

Ich haute aufs Land ab, melkte eine Kuh, und backte einen Kuchen: Regularization of strong verbs in modern German

Conni Covington
University of Georgia

Abstract Regularization of Germanic strong verbs is an expected process. Lieberman et al. (2007) and Carroll et al. (2012) showed that less common verbs are more likely to regularize than more common verbs. Edited texts represent a more formal register than unedited texts, and change in progress is more likely to be seen in informal texts (Weerman et al. 2013). Three verbs currently undergoing regularization are *backen* ‘to bake,’ *abhauen* ‘to go away,’ and *melken* ‘to milk.’ This corpus-based study supports the findings of Lieberman et al. (2007), Carroll et al. (2012), and Weerman et al. (2013).

1 Introduction

The following paper is a corpus-based pilot study of the regularization of three strong verbs in modern German: *backen*, *abhauen*, and *melken*. Building on the studies by Lieberman et al. (2007) and Carroll et al. (2012), I employ quantitative analysis to test their findings of regularization rates of three verbs currently undergoing change, comparing data from formal and informal text types. Consistent with previous literature, I found that the least frequent verb is the most likely to regularize, and that the most common verb is least likely to regularize. The broader implication of this study is a greater understanding of language change in progress. The structure of this paper is as follows: a discussion of Germanic verbs and their regularization, register as a factor in the process, method, data, discussion, and a conclusion.

1.1 Germanic verbs

Germanic verbs are classified as ‘strong’ (i.e. irregular) and ‘weak’ (i.e. regular). Weak verbs form the preterite using a dental suffix, which is an innovation in the Germanic languages and persists as *-te* in modern German and *-ed* in modern English (Salmons 2012: 78).

Strong verbs in the e-group derive from Proto-Indo-European (PIE) stems and continue to follow the PIE Ablaut pattern, which in Germanic terms is e-grade present stem, a-grade preterite singular, zero-grade preterite plural, and zero-grade (classes I–III) or lengthened e-grade (IV–V) past participle (Salmons 2012: 73). The verbs in the e-group can be further classified into five classes based on the shape of their root, as shown in table 1. The preterite is formed in this group by changing the <e> in the stem to an <a>, and the past participle is formed by altering vowel quality, as in *singen/sang/gesungen* (‘sing, sang, sung’ class IV) or *helfen/half/geholfen* (‘help, helped’ class III). Classes III and IV gain a <u> in the participle through anaptyxis. The origin of the ē in class V is a matter of debate. The consonant cluster /gb/ is not permitted in German, and faced with the zero-grade preterite plural form *gbun, a speaker of Proto-Germanic would have added a vowel. It is possible that /ē/ spread from the /e/ in the past participle, which itself spread from the root stem, or as analogy from the Class VI verbs, which lengthen the vowel (Van Coetsem 1994: 124). Germanic has two additional verb classes in the a-group, only one of which, class VI, persists today¹. These class VI verbs follow the pattern a-grade present stem, ō-grade preterite singular and plural stem, and a-grade past participle (Salmons 2012: 75); that is, the preterite is formed by changing the stem vowel to an ō, and the participle has the same vowel as the stem, as in OHG *faran/fōr/gefaran* (‘travel, traveled’). Class VII formed the preterite by reduplication (Salmons 2012: 75), as in Gothic *haitan/haihait/haitan* (‘to be called,’ ModGm *heißen/hieß/geheißen*). Class VII had several stem shapes, which were similar to classes I–III (Van Coetsem 1994: 125–127). Modern German does not have reduplication as a strategy to form the past tense, and the class VII verbs developed an ablaut pattern in the past tense through reanalysis and restructuring (Jasanoff 2007: 261). These a-group verbs are also shown in Table 1.

