Prototype-driven alternations: the case of German weak nouns

To appear in *Corpus Linguistics and Linguistic Theory*

Roland Schäfer

Abstract: Over the past years, multifactorial corpus-based explorations of alternations in grammar have become an accepted major tool in cognitively oriented corpus linguistics. For example, prototype theory as a theory of similarity-based and inherently probabilistic linguistic categorization has received support from studies showing that alternating constructions and items often occur with probabilities influenced by prototypical formal, semantic or contextual factors. In this paper, I analyze a low-frequency alternation effect in German noun inflection in terms of prototype theory, based on strong hypotheses from the existing literature that I integrate into an established theoretical framework of usage-based probabilistic morphology, which allows us to account for similarity effects even in seemingly regular areas of the grammar. Specifically, the so-called *weak* masculine nouns in German, which follow an unusual pattern of case marking and often have characteristic lexical properties, sporadically occur in forms of the dominant *strong* masculine nouns. Using data from the nine-billion-token DECOW12A web corpus of contemporary German, I demonstrate that the probability of the alternation is influenced by the presence or absence of semantic, phonotactic, and paradigmatic features. Token frequency is also shown to have an effect on the alternation, in line with common assumptions about the relation between frequency and entrenchment. I use a version of prototype theory with weighted features and polycentric categories, but I also discuss the question of whether such corpus data can be taken as strong evidence for or against specific models of cognitive representation (prototypes vs. exemplars).

1. Overview

On July 23, 2012, the German online newspaper *Spiegel Online* published an article about a controversial statement made by Philipp Rösler, the former leader of the German Liberal Democratic Party. The article includes a comment from a fellow party member quoted in a headline as (1a), but repeated in the text body as (1b).[1]

(1) a. Auf welchem **Planeten** lebt er?
          on   which   planet   lives he
          'What planet does he live on?'
     b. Auf welchem **Planet**   lebt er?

In (1a), the so-called weak masculine noun *Planet* 'planet' takes the inflectional marker *-en* in the dative singular, but it does not in (1b). The form in (1b) represents a non-standard alternation because dative singular forms without a suffix are characteristic of the much more productive strong masculine declension class, to which *Planet* predominantly does not belong. While it is impossible to know (and irrelevant) which variant was originally uttered in this specific case, many native speakers would agree that dropping the *-en* does not lead to full unacceptability. Similar examples of an accusative singular and a genitive singular of a weak noun (*Mensch* 'human') in strong forms (accusative *Mensch* instead of *Menschen* and genitive *Mensches* instead of *Menschen*) are shown in (2) and (3).[2,3] The two paradigms (standard and alternative) are shown in Table 1. Notice that the nominative singular and the whole plural are not affected. The remainder of the paper deals exclusively with the three non-nominative singular cases.

(2) Gibt es   einen Mensch,   der   stetig   wächst?
     gives it   a   human   who   constantly grows
     'Is there any person who grows constantly?'
(3) Das Leben   eines Mensches wird   zu   politischen Zwecken
     the life   of.a human   becomes   for political   reasons
     aufs   Spiel gesetzt.

on.the   game put

'A person's life is put at risk for political reasons.'

Table 1. The canonical weak and the alternative strong forms of weak nouns with the appropriate form of the masculine definite article *der* 'the'.

		Weak (canonical)	Strong (alternate)
Singular	**Nominative**	der Linguist	der Linguist
	Accusative	den Linguist-**en**	den Linguist
	Dative	dem Linguist-**en**	dem Linguist
	Genitive	des Linguist-**en**	des Linguist-**s**
Plural	**Nominative**	die Linguist-en	
	Accusative/Dative	den Linguist-en	
	Genitive	der Linguist-en	

The (simplex) weak nouns form a small class of just over 450 masculine nouns. Interestingly, many of them, according to Köpcke (1995), have prototypical semantic and phonotactic properties such as human denotation or non-final accent (cf. Section 2.4). This makes the alternative forms, as seen in the sentences (1) to (3), particularly interesting. Under classical Aristotelian theories of linguistic categorization, we could only classify the alternate forms as the results of *performance error*, while otherwise adhering to a notion of fully discrete grammatical categories not affected by prototypicality, similarity effects, or fuzziness. However, such an interpretation is revealed to be inappropriate if we can show that the presence or the absence of the prototypical semantic and phonotactic features influences the alternation strength of weak nouns.[4] After all, the term *error* (as in *performance error*) can only refer to the unmodeled random component of an alternation phenomenon, which is essentially irrelevant from a theoretical perspective. To assume something like irrelevant effects which are not randomly distributed (or relevant effects which are randomly distributed) would be contradictory. The guiding hypothesis for this study is therefore clear: if there is a cognitively real prototype, then a weak noun's probability of occurring in a strong form should be inversely

correlated with its degree of prototypicality. In the spirit of similar corpus-linguistic approaches to alternations (e.g., Gries 2003; Bresnan et al. 2007; Nesset and Janda 2010; Divjak and Arppe 2013), I assume that corpora are a valid source of data to test such hypotheses, at least as long as we can find enough occurrences of both alternatives to allow for an appropriate type of statistical inference. Despite its apparent relevance for cognitively oriented corpus-based linguistics, the alternation strength of weak nouns in German has not yet been examined in a large-scale, methodologically sound corpus study. This situation might be partially attributed to the lack of appropriate corpora, and the present paper aims to remedy it by using a very large web corpus of German and by applying fully automatic methods of data retrieval and analysis.

The paper is structured as follows. In Section 2, I discuss the theoretical background in detail, i.e., relating to prototype theory as a theory of similarity-based categorization in the broad sense. I also argue for the validity of using corpora in prototype research, and recapitulate and extend the existing analyses of the weak-noun prototype in order to derive hypotheses for the corpus study. The corpus study is reported in Section 3, including a detailed description of the corpus resources and the methods applied to extract the data, as well as a report of an appropriate Generalized Linear Model and an interpretation of the results. Finally, I summarize the findings in Section 4.

2. On prototypes, cognitive corpus-based morphology, and weak nouns

2.1. Prototype theory

As mentioned in the previous section, the main hypotheses for this study have been previously formulated within prototype theory (Köpcke 1995). Therefore, I will briefly introduce the relevant ideas from prototype theory in this section.[5] According to formal approaches to linguistic categorization inspired by Aristotelian logic, category membership is determined by rules and defining features, and it is consequently not viewed as a matter of degree (Sutcliffe 1993; Murphy 2002: 11–16). However, based on evidence suggesting that humans often categorize objects (standard examples include types of birds or furniture, or similar semantic categories) by similarity and often with varying degrees of fuzziness, cognitive theories like *prototype theory* (Rosch 1973) and its di-

rect rival *exemplar theory* (Medin and Schaffer 1978; Hintzman 1986) were developed. Prototype theory assumes that categories are defined by the similarity of their members to a mentally stored abstraction: the most prototypical member, or, in later versions of prototype theory, weighted features defining the prototype in a non-discrete fashion. Exemplar theory assumes that a category is simply a set of memory traces of previously encountered similar items, and that no abstraction takes place.[6] Both theories can deal with certain members of categories being *better* or more central than others, as well as category membership being determined to varying degrees based on similarity rather than on a fixed set of necessary features defining strictly discrete categories (cf. also Divjak and Arppe 2013: 224).[7] The major difference between them is how they model the cognitive representations of categories, as either abstractions (prototypes) or memory traces of concrete exemplars. The two theories are supported by different sets of experimental methods, such as similarity ratings, typicality ratings, and category naming. See Storms et al. (2000) as an example of those methods applied to semantic categories in natural language, including a discussion of the relevance of the respective findings for the two theories. It has been suggested by Barsalou (1990: 72–77) that the theories might be informationally equivalent. In the same paper, Barsalou (1990: 63) argues that behavioral experiments fail to produce conclusive evidence for or against either model of cognitive representation, because mental representations cannot be observed directly and experiments always show effects of both the representation and the processes involved in producing participants' reactions. Barsalou states that

> we can not say whether category knowledge is distributed in exemplars or centralized abstractions. But we do know that any account of knowledge that excludes idiosyncratic information, cooccurrence information, or dynamic representation is inadequate. (Barsalou 1990: 84)

With this argument in mind, which is even more applicable to corpus-based methods where we can intrinsically only observe the results of categorization made by writers (cf. also Section 2.2), I adopt a strictly prototype-theoretical formulation in the remainder of this paper, though I consider my findings to be compatible with alternative theories.[8]

While in her earlier work Rosch conceptualized the representation or the prototype as the most prototypical member of a category (Taylor 2008: 42–44), in her later work (Rosch et al. 1976; Rosch 1978) and in most of the subsequent literature (summarized in Taylor 2008: 44–46, 2015: 564), prototypicality is modeled in terms of weighted features. Under this view, an object belongs to a specific category to the extent that it has a sufficiently large number of diagnostic properties, each of which is (at least potentially) not strictly required for membership in that class. To account for effects of similarity and fuzziness of membership, the features are weighted as being more or less characteristic. This is formulated in terms of a high or low *cue validity* of a feature (Rosch et al. 1976: 384–385). Ideally, the cue validity would specify the conditional probability $p(F|C)$ that an item which exhibits a feature F is a member of category C.[9] Obviously, exact cue validities and exhaustive feature sets are hard to establish experimentally; estimating the *relative* importance in a set of relevant features is a more realistic goal. The concept of cue validity is crucial for this study, and it comes with two corollaries which are equally important. The first one is related to the distinction between rule-based and similarity-based categorization: if some prototypical feature is weighted high enough compared to all other features, it might become a de-facto requirement for category membership. Technically, prototype theory based on weighted features thus encompasses the classical Aristotelian approach as a limiting case (Taylor 2008: 45, 2015: 564). This is highly important when we apply prototype theory to effects in grammar, where, arguably, categories (such as a word's membership in an inflectional class) have a much more discrete character than in lexical semantics (see Section 2.3). Modeling a continuum between rule-based approaches and similarity-based approaches is thus technically possible within a feature-based prototype theory.[10] The second corollary is that there is no need to commit to a *single* prototype as the representation of a category, or, alternatively, a *single* cluster of feature values defining the center of the category. Given the relevant features and their weights, there might be multiple more or less dissimilar prototypical subsets of features instantiated in corresponding prototypical representatives, resulting in a polycentric category.[11]

As mentioned above, prototype theory was first developed with a focus on noun semantics. However, the present study stands in a long tradition of applying prototype

theory to grammatical categories (Taylor 2008: 51–60, 2015: 568–576). If we regard it as a general model of the organization of cognitive representations, it should be expected to be applicable to grammatical categories. To name just a few examples, prototype effects have been detected in the categorization of strong verbs in English (Bybee and Moder 1983), in part-of-speech classification (Uehara 2003), and in the selection and/or classification of syntactic constructions and the syntax-lexis interface (Winters 1990; Gries 2003; Divjak and Arppe 2013; Dobrić 2015).[12] Especially in morphosyntax, large feature sets pertaining to all levels of linguistic description, including features referring to larger linguistic (and even extra-linguistic) contexts can be taken into account, either defining prototypical constructions in which items appear (cf. especially Divjak and Arppe 2013: 228 on the importance of the contextual element) or defining prototypical choices of competing constructions which are prototypically instantiated in certain contexts and given certain lexical choices. Some of the aforementioned studies and the study presented here use corpus data, and I turn to the validity of corpus linguistic methods in prototype theory in Section 2.2.

