Roland Schäfer (Humboldt-Universität zu Berlin)
Quantitative Approaches to the Study of German (Draft)

1. The scope of this chapter: What is a quantitative approach?

Each method (new or old) should always be seen in the context of the theory it is supposed to assess. In this chapter, I focus mostly on methods that are linked to theories which embrace quantitative reasoning themselves. When we analyse the productivity of a process (Section 2), when we calculate the tendency of specific verbs to occur more or less often in a syntactic construction (Collostructional Analysis, Section 3), and when we specify a statistical model of the factors influencing a pair of alternating near-synonymous constructions (Section 4), we often make a commitment to a quantitative or probabilistic model of language itself, not just to a quantitative method. Under such models, it is assumed that probabilities play a role in linguistic competence, and the methods provide us with ways of modelling probabilistic competence more or less directly. A second constraint on the scope of this chapter results from the fact that a full introduction to all quantitative methods and the associated theoretical frameworks would be impossible.

Therefore, I focus on methods used in Probabilistic Linguistics (PL; Baayen & Hay 2005, Bresnan et al. 2007, Grafmiller et al. 2018), Usage-based Linguistics and Cognitively Oriented Linguistics (UBL and COL; Tomasello 2003, Bybee & Beckner 2009, Divjak 2016, Kapatsinski 2014).[1] Furthermore, the methods discussed here are typical or exclusive of corpus-based work, as corpora are prominently used in PL and UBL. I will now briefly mention strands of research that had to be excluded although they could very well fit under the title of this chapter.

A lot of research on German in psycholinguistics is not featured here. For example, Bader & Portele (2019) test Kehler et al.'s (2008) Bayesian model of pronoun resolution in a number of experiments. Similarly, clearly processing-oriented studies like Marusch et al. (2019)—a priming study on regular and irregular verbal inflection—fell outside of the scope of this chapter. Additionally, I discuss only well-established methods. As a result, Baayen & Smolka (2020)—while providing interesting insights into German morphology—is an example of a paper I could not discuss in detail, because (i) it is processing-oriented, and (ii) it uses Naïve Discriminative Learning, a bevahiouristic learning algorithm that is not in the linguistic mainstream.

Some quantitative methods that have been occasionally applied to German but are not mainstream methods also do not receive much attention here. For example, Wulf (2002) models German noun plurals using Analog Modelling (AM, Skousen 1989:!), a similarity-/exemplar-

[1] Not all of the works included in this overview commit to PL, UBL, or COG. However, I consider all of them relevant to the probabilistic programme and their results interpretable under a probabilistic perspective.

based algorithm that does not learn rules or abstractions but classifies unseen exemplars by comparing them to known exemplars. Since AM is not embraced by the linguistic mainstream, I do not discuss it here. The same goes for work in Computational Linguistics. While linguistically informed approaches in Computational Linguistics could be of interest to linguists, the technical nature of many papers renders them incompatible with the conception of this chapter. See, for example, Köper & Schulte im Walde (2017), Köper (2018), Schulte im Walde et al. (2018) on the semantic classification of German complex verbs.

There is little exchange between quantitatively-oriented COL, UBL, and PL and so-called Quantitative Linguistics (QL) as established in Eastern European countries and Germany in the second half of the 20th century (Köhler 2004). QL relies heavily on 'counting and measuring' (Köhler 2004: 2), early mathematisation, and the formulation of mathematical laws. A famous example is the Altmann-Menzerath Law, which states that, the longer a linguistic unit is, the shorter its constituents will be (Altmann 1980, Cramer 2004). Reasons for the very limited exchange between the camps likely include a difference in epistemological philosophies. While Köhler (2004: 6) rejects the idea, QL has strong positivist and inductivist tendencies, which are incompatible with the far-reaching deductive inferences desired in cognitive/usage-based circles. Due to these incompatibilities, works on German in QL cannot be taken into account.

Finally, while diachronic research features prominently in this chapter, quantitative work on the history of German that does not fit into the three main areas covered in this chapter, had to be left out. For example, Neels & Hartmann (2017) perform a quantitative analysis of the grammaticalisation paths of (*ein*) *wenig* and (*ein*) *bisschen* (both 'a little') by tracking their token frequencies across centuries as well as their affinity to combine with concrete and abstract nouns and their increasing tendency to combine exclusively with the indefinite article *ein*. Unfortunately, the paper does not fit into any of the major sections of this chapter.

The scope of this chapter thus outlined, I focus on three main areas of research. In Section 2, I discuss the analysis of productivity. Section 3 is dedicated to the quantification of co-occurrence tendencies. Section 4 deals with the probabilistic modelling of alternations.

2 Quantifying productivity

2.1 How productivity is quantified

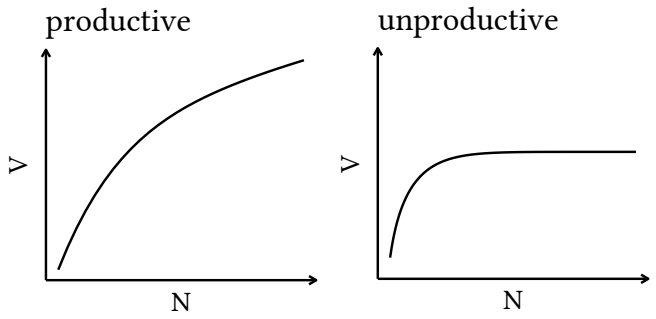
Baayen (2009: 900) defines productive morphological categories as those with growing membership. A suffix that readily combines with new bases is productive. A suffix that never or rarely combines with new bases is unproductive. Baayen argues that a strict distinction between fully productive and fully unproductive rules or processes is unreasonable in the light of findings from cognitively oriented studies that show that individual linguistic experience, processing constraints, communicative contexts, social styles, registers, etc. shape the grammar of speakers and communities (also Baayen 1993: 181).

Mathematically, three measures of productivity are popular (Baayen 2001, Baayen 2009). *Realised Productivity* is the number of types of a category C in a corpus. A category that accounts

for many types has achieved greater productivity in the past than a category that accounts for less types. *Expanding Productivity* is the number of hapax legomena in a corpus (words occurring exactly once) which belong to C divided by the total number of hapax legomena. A high number of hapax legomena means that the category is in the process of expanding.[2] Finally, *Potential Productivity* $\mathcal{P}$ is calculated as the number of hapax legomena of C divided by the total number of members in C. It ranges from 0 (totally unproductive process, no members of C are hapax legomena) to 1 (fully productive process, all members of C are hapax legomena). All of these measures are sensitive to corpus size and should not be compared across corpora.

$\mathcal{P}$ can be derived from so-called vocabulary growth curves (VGCs). As one goes through a corpus token by token and counts the types (V for *vocabulary*) and tokens (N) of a productive process (or construction), new unseen types will be found since the process generates hapax legomena. The vocabulary associated with the process keeps growing. As one does the same for an unproductive process, the vocabulary stops growing at some point. The resulting VGCs resembles the idealised curves in Figure 1 from Pankratz (2019).

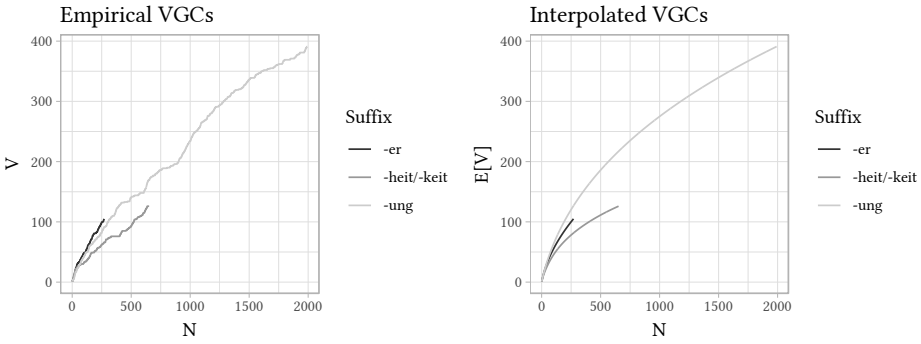
Figure 1. Idealised VGCs



In reality, VGCs are not as smooth as in Figure 1, and Baayen (2001) introduces an interpolation method. Figure 2 from Pankratz (2019) shows actual and interpolated VGCs for the derivational affixes *-er*, *-heit*/*-keit*, and *-ung* in the RIDGES corpus (Odebrecht et al. 2017). Baayen (2001) shows that $\mathcal{P}$ is the slope of the tangent of the VGC at its endpoint.

[2] Following Baayen (1989), the count of hapax legomena is used as an estimate of the count of actual neologisms. This is a just a rough approximation as corpora of limited size also contain many hapaxes which are not actually neologisms.