1.2 Verb tenses in German

Modern German has two forms of the past tense: a periphrastic perfect and an analytic preterite. The perfect takes *haben* ‘to have’ or *sein* ‘to be’ as the auxiliary verb and pairs with the past participle, generally formed with a circumfix of *ge-* and *-en* for strong verbs (example 1a) or *-t* for weak verbs (1b). The preterite is

¹Some scholars argue that class VII persists in Dutch, but the main feature that sets class VII apart, reduplication, is no longer productive in Germanic languages

Class	Template	Present	Pret. Sg.	Pret. Pl.	Past part.	Meaning
I	CejC	<i>dreiban</i>	<i>draib</i>	<i>dribun</i>	<i>driban</i>	‘to drive’
II	CewC	<i>leugan</i>	<i>laug</i>	<i>lugun</i>	<i>lugan</i>	‘to lie’
IIIa	CeLC	<i>helpan</i>	<i>halp</i>	<i>hulpun</i>	<i>hulpan</i>	‘to help’
IIIb	CeNC	<i>drinkan</i>	<i>drank</i>	<i>drunkun</i>	<i>drunkan</i>	‘to drink’
IV	CeL	<i>stelan</i>	<i>stal</i>	<i>stēlun</i>	<i>stulan</i>	‘to steal’
V	CeC	<i>geban</i>	<i>gab</i>	<i>gēbun</i>	<i>geban</i>	‘to give’
VI	CaC	<i>faran</i>	<i>fōr</i>	<i>fōrun</i>	<i>faran</i>	‘to drive’
VII		<i>haitan</i>	<i>haihait</i>	<i>haihaitun</i>	<i>haitan</i>	‘to be called’

Table 1 Germanic strong verbs by class. Adapted from [Salmons \(2012: 74–75\)](#)

formed as discussed above, either with a stem vowel change for strong verbs (2a) or the dental suffix *-t* for weak verbs (2b).

- (1) a. *Meine Schwester hat mich gesehen.*
 my-NOM-FEM sister-NOM has-3.SG.PRES.IND me seen-PPART
 ‘My sister saw/has seen me.’
- b. *Meine Schwester hat das gekauft.*
 my-NOM-FEM sister-NOM has-3.SG.PRES.IND that bought-PPART
 ‘My sister bought/has bought that.’
- (2) a. *Meine Schwester sah mich.*
 my-NOM-FEM sister-NOM see-3.SG.PRET.IND STRONG
 My sister saw me.
- b. *Meine Schwester kaufte das.*
 my-NOM-FEM sister-NOM buy-3.SG.PRET.IND WEAK
 My sister bought that.

The German periphrastic perfect, despite being constructed in a parallel manner to English, is not used in the same manner. In German, the perfect is typically used in spoken language for narration, and it also has a resultative use, conveying that the action under consideration has been completed ([König & Gast 2012: 87](#)).

For example, if an event is ongoing, where English can use the present perfect ‘I have lived in Athens for two years,’ German uses the present ‘Ich wohne in Athens seit zwei Jahren,’ literally ‘I live in Athens since two years.’ The preterite is used primarily in written language and is considered more formal, although the two tenses are not completely interchangeable (König & Gast 2012: 88, 92).

1.3 Regularization of strong verbs

Over time, some strong verbs have been reclassified as weak (i.e., have been regularized). If it is assumed that all verbs in PIE followed the ablaut pattern, then pre-Proto-Germanic would have had a predominance of strong ablauting verbs, which would change as the dental preterite began to become established. Carroll et al. find around 190 strong verbs in Modern German, and thus they comprise only around 2% of all verbs in modern German (2012: 157).

Strong verbs have weakened in English as well as in modern German. Lieberman et al. found that for English, strong verbs ‘regularize at a rate that is inversely proportional to the square root of their usage frequency’ (2007: 714), that is, infrequently used strong verbs are more likely to become regular than commonly used ones, and there is a quantifiable rate at which this occurs, depending on the frequency of use. The notion of a fixed rate of linguistic change should be viewed with skepticism, because the field of glottochronology was inspired in part ‘by the use of Carbon-14 dating in archeology’ (Carroll et al. 2012: 154), and language does not behave the way that radioactive elements do.