2.2. Prototypes, alternations, and corpus data

With regard to using corpus data to support prototype approaches, two major questions need to be clarified. The first and rather technical question is that of the appropriateness of concrete methods. The second question is more fundamental and concerns the validity of corpus-based experiments in prototype theory and cognitive linguistics in general. I briefly discuss these two questions in this section.

Methodologically, there are diverse approaches to corpus-based prototype research; I provide only an exemplary overview. For example, Gilquin (2006) examines raw and relative frequencies in order to determine how often prototypical properties of English periphrastic causative constructions described in the existing literature are instantiated in corpus exemplars. She sees this as a test of the idea that prototypicality implies salience, which in turn implies high frequency (Gilquin 2006: 169, 178), and as a test of the existing theoretical models of prototypical causation (Gilquin 2006: 178–180). Dobrić (2015) models the prototypicality of the different senses of the English verb *look* by, among other things, counting the number of corpus sentences in which they are instanti-

ated (also under the hypothesis that prototypicality implies high frequency) and by determining how restricted the senses are with respect to the different contexts in which they appear. Another approach does not rely on assumptions about a direct correspondence between prototypicality and absolute frequency and focuses on situations where two or more lexical items or two or more constructions compete (i.e., where they alternate). Some form of multifactorial modeling is then applied to determine how strongly the relevant features influence the choice of each of the competing elements. Under this approach, the features of the multifactorial model define the prototypicality of the competing items, and the calculated probabilities in favor of a specific choice can be seen as an approximation of the cue validity of the features. For example, Gries (2003) examines how a large number of features influences the choice between ditransitive and competing prepositional constructions in English. Similarly, Divjak and Arppe (2013) model the choice among larger sets of Finnish and Russian *try* verbs, determining prototypical properties for each verb (Divjak and Arppe 2013: 260–267) by applying polytomous regression using a large set of grammatical, lexical, and contextual features extracted from corpus exemplars.[13] The present study stands in this line of research.[14]

Finally, I turn to the important question of the validity of corpus-based findings in cognitive linguistics in general and prototype theory in particular. Some general objections against corpus data, which could be seen as applicable here, have been dealt with in Bresnan et al.'s (2007) paper on the dative alternation in English.[15] First of all, Bresnan et al. (2007: 82–84) show that the factors which influence the choice between two alternative constructions are stable across individual speakers represented in the corpus. Corpus data, which pool a large number of individual speakers, can thus still be used to infer regularities of individual speakers' mental grammars (contra Newmeyer 2003, especially p. 696). Similarly, Bresnan et al. (2007: 87–90) show that cross-corpus differences (which might, at least partially, be attributed to different distributions of register or genre) do not per se invalidate grammatical inferences from corpus data. Furthermore, the contributions in Glyn and Fischer (2010) (especially Gries and Divjak 2010), all argue implicitly or explicitly for the applicability of corpus-based methods in cognitive linguistics, including references to work where corpus-based and experimental work (which offers more direct access to cognitive representations and processes) have

been successfully correlated. Thus, there clearly exists a successful research paradigm for corpus linguistics as "psychologically informed (cognitively-inspired) usage-based linguistics" (Gries 2010: 334). For a concrete corpus study, however, it is still important to evaluate the *construct validity* of the specific constructs used.[16] In prototype theory in the narrow sense, the construct is a cognitive representation of a linguistic category such as the meaning of a word, either in the form of a maximally prototypical exemplar or a set of weighted features (cf. Section 2.1). Gries (2003) comments:

> Frequently, Rosch's results were [...] interpreted as if they were statements on the structure of mental representations as such; cf. the *effects=structure* fallacy and the *prototype=representation* fallacy. I do not wish to support such interpretations. [...] Still, even if the form of analysis does not translate into statements on mental representations, the high predictive power [...] shows that the cognitive factors underlying the choice of construction have been identified properly and weighted in accordance with their importance for actual usage. (Gries 2003: 22)

Making a significantly stronger claim, Divjak and Arppe (2013) state:

> The objectives of this study are, first, to explore how the prototype and exemplar models of categorization manifest themselves in corpus data [...]. Although corpus data do not reflect the characteristics of mental grammars directly, we do consider corpus data a legitimate source of data about mental grammars. (Divjak and Arppe 2013: 229–230)

However, as Divjak and Arppe (2013: 224) themselves point out (citing Barsalou 1990), prototype and exemplar models can be seen as informationally equivalent (cf. also Section 2.1), and the relevant difference between them is the level of abstraction assumed in mental representations.[17] I therefore follow Gries (2003), taking the safe route and suggesting that if corpus data show the result of similarity-based categorization and fuzziness of categories, they provide support for cognitive linguistic theories of categorization and against traditional linguistic theories. Thus, a lot is achieved if we make sure "that the cognitive factors underlying the choice of construction have been

identified properly and weighted in accordance with their importance for actual usage"
(Gries 2003: 22).

To sum up, corpus data have been established as a reliable testing ground for cognitive linguistic theory. Prototype theory is well suited to model cases where forms or constructions alternate if we regard the features which control the choice of alternative forms and their weights as defining prototypes or prototypical contexts for their use. The notion of cue validity from prototype theory correlates with the contribution made by single feature values to the probability of a specific alternate being chosen in multifactorial analyses.

2.3. Prototypes, paradigms, and probabilistic morphology

As stated above, the focus of this paper is on alternation between two inflectional classes. In this section, I propose to treat this alternation under a prototype view of inflectional morphology, based on Sections 2.1 and 2.2. I also relate the present study to a larger body of work which suggests that a probabilistic usage-based approach to inflectional morphology is appropriate.

The prototype approach to paradigm structure has been followed, for example, by Nesset and Janda (2010). They argue that paradigms have a theoretical status (which in cognitive linguistics means a cognitive reality), which is denied in traditional Item-and-Arrangement or Item-and-Process approaches to morphology (see the discussion in Bybee 1985). Even more importantly, Nesset and Janda (2010: 705–707) argue that paradigms are structured as *radial categories*, with highly central and prototypical, but also marginal exponents. Within verbal paradigms, for example, singular exponents are assumed to be more prototypical than plural exponents, finite exponents more prototypical than infinite ones, etc. Nesset and Janda (2010: 172–173) formulate this as the *Paradigm Structure Hypothesis* and argue that prototypical (and thereby unmarked) exponents within a paradigm are "more resistant to [...] change". They show that this assumption is supported by examining ongoing changes in Russian verbal inflection. I will return to this in Section 2.4 and Section 3 when looking at differences in case forms of weak nouns in German.

Nesset and Janda (2010) use prototype theory as their specific background, but the general cognitive framework detailed in Sections 2.1 and 2.2 extends to a more general probabilistic view of morphology supported by the data presented here. I do not go into detail on the concrete theoretical or computational implementations of probabilistic, gradient morphology, following Hay and Baayen (2005):

> [...W]e separate the questions of subsymbolic versus symbolic implementation on the one hand, from discrete versus gradient structure on the other. The results [...] could potentially be modelled both by symbolic and non-symbolic approaches. However, they resist modelling by strictly deterministic, non-probabilistic approaches. (Hay and Baayen 2005: 342–343)

There is ample support for the view that grammar is inherently probabilistic (overviews in Hay and Baayen 2005 or Kapatsinski 2014), and it is not necessary to review that evidence extensively here. The one aspect I want to point out is that probabilistic analogical effects in inflectional morphology have been diagnosed not just for irregular forms (like strong verbs in English), but also in the formation of completely regular forms (Ernestus and Baayen 2004; Tabak et al. 2010; Ramscar et al. 2013). The emerging picture shows that phonology, semantics, frequency, and similarity to other items drive analogical processes in both irregular and regular inflection. This is relevant for the present study because weak nouns in German cannot be called irregular, but their inflection is nonetheless affected by probabilistic effects.

Finally, the token frequency of specific items and the type and token frequencies of a pattern (such as an inflectional class) are assumed to play a role under cognitively oriented usage-based theories, although modeling exact effects and interactions is complex (cf. for example, Bybee 1985; Langacker 1987; Schmid 2010a; Divjak and Caldwell-Harris 2015; a discussion with a focus on corpus data is provided in Schmid 2010b). In general, it is assumed that high token frequency of an individual item leads to entrenchment of that item (Langacker 1987: 59–60). In other words, what is frequently repeated individually gradually becomes stored individually, and is easier to access and less susceptible to analogy (Bybee and Hopper 2001). On the other hand, high type frequency (i.e., the frequent occurrence of different items in a certain structure) favors the

storage of an abstract and productive pattern of that structure (Bybee and Thompson 2000). Divjak and Caldwell-Harris (2015) summarize further developments and the experimental techniques being used to examine frequency effects, as well as the ongoing discussion about the problems involved in the interpretation of results. I include only simple lemma token frequencies in the statistics in Section 3 (cf. especially Section 3.3), and therefore do not go into further detail here.

2.4. The German weak-noun prototype(s)

In this section, I provide an analysis of the relevant aspects of the German weak-noun inflection. The aim is to derive testable hypotheses for the corpus study presented in Section 3 considering the theoretical background specified in Sections 2.1 through 2.3. Some of the details discussed by Köpcke (1995) – especially diachronic aspects – are not dealt with prominently here.

Table 2. The two weak-noun prototypes and their features with example nouns from Köpcke (1995:178).

	ACCENT	FIN. SCHWA	**Protoype I**	SYLL.	SEM.	**Prototype II**	ACCENT
→**higher prototypicality**→	PENULT (or N/A)	+	*Matrose*	POLY	HUM	*Artist*	ULT
			Kurde				
			Schimpanse		ANI	*Leopard*	
			Falke				
			Gedanke		INANI	*Trabant*	
			Mensch	MONO	HUM		
			Bär		ANI		

Köpcke (1995) establishes two major prototypes for weak nouns, thus effectively defining a polycentric (bifocal) category (cf. Section 2.1). The two types of prototypical weak nouns are analyzed as bearing the semantic and phonotactic features summarized in Table 2 (derived from Köpcke's 1995: 178 diagram), where the degree of prototypicality is encoded vertically. Semantically, weak nouns denote humans (highest cue va-

lidity), animate objects, or inanimate objects (lowest cue validity). The other features are phonotactic, and their weightings are more complicated, as should become clear immediately when looking at Table 2. Final schwa is highly characteristic of prototype I.[18] Furthermore, prototype I has penultimate accent, and prototype II has ultimate accent except for the words ending in schwa. This falls out because schwa can never be accented in German. Nearly all weak nouns are polysyllabic, with the exception of the Germanic monosyllables in prototype I (those without final schwa). Clearly, the phonotactic features are not independent, and they will consequently be coded as one factor with the appropriate levels in the corpus study in Section 3. Interestingly, the semantic property of denoting humans (or animals) is not even remotely exclusive to weak nouns. On the other hand, polysyllabic nouns with non-initial accent are quite atypical and relatively rare because the dominant phonotactic pattern for nouns is monosyllabic or trochaic (Eisenberg 2012: 18–19, Eisenberg 2013: Ch. 5). I therefore suggest that the phonotactic features must be expected to be more salient than the semantic features as cues for the class of weak nouns.