Figure 2. Empirical and interpolated VGCs for derivational suffixes in RIDGES



Finally, if one makes the reasonable assumption that vocabulary is finite, then interpolated VGCs will always flatten out at arbitrary sample sizes, but VGCs for productive processes will flatten out much later and at higher values for V compared to less productive processes. Once they have flattened, the curves are comparable. To overcome the problem that corpus samples can never be obtained at arbitrary sizes, extrapolations in the form of Large Number of Rare Events models (LNRE) are available (Baayen 2001), the most widely used being the finite Zipf-Mandelbrot model (fZM; Evert 2004).[3] As already mentioned in Baayen & Tweedie (1998: 145) and demonstrated in Pankratz (2019) for diachronic German data, even such models depend on the sample size used for the interpolation.

2.2 Morphological productivity in German

Gaeta & Zeldes (2017) look at a kind of synthetic noun + noun compound in German where the second noun is an agent noun derived from a verb using *-er*, as in *Vögelbeobachter* (*Vögel* 'birds' + *Beobachter* 'watcher' from *beobachten* 'watch'), and where in many cases a syntactic object–verb construction is also available (*Vögel beobachten* 'to watch birds'). They find in a corpus study that there are three classes of such constructions: (i) those where the syntactic realisation exists but not the compound, (ii) those where both options occur, and (iii) those where only the compound exists but not the syntactic realisation. First, to test whether verbs that productively combine with objects in syntax also combine productively with nouns in synthetic compounds, the authors calculate the realised productivity for the verb lemmas that occur in compounds and the objects in VP constructions. However, they find only a weak correlation with many outliers and reject the hypothesis of a direct syntactic motivation for the productive formation of the compounds. They then look at the very productive verbs in the compounds, finding that only few of the hapax legomena (between 1.5% and 38.2% depending on the verb) are also realised as VPs. The authors conclude that the morphological pattern is independent of VP formation and provide a model of the phenomenon in construction morphology.

Hein & Engelberg (2017) is an exploratory attempt to structure the morphologically complex lexicon by using productivity measures. They show that colour terms like *blau* 'blue' and

[3] The Zipf-Mandelbrot Law is omnipresent in linguistic and other phenomena. See its application in German diachronic linguistics in Koplening (2018).

orange ‘orange’ have different Realised and Expanding Productivity in nouns + adjective compounds such as *kirschrot* ‘cherry red’. However, the order of of the adjectives is the same across both measures. For example, *blau* has the highest Realized Productivity and the highest Expanding Productivity. Thus, *blau* has historically been the most productive compound-forming colour term, and it still is the most productive one. They also find that *orange* surpasses *blau* in $\mathcal{P}$. This is plausible, as there are relatively few compounds with *orange* already in use (low Realised Productivity), and *orange* thus has more potential to form unseen compounds. In a broader perspective, the authors are interested in causes of such differences in productivity. To test the hypothesis that the semantics of the head might be a relevant factor, they analyse the productivity of near-synonyms expressing affective states like *Angst* ‘fear’ / *Furcht* ‘dread’ and *Wut* ‘anger’ / Zorn ‘wrath’ in another corpus study. They find that the productivity of the whole group of affective terms is nowhere near converging, nor do the near-synonymous pairs group together in terms of productivity. Similar results are reported to obtain for colour compounds. The authors suggest the use of other types of analysis where the productivity of semantic subclasses of such compounds is used instead of single colour/affective terms (such as colour compounds with an intensifier).

Hein & Brunner (2020) follow up und Hein & Engelberg (2017). The authors look at the morphological complexity of the head noun as a factor influencing that head noun’s productivity in compound formation. The core hypothesis is that derived heads form new compounds less productively than non-derived heads. Based on automatic morphological analyses, they divide their data set into three categories: (i) compounds with simplex heads, (ii) compounds with morphologically derived heads, and (iii) compounds with recursively compounded heads. $\mathcal{P}$ is used with resampling to resolve the issue that comparisons of two measures of $\mathcal{P}$ is only possible between corpora of the same size. They repeatedly take samples of 2,000 tokens 1,000 times and compare the distribution of the results. After correcting for some biases in their sample, compound heads turn out to be the least productive group (averaging around 0.725), followed by simplex heads (averaging around 0.775), and derived heads are most productive (averaging around 0.825).

In Schäfer & Pankratz (2018), we measured productivity to allow for an informed selection of candidates for subsequent studies. The paper deals with so-called linking elements in German noun + noun compounds. Such compounds are almost always written as one word, and in approximately 40% of all types, a morphological element without a clearly specifiable function is inserted in between the nouns. In many cases, this linking element can be omitted: *Briefesammlung* ‘letter collection’ (from *Brief* ‘letter’, *Sammlung* ‘collection’ with an *-e* link) versus *Briefsammlung*. Many linking elements are formally identical to the plural of the first noun (the non-head), and we investigated whether plural meaning (induced via collective head nouns like *Sammlung*) had an effect on the presence or absence of such plural-like linking elements. Since previous accounts denied that there are nouns that alternate between a plural-like link and a non-plural-like link, we used the full compound analysis available in DECOW16 (Schäfer & Bildhauer 2012, Schäfer 2015) to extract all non-heads from compounds. Then, we calculated $\mathcal{P}$ for each of them with and without the plural-like link, showing that actually

hundreds of non-heads are in fact productive in both constructions. Further corpus analyses and a behavioural experiment were conducted informed by these productivity scores (Section 3.2).

2.3 Productivity in German syntax

Syntactic constructions are also subject to graded productivity. While sometimes raw type and token frequencies are used to measure the productivity of constructions (see Madlener 2018 on a German participle construction), advanced productivity analyses are also applied. Zeldes (2012) examines the German adposition *wegen* 'because of'. It comes in three syntactic variants: (i) as a preposition governing genitive case (*wegen des*$_{\text{Gen}}$ *Wagens*$_{\text{Gen}}$ 'because of the car'), (ii) as a postposition governing genitive case (*des*$_{\text{Gen}}$ *Wagens*$_{\text{Gen}}$ *wegen*), and (iii) as a preposition governing dative case (*wegen dem*$_{\text{Dat}}$ *Wagen*$_{\text{Dat}}$). The variants differ in diachronic and stylistic status: (ii) is the oldest variant and sounds distinctly archaic, (i) is the current norm, and (iii) is the newest and sounds colloquial to many speakers. Under such circumstances, it is expected that the archaic postposition should combine less productively with nouns than the other two variants. Indeed, the author finds that the postposition has a lower $\mathcal{P}$ (0.37) than the preposition with the genitive (0.58). Furthermore, the (extrapolated) VGC corroborates this result in that the gap between the preposition and the postposition grows constantly. After a discussion of the reliability to the VGCs (involving repeated sampling), Zeldes looks at all three alternants and finds that the preposition with the genitive is most productive (0.60), followed by the preposition with the dative (0.56), and the postposition (0.37). Again, the extrapolated VGCs paint a similar picture. While all variants are productive, "usage of the different variants is not identical, as postpositional *wegen*, in keeping with intuitive predictions, is less frequent, has a substantially smaller vocabulary and is markedly less likely to manifest hapax legomena at an equal sample size" (Zeldes 2012: 113).

2.4 Productivity in the history of German

Several problems plague diachronic quantitative analysis of productivity. First and foremost, the data are often sparse, which leads to massive fluctuations in any counts across the diachronically ordered corpora. Second, productivity measures are sensitive to corpus size. In this section, advances in the analysis of productivity in diachrony are discussed, all of which try to solve these problems.

Hartmann (2016) looks at deverbal action nouns with *-ung*, comparing them to nominalised infinitives, the closest alternative to *-ung* nominalisations. The two morphological constructions are viewed as having undergone conceptual changes. While *-ung* nominalisations acquire a stronger nominal character and are used in explicit naming, the nominalised infinitive retains the ability to form ad hoc nominalisations and remains process-oriented. Hartmann divides the GerManC corpus of texts from between 1650 and 1800 (Durrell et al. 2007) and the Mainzer Frühneuhochdeutschkorpus (MzFnhd) into slices and calculates a range of productivity measures for each slice in order to observe the development over time. Although GerManC is balanced for three major time slices, Hartmann partitions it further into smaller slices. The

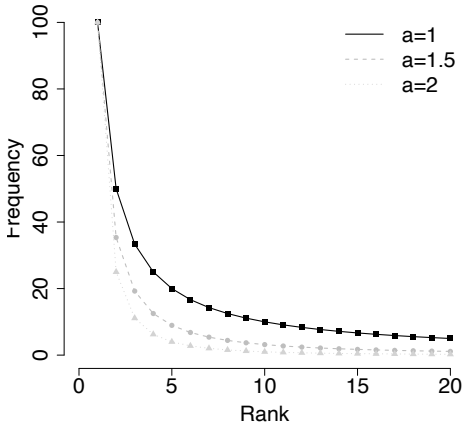
slices thus created are not balanced in terms of their (sub)corpus size, which is a problem for $\mathcal{P}$ due to its dependence on sample size. The author acknowledges this but cavalierly suggests that for an examination of rough trends the problem is irrelevant, and that careful interpretation solves the problem. He observes that the raw token frequency of all examined nominalisations rises over the time periods covered in GerManC and MzFnhd, which is interpreted as a function of a stylistic change which favours the expression of abstract concepts and reification more and more (*Nominalstil* 'nouny style'). The Realised Productivity of *-ung* nominalisation rises over time in both corpora. However, a significance test for Kendall's τ does not produce a significant result for GerManC. For nominalised infinitives, the picture is even less clear, especially given the strong fluctuations due to varying sizes of the quite small corpus slices. It appears as if nominalised infinitives remain at a stable level of Realised Productivity. When measuring diachronic tendencies in $\mathcal{P}$, Hartmann first reports that across the seven slices covering 200 years in the MzFnhd corpus, there is no clear tendency for either *-ung* or infinitives. Then, he creates slices covering decades and even individual years in GerManC. With such unrestricted shifting of the parameters of his analysis, he manages to show that—as expected—nominalisations with *-ung* lose some potential productivity, while nominalised infinitives remain at a fixed level. The problems with strong randomness inherent in the data as encountered by Hartmann might very well be the result of the incomparability of productivity measures between samples of different sizes. It remains to be seen whether the methodology in Hartmann's pioneering work is sound.