Carroll et al. sought to replicate Lieberman et al.’s study for German, and they found a similar result: less frequent verbs were more likely to regularize than the most frequent (2012: 159). However, they did not find a constant rate of change, as Lieberman et al. found in English. Their data suggest that regularization does not occur at a constant rate in German, and that English regularizes verbs to a greater extent than German (Carroll et al. 2012: 162), although this could be a result of the coarser division of English into three periods as opposed to German’s four (Carroll et al. 2012: 163). They also investigated whether the frequency of each verb class had an effect on rate of regularization and found that the verbs in classes I–V were more common and less likely to regularize than the verbs in classes VI and VII. The regularization rate of classes I–V ranged from 6.3% (V) to 10.5% (III), with class IV having the highest rate in that group at 20%, close to the 26.7% regularization rate of class VI, while class VII regularized at a

rate of 41.7% (Carroll et al. 2012: 164). Regularization is an expected process in when morphological patterns are irregular, because the regular (here, weak) form is typically the default: new verbs in both German and English form their preterites with the dental ending. Both Lieberman et al. (2007: 713) and Carroll et al. (2012: 157) use ‘to google’ or ‘*googeln*’ as an example: he googled it, or *er googelte es*.

Newberry et al. (2017) analyzed verbs with multiple preterite forms in the Corpus of Historical American English and found that some verbs regularized while others became irregular. They suggest that frequency has an effect on this as well: weak *dive/dived* starts showing a preference for *dive/dove* when there is ‘a marked increase in the use of the irregular verb *drive/drove* in the corpus’ as cars become more common (Newberry et al. 2017: 124). In this case, phonological form affects morphological categorization, mimicking the PIE verb classes, and, rarely, weak verbs can be drawn into the strong classes². Frequency and phonological form are two factors in competition with each other, both of which can determine membership in a given strong verb class, or shift to (or from) it.

1.4 Change in progress

Three verbs which appear to be currently undergoing regularization are *backen* (‘to bake’), for which *buk* alternates with *backte*; *melken* (‘to milk’), for which *molke* alternates with *melkte*; and *abhauen*, for which *hiebt* alternates with *haute* (Carroll et al. 2012, Kubczak 2016). For all three verbs, the past participle also shows signs of regularization (Kubczak 2016: 27, 28, 30).

The existence of multiple past tense forms is subject to social evaluation and metalinguistic discussion, which indicates that this change is above the level of consciousness. Popular language columnist Bastian Sick in his column ‘Zwiebel-fisch’ discusses the seeming flexibility of past tense formations in spoken German, beginning with an anecdote about having milked a cow for the first time and telling an interviewer that he ‘hätte die Kuh ‘gemelkt’. Zum Glück hat die Zeitung das nicht gedruckt [sic]³’ (Sick 2008). Throughout the column, he intentionally uses the prescriptively wrong inflectional forms, and he makes reference to the rhyming patterns: ‘ich singe, ich sang, ich habe gesungen. Ich klinge, ich klang, ich habe

²For a comparative discussion of the effects of frequency and analogy in irregularization across the Germanic languages, see Nübling 2000 and 2010.

³‘had milked [weak pp for strong verb] the cow. Luckily the newspaper didn’t print [strong pp for weak verb] it.’

geklungen. Ich bringe, ich brang, ich habe. . . Nun ja, mit den Analogien ist das so eine Sache⁴ (Sick 2008).

1.5 Register as a factor

Some styles of text, such as literature and newspapers, are more formal, and they are usually composed and/or edited to follow a prescriptive norm (Weerman et al. 2013: 354). Because of this, language change is reflected more slowly in this kind of medium than in informal styles of writing (Weerman et al. 2013: 353, Salmons, 2012: 83), because an editor is able to force a formal text to adhere to the prescriptive standards, rather than allowing it to reflect language as it is spoken.

Intended audience also plays a role in the register that a speaker chooses. A speaker typically opts for a different style when talking with peers than with their boss, a parent, or a member of the clergy. This selection of register can carry over into written language, in that a personal blog may use a more informal style because the target audience is peers, or the writer could choose a more formal style because the target audience is an unknown public.