Some weak nouns have become fully or mostly strong (Wurzel 1985; Joeres 1996) over time, and Köpcke argues, mostly based on diachronic data, that the assimilation to the more productive strong inflection was controlled mainly by the semantics of the respective nouns. Nouns denoting humans or animals (especially animals strongly linked to humans such as farm animals or pets) are more immune to becoming strong. However, while some individual weak nouns have become strong, Köpcke does not consider the class of weak nouns to be generally eroding:

> Mit zunehmender Entfernung von den Prototypen nehmen Übergangsprozesse von schwacher zu starker Deklination zu. Abbautendenzen für die schwachen Maskulina finden aber ausschließlich in der Peripherie des Kontinuums statt; der prototypische Bereich ist hiervon nicht betroffen. Im Gegenteil: Hier lassen sich Aufbautendenzen beobachten. [With a larger distance from the prototype, it is more likely that a weak noun shows transitional phenomena towards the strong declension. However, full transition of weak nouns to the strong declension only occurs in the periphery of the continuum, and pro-

totypical members are not affected. On the contrary: here we even find the
opposite, i.e., strong nouns becoming weak.] (Köpcke 1995: 159–160)

In Section 3, I will therefore seek first of all to determine how stable is the class of
weak nouns in a corpus of contemporary German, and argue that a purely synchronic
view on the phenomenon is justified. If morphology is inherently probabilistic (Section
2.3) and similarity-based (Section 2.1 and 2.2), and if the weak-noun prototype does
have a cognitive reality, then we should expect to find a certain number of strong forms
of weak nouns. It has already been demonstrated in Section 1 that this is the case. If the
class is stable, however, the strong forms should be rare. Furthermore, the more proto-
typical a weak noun, the lower should be its probability of occurring in a strong form.
Under this hypothesis, the alternation strength is an indirect indicator of the cue validity
of the features summarized in Table 2, and a corpus study can provide an evaluation of
the relative cue validities of the features. These hypotheses constitute the theoretical
core of Section 3.

An additional factor is contextual in the sense of Section 2.2, namely the grammatical
case of the nouns. Thieroff (2003: 113–115) argues that it is natural for weak nouns to
occur in strong forms from a paradigmatic perspective. He derives this conclusion from
the fact that weak nouns violate very productive generalizations of case inflection of
German nouns in general and masculine nouns in particular. First of all, the accusative
and dative singular of all other nouns (masculine or otherwise) are never suffixed and
are thus formally identical to the nominative singular base form.[19] This is the most pro-
ductive generalization because it concerns all German nouns, except for the roughly 450
weak nouns. Secondly, the genitive singular of masculine nouns is otherwise always
distinguishable from the accusative and the dative because the genitive is marked (i.e.,
not identical to the nominative base form), and the accusative and dative are not. Third-
ly, the genitive marker (if there is one) is otherwise always -*es*.[20] These generalizations
are relevant to this study because I assume that paradigms have an internal structure and
a cognitive reality (cf. Section 2.3 and the detailed overview in Nesset and Janda 2010:
704–705), and that it consequently matters which formal oppositions within the para-
digm are salient. Clearly, the most productive and therefore most salient of the relevant

generalizations about nominal paradigms in German (i.e., the identity of all accusative and dative singular forms in relation to the respective nominative singular base form) as well as the second (the formal distinguishability of the masculine genitive singular from the accusative and dative singular) are satisfied by making only the accusative and the dative strong and leaving the genitive weak. Only the third generalization (masculine genitive singular forms have *-es*) requires even the genitive of weak nouns to become strong. I therefore expect that the probability of finding strong forms of weak nouns is lower in the genitive than in the accusative and dative.

To summarize the hypotheses for the corpus study in Section 3: (i) I expect to find a low-frequency alternation effect in the sense that weak nouns occur in strong inflectional forms, but with low frequency. In other words, there is no general erosion of the weak nouns, and the class is stable, especially because it is associated with a strong prototype. (ii) Furthermore, I expect the probability of the alternation to be higher in the absence of the prototypical lexical features of weak nouns, namely semantics (human – animate – inanimate) and phonotactics. (iii) The grammatical case of the noun is also expected to affect the alternation strength because the analogical attraction exerted by the more productive strong paradigms is stronger in the accusative and dative than in the genitive. (iv) In accordance with what was said about frequency in Section 2.3, a higher token frequency is expected to gradually help prevent the alternation. If supported by corpus data, these hypotheses strengthen the general research program of corpus-based cognitive linguistics and provide welcome support for similarity-based models of categorization.

3. Corpus study

3.1. Challenges and corpus choice

For a corpus-based account of the factors that influence the strength of the alternation of weak nouns towards the strong pattern, the source of data has to satisfy two requirements, which are not satisfied by most available corpora. First, the corpus has to contain unedited spontaneous writing because the strong forms of weak nouns are very rare in standard written German. Occurrences in professionally written media like (1b) in Sec-

tion 1 are exceptional. Secondly, as the phenomenon is rare, an appropriate corpus should be as large as possible. The size requirement is even more vital because the strong forms turned out to be so rare that we cannot locate them in an acceptable time by going through large concordances manually.[21] This means that only co-occurrences of weak nouns with the singular indefinite determiner can be used, as I will argue now.

Consider the forms whose frequencies are to be compared, shown in Table 1.1. German masculine determiners have unambiguous case forms in the singular, so it appears that we could just search for sequences of an appropriately inflected determiner, optionally followed by zero, or more appropriately inflected adjectives followed by a weak-noun lemma. However, NPs containing the canonical accusative singular are (as a whole phrase) homonymous with those containing the accusative or dative plural, as can be seen in Table 1.1. This makes it impossible to search for canonical forms with determiners that have a plural without having to fall back on the manual inspection of huge concordances. The only determiner without a plural is *ein* 'a' because indefinite plural NPs are determinerless in German. As a quantifier, *ein* 'one' is semantically blocked from occurring in the plural. This reduces the unambiguous NP configurations to those with the determiner *ein*.

Because of the problems of data sparseness, the DECOW12A web corpus was chosen.[22] It is part of the COW collection of web corpora and contains 9.1 billion tokens, making it the largest linguistically annotated corpus of German available at the time this study was conducted.[23] Furthermore, it can be downloaded in sentence-shuffled form for local processing. Since not much is known about the properties of large web corpora from a linguistic perspective, I briefly argue for the use of this resource here. The DECOW corpus was described by its creators in Schäfer and Bildhauer (2012) and Schäfer and Bildhauer (2013). Schäfer and Bildhauer (2012: 492–493) and Biemann et al. (2013: 48) present an assessment of the text types and registers contained in the corpus based on manually annotated samples, showing that the corpus contains an estimated 22.5% of documents written in a *quasi-spontaneous* mode, i.e., mostly forum discussions. This is the kind of text where we can hope to find most non-standard strong forms of weak nouns. Biemann et al. (2013: 37–46) also show that the COW corpora compare favorably to traditional and web corpora in extrinsic evaluations (such as collocation

extraction tasks). In Schäfer and Sayatz (2014), DECOW12A was used for similar reasons as in this study, and Van Goethem and Hiligsmann (2014) base part of their corpus study on the Dutch COW corpus (NLCOW12A) for comparable reasons, too. The use of similar, earlier web corpora like the WaCky corpora (Baroni et al., 2009) has been argued for convincingly by, among others, Zeldes (2012: 96–98). The corpus I have chosen here can therefore be considered a recent but adequately tested source of data.

In the next section, I describe my method of obtaining reliable counts from these data sources by formulating precise queries in a spirit comparable to that of Zeldes, who states (writing about his work with the WaCky corpora) that

> [t]his type of automatically retrieved data is rather heterogeneous and possibly error-prone, so a main priority in searching through such corpora is to ensure high accuracy of results by formulating precise queries and manually evaluating error rates as required. (Zeldes 2012: 97)

However, heterogeneity in the data is also a blessing because it increases the external validity of a study, i.e., the probability that the findings generalize well (Maxwell and Delaney 2004: 30). Considering that corpora like DECOW allow us to take very large samples, the error-prone or noisy nature of the data is also less of a problem because the error is distributed randomly, at least if there is no systematic sampling error.

3.2. Sampling and descriptive statistics

To derive a sample for the analysis of the counts of weak and non-canonical strong uses of weak nouns, I proceeded as follows. First, a list of weak nouns was bootstrapped from DECOW12A by searching for word forms ending in *en* and preceded by a determiner in the genitive (such as *des Planeten* 'of a planet'). Using this bootstrapping method requires each selected noun to occur at least once in a canonically inflected genitive form with a determiner and without an adjective. This is a desirable by-effect because it ensures that no extremely infrequent nouns make it into the final sample. The list was sifted by hand in order to eliminate erroneous matches, including strong nouns in weak forms (however, see the discussion in Section 3.4). In total, 451 noun lemmas

were found, which is highly consistent with Bittner (2003: 98), who states that 40 native and more than 400 non-native lemmas fall into the class of weak nouns.

Then, the final sample was taken by querying inflected non-nominative singular forms of *ein* (accusative *einen*, dative *einem*, genitive *eines*), optionally followed by properly inflected adjectives, followed by either the weak or the strong form of each of the bootstrapped lemmas. In total 451 lemmas · 3 cases · 2 alternative forms = 2,706 queries were executed.[24] In the resulting sample (n = 953,117), there were 26,667 token occurrences of the strong inflectional pattern and 926,450 token occurrences of the weak inflectional pattern. Accounting for a mere 2.8% of all occurrences, the use of strong inflection can thus be considered a rare event. A list of the noun lemmas and their frequencies in the final sample can be found in the appendix. The confidence intervals for the proportion of strong forms specified there (per lemma) are large, which indicates that analyzing the alternation strengths of single lemmas is not a good idea from a statistical point of view.[25]

Some of the lemmas turned out to be problematic and were removed as follows. Some have homonyms which are frequent proper names, which always inflect according to the strong pattern. German proper person names also occur with determiners, not just in dialects, but also in certain styles of writing (4).[26]

(4) Ich   werde   lieber   mit   einem   Götze   Zweiter
     I   become   rather   with a   Götze   second
     als   mit   einer   Amoroso-Diva   Erster!
     than   with an   Amoroso diva   first
     'I would rather finish second with a guy like Götze than first with a diva like
       Amoroso!'