Hartmann (2018) extends Hartmann (2016). Still working on *-ung* nominalisations, he uses six balanced 50-year slices from between 1600 and 1900 from the *Deutsches Textarchiv* (DTA; Geyken et al. 2018). Again, he calculates Potential and Expanding Productivity measures. Up until the eighteenth century, he observes an increase in type and token frequency of *-ung* nominalisations and a corresponding drop in $\mathcal{P}$ as the suffix loses its potential to combine with verbs with which it previously had not combined. The fact that 'the nineteenth century sees an—albeit very slight—increase in the potential productivity of the pattern' (Hartmann 2018: 86) should not be interpreted due to the problems of $\mathcal{P}$ and sample size. Therefore, he re-analyses the data in just three slices covering one century each. A fZM model (with bootstrapping to do justice to the large variance in the sample and its small size) is used to calculate extrapolated $\mathcal{P}$, which is then found to decrease monotonically across the three centuries as expected. Thus, there is evidence that a DTA subcorpus with century-wide slices and fZM extrapolation results in the expected outcome.

Pankratz (in prep.) discusses a new solution to the problem of different sample sizes in comparisons of productivity. Type frequencies of the instances of a morpheme (or any linguistic unit) follow a known power law frequency distribution first described by Zipf (1949). The law states that for the types ranked by most to least frequent in decreasing order, the absolute frequency drops at a very high rate. Very few types have very high frequency and most types have a very low frequency with a varying number of types in the intermediate frequency range. A parameter denoted *a* in the formulation of the law determines how sharply the frequency drops as the rank goes down. See Figure 3 for a simplified illustration of the 20 most frequent types of three fictional processes with different values for *a*. As the author argues, the higher the

productivity of a process is, the slower the frequency drops towards the lower-ranked types. This is because for an unproductive process there are a few established types, and no significant number of neologisms is produced, some of which could come into use and gain intermediate frequency. To quantify this, Pankratz uses the corresponding probability distributions' entropy (Shannon 1948). The entropy of a distribution measures (in bits) how much information is needed on average to encode random events drawn from that distribution. Since the entropy of a spread-out distribution (high productivity) is higher than the entropy of a distribution that drops sharply (low productivity), Shannon entropy can be used as a measure of productivity. Even more importantly, rank/frequency distributions are self-similar, meaning they have the same shape at any scale. It follows that once the samples are large enough to produce a stable measure of entropy, sample size does not matter anymore. Bits as a measure of productivity can thus be compared across samples of different size. Pankratz demonstrates this new approach in two studies. First, she compares the productivity of 35 derivational suffixes in German, finding that obviously productive suffixes like *-ung* (8.84 bits) and *-er* (7.84 bits) are much more productive than ones only occurring with loanword bases like *-end* (1.57 bits). For tracking changes in productivity over time, samples might be too small to produce stable entropy measures. However, the actual entropy can never be lower than measured, even in small samples, and Pankratz suggests treating them as lower bounds, significantly restricting the potential slope of the diachronic development. She finds that in the RIDGES corpus (see Section 2.1) *-er* and *-heit* show no clear tendency, but *-ung* likely gains in productivity.

Figure 3. Toy rank/frequency plot of three processes with different values for *a* with the frequency of the most frequent type fixed at 100



3 Quantifying item-specific effects: collocations and collostructions

3.1 How collo-phenomena are quantified

Individual linguistic items co-occur with other linguistic items within windows of words, within sentences, within texts, etc. Collo-research deals with the frequencies of such co-occurrences and devises methods to find pairs that co-occur with unusually high frequency (positive association) and with unusually low frequency (negative association). Associations between

lexemes are called *collocations*, and associations between lexemes and constructions are called *collostructions*. Whether positive or negative associations have a meaningful linguistic interpretation or are simply an epiphenomenon is disputed (see Evert 2008 on collocations). However, Collostructional Analysis (CA) was explicitly introduced to capture item-specific associations under a constructivist usage-based view where prototype effects influence frequencies, and statistical associations between words and constructions might be part of the schematic knowledge associated with the constructions (Stefanowitsch & Gries 2003, Gries & Stefanowitsch 2004, Stefanowitsch & Gries 2005). Furthermore, CA has led to a revised status of negative evidence in corpus linguistics (Stefanowitsch 2006). It had been claimed previously that the absolute absence of some linguistic item from a corpus is intrinsically meaningless because it could be a mere accident as corpora of finite sizes never contain the whole spectrum of potentially rare linguistic items. Under CA, he notion of absolute absence from a corpus is replaced with a relative notion of absence. If a specific verb occurs very rarely in the passive in a corpus, CA reveals whether the verb would be *expected* to occur more often in the passive given its frequency and the frequency with which other verbs occur in the passive.

All collo-type analyses compare observed frequencies O as counted in a corpus with their expected frequencies E. In collocation research (Evert 2004, 2008), O is the count of two words w_1 and w_2 co-occurring in a window of n words (often $n = 5$), in a sentence, or in a syntactic relation like verb and object. E is a function of the overall frequencies of these words in the corpus (f_1 and f_2) and the total number of spans in the corpus, which is essentially the corpus size. If N is the corpus size in tokens, $E = f_1 \cdot f_2 / N$. Arguably the simplest comparison of O and E is achieved by calculating O / E, which is between 0 and 1 if the words co-occur less frequently than expected and above 1 if they co-occur more frequently than expected. A logarithmised variant of this quotient which is easier to interpret is the Mutual Information score $MI = log_2 O / E$. It can be used to rank collocations relative to one another. However, there is a host of other such simple measures of comparing observed and expected frequencies such as modifications of MI, the z- and t-scores, and log-likelihood ratios.

More advanced collocation measures and CA calculate statistics over contingency tables of frequencies. These statistics quantify the divergence between observations and expectations. Using Collostructional Analysis for illustration, the contingency table is constructed is as in Table 1 for a word **w** and a construction **c**.

Table 1. Basic contingency table as used in CA

	c	**¬c**	
w	o_{11}	o_{12}	R_1
¬w	o_{21}	o_{22}	R_2
	C_1	C_2	T

Spelled out, o_{11} is the corpus frequency of **w** occurring in **c**, o_{12} is the frequency of **w** occurring in constructions other than **c**. Analogously, o_{21} is the frequency of words other than **w** occurring in **c**, and o_{22} the frequency of words other than **w** occurring in constructions other than **c**. R_1 is the overall frequency of **w** in the corpus, R_2 the frequency of all words except **w**, C_1 is the

number of instances of **c** and C_2 the number of the other constructions. From the marginal sums labelled R and C we can calculate expected values for the four cells under the assumption that **w** is neither attracted nor repelled by **c** using essentially percentage calculation such as $e_{11} = R_1 \cdot C_1 / T$. All statistical collo-approaches quantify the total divergence between the observed values and the expected values to see whether the combination of **w** and **c** is more or less frequent than would be expected. The original papers on CA such as Stefanowitsch & Gries (2003), Gries & Stefanowitsch (2004), and Stefanowitsch & Gries (2005) used p-values from Fisher Exact Tests for ordering collexemes by association strength. Schmid & Küchenhoff (2013) and Küchenhoff & Schmid (2015) point out that p-values are measures of evidence, not effect strength. Hence, measures of effect strength like Cramér's v from a χ^2 test or odds ratios are more appropriate. Gries (2015) replies that p-values appear to do the job on the practical side and that CA is also used with other statistics. Schäfer (2018) and Schäfer & Pankratz (2018) to be discussed below are examples where Cramér's v was used.

Various types of CA have been developed which differ primarily in how they define **w**, ¬**w**, **c**, and ¬**c** (Stefanowitsch & Gries 2009). In *collexeme analysis*, **w** and **c** are understood as coming from a class of lexemes **W** such as verbs and constructions **C** where elements of **W** can fill a slot (potential collexemes). The ¬ operator simply forms the complement of **w** in **W** and of **c** in **C**. This allows for the broadest measurement of how strongly individual words are attracted or repelled by **c**. In *distinctive collexeme analysis*, ¬**c** is simply a single alternative construction, not a whole class of constructions. This provides a quantification of the affinity of **w** towards **c** and ¬**c**. *Covarying-collexeme analysis* is also described as a variant of CA. However, it is quite different as it compares two words $\mathbf{w_1}$ and $\mathbf{w_2}$ occurring in two slots of a specific construction, for example a verb-direct object construction. In such a construction, $\mathbf{w_1}$ could be the verb *ask* and $\mathbf{w_2}$ the noun *question*. covarying-collexeme analysis quantifies how strongly these two attract each other in the verb-object construction by comparing their frequencies to those of other transitive verbs ($\neg\mathbf{w_1}$) and other direct objects.