Informal written language is not necessarily equivalent to spoken language (Weerman et al. 2013: 356), but it can be compared to formal written language in order to ‘evaluate the linguistic status of a phenomenon [. . .] without having to rely on assumptions about relations between written and spoken language’ (2013: 356). Weerman et al. investigated synchronic variation in case marking in 17th-century Dutch using texts ranging from formal (a written history of the Netherlands) to informal (personal letters). They selected genitive and dative case marking to test whether it reflected a change in grammar as it related to register. They found that genitive marking was almost entirely absent from the informal documents and present only around a third of the time in the formal documents (2013: 370), and that dative was not marked on indirect objects in any of the less formal documents, while it was marked in all instances in the formal documents (2013: 372). This is highly suggestive that ‘genitive case marking had fallen out of use in the informal language, which we assume more closely reflects the spoken language at that time’ (2013: 374).

⁴‘I sing, I sang, I have sung. I ring, I rang, I have rung. I bring, I brang, I have. . . yeah, analogies are like that.’

1.6 The current study

The internet provides a wealth of written language, including both more and less formal styles, and internet-based corpora allow researchers to mine data from sources ranging from blogs to newspapers. If informal styles of writing potentially offer a better reflection of language change in progress than formal styles do, using corpora that include internet writing should allow researchers to compare styles in a variety of formality levels. It should be noted that writing on the internet is not equivalent to spoken language, but, like Weerman et al.'s letters, it presents an informal source.

This paper investigates the regularization of *backen*, *abhauen*, and *melken* in formal and informal written language. These verbs were chosen because they appear to be currently undergoing regularization. These three verbs belong to the e-group, although *hauen* was historically in class VII and has been reanalyzed to belong in class I. The verb *abhauen* 'to go away (colloquial)' was chosen over *hauen* 'to heave' because of its greater frequency in colloquial language. The preterite is primarily used in written language, such as novels and newspapers, and uncommonly used in speech, while the periphrastic past, which uses the past participle, is preferred in speech. A less frequently used form is predicted to regularize more quickly, as discussed above. It is hypothesized that informal media will show higher usage rates of the weak preterite forms for all three verbs and that the strong preterite will be preferred for all verbs in formal media.

2 Method

Frequency of strong and weak forms of each verb was investigated using corpora hosted at the Digitales Wörterbuch der deutschen Sprache (DWDS). The formal corpus was the combined DWDS Kernkorpus 1900–1999 and Kernkorpus 21 (2000–2010)⁵, which includes literature, technical literature, and newspapers (137 million tokens), and the informal corpora were Blogs (102 million tokens, 2003–2014) and the Webkorpus 2016 (3 billion tokens, 2001–2016). The web corpus contains both professional and personal websites⁶, so it contains both formal and informal styles. Each verb was searched for the exact wordform using the search

⁵This represents a change from an earlier version of this paper, which used data back to 1600, in order to focus on a narrower time window to allow for a more equal comparison.

⁶See www.dwds.de/d/k-spezial#ibk_web_2016 for details (in German)

string ‘backte’, etc., in each corpus, which excludes other forms of the lemma that would be found in a lemma-based search. For *abhauen*, the search strings ‘haute && ab’ and ‘hieb && ab’ were used to get the exact form of the verb and its prefix, separated by any number of words.

Time curves were generated using the generator at DWDS.de⁷. This tool only allows users to choose from the Reference corpus (1600–1999), the Newspapers corpus (since 1945), or the aggregated Reference and Newspapers corpus. The aggregated corpus was chosen because it includes the Kernkorpus 1900–1999 and Kernkorpus 21 as well as the German Text Archive (1600–1900) and the daily newspapers *Der Tagesspiegel* (1996–2005) and *Die Zeit* (1946–present). The search was limited from 1900 to the present to match the earlier data. There is no analogue in this tool for the Web or Blogs corpora. The time curves, therefore, primarily reflect the standard language, as the source material is subject to prescriptive editorial standards.