Other problematic nouns are those few for which it is known that they have been alternating strongly between weak and strong variant for a long time, some even having re-analyzed alternative stems ending in *-en* (e.g., strong *Frieden* 'peace' with the genitive *Friedens* versus the older but co-existing weak *Friede* with the genitive *Frieden*). In some cases, the differentiation of stems correlates with a semantic differentiation (Köpcke 1995: 173; also Wurzel 1985 and Joeres 1996). For the corpus study, this

means that these nouns introduce an ambiguity between accusatives and datives of the reanalyzed stem without a suffix (e.g., *dem Frieden*) and accusatives and datives of the older stem with a suffix (e.g., *dem Friede-n*). This ambiguity cannot be resolved automatically. They also have a highly increased alternation strength compared to the other weak nouns, for example 13% strong forms for *Friede*. These facts indicate that they do not belong to the otherwise homogeneous group of weak nouns without reanalyzed stems and but with low alternation strength. Therefore, I decided to remove them from the sample.[27] A small class of nouns was removed because the sample was contaminated with very recent English loans (always strong) such as *Artist* 'artiste' (old loan, weak) vs. 'artist' (recent loan, strong). The noun *Steinmetz* 'mason' was also removed because it has almost completed a development into a strong noun already (without stem reanalysis). The noun *Mensch* 'human' with 221,210 occurrences is extremely frequent and occurs over eight times more often than the next most frequent weak noun *Kunde* 'customer', with 26,576 occurrences. I removed it because otherwise the sample would have contained an extremely large proportion of token occurrences of this single noun. The nouns which were removed are still included in the list in the appendix.

The reduced sample contains 433 lemmas and 466,922 token occurrences, of which 10,488 are strong and 456,434 weak. The proportion of strong forms drops by 0.6% to 2.2%, which is mostly due to the removal of the nouns with two stems. However, there remains a very large sample of nouns showing the low-frequency alternation effect. Figure 1 shows a bagplot of the reduced sample, where each dot represents a weak noun, and the coordinates correspond to the total number of occurrences in the corpus (x-axis for strong, y-axis for weak occurrences). Bagplots (Rousseeuw et al. 1999) are bivariate boxplots. The inner polygon (*bag*) contains half of the points. The outer polygon (*fence*) is the bag grown by a factor of three. Data points outside the fence are considered outliers. This bagplot shows that there are many infrequent and few frequent nouns in the sample, which is trivially expected given the power-law distribution of word frequencies. However, we can also tell from the scales of the axes that the strong forms (x-axis) are very rare compared to the weak forms (y-axis). Outliers are few, and I decided against further removal of them.

Distribution of weak an strong forms per lemma
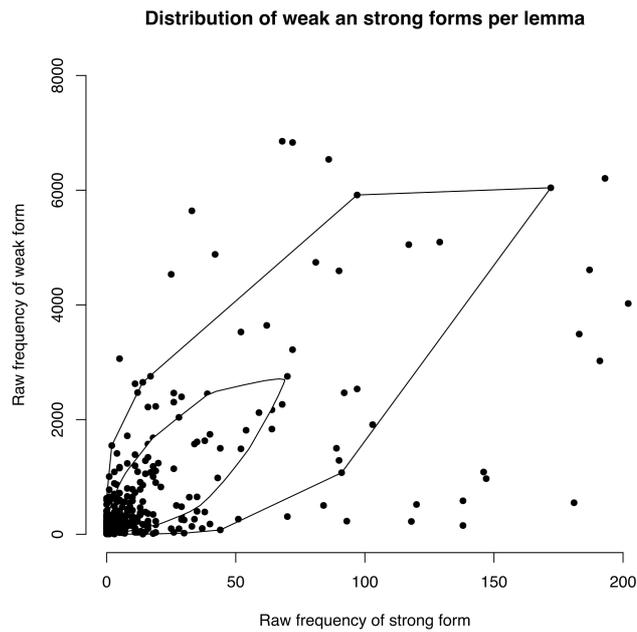
Figure 1. Bagplot of the distribution of the weak and the strong forms. Each dot represents one lemma. Seven outliers are not shown.

Because the influence of grammatical case was introduced as an additional factor on top of the features established by Köpcke (Section 2.4), a monofactorial descriptive analysis was performed to confirm that case was at least likely to have an effect. It was quite obvious that in the three grammatical cases there were remarkable distributional differences between strong and weak forms. See Table 3 for an overview of the proportions in the three cases. In the accusative and dative, the strong inflection is much more frequent.[28]

Table 3. Proportions of strong occurrences of weak nouns by grammatical case with 99% confidence intervals and sample size.

Case	Count (strong)	Proportion (strong)	99% CI	n
accusative	5,403	0.0247	±0.0009	218,740
dative	4,299	0.0278	±0.0011	154,623
genitive	1,048	0.0112	±0.0009	93,559

3.3. A statistical model of the prototype

In this section, I present inferential statistics which show that the strength of the alternation of weak nouns towards the strong inflection is both influenced by prototypical features of weak nouns and by preferences for case marking in the three grammatical cases. I estimate a Generalized Linear Model (GLM) in order to test for the significance of factors and quantify the strength of the significant factors.[29] Following the argumentation from Section 2.4, the odds associated with the different features can be interpreted as related to their cue validity, and the prototypicality of concrete nouns can be specified by calculating their cumulated odds.

Multifactorial modeling is required since the factors which play a role by hypothesis (cf. Section 2.4) are, at least, as follows: (i) the semantic class of the noun (human, animate non-human, inanimate), (ii) final schwa (yes, no), (iii) accent (final, non-final), and (iv) case (accusative, dative, genitive), and (v) token frequency. A binomial Generalized Linear Model based on token occurrences of weak nouns with the aforementioned factors as regressors and weak/strong as the response variable is a reasonable choice. With regard to the coding of the phonotactic variables, it should be noticed (cf. Section 2.4) that schwa syllables never take the accent, that monosyllabic words never contain schwa syllables, and that monosyllabic words offer no choice as to where the accent is placed. Consequently, atomic phonotactic variables are not independent. I

therefore coded the nouns for a single phonotactic regressor variable (*Pt*) with the levels *PolyUlt* for polysyllabic words with final accent, *PolyNult* for polysyllabic words with non-final accent, *PolySchwa* for polysyllabic words ending in schwa (and consequently with non-final accent), and *Mono* for monosyllabic words. Table 4 summarizes the response variable, the regressors and their levels, including the simple (log-transformed) token frequency of the lemma (Section 2.3 and Section 2.4).

Table 4. Variables used in the GLM reported in Section 3.3.

Variable	Level	Description
Case		**grammatical case:**
	Nom	nominative
	Acc	accusative
	Dat	dative
Sem		**semantic class of noun:**
	Hum	human
	Ani	animate non-human
	Ina	inanimate
Pt		**phonotactics:**
	PolySchwa	polysyllabic with final schwa
	PolyUlt	polysyllabic with final accent
	PolyNult	polysyllabic with non-final accent
	Mono	monosyllabic
LogFreq	[0, 4.393]	lemma log-frequency per million tokens, moved into the positive range

As was shown in Section 3.2, strong forms of weak nouns are rare (2.2% of all remaining forms). This makes logistic regression problematic, and special procedures exist for so-called *rare-events regression*. I opted for a straightforward approach to this, taking a stratified sample where both outcomes are represented in equal proportions. Models estimated based on such artificially stratified samples are not useful for making

predictions about the actual occurrence of events. This means that the coefficients and odds of the GLM can be interpreted only as indicators of the relative cue validities of the different features, and not directly as cue validities. The final sample used to estimate the GLM parameters contained 10,000 data points: 5,000 weak-noun exemplars in weak form and 5,000 exemplars in the strong form.

Table 5. Summary of the binomial GLM reported in Section 3.3. Response: strong inflection (non-canonical, positive coefficients) vs. weak variants (canonical, negative coefficients). β is the estimated coefficient and O the odds ratio. n = 10,000. VIF(Case) = 1.020, VIF(Sem) = 1.164, VIF(Pt) = 1.157, VIF(LogFreq) = 1.077. Analysis of deviance (for fully specified model df = 12,170, for intercept-only model df = 13,863) with χ^2: $p < 0.001$. Nagelkerke R^2 = 0.21.

Regressor	β	O	std. err.	z	p
Intercept (SemHum, PtPolySchwa, CaseGen)	−0.502	0.606	0.1293	−3.879	0.0001
CaseAcc	0.553	1.738	0.0676	8.174	< 0.0001
CaseDat	0.805	2.237	0.0702	11.471	< 0.0001
SemAni	0.664	1.943	0.0726	9.152	< 0.0001
SemIna	0.585	1.795	0.0664	8.812	< 0.0001
PtPolyUlt	1.073	2.925	0.0552	19.451	< 0.0001
PtMono	2.066	7.891	0.0778	26.540	< 0.0001
PtPolyNult	2.831	16.963	0.2309	12.260	< 0.0001
LogFreq	−0.381	0.683	0.0335	−11.373	< 0.0001

I ran standard model diagnostics (analysis of deviance, pseudo-R^2), and ensured that there was no severe collinearity by calculating the generalized variance inflation factors (VIF) according to Fox and Monette (1992). Table 5 summarizes the model. All regres-

sors remained in the model after running an AIC-based step-down procedure.[30] The proportion of variance explained by the model is fairly substantial with $R^2_{\text{Nagelkerke}} = 0.21$, but not very high. Readers might speculate that a mixed model (GLMM) could attribute some of this unexplained variance to lemma-, writer-, or even genre-specific random effects (cf. Kupermann and Bresnan 2012: 594; Gries 2015). First of all, it should be noticed that *Sem*, *Pt*, and *LogFreq* are lemma-level regressors anyway. I tried passing *lemma* as a random effect using the *glmer()* function from the *lme4* package in *R* (Bates et al. 2015), which would effectively to turn the model into a true multilevel model (Gelman and Hill 2006: 266), but the high number of different lemmas and the low counts for strong occurrences per lemma made it impossible to achieve convergence. *Speaker* or rather *writer* as a random effect are not available because authorship is usually unknown in web corpora. Furthermore, the high number of writers in very large samples from corpora as huge and diverse as DECOW makes it very likely that writer-specific random effects cannot be calculated due to extreme numbers of levels. In other words, virtually each sentence is written by a different person.