3.2 Collocation analyses of German

Collocations have been shown to be useful in the data-oriented extraction of word pairs that co-occur because some interpretable linguistic relation holds between them. As the status of surface word associations is unclear, the evaluation of such extraction results is of high importance. Krenn (2000) and Evert & Krenn (2001) compare collocation measures with respect to their potential for extracting German verb-preposition collocations. The same was shown for adjective-noun pairs in Evert & Kermes (2003). In such evaluations, collocations are extracted without analysing linguistic structure. The results are compared to gold standards of interpretable collocations extracted from digital dictionaries, manually parsed treebanks, or manual annotations. In some tasks, gold standards can be recovered quite well from surface collocations (Evert 2004). Collocations have also been used in the extraction and analysis of idiomatic language and multi-word units. See the contributions in Fellbaum (2007) for applications to German. Digital lexicography is another area where collocations are used (e.g., Lemnitzer & Geyken 2015). German online dictionaries like the *Digitales Wörterbuch der Deutschen Sprache* (DWDS) offer collocation analyses for general use (Geyken 2011).

3.3 Collostructional Analysis of present-day German

In Schäfer (2018), I analysed a morphosyntactic alternation in present-day German. In measure constructions semantically similar to the English *a glass of good wine*, the case of the kind-denoting noun (*Wein* 'wine') either is in the genitive as in (1a), or it agrees in case with the measure noun (*Glass* 'glass') as in (1b).

(1) a. Wir trinken [[ein Glas]$_{\text{Acc}}$ [guten Weins]$_{\text{Gen}}$]$_{\text{Acc}}$.

we drink a glass good wine

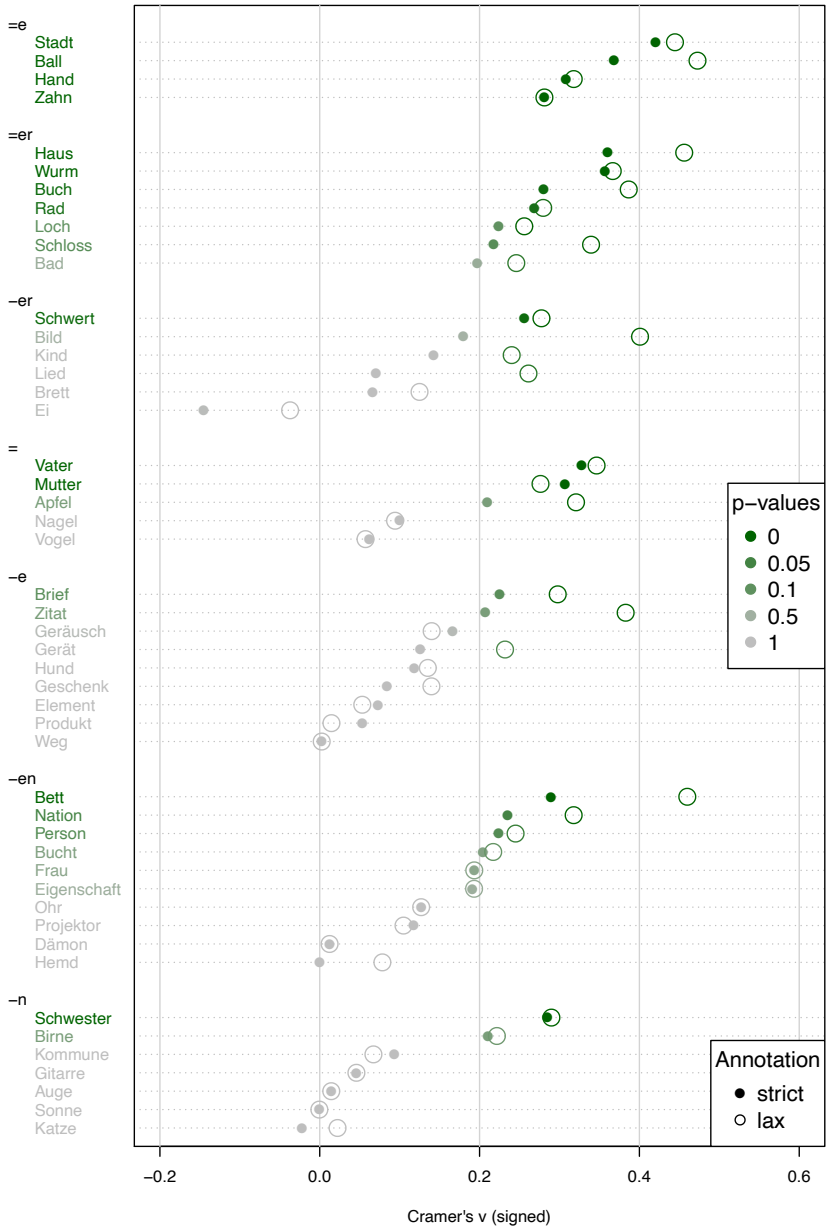
'We drink a glass of good wine.'

b. Wir trinken [[ein Glas]$_{\text{Acc}}$ [guten Wein]$_{\text{Acc}}$]$_{\text{Acc}}$.

I quantified how strongly the two constructions attract the individual kind lemmas and measure lemmas. The resulting values were used in a multifactorial model to control for item-specific effects. While simple quotient $f(\mathbf{w} \text{ in } \mathbf{c}_{genitive}) / f(\mathbf{w} \text{ in } \mathbf{c}_{agreement})$ had high predictive power, p-values from a distinctive collexeme analysis did not. This demonstrates that the usefulness and interpretation of methods like CA and the concrete measures used need constant testing and elaboration.

In Schäfer & Pankratz (2018) already discussed in Section 2.2, we used distinctive collexeme analysis of 48 nouns appearing as first (non-head) nouns in noun + noun compounds which can be used with a plural-like linking element. The purpose of the analysis was to quantify how strongly they are attracted by the construction with the linking element. In the analysis we compared each noun as **w** with the other 48 nouns as ¬**w** in the construction with the linking element as **c** and without the linking element as ¬**c**. We used (signed) Cramer's *v* as a measure of effect strength and amended them with Šidák-corrected p-values from (Monte Carlo-simulated) χ^2 tests for illustration. The results are shown in Figure 4. The nouns most strongly attracted to the construction with the plural-like linking element are the ones where the plural marking is formally most salient and in many cases uniquely associated with nominal plural, for example those with the umlaut such as *Städte* 'cities' from singular *Stadt*. In the interpretation, we link these results to cognitive principles.

Figure 4. Effect strengths from Schäfer & Pankratz (2018: 347)



Dekalo & Hampe (2017) focus on the interpretation of lists of collexemes as produced by CA, and they use cluster and network analysis to help with the interpretation. The paper deals with *vermögen* and *bekommen* in a modal reading allegedly in competition with *können* expressing possibility or capability. See simplified examples in (2).

(2) a. Wir vermögen zu feiern.

we can to celebrate

We know how to party.

b. Sie bekommen das Lied zu hören.

they get the song to hear

They can listen to the song. / They get to listen to the song.

These emerging auxiliaries require infinitives with the particle *zu* as seen in (2).[4] Dekalo & Hampe's goal is to profile the semantics of these modal constructions with *vermögen* and *bekommen* by analyzing their collexemes. To this end, they first perform a collexeme analysis in order to find 'significant' collexemes of the two constructions. They acknowledge that the analysis is insensitive to verbal polysemy, and annotate the tokens of the significant collexemes of each construction using a coarse taxonomy of verb classes from the GermaNet lexical database (Hamp & Feldweg 1997). Based on this, they find that *vermögen* predominantly attracts verbs of cognition, social interaction, communication, etc., whereas *bekommen* attracts verbs of perception, consumption, cognition, etc. GermaNet also provides word sense similarity networks, and by calculating pairwise similarity distances within the two sets of collexemes (for *vermögen* and *bekommen*), the authors add another layer of interpretation to the collexeme analysis. First, they cluster the two sets based on the similarities. The clusters found for each construction shed light on the verbs it attracts and thus its semantics. Clusters for *bekommen* contain verbs of consumption, perception, etc. A network analysis of the same data provides additional visualisation. In the discussion, they conclude that *vermögen* is a more strongly grammaticalised modal verb and *bekommen* is not, mostly because *bekommen* attracts a much narrower and more specific range of verb classes. The paper has shortcomings in that it ignores previous research from German linguistics and consequently has unusual premises. Jäger (2013: 98–115) provides a detailed account of modal constructions with *bekommen*. It is known that *bekommen* (as opposed to the highly archaic *vermögen*) itself is only weakly modal, but that it typically combines with other modals like *können* 'can' such as in (3).