3 Data

	<i>backte</i>	<i>buk</i>	<i>haute ab</i>	<i>hieb ab</i>	<i>melkte</i>	<i>molk</i>
<i>Kernkorpus</i>	0.102	0.241	0.0731	0.0658	0.0731	0.0658
<i>Webkorpus</i>	3.87	5.75	1.34	0.721	0.244	0.293
<i>Blogs</i>	0.00365	0.00365	0.000332	0.000332	0	0.000664
<i>Total</i>	0.130	0.195	0.0455	0.0258	0.0108	0.0126

Table 2 Number of tokens per million of each form of the verbs under consideration in each corpus and in all corpora combined

3.1 Backen

Backen is the most commonly occurring verb of the three across all corpora (Table 2). Forms of the verb *backen* occur most frequently in the web corpus. The strong preterite occurs more frequently in two of the three corpora, over twice as often as the weak preterite in the Kernkorpus (0.241/million versus 0.102/million) and 1.5 times as often in the web corpus (5.75/million versus 3.87/million). Both forms

⁷www.dwds.de/t/plot

occur with equal frequency in the blog corpus (0.00365/million). Overall, the strong preterite is favored.

3.2 *Abhauen*

Abhauen is the second most commonly occurring verb across all corpora (Table 2). Forms of this verb also occur most frequently in the web corpus. The weak preterite occurs more frequently in two of the three corpora, slightly more in the Kernkorpus (0.0731/million versus 0.0658/million) and nearly twice as often in the web corpus (1.34/million versus 0.721/million). Both forms occur with equal frequency in the blog corpus (0.000332/million). Overall, the weak preterite is favored.

3.3 *Melken*

Melken is the least frequent verb overall (Table 2). Forms of this verb also occur most frequently in the web corpus. The weak preterite is slightly more common in the Kernkorpus (0.0731/million versus 0.0658/million, but the strong preterite is slightly more common in the web corpus (0.293/million versus 0.244/million). The blog corpus contains only one usage of the term. Overall, the strong preterite is slightly favored.

3.4 Frequency

The frequency of the three verbs in the aggregated corpus, which includes literature, scientific literature, and newspapers, from 1900 to 2010 can be seen in Figure 1. *Backen* is the most common of these three, ranging from four to five tokens per million words. *Abhauen* begins the twentieth century at 1.75 tokens per million then increases slowly to just under two tokens per million in the 1960s. *Melken* decreases slightly in use over the twentieth century, from just under two tokens per million to 1.6 tokens per million.

At the start of the twentieth century, both forms of the preterite of *backen* were in equal distribution (Figure 2, top panel), but *buk* remains steady around 0.17 tokens per million until the 1980s, when it begins to decline to 0.11 per million in the 2010s and *backte* increases from 0.06 per million in the 1980s to 0.07 per million in the 2010s.

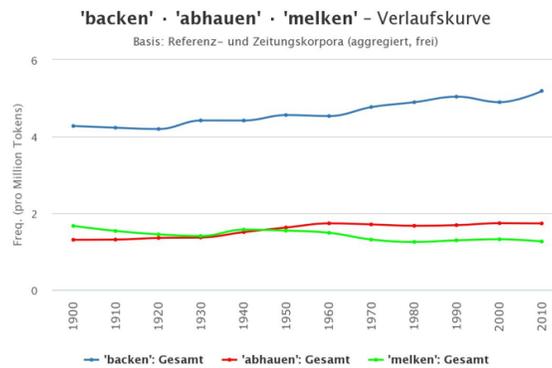


Figure 1 Time course of the frequencies of the lemmas *backen*, *abhauen*, and *melken*, 1900–2018, generated by DWDS, based on the aggregated newspaper and reference corpus.

In 1900, the strong preterite (0.11/million) of *abhauen* was more common than the weak (0.07/million) (Figure 2, middle panel), and the time curves cross in the 1940s, when the weak preterite begins to predominate. The weak form peaks at 0.10 tokens per million in the 1950s then decreases to a steady 0.07 tokens per million, while the strong form declines steadily to 0.02 tokens per million in the 2010s.

Both forms of the preterite of *melken* were in competition throughout the twentieth century and into the twenty-first (Figure 2, bottom panel). The weak preterite predominated from 1920 to 1960 and in the 1980s. Since the 1980s, the strong preterite has predominated, although both forms have decreased steadily in frequency since the 1950s.