The intercept models those factor levels which most strongly favor the weak forms by hypothesis over the non-canonical strong forms by hypothesis (*SemHum*, *PtPolySchwa*, *CaseGen*). The accusative (O = 1.738) and the dative (O = 2.237) both cause a significant tendency towards the strong inflection compared to the genitive, which is modeled by the intercept. The same is true if the noun denotes inanimate (O = 1.795) or non-human animate objects (O = 1.943), compared to humans. There is no interpretable difference between animal-denoting nouns and nouns denoting inanimate objects, however. The phonotactic effects are more extreme than the effects of case and semantics. Non-final accent in polysyllabic words (O = 16.963) favors the strong pattern most clearly, followed by monosyllabicity (O = 7.891) and final accent in polysyllabic words (O = 2.925). The stronger contrasts between the levels of the phonotactic variable are expected because the phonotactic prototypical features are much closer to being exclusive of weak nouns compared to humanness of denotation. In other words, the hypotheses from Section 2.4 are fully supported by the data.[31] We can now use the coefficients or odds to amend Table 2. Specifically, for the sample nouns ordered according to their

assumed prototypicality by Köpcke (1995: 178), we can specify the overall alternation strength as an inverse indicator of their prototypicality by calculating summed odds (cf. Table 6).[32] With one negligible deviation (*Gedanke*), the predictions derived from Köpcke (1995) are confirmed, and the alternation strength increases towards the bottom of the table.

Table 6. The nouns from Table 2 with the estimates of summed odds from the GLM reported in Table 5 as indicators of their prototypicality. The nouns are merely examples of their respective sub-types.

	Protoype I	Odds	Prototype II	Odds
	Matrose	0.605		
	Kurde	0.605		
	Schimpanse	1.176		
	Falke	1.176		
	Gedanke	1.087		
			Artist	1.770
			Leopard	3.438
			Trabant	3.177
	Mensch	4.778		
	Bär	9.281		

↓ higher alternation strength ↓   ↑ higher prototypicality ↑

Finally, a higher token frequency of a weak noun (measured as log-transformed token frequency per million tokens, moved into the positive range) disfavors its occurrence in strong forms (O = 0.683). Again, this is in line with the predictions because frequent individual items are expected to be more resistant to analogy through entrenchment (Section 2.3). The high relative type frequency of the strong nouns compared to the weak nouns, on the other hand, makes it plausible that an otherwise stable class (Section 2.4) shows a consistently measurable alternation effect in the first place. While being less productive than the strong pattern, the weak pattern should still be marginally pro-

ductive when it is associated with an underlying prototype. In Section 3.4, I show how such a marginal productivity can be observed as a kind of reverse alternation effect.

3.4. Weak nouns as a productive pattern: strong nouns in weak forms

If predominantly strong nouns with the prototypical properties of weak nouns could be found in weak forms, this would lend further support to the conceptualization of weak nouns as a prototypically structured category. To show that such noun forms are indeed extant, I now look at a subclass of strong loan nouns which occurred very often in weak forms in the bootstrap sample described in Section 3.2, namely masculine Latin loans ending in *-or*. In line with the findings from Section 3.3, all these words qualify as good weak nouns because they have prototypical phonotactic features (polysyllabicity and penultimate accent, Köpcke's prototype II), and we do in fact find non-standard weak dative forms like *dem Autoren* 'the author' instead of *dem Autor* (strong, standard). Interestingly, the corollary that prototype theory with weighted features allows single features to be so highly weighted that they become a de-facto requirement for class membership (discussed in Section 2.1) is supported because strong nouns which do not fulfill the phonotactic requirements simply cannot occur in weak forms. Native speakers will very likely agree that accusatives such as *den Wegen* 'the path' instead of *den Weg* are beyond even peripheral acceptability.

Table 7. Counts of strong and weak occurrences of 62 strong nouns ending in *-or* by humanness of denotation. n = 114,486. Considering the size of the sample, significance testing is not required. Odds ratio O(weak|human,weak|non-human) = 18.868.

	strong	weak	proportion weak
human	53,596	2,274	0.0407
non-human	58,482	134	0.0023

Interestingly, some of these nouns, such as *Autor* 'author', denote humans, while others, such as *Transistor* 'transistor', do not. I therefore counted the occurrences of 32

human-denoting and 30 non-human-denoting nouns in strong and weak forms. Table 7 sums up the results. Clearly, the results from the earlier sections are confirmed, as human-denoting -*or* nouns have a much stronger tendency to occur in weak forms. The odds ratio reveals that the chance of seeing a weak form of a human-denoting noun is 18.868 times higher than the chance of seeing a weak form of a non-human-denoting noun. The results for strong and weak forms in the three grammatical cases also confirm the results from earlier sections (cf. Table 8). The genitive slightly favors the weak pattern compared to the accusative (O = 1.557) and the dative (O = 1.511). A higher alternation strength of originally strong nouns towards the weak pattern corresponds with a stronger resistance against the alternation towards the strong pattern for weak nouns.

Table 8. Counts of strong and weak occurrences of 62 strong nouns ending in -*or* by case. n = 114,486. Considering the size of the sample, significance testing is not required. Pairwise odds ratios (O) are as follows: O(weak|dat, weak|acc) = 1.030, O(weak|gen, weak|acc) = 1.557, O(weak|gen, weak|dat) = 1.511.

	strong	weak	proportion weak
accusative	54,556	1,053	0.0189
dative	36,755	731	0.0195
genitive	20,767	624	0.0291

4. Conclusions

In Section 2, I made it clear that I do not wish to argue for specific models of mental representation of linguistic categories, but merely for a class of theories which can account for effects of similarity and fuzziness in linguistic categorization. The hypotheses and the predictions for the corpus study developed in Section 2.4 were nevertheless very specific and could easily have been refuted if (i) cognitive theories of categorization were incorrect, (ii) corpus data were not an appropriate testing ground for such theories, or (iii) the chosen corpus were inappropriate. Since the predicted effects could all be

demonstrated, all these scenarios are unlikely. Specifically, I have shown that the proto-typical semantic and phonotactic properties of weak nouns as described by Köpcke (1995) influence the alternation strength of those nouns in exactly the direction suggested by the theory: the more prototypical a weak noun is, the lower its probability of occurring in a strong form is. The phonotactic properties, which are more exclusive to weak nouns than the semantic properties, have proven to have a higher cue validity than the semantic properties. Furthermore, the demonstrated significant influence of grammatical case on the alternation strength was expected under a theory which assigns a cognitive status to paradigms and their structure, because the alternation is stronger in those forms of the weak nouns (accusative and dative) which violate most clearly the most productive generalizations about paradigms of German nouns. Finally, the fact that higher lemma token frequencies tend to prevent the alternation from occurring strengthens a usage-based interpretation of the phenomenon. The effects of the noun's semantics and grammatical case were cross-examined by showing that strong nouns with properties prototypical of weak nouns do in fact occur in weak forms.

From a corpus linguistic perspective, the challenge lay in the fact that the alternative forms occur mostly in non-standard language and are rare, such that a corpus was needed which contained at least some non-standard language, and which was very large. The DECOW12A web corpus was demonstrated to have those properties; in Section 3.2 I showed how the size of the corpus allows researchers to use high-precision but low-recall techniques of automatic non-manual querying and data processing, while still retrieving large-enough samples. In line with Section 3.1, I strongly suggest that this proves the necessity of further development and exploration of freely available web corpora as an invaluable source of insight for quantitative corpus linguistics.

While further work specifically on the weak nouns might include a more systematic look at the productivity of the class in terms of its power to attract strong nouns like the -*or* subclass described in Section 3.4, a more general look at similar alternations in the morphosyntax of German under a cognitive perspective should also be highly fruitful. German is famous for its grammatical *Zweifelsfälle* 'cases of doubt', including many instances of co-existing, alternating forms in nominal and verbal morphology and blurred boundaries between standard and non-standard language. While the rich and in-

sightful existing literature is largely descriptive, diachronic, structuralist, or even normative (cf. Klein 2009), the phenomena are ideal for cognitive explorations such as that undertaken in this paper, and they have great potential for strengthening the usage-based linguistics paradigm in general, but also the position of corpus linguistics within it.

Appendix: weak nouns in the full sample[33]

Abiturient (398/0.013±0.011), Abonnent (706/0.004±0.005), Absolvent (479/0.013±0.01), Adept (305/0.039±0.022), Adjutant (140/0.021±0.024), Adressat (585/0.024±0.012), Adventist (66/0±0), Advokat (262/0.038±0.023), Affe (3579/0.015±0.004), Afghane (254/0±0), Agent (3214/0.059±0.008), Ahn (315/0.162±0.041), Akrobat (68/0.015±0.029), Aktivist (613/0.007±0.006), Alchimist (193/0.005±0.01), Alemanne (83/0.024±0.033), Ammonit (77/0.013±0.025), Analphabet (196/0.005±0.01), Analyst (310/0.032±0.02), Anarchist (326/0.009±0.01), Anästhesist (351/0.011±0.011), Anatom (45/0.267±0.129), Anglist (9/0±0), Antagonist (184/0.016±0.018), Anthropologe (117/0.017±0.024), Antichrist (60/0.217±0.104), Apologet (37/0.027±0.052), Archäologe (598/0±0), Architekt (5168/0.023±0.004), Archont (14/0.071±0.135), Argonaut (7/0±0), Aristokrat (198/0±0), Artist (297/0.444±0.056), Asiate (419/0.002±0.005), Asket (230/0.004±0.008), Aspirant (62/0±0), Assistent (2489/0.016±0.005), Asteroid (1543/0.029±0.008), Ästhet (60/0.017±0.032), Astronaut (556/0.016±0.01), Astronom (359/0.081±0.028), Aszendent (69/0.058±0.055), Atheist (921/0.014±0.008), Athlet (1070/0.014±0.007), Autist (331/0.003±0.006), Autokrat (59/0.034±0.046), Automat (3674/0.05±0.007), Baptist (34/0.029±0.057), Bär (5551/0.089±0.008), Barbare (571/0.012±0.009), Barde (573/0.01±0.008), Baske (23/0±0), Bassist (1024/0.018±0.008), Beduine (214/0±0), Biograph (72/0±0), Biologe (581/0.003±0.005), Böhme (24/0.292±0.182), Bolide (337/0±0), Borusse (77/0±0), Bote (2824/0.025±0.006), Brahmane (283/0±0), Brillant (1378/0.065±0.013), Brite (628/0.005±0.005), Bube (724/0.007±0.006), Buchstabe (5818/0.1±0.008), Bulgare (93/0±0), Bulle (1261/0.016±0.007), Bürge (660/0.015±0.009), Bürokrat (150/0.013±0.018), Bursche (1159/0.004±0.004), Cellist (138/0±0), Chaot (189/0.074±0.037), Chilene (83/0±0), Chinese (1783/0.022±0.007),