(3) In dieser Kneipe kann man einiges zu hören bekommen.

in this pub can one a.lot to hear get

One can/gets to listen to a lot of stories in this pub.

3.4 Collostructional Analysis in diachronic linguistics

Hilpert (2008: 131–155) analyses the German analytical future construction with the auxiliary *werden* 'become' and an infinitive. Apart from the future meaning, the construction can also be used to express an epistemic modality. Hilpert's overarching research goal is a diachronic analysis of grammaticalization paths and the corresponding functional/semantic changes in Germanic languages. First, he uses collexeme analysis to find the verbs most strongly attracted

[4] The authors do not mention that proper grammaticalised modal verbs in German always require the bare infinitive without *zu* (Eisenberg 2020: 93–97). Even semi-grammaticalised emerging modals like *brauchen* 'need', which used to combine with *zu* infinitives are losing the *zu* particle (Askedal 1997, Reis 2005).

by the *werden* future in present-day German. In the qualitative analysis of the resulting list of collexemes, he finds that existential *geben* 'exist' (literally 'give') is the strongest collexeme. Furthermore, continuative verbs like *dauern* 'take', verbs denoting abstract processes like *zunehmen* 'increase', and speech act verbs like *entscheiden* 'decide' are strong collexemes. It appears that most of the verbs attracted by the construction select subjects with inanimate referents, and there are only few stative verbs among them. The author argues that the analysis shows that in language use, the construction predominantly expresses future tense, and the epistemic interpretation is secondary. Second, Hilpert performs *diachronic distinctive collexeme analysis*. This method also uses contingency tables, but it compares **w** (some verb in the *werden* construction) with ¬**w** (all other verbs in the same construction) across different time periods $\mathbf{t_1}$, $\mathbf{t_2}$, etc. The compared time periods are the 15th and 16th century (as one period) as well as the 18th, 19th, and 20th centuries. In the analysis Hilpert argues that the construction did not develop from one expressing intentions to a temporal construction (as previously claimed by others). There is some indication that the epistemic interpretation is secondary because static verbs as clear markers of the epistemic interpretation become more strongly attracted in the 18th century and later.

Hartmann (2016, 2018) (see Section 2.4) uses CA for tracking the diachronic development of the *-ung* nominalisation. I focus on Hartmann (2018: 100–110). His method is called *morphological cross-tabulation analysis* (yet another contingency table approach). It is used to quantify how strongly individual verbal bases are attracted by the *-ung* construction. In the contingency table, the rows correspond to the frequency of the verbal base **v** with the affix ($\mathbf{v_a}$) and the base without the affix ($\mathbf{v_{\neg a}}$). The columns correspond to all nouns **n** derived with the affix ($\mathbf{n_a}$) and all nouns not derived with the affix ($\mathbf{n_{\neg a}}$). Log-likelihood ratios are used to rank the bases attracted and repelled by *-ung*. Separate analyses are run for the 17th, 18th, and 19th centuries, followed by a comparison of the three. Hartmann finds that a well-known constraint of present-day German emerges across the period covered in his corpus, namely that durative and inchoative verbs are ineligible for *-ung* nominalisation (Demske 2000). The fact the quantitative analysis is able to reproduce these well-established results adds support to the validity of the method.

4 Describing and predicting outcomes of probabilistic processes: alternations

4.1 Methods in research on non-discreteness in language

Alternations are cases where two or more forms or constructions are available with only minor or subtle differences in acceptability, function, or meaning. Many studies of alternations have been published (Bresnan et al. 2007, Bresnan & Hay 2010, Bresnan & Ford 2010, Divjak & Arppe 2013, Gries 2003, Gries 2015, Gries 2017, Nesset & Janda 2010, Levshina 2016, Wulff 2003), and many of them are multifactorial: they model the influence of various factors on language users' decisions. A standard tool used in this type of analysis are Generalised Linear Models (GLMs) with a logit link function (essentially synonymous with *logistic regression*). A GLM models the variance in a response variable (which adjective comes first in a sequence of two, which suffix forming action nouns from verbs is chosen, etc.) that can be attributed to

fixed population effects (discourse status of referents, polarity of the embedding clause, etc.). It can then predict outcomes based on those influencing factors. Hierarchical GLMs (also called *mixed models* or GLMMs) have also been used to model additional item-specific effects such as tendencies of individual lemmas or genre- or speaker effects (Gries 2015). However, Schaefer (2020: 545) has some technical warnings about mechanical decisions to model certain variables always as random effects.

4.2 Studies on alternations and similar non-discrete phenomena in German

Schäfer (2018) provides an analysis of a case alternation in measure noun phrases (see Section 3.3 for an introduction). To understand the alternation, we have to look at neighbouring constructions. If there is no adjective and no determiner, the kind noun is always in agreement ([*ein Glas* [*Wein*]] 'a glass of wine'). If there is only a determiner and no adjective, the kind noun is always in the genitive ([*ein Glas* [*des Weins*]] 'a glass of the wine'). I derive prototypes for the agreement and the genitive variant from theoretical insights: (i) the genitive is preferred in higher registers (operationalised via two proxy variables), (ii) the agreement construction is more appropriate for highly grammaticalised measure nouns because it cannot be interpreted as a pseudo-partitive. For (ii), measure nouns are classified into groups of nouns denoting physical measures, containers, amounts, and portions (in descending order of grammaticalisation). Furthermore, it is hypothesised that non-cardinal determiners in the matrix (*einige Gläser leckerer Wein* 'some glasses of tasty wine') attract the agreement construction more than cardinal determiners (*zwei Gläser leckeren Weins* 'two glasses of tasty wine'). While these features define the prototypes, it is also assumed that item-specific effects play a role. The attraction of individual measure nouns and kind nouns towards the genitive construction is therefore measured independently (Section 3.3), and the attraction strength is included as a regressor in the GLMMs as second-level effects. In the corpus study of the alternation, evidence in favour of all hypotheses is found. Figure 5 shows effect plots for the grammaticalisation-related prototype features. The probability for the genitive rises across measure noun classes, and cardinal determiners favour the agreement construction. The item-specific attraction effects are also substantiated (Figure 6): the higher the independently measured attraction of an item to the genitive construction, the higher its probability of occurring in the genitive construction in the alternating case. The results are validated in a reading time experiment and a decision experiment.

Figure 5. Effect plots for the main grammaticalisation-related effects (Schäfer 2018: 753)

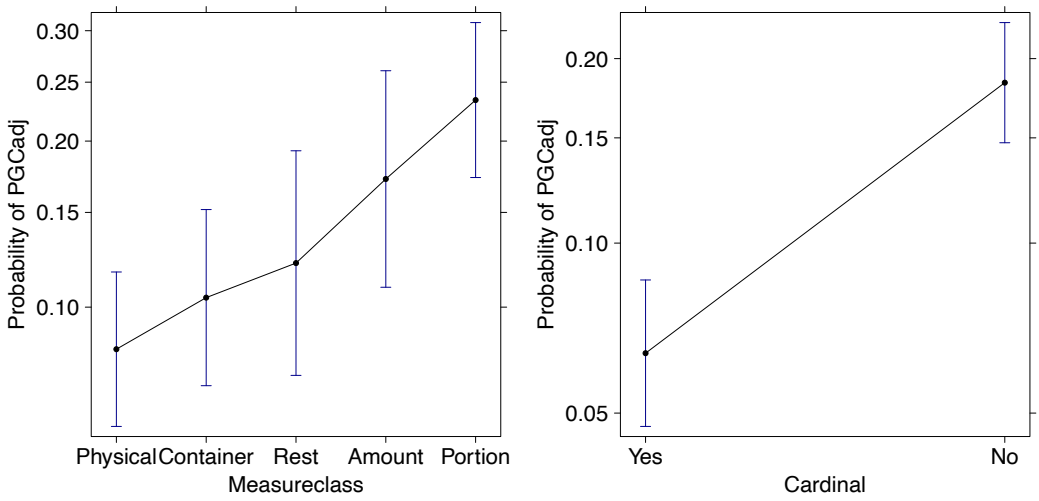
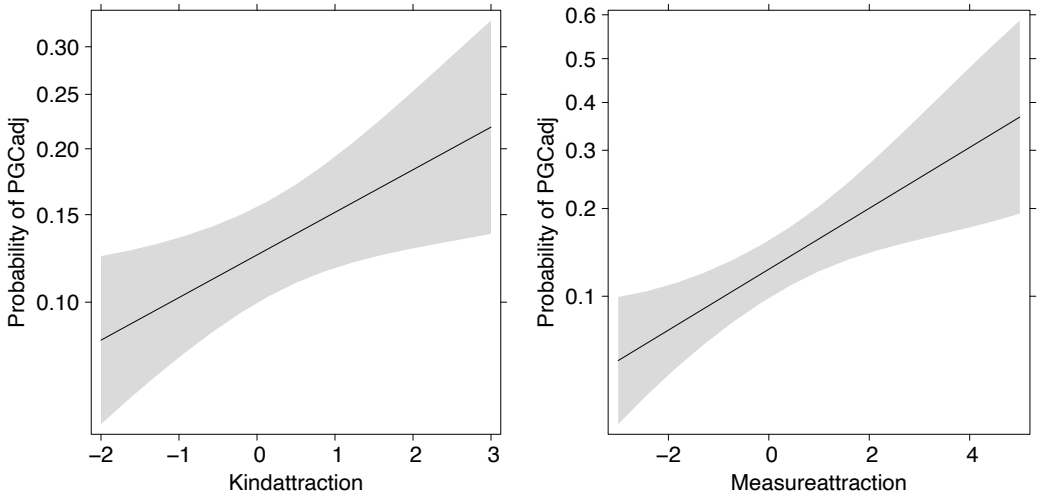