4 Discussion and conclusion

The majority of all tokens come from the web corpus, and the quantity from the blog corpus is negligible (Table 2). Greater than 97% of tokens of *backte* and around 95% of tokens of *buk* came from the web corpus (Figure 3). The slight difference in ratio by register suggests that these two forms are in competition, with the less formal register marginally preferring the weak preterite. Around 95% of tokens of *haute ab* and 91% of *hiebt ab* occur in the web corpus (Figure 3). There is a slightly larger difference in rate of occurrence for this verb, with the less formal register somewhat preferring the weak preterite. Around 78% of tokens of *melkte*

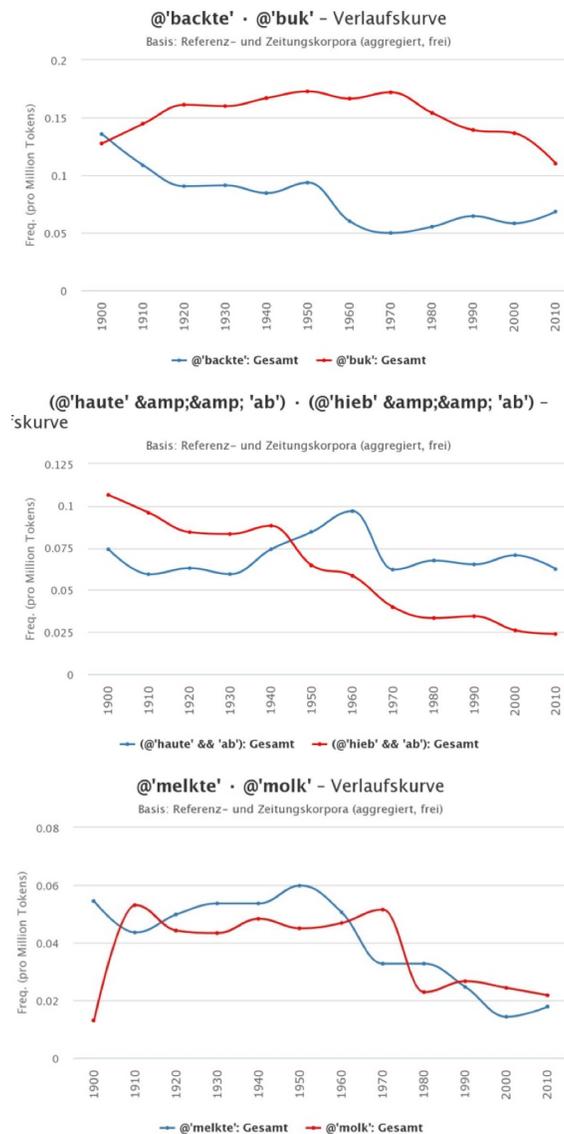


Figure 2 Time curves of the weak vs strong preterite forms, 1900-2017, generated by DWDS, based on the aggregated reference corpus.

and 81% of *molk* occurred in the web corpus (Figure 3). Interestingly, the more formal corpus shows a marginal preference for the weak preterite.

Two trends can be observed from the data: informal media tends to be more likely to use the weak preterite with the more common verbs, but the formal media

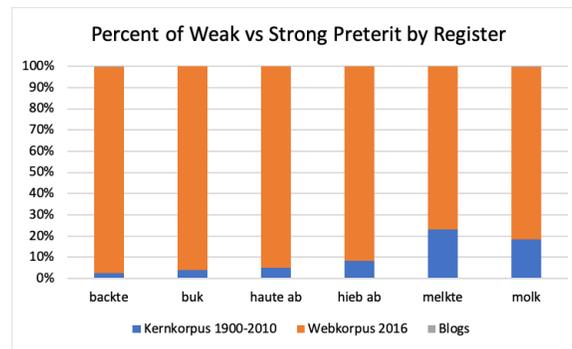


Figure 3 Percent of weak vs strong preterite by register. The *Kernkorpus* represents formal language, *Webkorpus* an intermediate level of formality, and blogs the least formal register

have a slight preference for the weak form, while the informal medium slightly prefers the strong form. In raw numbers, however, there are 25 instances of *melkte* and 30 of *molk* in the Blogs corpus, so there may not be a real difference. The latter trend parallels Carroll et al.'s findings, that less frequent verbs regularized more often, and the former parallels Weerman et al.'s findings, that the change in progress is more visible in the less formal medium. Evidence of language change appears more slowly in formal texts, because these typically have editors who adhere to prescriptive norms. In order to observe change in progress, one needs to examine informal texts as well, because these better reflect language as it is used, although it is not strictly equivalent to spoken language.