Chirurg (2015/0.051±0.01), Choreograph (107/0.056±0.044), Christ (5224/0.025±0.004), Chronist (249/0.008±0.011), Chronograph (185/0.103±0.044), Dadaist (9/0.111±0.205), Däne (346/0±0), Delinquent (137/0.007±0.014), Demagoge (163/0±0), Demiurg (65/0.046±0.051), Demokrat (491/0.006±0.007), Demonstrant (593/0.017±0.01), Dendrit (35/0.171±0.125), Denunziant (95/0.011±0.02), Depp (722/0.191±0.029), Despot (446/0.018±0.012), Dilettant (132/0±0), Diözesan (45/0±0), Dirigent (1298/0.012±0.006), Disponent (146/0.041±0.032), Dissident (192/0±0), Doge (34/0.147±0.119), Dozent (1646/0.021±0.007), Dramaturg (88/0.08±0.056), Drogist (57/0.035±0.048), Druide (868/0.005±0.004), Egoist (231/0.026±0.02), Egomane (126/0.04±0.034), Elefant (6214/0.028±0.004), Elektrostat (59/0.102±0.077), Emigrant (238/0.004±0.008), Epigone (29/0±0), Eremit (235/0.051±0.028), Erotomane (10/0±0), Essayist (19/0±0), Este (35/0.029±0.055), Ethnologe (87/0±0), Eunuch (122/0.025±0.028), Evangelist (190/0.005±0.01), Exeget (21/0±0), Exorzist (252/0.016±0.015), Exot (530/0.051±0.019), Experte (9299/0.002±0.001), Exponent (216/0.037±0.025), Extremist (180/0.006±0.011), Fabrikant (331/0.018±0.014), Falke (920/0.021±0.009), Faschist (251/0.016±0.016), Favorit (2559/0.036±0.007), Feuilletonist (31/0±0), Filialist (57/0.018±0.034), Finalist (100/0.06±0.046), Fink (51/0.373±0.133), Finne (227/0.009±0.012), Florist (145/0.014±0.019), Flötist (53/0±0), Fotograf (4797/0.039±0.006), Franke (478/0.027±0.015), Franzose (2483/0.005±0.003), Friede (8063/0.13±0.007), Friese (435/0.08±0.026), Fundamentalist (263/0.004±0.007), Fürst (2333/0.029±0.007), Galerist (184/0.005±0.011), Ganove (230/0.013±0.015), Garant (342/0.345±0.05), Gardist (209/0.019±0.019), Gatte (374/0.011±0.01), Geck (59/0.441±0.127), Gedanke (19068/0.065±0.004), Gehilfe (573/0.002±0.003), Gendarm (171/0.193±0.059), Generalist (89/0.034±0.038), Genosse (779/0.001±0.002), Geograph (82/0.085±0.06), Geologe (306/0±0), Gepard (322/0.289±0.05), Germane (154/0±0), Germanist (185/0±0), Gigant (403/0.037±0.018), Gitarrist (2067/0.014±0.005), Glaube (11857/0.149±0.006), Götze (635/0.09±0.022), Graf (1589/0.056±0.011), Graph (1117/0.132±0.02), Gratulant (18/0±0), Grieche (1091/0.003±0.003), Grossist (31/0±0), Gymnasiast (303/0.023±0.017), Gynäkologe (571/0.007±0.007), Hanseat (36/0.028±0.054), Hase (4683/0.019±0.004), Havarist (39/0±0), Heide (1400/0.008±0.005), Held (13473/0.042±0.003), Hermaphrodit (50/0.06±0.066), Herr

(3461/0.146±0.012), Hesse (121/0.091±0.051), Hirte (3067/0.002±0.001), Humanist (172/0.006±0.011), Humorist (73/0±0), Hüne (125/0±0), Husar (74/0.068±0.057), Hydrant (510/0.057±0.02), Idealist (280/0.018±0.016), Ideologe (83/0±0), Idiot (2180/0.027±0.007), Ignorant (249/0.004±0.008), Illuminat (51/0.275±0.122), Immigrant (170/0.047±0.032), Imperialist (23/0.043±0.083), Individualist (110/0.045±0.039), Informant (1095/0.016±0.007), Insasse (341/0.006±0.008), Inserent (46/0±0), Instrumentalist (60/0±0), Intendant (220/0.009±0.012), Interessent (2772/0.006±0.003), Internist (795/0.016±0.009), Interpret (846/0.025±0.01), Ire (362/0.011±0.011), Israelit (126/0.008±0.016), Jesuit (252/0.008±0.011), Journalist (6904/0.01±0.002), Jude (4558/0.005±0.002), Junge (79464/0.006±0), Jurist (2489/0.01±0.004), Kabarettist (1549/0.001±0.002), Kalif (158/0.032±0.027), Kalligraph (27/0.037±0.071), Kamerad (3704/0.017±0.004), Kandidat (8728/0.012±0.002), Kannibale (148/0.014±0.019), Kapitalist (414/0.01±0.009), Karikaturist (107/0.009±0.018), Kartograph (54/0.037±0.05), Kasache (43/0±0), Katalane (28/0±0), Katholik (738/0.015±0.009), Kaukase (49/0±0), Kelte (65/0±0), Kentaur (45/0.244±0.126), Kirgise (22/0±0), Kleptomane (15/0±0), Klient (1700/0.011±0.005), Klingone (655/0.002±0.003), Knabe (2249/0.008±0.004), Knappe (15073/0.002±0.001), Koeffizient (224/0.031±0.023), Kollege (25536/0.009±0.001), Kolonist (109/0.018±0.025), Kolumnist (86/0.035±0.039), Komet (1669/0.023±0.007), Kommunist (808/0.01±0.007), Komponist (2636/0.004±0.002), Kongolese (38/0±0), Konkurrent (6921/0.01±0.002), Konsonant (427/0.089±0.027), Konsument (669/0.012±0.008), Kontrahent (890/0.003±0.004), Konvertit (116/0.009±0.017), Kopist (66/0±0), Korrespondent (323/0.009±0.01), Korse (17/0±0), Kosmonaut (61/0.049±0.054), Krake (229/0.048±0.028), Kroate (182/0±0), Kryostat (41/0.146±0.108), Kunde (26576/0.007±0.001), Kurde (241/0±0), Laie (5673/0.006±0.002), Leopard (929/0.238±0.027), Lette (30/0.033±0.064), Libanese (147/0±0), Ligist (142/0.042±0.033), Linguist (114/0±0), Literat (284/0.004±0.007), Lithograph (21/0±0), Liturg (7/0.143±0.259), Lobbyist (220/0.023±0.02), Lombarde (5/0.2±0.351), Lotse (399/0±0), Löwe (4823/0.017±0.004), Lymphozyt (10/0.3±0.284), Magnet (3761/0.147±0.011), Marxist (186/0±0), Maschinist (149/0.054±0.036), Materialist (85/0.012±0.023), Matrose (649/0.005±0.005), Mensch (221210/0.017±0), Meteorit (1026/0.042±0.012), Mime (54/0.13±0.09), Ministrant (112/0.027±0.03), Minotaur

(57/0.263±0.114), Misanthrop (87/0.08±0.057), Mohr (123/0.203±0.071), Monarch (687/0.051±0.016), Monolith (379/0.185±0.039), Monopolist (344/0.012±0.011), Moralist (178/0.006±0.011), Mutant (364/0.049±0.022), Name (107348/0.028±0.001), Narr (2235/0.029±0.007), Nationalist (178/0.006±0.011), Neffe (795/0.013±0.008), Nekromant (350/0.037±0.02), Neophyt (33/0.182±0.132), Neurologe (1725/0.005±0.003), Nomade (157/0.013±0.018), Obelisk (640/0.188±0.03), Obrist (42/0.071±0.078), Ochse (1358/0.012±0.006), Ökonom (287/0.101±0.035), Oligarch (69/0.014±0.028), Opponent (74/0.081±0.062), Opportunist (136/0±0), Optimist (391/0.02±0.014), Organist (390/0.008±0.009), Orientale (131/0.008±0.015), Ostheopath (51/0.039±0.053), Pädagoge (629/0±0), Päderast (43/0±0), Page (350/0.417±0.052), Paragraf (321/0.019±0.015), Parasit (679/0.047±0.016), Partisan (180/0.089±0.042), Pate (1175/0.004±0.004), Pathologe (408/0±0), Patient (18232/0.01±0.002), Patriot (297/0.114±0.036), Pazifist (169/0.024±0.023), Perfektionist (150/0.013±0.018), Pessimist (211/0.047±0.029), Pfaffe (146/0.007±0.013), Phantast (101/0.02±0.027), Pharmazeut (49/0±0), Philanthrop (28/0.036±0.069), Philologe (116/0±0), Philosoph (2330/0.011±0.004), Pianist (1124/0.017±0.008), Pilot (4226/0.048±0.006), Planet (19683/0.018±0.002), Planetoid (166/0.012±0.017), Poet (428/0.023±0.014), Pole (1013/0.139±0.021), Politologe (88/0±0), Polizist (10166/0.018±0.003), Polyp (219/0.183±0.051), Populist (117/0.009±0.017), Portugiese (189/0.005±0.01), Posaunist (56/0±0), Potentat (91/0.022±0.03), Präfekt (107/0.028±0.031), Praktikant (1542/0.034±0.009), Prälat (117/0.009±0.017), Präsident (6013/0.016±0.003), Prinz (3152/0.085±0.01), Proband (491/0.018±0.012), Produzent (2426/0.012±0.004), Prokurist (341/0.003±0.006), Prolet (99/0.01±0.02), Propagandist (50/0.02±0.039), Prophet (3291/0.022±0.005), Proselyt (29/0±0), Protagonist (1103/0.011±0.006), Protestant (267/0.004±0.007), Protokollant (80/0.012±0.024), Psychopath (873/0.016±0.008), Publizist (156/0.006±0.012), Purist (42/0.024±0.046), Pyromane (51/0.02±0.038), Quadrant (285/0.021±0.017), Querulant (142/0.007±0.014), Quotient (247/0.077±0.033), Rabe (1169/0.022±0.008), Radiologe (273/0.007±0.01), Rassist (378/0.042±0.02), Realist (210/0.024±0.021), Rebell (587/0.143±0.028), Referent (1608/0.021±0.007), Regent (335/0.024±0.016), Rekrut (302/0.013±0.013), Renegat (51/0.02±0.038), Reservist (93/0.011±0.021), Revisionist (30/0±0), Rezensent (325/0.009±0.01), Rezipient