Figure 6. Effects plots for the collexeme attraction effect (Schäfer 2018: 752)



Schäfer (2019) analyses the inflectional alternation of weak masculine nouns. Weak nouns are a small class which does not inflect according to the same patterns as all other nouns. Köpcke (1993) argued that the weak nouns instantiate semantic and prosodic prototypes. The first prototype is either monosyllabic (*Bär* 'bear') or polysyllabic with penultimate accent and a final schwa syllable (*Matrose* 'sailor'). The second prototype is polysyllabic with ultimate accent (*Leopard* 'leopard'). Furthermore, most weak nouns denote humans or animals. I analyse an alternation in terms of these prototypes. Weak nouns are astonishingly acceptable when they inflect as strong nouns, which happens occasionally even in written German.[5] By hypothesis, the more prototypical of a weak noun a noun is, the less likely it should be to occur in the strong inflection. Table 2 shows the coefficients for all regressors used in the GLM analysis of the corpus data. The so-called intercept in regression models specifies a baseline outcome for

[5] The opposite (strong nouns inflecting like weak nouns) is completely inacceptable for most nouns.

defaults of the predictors (human denotation, polysyllabic words with schwa final, genitive case). All other coefficients (β) are positive, and the Odds ratios (O) are higher than 1, meaning that the corresponding values increase the chance of the alternation compared to the values on the intercept: animate or inanimate denotation lead to more cases inflected according to the strong pattern than human denotation. Also, the prosodic features corroborate the hypotheses derived from previous research.

Table 2. Estimates for the regressors of the GLM (abbreviated from Schäfer 2019: 404)

Regressor	**β**	**O**	**p**
Intercept (Human, Polysyllabic with final schwa)	−0.502	0.606	0.0001
Animate	0.664	1.943	< 0.0001
Inanimate	0.585	1.795	< 0.0001
Polysyllabic, ultimate accent	1.073	2.925	< 0.0001
Monosyllabic	2.066	7.891	< 0.0001
Polysyllabic, non-ultimate accent	2.831	16.963	< 0.0001

In Schäfer & Sayatz (2014) we apply similar methodology to corpus-based studies of a cliticised form of the indefinite article (the stem *ein* 'a', cliticised to *n*) which is typical of colloquial writing. Multiple corpus studies and multifactorial modelling are used to show that *n* is not just a general drop-in replacement for *ein* in colloquial writing, but that its occurrence is blocked and favoured under specific circumstances. Sentence-initially (with no host to cliticise onto) and after prepositions *n* is disfavoured. However, when the alternative would be bisyllabic, the clitic is preferred (*nem* instead of *einem* in the masculine dative singular).

In Schäfer & Sayatz (2016), we analyse *obwohl* 'although' and *weil* 'because' clauses with verb-second order based on the use of punctuation marks with this non-standard structure. Verb-second order is used only in independent sentences, and in standard written language *obwohl* and *weil* are subordinating particles triggering verb-last order as found in embedded clauses. In spoken German and colloquial writing, they have developed a non-standard variant which embeds verb-second order. By analysing the punctuation before and after the particles in a series of corpus studies, we show that *obwohl* with verb-second order has the status of a discourse particle that mostly introduces independent sentences. On the other hand, *weil* still tends to form subordinate clauses, even with verb-second order. In the paper, we introduce usage-based graphemics as a framework where the process of drawing inferences from the usage of graphemics means (in texts written with low normative pressure) to the grammatical system is systematised.

Heylen (2005) is a very early study of constituent order variation in the so called middle field where multifactorial statistics are used. In the middle field subject (S) and object (O) can occur in both SO and OS order. Heylen (2005) focusses on cases with a pronominal object and a full NP subject. Some influencing factors are the case of O, the thematic role of S, length differences between S and O, and discourse status of S. The author uses corpus data and logistic

regression, finding that all factors have at least some influence on ordering. A regression tree analysis complements the regression.

Bader (2020) goes back to Heylen (2005) and examines the order of pronominal objects and non-pronominal NP subjects. He apparently draws a clear distinction between a (likely non-probabilistic) grammar and performance (Bader 2020: 1065), but his results are relevant for probabilistic approaches. In a corpus study Bader (2020) tests the influence of various factors on constituent order. Logistic regression with stepwise elimination of factors (including a maximal interaction structure) is used. The results confirm and refine Heylen's (2005) findings. The length of the subject NP emerges as a robust predictor, further analysis showing that length cannot be reduced to other factors and vice versa.

Ellsiepen & Bader (2018) look at the order of nominal constituents in the middle field. They use a Harmonic Grammar (HG) and probabilistic Optimality Theory (OT) approach. From HG and OT, they capitalise on the idea that constraints in grammar are violable, and that the Harmony of a sentence is determined by the number of violated constraints and their weights. The paper aims at examining the relation between Harmony and acceptability. From previous approaches, the authors derive a number of potential constraints influencing acceptability depending on constituent order and retrieve Magnitude Estimation judgements from participants in four experiments.[6] The experiments are analysed using Linear Mixed Models, and the authors derive potential constraint rankings from the results. They find that most constraints discussed in earlier works play a role: animacy (animate before inanimate), constraints of default case ordering (nominative before accusative and dative), and definiteness (definite before indefinite).

Engelberg (2018) is a study of alternations in the argument structure (AS) of psych verbs. The approach is not fully multifactorial in that it looks at factors in isolation. German psych verbs like *ärgern* ('be/make angry') alternate between AS patterns such as $NP_{1/nom}$ $NP_{2/acc}$ ('NP_1 makes NP_2 angry'), CP NP_2 ('CP makes NP_2 angry'), $NP_{1/nom}$ $Refl_1$ [P NP_2] ('NP_1 is angry about NP_2'), etc. The author looks at four factors influencing the choice of such AS. He uses a method called *verb profiles*, which for each verb counts the argument realisation patterns (ARPs). ARPs are lists of thematic roles mapped onto their formal realisation (such as Experiencer→NP_{nom}). Corpus studies in newspaper text and a varying-genre corpus show diverse structural regularities (equal preference for complement clause realisation in stimulus-as-subject and stimulus-as-object constructions, dispreference for passivisation in experiencer-as-subject constructions) as well as genre-specific tendencies. Finally, he uses the ΔP measure (Baayen 2011) in another very large corpus study and shows that the entrenchment of ARPs (as constructions) is likely facilitated because they tend to be strongly associated with specific lexical verbs.

Willems et al. (2108) looks at two-way prepositions traditionally described as having a local meaning with the dative (*auf dem Stein* 'on the stone') and a directional meaning with the

[6] Magnitude Estimation is an experimental paradigm where participants judge the relative acceptability of stimuli.

accusative (*auf den Stein* 'onto the stone'). They closely examine two verbs occurring with such prepositions in a corpus study. The multifactorial analyses uses classification trees, an exploratory algorithm that creates a decision tree which models speakers' decisions to use either the dative or the accusative. The decision nodes represent individual influencing factors. Willems et al. show that factors like transitivity, grammatical voice, and perfect tense influence the choice of variants. Thus, the choice is thus not as simple as locality vs. directionality. The study also provides evidence that the preposition-case patterns are argument structure constructions in their own right.

Patil et al. (2020) discuss two constraints on the choice of the German demonstrative pronouns *d-* (forms like *der*, *die*, *das*) and *dies-* (forms like *dieser*, *diese*, *dieses*), both 'this'. They examine two factors influencing the choice: (i) formal and informal register, (ii) subject antecedent avoidance. Methodologically, the authors use forced-choice experiments and regression analysis. They show that there is a clear preference for *dies-* pronouns in formal registers, but that both types of pronouns have equally strong tendencies for subject avoidance.

Hartmann (2018) analyses the usage of two competing future constructions in German. As an alternative to the analytical future with the *werden* auxiliary, the present tense form can also be used to express references to the future. In a corpus study Hartmann (2018) looks at potential factors influencing the choice between the two constructions: genre, negation, active vs. passive, temporal distance (near vs. distant), and clause type (independent vs. subordinate). The data are analysed with random forests, a method which constructs a large number of decision trees from random subsamples of the input data and the predictors. The prediction of the forest is averaged over the predictions of the single trees. It is found that *werden* is preferred for events in the distant future and dispreferred in passives and negated clauses. The author fails to corroborate the corpus study in an experiment where participants rated the temporal distance of events expressed in the two constructional variants. He offers an explanation of this failure in terms of the design of the stimuli.