The higher rates of *molk* in the internet-based corpora could be related to prescriptive tendencies in, for example, spell check software, or may suggest that a person writing a blog post could check the Duden dictionary online before publishing. That is, written text, even in informal media such as blogs and the web, does not perfectly imitate spoken language, and certain types of formal media do not perfectly represent formal written text. In the late 20th century, computer mediated communication allowed people to communicate more readily and created a space where people could use informal language in a written medium, and the data suggests that the informal medium reflects to some extent language as it is used, rather than a prescriptive standard.

It is consistent with Carroll et al. (2012) that an infrequently used strong verb, like *melken*, should weaken over time, as speakers are not exposed to the strong

preterite form and fall back on the default weak (i.e. regular) paradigm. It is also consistent with their findings that as a strong verb becomes less common, its weak form should enter into competition with the strong form. Three German strong verbs currently undergoing weakening are *backen*, *abhauen*, and *melken*. Of the three, *melken* is the least common, and it appears to be furthest along the path of weakening, with the weak form predominating even in formal texts (see Table 2). *Backen* is the most common, and the strong form predominates in all media types. The majority of tokens of the weak form occur in the less formal medium. *Abhauen* is slightly more common than *melken*, and its weak preterite predominates overall, although the more formal texts show a slight preference for the strong form.

Acknowledgements I would like to acknowledge the anonymous reviewers and Joshua Bousquette, whose suggestions have improved my paper.

References

- Carroll, Ryan, Ragnar Svare & Joseph Salmons. 2012. Quantifying the evolutionary dynamics of German verbs. *Journal of Historical Linguistics* 2(2). 153–172.
- Jasanoff, Jay H. 2007. From reduplication to ablaut: The class VII strong verbs of Northwest Germanic. *Historische Sprachforschung* 120. 241–284.
- König, Ekkehard & Völker Gast. 2012. *Understanding English-German contrasts*. Berlin: Erich Schmidt Verlag.
- Kubczak, Jacqueline. 2016. Backte oder buk, haute oder hieb? schwache oder starke flexion. *Sprachreport* 32(4). 24–31.
- Lieberman, Erez, Jean-Baptiste Michel, Joe Jackson, Tina Tang & Martin Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature* 449. 713–716.
- Newberry, Mitchell G., Christopher A. Ahern, Robin Clark & Joshua B. Plotkin. 2017. Detecting evolutionary forces in language change. *Nature* 551. 223–226.
- Nübling, Damaris. 2000. *Prinzipien der Irregularisierung: Eine kontrastive Untersuchung von zehn Verben in zehn germanischen Sprachen*. Tübingen: Niemeyer.
- Nübling, Damaris. 2010. How do exceptions arise? on different paths to morphological irregularity. In Heike Wiese & Horst J. Simon (eds.), *In expecting the unexpected: Exceptions in grammar*, Berlin: De Gruyter Mouton.
- Salmons, Joseph. 2012. *A history of German: What the past reveals about today's language*. Oxford: Oxford University Press.
- Sick, Bastian. 2008. Heute schon gekrischen? Spiegel Online.
- Van Coetsem, Frans. 1994. *The vocalism of the Germanic parent language: Systemic evolution and sociohistorical context* Untersuchungen zur vergleichenden Grammatik der germanischen Sprachen. Heidelberg: Universitätsverlag C. Winter.

- Weerman, Fred, Mike Olson & Robert A. Cloutier. 2013. Synchronic variation and loss of case: Formal and informal language in a Dutch corpus of 17th-century Amsterdam texts. *Diachronica* 30(3). 353–381.