(128/0.008±0.015), Riese (15717/0.002±0.001), Rivale (1007/0.001±0.002), Romanist (18/0±0), Rüde (4923/0.009±0.003), Rumäne (224/0.004±0.009), Russe (1244/0.006±0.004), Sachse (240/0.042±0.025), Sadist (178/0.011±0.016), Same (998/0.063±0.015), Sarde (15/0.067±0.126), Satellit (2631/0.037±0.007), Schamane (1589/0.01±0.005), Schenk (247/0.065±0.031), Scherge (100/0±0), Schiit (36/0±0), Schimpanse (603/0.007±0.006), Schöffe (191/0±0), Schotte (387/0.034±0.018), Schultheiß (120/0.367±0.086), Schurke (1202/0.015±0.007), Schwabe (408/0±0), Schwede (635/0.006±0.006), Seismograph (156/0.064±0.038), Semit (7/0±0), Senegalese (37/0±0), Separatist (22/0.136±0.143), Serbe (215/0.005±0.009), Sklave (2664/0.005±0.003), Slawe (21/0±0), Slowake (69/0±0), Slowene (58/0±0), Solipsist (18/0±0), Solist (336/0.015±0.013), Sophist (52/0.019±0.037), Sorbe (12/0±0), Sozialist (263/0.027±0.02), Soziologe (303/0±0), Spekulant (152/0.02±0.022), Spezialist (8607/0.011±0.002), Statist (165/0.055±0.035), Stipendiat (115/0.139±0.063), Stratege (222/0±0), Student (6399/0.03±0.004), Stylist (153/0.033±0.028), Sudanese (38/0±0), Sunnit (31/0.032±0.062), Sympathisant (113/0.053±0.041), Szenarist (11/0±0), Technokrat (38/0±0), Telefonist (20/0.1±0.132), Telegraph (83/0.024±0.033), Terrorist (1899/0.034±0.008), Theologe (1414/0.003±0.003), Therapeut (6623/0.013±0.003), Tomograph (31/0±0), Tourist (1868/0.029±0.008), Transvestit (255/0.043±0.025), Trilith (22/0.136±0.143), Trotzkist (27/0±0), Tscheche (263/0.008±0.01), Tscherkesse (8/0±0), Türke (2235/0.007±0.004), Tyrann (1165/0.078±0.015), Untertan (291/0.474±0.057), Urologe (1206/0.009±0.005), Vasall (280/0.107±0.036), Veteran (731/0.248±0.031), Virtuose (520/0±0), Waise (74/0.014±0.026), Welsche (26/0±0), Westfale (100/0.04±0.038), Wille (11581/0.132±0.006), Zar (125/0.224±0.073), Zentaur (139/0.266±0.074), Zeuge (11613/0.005±0.001), Zionist (57/0±0), Zivilist (604/0.008±0.007), Zoologe (87/0.011±0.022), Zyklop (229/0.031±0.022), Aeronaut (17/0.824±0.181), Bauer (13842/0.068±0.004), Bayer (1232/0.119±0.018), Centaur (50/0.6±0.136), Doktorand (319/0.031±0.019), Infant (8/0.375±0.336), Infanterist (234/0.021±0.018), Resident (171/0.567±0.074), Steinmetz (491/0.839±0.032), Titan (528/0.57±0.042), Zeolith (30/0.6±0.175)

Endnotes

1   http://www.spiegel.de/politik/deutschland/philipp-roesler-empoert-fdp-freunde-mit-griechenland-aeusserung-a-845980.html

2   http://www.flegel-g.de/wachstum-wachstum.html

3   http://www.mumia.de/doc/aktuell/991201ai00.html

4   By *alternation strength*, I mean the probability that a specific weak noun occurs in a strong form, possibly in a specific context.

5   Convenient overviews can be found, e.g., in Taylor (2003a), Taylor (2008), Taylor (2015).

6   It should be noticed that, technically, the popular schema-based approach in cognitive linguistics (Langacker 1987; for an overview, see Tuggy 2007) is close to the Aristotelian tradition w.r.t. discreteness of class membership (cf. also Divjak and Arppe 2013: 225–226). However, it has been argued that under the appropriate interpretation, prototypes and schemas are compatible, albeit fulfilling different roles in the cognitive linguistic architecture (Langacker 1987, summarized in Taylor 2003a: 69–72).

7   Close relatives of exemplar models are so-called analogical approaches. Prominently, Skousen (1989) introduced a model which does without abstraction. Daelemans and van den Bosch (2005) include feature and feature weights, and thus a certain level of abstraction. See Kapatsinski (2014: 4–14; esp. 11–12) for an overview, a comparison and an argumentation in favor of including weighted features in analogical models.

8   Cf. also Murphy (2002: 39–65) on how the two theories both overcome the problems of the Aristotelian view in different ways, but with similar results.

9   Mathematically different formulations have been proposed within prototype theory and similar cognitive theories of categorization and category learning (e.g., Busemeyer et al. 1993 or Kruschke and Johansen 1999), but the differences are of little importance for the present study.

10 On the desirability of such a continuum, cf. also Divjak and Arppe (2013: 224–229), citing Goldberg (2006), Langacker (2010), etc.

11 This is reminiscent of polysemy in lexical semantics, i.e., the association of multiple senses with one form. The difference to classical polysemy, which assumes clearly distinct senses, is that the weighted-feature approach to prototype theory allows for diverse equally prototypical manifestations with varying degrees of similarity. In other words, nothing forces us to assume clearly distinct and/or hierarchically ordered centers of the category, and polycentric categories do not pose a problem. Cf. especially Taylor (2003b, 2006) for a discussion in favor of such a view, although not in the context of prototype theory.

12 There are also studies which, although they do not use prototype-theoretical terminology, can be seen as supporting the application of similarity-based theories to grammar, for example Dąbrowska's (2008) exemplar-based study of Polish-speakers' ability to classify novel nouns with regard to their inflectional class based on, among other things, phonological and morphological similarity to known nouns.

13 There is more work along these lines which does not necessarily use prototype-theoretical terminology. For example, Barth and Kapatsinski (2014, ahead of print) use regression techniques to model the circumstances under which English auxiliaries are and are not contracted (*am* vs. *'m* etc.) in corpus data. Similarly, Baayen et al. (2013) model four grammatical choices in Russian, using corpus data for three of them.

14 For an overview of the available and most popular methods for such tasks, cf. Gries (2014).

15 Of course, the argumentation in Bresnan et al. (2007) is specific to their research question and the corpus data they used. There is no reason, however, to doubt that the findings generalize to other corpora and research questions.

16 *Construct validity* specifies how well a theoretical construct, which cannot be operationalized directly but is decomposed into a number of operationalizable indicator variables, is measured in an experiment. On the subject of construct validity and test validity in general, see Cronbach and Meehl (1955), Campbell and Fiske (1959), or the convenient summary in Maxwell and Delaney (2004: Ch. 1).

17 Kapatsinksi (2014: 15) takes this even further with his hypothetical statement that "[i]n the extreme, some speakers' heads could host exemplar models, and some could contain fairly abstract grammars, and the produced output would be essentially identical."

18 Köpcke (2000: 119) even goes so far as to call the final schwa a marker of humanness or even agentivity for masculine nouns of prototype I. Indeed, there are only a few masculine (weak) nouns ending in schwa which do not denote humans (*Affe* 'ape', *Löwe* 'lion', etc.). This interesting quasi-causal relationship between form and lexical meaning is a far-reaching interpretation which cannot be tested in a corpus study, and is therefore left aside here.

19 The dative singular marker -*e* is now virtually extinct in contemporary German.

20 These generalizations are implicitly or explicitly accepted by most German grammarians, e.g., Eisenberg (2013: Ch. 5) or Schäfer (2015: Ch. 8).

21 This can be illustrated using a slightly simplified calculation. In Section 3.2, I show that only 2.8% of the occurrences of weak nouns are strong. This means that in order to find one strong exemplar by hand, we have to look at 36 concordance lines on average. There are 451 weak noun lemmas in my sample. So, even in order to find one strong exemplar of each noun, we would have to go through $36 \cdot 541 = 16{,}107$ concordance lines. A manual search is clearly not feasible, and an automatic approach is required, even if it has high precision but low recall.

22 The corpus was built and is hosted by Freie Universität Berlin. Information can be found at: http://corporafromtheweb.org/decow14/

23 In the meantime, it has been superseded by the 20.5-billion-token DECOW14A, which is a superset of DECOW12A. There are other COW corpora available in Dutch, English, French, Spanish and Swedish.

24 A sample query as performed in the IMS OCWB (Evert and Hardie 2011): [word="einen"%c] [pos="ADJA" & word=".+en"]* [word="Abiturient"%c]

25 Unfortunately, this also stands in the way of using the lemma as a random effect (cf. Section 3.3).

26 http://www.breitnigge.de/2011/01/14/so-schnell-sind-23-tage-winterpause-rum/

27 Recall from Section 2.4 that I do not attempt a diachronic interpretation. Excluding the very few nouns which might actually be undergoing diachronic change is in line with this decision.

28 Notice how the large sample sizes allow estimation of the population proportions with very high confidence.

29 All calculations were performed using the R statistics software (R Core Team 2014), specifically the *glm()* function to fit the GLM. My approach to regression is guided by Fahrmeir et al. (2013) and Zuur et al. (2009).

30 A word on model selection is in order at this point. Consider the facts that collinearity is low, that the number of regressors is low, and that the regressors are selected based on strong theoretical assumptions. More advanced methods of model selection, in particular multimodel inference (introduced in Burnham and Anderson 2002 and used in the linguistics literature, for example, in Kuperman and Bresnan 2012 and Barth and Kapatsinski 2014 [ahead of print]) are simply not required here. Notice that Burnham and Anderson (2002: 4) state that "[s]imple models with only 1–2 parameters are not the central focus of this book; rather, we focus on models of more complex systems."

31 I also tried to estimate a GLM with an interaction between *Pt* and *Sem*. However, while animateness and non-ultimate accent showed a significant interaction, Nagelkerke's pseudo-R^2 did not increase, while the standard errors and variance inflation factors increased dramatically. I therefore decided to work with the simpler model as reported above.

32 Since a coefficient P for dummy-coded binary variables can be interpreted *ceteris paribus*, we can calculate the effect of two coefficients P_1 and P_2 simply as $P_{1+2} = P_1 + P_2$, or in terms of odds as $O_{1+2} = \exp(P_1 + P_2)$.

33 For each noun, the total count in the full sample and the proportion of strong forms with half a 95% confidence interval (rounded to three decimal digits) is specified.

References

Baayen, R. Harald, Anna Endresen, Laura A. Janda, Anastasia Makarova & Tore Nesset. 2013. Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics* 37. 253–291.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky Wide Web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226.

Barsalou, Lawrence W. 1990. On the indistinguishability of exemplar memory and abstraction in category representation. In Thomas K. Srull & Robert S. Wyer (eds.), *Advances in social cognition, Volume III: Content and process specificity in the effects of prior experiences*, 61–88. Hillsdale: Lawrence Erlbaum Associates.

Barth, Danielle & Vsevolod Kapatsinski. 2014, ahead of print. A multimodel inference approach to categorical variant choice: Construction, priming and frequency effects on the choice between full and contracted forms of "am", "are" and "is". *Corpus Linguistics and Linguistic Theory*. DOI 10.1515/cllt-2014-0022

Bates, Douglas, Martin Maechler, Ben Bolker, Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48.

Biemann, Chris, Felix Bildhauer, Stefan Evert, Dirk Goldhahn, Uwe Quasthoff, Roland Schäfer, Johannes Simon, Leonard Swiezinski & Torsten Zesch. 2013. Scalable construction of high-quality web corpora. *Journal for Language Technology and Computational Linguistics* 28(2). 23–60.