Finally, it should be noted that the grammar department at the Leibinz-Institute for the German Language (Mannheim) is currently working on a larger descriptive grammar of German that is based on corpus data (Bubenhofer et al. 2016) and applies the kind of quantitative modelling discussed in this section to alternation phenomena. While only preliminary studies have been published (Münzberg & Bildhauer 2020) the resulting work will likely make such analyses and techniques more accessible to a wider audience in German linguistics.

5 Conclusion and outlook

This chapter could only provide a short overview of some quantitative methods used in the analysis of German. However, the publications discussed show that there is innovation in the application and the development of quantitative methods coming from within German linguistics, both synchronic and diachronic. A major problem to solve in the future will be the interpretation of the results. Hopefully, this chapter has demonstrated that measures of productivity, measures of collocational and collostructional attraction, and the modelling of alternations—

much like any quantitative method—require substantial amounts of validation with respect to the theories they are supposed to support. Do these measures correspond to cognitive and social constructs, or are they merely artefacts of the corpora used and the calculations applied? Do the models discussed in Section 4 *strongly* support a probabilistic view of grammar, or could the effects also be interpreted as supporting a discrete conception of grammar? Most likely, even the authors discussed in this chapter would disagree on such matters. It is only through repeated testing using a variety of language data that we will eventually reveal what the appropriate interpretation of the quantitative effects apparent in the data might be. The German linguistics tradition with its deep centuries-old tradition of describing and analysing the German language can and should contribute further to this endeavour.

References

Altmann, Gabriel. 1980. Prolegomena to Menzerath's law. *Glottometrika* 2, 1–10.

Askedal, John Ole. 1997. Brauchen mit Infinitiv. Aspekte der Auxiliarisierung. *Jahrbuch der ungarischen Germanistik*, 53–68.

Baayen, R. Harald. 1993. On frequency, transparency, and productivity. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of morphology*, 181–208. Dordrecht: Kluwer.

Baayen, R. Harald. 2001. *Word frequency distributions*. Dordrecht: Kluwer.

Baayen, R. Harald. 2009. Corpus linguistics in morphology: morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. An international handbook*, 900–919. Berlin: Mouton De Gruyter.

Baayen, R. Harald. 2011. Corpus linguistics and naive discriminative learning. *Brazilian Journal of Applied Linguistics* 11, 295–328.

Baayen, R. Harald & Eva Smolka. 2020. Modeling morphological priming in German with Naive Discriminative Learning. *Frontiers in Communication* 5(17).

Baayen, R. Harald & Fiona J. Tweedie. 1998. Sample-size invariance of LNRE model parameters: problems and opportunities. *Journal of Quantitative Linguistics* 5(3), 145–154.

Bader, Markus. 2020. The position of object pronouns in the German middlefield. *Linguistics* 58(4), 1059–1115.

Bader, Markus & Yvonne Portele. 2019. The interpretation of German personal pronouns and d-pronouns. *Zeitschrift für Sprachwissenschaft* 38(2), 155–190.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen.

Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: processing dative constructions in American and Australian varieties of English. *Language* 86(1), 168–213.

Bresnan, Joan & Jennifer Hay. 2008. Gradient grammar: an effect of animacy on the syntax of 'give' in New Zealand and American English. *Lingua* 118, 245–259.

Bubenhofer, Noah, Marek Konopka & Roman Schneider. 2016. *Präliminarien einer Korpusgrammatik*. Tübingen: Narr.

Bybee, Joan L. & Clay Beckner. 2009. Usage-based theory. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 827–856. Oxford: Oxford University Press.

Cramer, Irene M. 2005. Das Menzerathsche Gesetz. In Reinhard Köhler, Gabriel Altmann & Rajmund G. Piotrowski (eds.), *Quantitative linguistics: an international handbook*, 659–688. Berlin & New York: De Gruyter.

Dekalo, Volodymyr & Beate Hampe. 2017. Networks of meanings: complementing collostructional analysis by cluster and network analyses. In Stefan Hartmann (ed.), *Yearbook of the german cognitive linguistics association*, 143–176. Berlin & Boston: De Gruyter Mouton.

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.

Divjak, Dagmar. 2016. Four challenges for usage-based linguistics. In Jocelyne Daems, Eline Zenner, Kris Heylen, Dirk Speelman & Hubert Cuyckens (eds.), *Change of paradigms – new paradoxes – recontextualizing language and linguistics*, 297–309. Berlin/Boston: De Gruyter Mouton.

Divjak, Dagmar & Antti Arppe. 2013. Extracting prototypes from exemplars what can corpus data tell us about concept representation? *Cognitive Linguistics* 24(2), 221–274.

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

Eisenberg, Peter. 2020. *Grundriss der deutschen Grammatik: Der Satz*. 5th edn. Stuttgart: Metzler.

Ellsiepen, Emilia & Markus Bader. 2018. Constraints on argument linearization in German. *Glossa* 3(1), 1–36.

Engelberg, Stefan. 2018. The argument structure of psych-verbs: a quantitative corpus study on cognitive entrenchment. In Hans C. Boas & Alexander Ziem (eds.), *Constructional approaches to syntactic structures in German*, 47–84. Berlin & New York: De Gruyter.

Evert, Stefan. 2004. *The statistics of word cooccurrences: word pairs and collocations*. Stuttgart: IMS, University of Stuttgart PhD dissertation.

Evert, Stefan & Hannah Kermes. 2003. Experiments on candidate data for collocation extraction. In *10th conference of the European chapter of the association for computational linguistics*, 83–86. Budapest.

Evert, Stefan & Brigitte Krenn. 2001. Methods for the qualitative evaluation of lexical association measures. In *Proceedings of the 39th annual meeting of the association for computational linguistics*, 188–195. Toulouse.

Fellbaum, Christiane (ed.). 2007. *Idioms and collocations: corpus-based linguistic and lexicographic studies*. Birmingham: Continuum Press.

Gaeta, Livio & Amir Zeldes. 2017. Between VP and NP: on the constructional types of German -er compounds. *Constructions and Frames* 9(1), 1–40.

Geyken, Alexander. 2011. Die dynamische Verknüpfung von Kollokationen mit Korpusbelegen und deren Repräsentation im DWDS-Wörterbuch. *OPAL - Online publizierte Arbeiten zur Linguistik* 2, 9–22.

Geyken, Alexander, Matthias Boenig, Susanne Haaf, Bryan Jurish, Christian Thomas & Frank Wiegand. 2018. Das Deutsche Textarchiv als Forschungsplattform für historische Daten in CLARIN. In Henning Lobin, Roman Schneider & Andreas Witt (eds.), *Digitale Infrastrukturen für die germanistische Forschung*, 219–248. Berlin & Boston: De Gruyter.

Grafmiller, Jason, Benedikt Szmrecsanyi, Melanie Röthlisberger & Benedikt Heller. 2018. General introduction: a comparative perspective on probabilistic variation in grammar. *Glossa* 3(1), 94.

Gries, Stefan Th. 2003. Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1, 1–27.

Gries, Stefan Th. 2015a. More (old and new) misunderstandings of collostructional analysis: on Schmid and Küchenhoff (2013). *Cognitive Linguistics* 26(3), 505–536.

Gries, Stefan Th. 2015b. The most underused statistical method in corpus linguistics: multilevel (and mixed-effects) models. *Corpora* 10(1), 95–126.

Gries, Stefan Th. 2017. Syntactic alternation research. taking stock and some suggestions for the future. In Ludovic De Cuypere, Clara Vanderschueren & Gert De Sutter (eds.), *Current trends in analyzing syntactic variation*, vol. 31 (Belgian Journal of Linguistics), 7–27. Amsterdam: Benjamins.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Co-varying collexemes in the into-causative. In Michel Achard & Suzanne Kemmer (eds.), *Language, culture, and mind*, 225–236. Stanford, CA: CSLI.

Hamp, Birgit & Helmut Feldweg. 1997. GermaNet – a lexical-semantic net for German. In *Proceedings of the ACL workshop automatic information extraction and building of lexical semantic resources for NLP applications*. Madrid.

Hartmann, Stefan. 2016. *Wortbildungswandel*. Berlin & Boston: De Gruyter.

Hartmann, Stefan. 2018a. Derivational morphology in flux: a case study of word-formation change in German. *Cognitive Linguistics* 29(1), 77–119.

Hartmann, Stefan. 2018b. Divergent theories, converging evidence: the constructional semantics of competing future constructions. In Beate Hampe & Susanne Flach (eds.), *Corpora, constructions, cognition: yearbook of the german cognitive linguistics association*, 91–114. Berlin & Boston: De Gruyter.

Hay, Jennifer B. & R. Harald Baayen. 2005. Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences* 9(7), 342–348.

Hein, Katrin & Annelen Brunner. 2020. Why do some lexemes combine more frequently than others? – an empirical approach to productivity in German compound formation. In Jenny Audring, Nikos Koutsoukos & Christina Manoulidou (eds.), *Rules, patterns, schemas, and analogy. online proceedings of the twelfth Mediterranean morphology meeting*, 28–41. Nicosia.

