf. Barðdal 2008: 34). Many of these instantiations are lexicalized, i.e. they exhibit a high degree of lexical strength (rather than strengthening the word-formation schema), which leads to a lower degree of semantic coherence of the schema (cf. Barðdal 2008: 27). As, in the case of productive constructions, the schema is very salient, instantiations with low lexical strength (hence, word-formation products which are not lexicalized) can be considered prototypical. In the case of unproductive constructions, by contrast, ad-hoc-formations are considered deviant constructs (cf. English *greenth* or *coolth*). While *ung*-nominalization is not as unproductive as English

lexical / contentful	grammatical / processual

constrained	unconstrained
unproductive	productive
extension by analogy	extension by schema reliance
high lexical strength of instantiations	low lexical strength of instantiations
ad-hoc formations as deviant constructs	ad-hoc formations as prototypical instantiations

Fig. 3: Continuum of lexical/contentful and grammatical/processual constructions in word-formation.

-th, it is far more constrained than Infinitival Nominalization. Therefore, it is not surprising that the repelled ad-hoc formations in Table 5 constitute prototypical examples of Infinitival Nominalization, whereas those in Table 4 seem ungrammatical today and were probably not perceived as prototypical instantiations of *ung*-nominalization by contemporary language users either.

Apart from the quantitative data discussed in Section 2, one clue suggesting that *ung*-nominalization has to be allocated further at the "lexical/contentful" pole of Figure 3 already in the period under discussion is its lack of semantic coherence. For *ung*-nominals, we can distinguish different constructional subschemas (Booij 2010). As Scherer (2006) observes, meaning variants such as 'Object' or 'Person' are reanalyzed from lexicalized *ung*-nominals, 'transferred' to the word-formation pattern itself, and henceforth used productively. Identifying these subschemas as well as their emergence and development in diachronic terms is a key desideratum for further research. The frequency analyses presented in this paper are an important first step toward this goal. In acquiring morphological schemas by means of making generalizations over actual usage events, the most pertinent clues language users rely on are the relative frequencies of the word-formation products and their respective base words as well as the contexts in which they are encountered. While the former are explicitly considered in the analyses presented above, the latter would most reasonably be addressed in language acquisition studies and multimodal discourse analysis.

Most importantly, then, the cross-tabulation analyses give us an impression of the linguistic input language users rely on when they learn the constructions

in question. In addition, they can help us understand how this input changes over time. This paper has only touched upon this topic, but comparing the three panels in Figure 2 makes a diachronic comparison of cross-tabulation results seem highly promising. However, larger corpora would be desirable to arrive at more reliable results for the individual periods. In addition, the [PREP N] construction mentioned in Section 3.1 merits further investigation as it seems to play a decisive role for the diachronic development of *ung*-nominalization. The case study presented here can therefore only be a first step towards widening the scope of morphological research from examining one construction in isolation to investigating the interaction between different morphological and syntactic constructions.

References

Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics. An international handbook. Vol. 2.* (HSK, 29.2), 899–919. Berlin & New York: De Gruyter.

Barðdal, Jóhanna. 2008. *Productivity. Evidence from Case and Argument Structure in Icelandic* (Constructional Approaches to Language, 8). Amsterdam & Philadelphia: John Benjamins.

Barz, Irmhild. 1998. Zur Lexikalisierungspotenz nominalisierter Infinitive. In Irmhild Barz & Günther Öhlschläger (eds.), *Zwischen Grammatik und Lexikon*, 57–68. Tübingen: Niemeyer.

Booij, Geert E. 2010. *Construction morphology*. Oxford: Oxford University Press.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10. 425–455.

Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Clahsen, Harald, Ingrid Sonnenstuhl & James P. Blevins. 2003. Derivational morphology in the German mental lexicon. A dual mechanism account. In Harald R. Baayen & Robert Schreuder (eds.), *Morphological structure in language processing*. (Trends in linguistics. Studies and monographs, 151), 125–155. 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.

Donalies, Elke. 2005. *Die Wortbildung des Deutschen. Ein Überblick*. 2nd edn. (Studien zur deutschen Sprache, 27). Tübingen: Narr.

Durrell, Martin, Astrid Ensslin & Paul Bennett. 2007. The GerManC Project. *Sprache und Datenverarbeitung* 31. 71–80.

Goldberg, Adele E. 2006. *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis. A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9. 97–129.

Hartmann, Stefan. 2014. 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. forthcoming. "Nominalization" taken literally: A diachronic corpus study of German word-formation patterns. To appear in: *Italian Journal of Linguistics, Special Issue: New Territories in Word-Formation*.

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.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.

Scherer, Carmen. 2006. Was ist Wortbildungswandel? *Linguistische Berichte* 205. 3–28.

Shin, Soo-Song. 2001. On the event structure of -ung-nominals in German. *Linguistics* 39. 297–319.

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 onteraction 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.

Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and constructional changes* (Oxford Studies in Diachronic and Historical Linguistics, 6). Oxford: Oxford University Press.