Bittner, Dagmar. 2003. Von starken Feminina und schwachen Maskulina: Die neuhochdeutsche Substantivflexion - Eine Systemanalyse im Rahmen der natürlichen Morphologie. Berlin: ZAS.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Kraemer & Joost Zwarts (eds.),

Cognitive foundations of interpretation, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Burnham, Kenneth P. & David R. Anderson. 2002. *Model selection and multimodel inference: A practical Information-Theoretic approach*, 2nd edn. New York: Springer.

Busemeyer, Jerome R., In Jae Myung & Mark A. McDaniel. 1993. Cue competition effects: empirical tests of adaptive network learning models. *Psychological Science* 4(3). 190–195

Bybee, Joan L. 1985. *Morphology*. Amsterdam, Philadelphia: Benjamins.

Bybee, Joan L. & Paul Hopper. 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.

Bybee, Joan L. & Carol Lynn Moder. 1983. Morphological classes as natural categories. *Language* 59. 251–70.

Bybee, Joan L. & Sandra Thompson. 2000. Three frequency effects in syntax. *Berkeley Linguistic Society* 23. 378–388.

Campell, Donald T. & Donald W. Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56. 81–105.

Cronbach, Lee J. & Paul E. Meehl. 1955. Construct validity in psychological tests. *Psychological Bulletin* 52. 281–302.

Dąbrowska, Ewa. 2008. The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language* 58. 931–951.

Daelemans, Walter & Antal van den Bosch. 2005. *Memory-based language processing*. Cambridge: Cambridge University Press.

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

Divjak, Dagmar & Catherine L. Caldwell-Harris. 2015. Frequency and entrenchment. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*, 53–75. Berlin: De Gruyter.

Dobrić, Nikola. 2015. Three-factor prototypicality evaluation and the verb "look". *Language Sciences* 50. 1–11.

Eisenberg, Peter. 2012. *Das Fremdwort im Deutschen*, 2nd edn. Berlin, New York: De Gruyter.

Eisenberg, Peter. 2013. *Grundriss der deutschen Grammatik: Das Wort*, 4th edn. edited by Nanna Fuhrhop. Stuttgart: Metzler.

Ernestus, Mirjam & R. Harald Baayen. 2004. Analogical effects in regular past tense production in Dutch. *Linguistics* 42(5). 873–903.

Evert, Stefan & Andrew Hardie. 2011. Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. *Proceedings of the Corpus Linguistics 2011 Conference*. Birmingham: University of Birmingham.

Fahrmeir, Ludwig, Thomas Kneib, Stefan Lang & Brian Marx. 2013. *Regression – Models, methods, and application*. Berlin: Springer.

Fox, John & Georges Monette. 1992. Generalized collinearity diagnostics. *Journal of the American Statistics Association* 87. 178–183.

Gelman, Andrew & Jennifer Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gilquin, Gaëtanelle. 2006. The place of prototypicality in corpus linguistics: Causation in the hot seat. In Stefan Th. Gries & Anatol Stefanowitsch (eds.), *Corpus-based approaches to syntax and lexis*, 159–191. Berlin: Mouton de Gruyter.

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

Glyn, Dylan & Kerstin Fischer (eds.). 2010. *Quantitative methods in cognitive semantics: Corpus-driven approaches*. Berlin: Mouton de Gruyter.

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. 2010. Corpus linguistics and theoretical linguistics – a love-hate relationship? Not Necessarily. *International Journal of Corpus Linguistics* 15(3). 327–343.

Gries, Stefan Th. 2014. Corpus and quantitative methods. In Jeanette Littlemore & John R. Taylor (eds.), *The Bloomsbury companion to cognitive linguistics*, 279–300. London, New York: Bloomsbury.

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

Gries, Stefan Th. & Dagmar S. Divjak. 2010. Quantitative approaches in usage-based cognitive semantics. In Dylan Glyn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches*, 331–354. Berlin: Mouton de Gruyter.

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

Hintzman, Douglas, L. 1986. Schema abstraction in a multiple-trace memory model. *Psychological Review* 93(4). 411–428.

Joeres, Rolf. 1996. "Der Friede" oder "der Frieden": ein Normproblem der Substantivflexion. Sprachwissenschaft 21. 301–336.

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

Klein, Wolf Peter. 2009. Auf der Kippe? Zweifelsfälle als Herausforderung(en) für Sprachwissenschaft und Sprachnormierung. In Marek Konopka & Bruno Strecker (eds.), *Deutsche Grammatik - Regeln, Normen, Sprachgebrauch*, 141–165. Berlin: de Gruyter.

Köpcke, Klaus-Michael. 1995. Die Klassifikation der schwachen Maskulina in der deutschen Gegenwartssprache – ein Beispiel für die Leistungsfähigkeit der Prototypentheorie. *Zeitschrift für Sprachwissenschaft* 14(2). 159–180.

Köpcke, Klaus-Michael. 2000. Chaos und Ordnung – zur semantischen Remotivierung einer Deklinationsklasse im Übergang vom Mhd. zum Nhd. In Andreas Bittner, Dagmar Bittner & Klaus-Michael Köpcke (eds.), *Angemessene Strukturen: Systemorganisation in Phonologie, Morphologie und Syntax*, 107–122. Hildesheim, Zürich, New York: Olms.

Kruschke, John K. & Mark K. Johansen. 1999. A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25(5). 1083–1119.

Kuperman, Victor & Joan Bresnan. 2012. The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language* 66. 588–611.

Langacker, Ronald. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford: Stanford University Press.

Langacker, Ronald. 2010. How not to disagree: The emergence of structure from usage. In Kasper Boye & Elisabeth Engberg-Pedersen (eds.), *Language usage and language structure*, 107–143. Berlin: Mouton De Gruyter.

Maxwell, Scott E. & Harold D. Delaney. 2004. *Designing experiments and analyzing data: a model comparison perspective*. Mahwa, New Jersey, London: Lawrence Erlbaum Associates.

Medin, Douglas L. & Marguerite M. Schaffer. 1978. Context theory of classification learning. *Psychological Review* 85(3). 207–238.

Murphy, Gregory. 2002. *The big book of concepts*. Cambridge: MIT Press.

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

Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. *Language* 79. 682–707.

R Core Team. 2014. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Ramscar, Michael, Melody Dye & Malte Hübner. 2013. When the fly flied and when the fly flew: How semantics affect the processing of inflected verbs. *Language and Cognitive Processes* 28(4). 468–497.

Rosch, Eleanor. 1973. Natural categories. *Cognitive Psychology* 4. 328–350.

Rosch, Eleanor. 1978. Principles of categorization. In Eleanor Rosch & Barbara B. Lloyd (eds.), *Cognition and categorization*, 27–48. Hillsdale: Lawrence Erlbaum Associates.

Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson & Penny Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology* 8. 382–439.

Rousseeuw, Peter J., Ida Ruts & John W. Tukey. 1999. The bagplot: A bivariate boxplot. *The American Statistician* 53 (4). 382–387.

Schäfer, Roland. 2015. *Einführung in die grammatische Beschreibung des Deutschen*. Berlin: Language Science Press.

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

Schäfer, Roland & Felix Bildhauer. 2012. Building large corpora from the web using a new effcient tool chain. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk & Ste-

lios Piperidis (eds.), *Proceedings of the eighth International Conference on Language Resources and Evaluation*, 486–493. Istanbul: ELRA.

Schäfer, Roland & Felix Bildhauer. 2013. *Web Corpus Construction*. San Francisco: Morgan & Claypool.

Schmid, Hans-Jörg. 2010a. Entrenchment, salience, and basic levels. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 117–134. Oxford: Oxford University Press.

Schmid, Hans-Jörg. 2010b. Does frequency in text really instantiate entrenchment in the cognitive system? In Dylan Glyn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches*, 101–133. Berlin: Mouton de Gruyter.

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

Storms, Gert, Paul De Boeck & Wim Ruts. 2000. Prototype and exemplar-based information in natural language categories. *Journal of Memory and Language* 42. 51–73.

Sutcliffe, John P. 1993. Concepts, class, and category in the tradition of Aristotle. In: Iven Van Mechelen, James A. Hampton, Ryszard S. Michalski, & Peter Theuns (eds.), *Categories and concepts: Theoretical views and inductive data analysis*, 35–65. London: Academic Press.

Tabak, Wieke, Robert Schreuder & R. Harald Baayen. 2010. Producing inflected verbs: A picture naming study. *The Mental Lexicon* 5(1). 22–46.

Taylor, John R. 2003a. *Linguistic Categorization*, 3rd edn. Oxford: Oxford University Press.

Taylor, John R. 2003b. Near-synonyms as co-extensive categories. *Language Sciences* 25. 637–655.

Taylor, John R. 2006. Polysemy and the lexicon. In Gitte Kristiansen, Michel Achard, René Dirven & Francisco J. Ruiz de Mendoza Ibáñez (eds.), *Cognitive*

linguistics: Current applications and future perspectives, 51–80. Berlin, New York: Mouton de Gruyter.

Taylor, John R. 2008. Prototypes in cognitive linguistics. In Peter Robinson & Nick C. Ellis (eds.), *Handbook of cognitive linguistics and second language acquisition*, 39–65. New York and London: Routledge.

Taylor, John R. 2015. Prototype effects in grammar. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*, 562–579. Berlin: De Gruyter.

Tuggy, David. 2007. Schematicity. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 82–116. Oxford: Oxford University Press.

Thieroff, Rolf. 2003. Die Bedienung des Automatens durch den Mensch. Deklination der schwachen Maskulina als Zweifelsfall. *Linguistik Online* 16(4). 105–117.

Uehara, Satoshi. 2003. A diachronic perspective on prototypicality: the case of nominal adjectives in Japanese. In Hubert Cuyckens, René Dirven & Taylor, John (eds.), *Cognitive approaches to lexical semantics*, 363–391. Berlin: Mouton de Gruyter.

Van Goethem, Kristel & Philippe Hiligsmann. 2014. When two paths converge: Debonding and clipping of Dutch "reuze". *Journal of Germanic Linguistics* 26(1). 31–64.

Winters, Margaret E. 1990. Toward a theory of syntactic prototypes. In Savas L. Tsohatzidis (ed.), *Meanings and prototypes. Studies in linguistic categorization*, 285–306. London, New York: Routledge.

Wurzel, Wolfgang U. 1985. Deutsch "der Funke" zu "der Funken": ein Fall für die natürliche Morphologie. *Linguistische Studien des Zentralinstituts für Sprachwissenschaft der Akademie der Wissenschaften der DDR* A(127). 129–145.

Zeldes, Amir. 2012. *Productivity in Argument Selection*. Berlin, Boston: De Gruyter.

Zuur, Alain F., Elena N. Ieno, Neil Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. Berlin: Springer.