Hein, Katrin & Stefan Engelberg. 2017. Morphological variation: the case of productivity in German compound formation. In Nikos Koutsoukos, Jenny Audring & Francesca Masini (eds.), *Morphological variation: synchrony and diachrony. online proceedings of the eleventh Mediterranean morphology meeting*, 36–50. Nicosia.

Heylen, Kris. 2005. A quantitative corpus study of German word order variation. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: empirical, theoretical and computational perspectives*, 241–263. Berlin & New York: Mouton De Gruyter.

Hilpert, Martin. 2008. *Germanic future constructions: a usage-based approach to language change*. Amsterdam: Benjamins.

Jäger, Anne. 2013. *Der Status von bekommen + zu + Infinitiv zwischen Modalität und semantischer Perspektivierung*. Frankfurt, Main: Peter Lang.

Kapatsinski, Vsevolod. 2014. What is grammar like? A usage-based constructionist perspective. *Linguistic Issues in Language Technology* 11, 1–41.

Kehler, Andrew, Laura Kertz, Hannah Rohde & Jeffrey L. Elman. 2008. Coherence and coreference revisited. *Journal of Semantics* 25(1), 1–44.

Köhler, Reinhard. 2005. Gegenstand und Arbeitsweise der Quantitativen Linguistik. In Reinhard Köhler, Gabriel Altmann & Rajmund G. Piotrowski (eds.), *Quantitative linguistics: an international handbook*, 1–16. Berlin & New York: De Gruyter.

Köpcke, Klaus-Michael. 1993. *Schemata bei der Pluralbildung im Deutschen: Versuch einer kognitiven Morphologie*. Tübingen: Narr.

Köper, Maximilian. 2018. *Computational approaches for German particle verbs: compositionality, sense discrimination and non-literal language*. Stuttgart: Universität Stuttgart dissertation.

Köper, Maximilian & Sabine Schulte im Walde. 2017. Applying multi-sense embeddings for German verbs to determine semantic relatedness and to detect non-literal language. In *Proceedings of the 15th conference of the European chapter of the association for computational linguistics*, vol. 2, 535–542. Valencia.

Koplenig, Alexander. 2018. Using the parameters of the Zipf–Mandelbrot law to measure diachronic lexical, syntactical and stylistic changes – a large-scale corpus analysis. *Corpus Linguistics and Linguistic Theory* 14(1), 1–34.

Krenn, Brigitte. 2000. *The usual suspects: data-oriented models for the identification and representation of lexical collocations*. Saarbrücken: DFKI & Universität des Saarlandes dissertation.

Küchenhoff, Helmut & Hans-Jörg Schmid. 2015. Reply to "More (old and new) misunderstandings of collostructional analysis: on Schmid & Küchenhoff" by Stefan Th. Gries. *Cognitive Linguistics* 26(3), 537–547.

Lemnitzer, Lothar & Alexander Geyken. 2015. Semantic modeling of collocations for lexicographic purposes. *Journal of Cognitive Science* 16(4), 200–223.

Levshina, Natalia. 2016. When variables align: a Bayesian multinomial mixed-effects model of English permissive constructions. *Cognitive Linguistics* 27(2), 235–268.

Madlener, Karin. 2018. Type and token frequency effects on developing constructional productivity: the case of the German sein 'be' + present participle construction. In Hans C. Boas & Alexander Ziem (eds.), *Constructional approaches to syntactic structures in German*, 327–366. Berlin & New York: De Gruyter.

Marusch, Tina, Lena A. Jäger, Leander Neiß, Frank Burchert & Lyndsey Nickels. 2019. Overt language production of German past participles: investigating (ir-)regularity. *Language, Cognition and Neuroscience* 34(3), 289–308.

Münzberg, Franziska & Felix Bildhauer. 2020. Reihenfolge attributiver Adjektive. *Bausteine einer Korpusgrammatik des Deutschen* 1, 131–156.

Neels, Jakob & Stefan Hartmann. 2017. Reduction or expansion? a bit of both: a case study on the development of German degree modifiers. In Evie Coussé, Peter Andersson & Joel Olofsson (eds.), *Grammaticalization meets construction grammar*, 137–168. Amsterdam & Philadelphia: Benjamins.

Nesset, Tore & Laura A. Janda. 2010. Paradigm structure: evidence from Russian suffix shift. *Cognitive Linguistics* 21(4), 699–725.

Odebrecht, Carolin, Malte Belz, Amir Zeldes, Anke Lüdeling & Thomas Krause. 2017. RIDGES Herbology: designing a diachronic multi-layer corpus. *Language Resources and Evaluation* 51(3), 695–725.

Pankratz, Elizabeth. 2019. *A study of diachronic changes in the productivity of several early new high German derivational morphemes using the ridges corpus*. Humboldt-Universität zu Berlin, Sprach- und literaturwissenschaftliche Fakultät MA thesis.

Pankratz, Elizabeth. In prep. An entropy-based approach to measuring morphological productivity. in preparation.

Patil, Umesh, Peter Bosch & Stefan Hinterwimmer. 2020. Constraints on German diese demonstratives: language formality and subject-avoidance. *Glossa* 5(1), 1–22.

Reis, Marga. 2005. "Wer brauchen ohne zu gebraucht...": Zu systemgerechten 'Verstößen' im Gegenwartsdeutschen. *Cahiers d'Études Germaniques* 48(1), 101–114.

Schäfer, Roland & Ulrike Sayatz. 2014. Die Kurzformen des Indefinitartikels im Deutschen. *Zeitschrift für Sprachwissenschaft* 33(2), 215–250.

Schäfer, Roland & Ulrike Sayatz. 2016. Punctuation and syntactic structure in "obwohl" and "weil" clauses in nonstandard written German. *Written Language and Literacy* 19(2), 212–245.

Schäfer, Roland. 2015. Processing and querying large web corpora with the COW14 architecture. In Piotr Bański, Hanno Biber, Evelyn Breiteneder, Marc Kupietz, Harald Lüngen & Andreas Witt (eds.), *Proceedings of challenges in the management of large corpora 3 (CMLC-3)*. Lancaster: IDS.

Schäfer, Roland. 2018. Abstractions and exemplars: the measure noun phrase alternation in German. *Cognitive Linguistics* 29(4), 729–771.

Schäfer, Roland. 2019. Prototype-driven alternations: the case of German weak nouns. *Corpus Linguistics and Linguistic Theory* 15(2), 383–418.

Schäfer, Roland. 2020. Generalized linear mixed models. In Stefan Gries & Magali Paquot (eds.), *The practical handbook of corpus linguistics*, 535–561. Berlin, Heidelberg: Springer.

Schäfer, Roland & Felix Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the eighth international conference on language resources and evaluation (LREC '12) international conference on language resources and evaluation (LREC 12)*, 486–493. Istanbul: European Language Resources Association (ELRA).

Schäfer, Roland & Elizabeth Pankratz. 2018. The plural interpretability of German linking elements. *Morphology* 28(4), 325–358.

Schmid, Hans-Jörg & Helmut Küchenhoff. 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24(3), 531–577.

Schulte im Walde, Sabine, Maximilian Köper & Sylvia Springorum. 2018. Assessing meaning components in German complex verbs: a collection of source–target domains and directionality. In *Proceedings of the 7th joint conference on lexical and computational semantics*, 22–32. New Orleans.

Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3), 379–423.

Skousen, Royal. 1989. *Analogical modeling of language*. Dordrecht: Kluwer.

Stefanowitsch, Anatol. 2006. Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory* 2(1), 61–77.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8(2), 209–243.

Stefanowitsch, Anatol & Stefan Th. Gries. 2005. Co-varying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1), 1–43.

Stefanowitsch, Anatol & Stefan Th. Gries. 2009. Corpora and grammar. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. An international handbook*, vol. 2, 933–952. Berlin: De Gruyter Mouton.

Tomasello, Michael. 2003. *Constructing a language: a usage-based theory of language acquisition*. Harvard: Harvard University Press.

Willems, Klaas, Ludovic De Cuypere & Jonah Rys. 2018. Case alternation in argument structure constructions with prepositional verbs: a case study in corpus-based constructional analysis. In Hans C. Boas & Alexander Ziem (eds.), *Constructional approaches to syntactic structures in German*, 85–130. Berlin & Boston: De Gruyter. Wulf, Douglas J. 2002. Applying Analogical Modeling to the German plural. In Dilworth B. Parkinson, Deryle Lonsdale & Royal Skousen (eds.), *Analogical modeling: an exemplar-based approach to language*, 110–122. Amsterdam: Benjamins.

Wulff, Stefanie. 2003. A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics* 8(2), 245–282.

Zeldes, Amir. 2012. *Productivity in argument selection. from morphology to syntax*. Vol. 260 (Trends in Linguistics: Studies and Monographs). Berlin/Boston: De Gruyter.

Zipf, George Kingsley. 1949. *Human behavior and the principle of least effort*. Cambridge: Addison-Wesley.