FREQUENCY AND THE GERMAN(IC) VERB: A HISTORICAL SOCIOLINGUISTIC STUDY OF CLASS VII

by

CONNI COVINGTON

(Under the Direction of Joshua Bousquette)

ABSTRACT

The frequency of Germanic strong verbs influences whether the preterites remain strong or become weak. Analysis of the data suggests that there is a threshold token frequency below which a strong verb will become weak. *Hauen* is considered as a case study, as a verb that shows variability in its preterite in contemporary German.

INDEX WORDS: type frequency, token frequency, Germanic strong verbs, reduplicating verbs, hauen

# FREQUENCY AND THE GERMAN(IC) VERB: A HISTORICAL SOCIOLINGUISTIC STUDY OF CLASS VII

by

CONNI COVINGTON

BS, Juniata College, 1998

MA, University of Georgia, 2018

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2019

# FREQUENCY AND THE GERMAN(IC) VERB: A HISTORICAL SOCIOLINGUISTIC STUDY OF CLASS VII

by

CONNI COVINGTON

| | |
|---|---|
| Major Professor: | Joshua Bousquette |
| Committee: | Jared Klein |
| | Chad Howe |

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2019

# DEDICATION

To Max and Bette Heinrich, for always believing in me, and to my friends in Seventh Prime.

# ACKNOWLEDGEMENTS

I would like to thank Stephanie Adam for providing a list of prefixed German verbs, and Joey Stanley for helping me make a pretty plot in R. I would also like to thank my advisor, Joshua Bousquette, for reading and commenting on the early drafts.

TABLE OF CONTENTS

Page

# LIST OF TABLES

Page

# LIST OF FIGURES

Page

# CHAPTER 1

# INTRODUCTION

This thesis is a corpus-based diachronic study which investigates the effects of token frequency and degree of formality on the weakening (regularization)[1] of class VII strong verbs from Old High German into Modern German. The primary goal of this study is to investigate empirical reasons for why a portion of these verbs have taken on a weak inflection. Analysis of the data suggests that there is a threshold frequency around two tokens per 100,000, below which verbs are more likely to become weak. Degree of formality of the text also influences the visibility of language change, as more formal texts retain older forms, while informal texts reflect ongoing change. Thus an informal text is more likely to show a weak form in the presence of ongoing change than a formal text. The secondary goal of this study is to describe larger trends in the disposition of strong verbs, such as those based on shape of the stem.

This study builds on research by Lieberman et al. (2007) and Carroll et al. (2012), in that it focuses on one subset of the strong verbs and allows for a detailed analysis, rather than a general trend. Whereas Lieberman et al. posit a half-life of strong verbs based on their frequency, Carroll et al. argue that there are more factors that affect weakening than frequency alone. This study argues that there is a usage-based pattern of token frequency that affects why some verbs weaken and other verbs do not.

The Germanic strong verbs are divided into seven classes based on shape of the verb stem, as will be discussed in section 2.1. These classes are divided into two groups based on the

[1] Throughout this thesis, Grimm's terminology "strong" and "weak" will be used to denote verbs called irregular and regular, respectively, in common parlance. See 2.1.

stem vowel, the *e-group* (I-V) and the *a-group* (VI-VII). In focusing solely on the class VII verbs, the confounding factor of type frequency, e.g. class III having a high type frequency and class VII low, is eliminated. Rather than collecting all attested strong verbs into frequency bins as both Lieberman et al. and Carroll et al. did, this study looks at raw frequency, which is possible because of the limited number of attested class VII verbs.

Furthermore, it is possible to test superordinate factors relating to the development of (Proto-)Germanic (PGmc) from Proto-Indo-European (PIE). The *e*-group verbs, inherited from PIE, are as a whole more frequent than the innovative *a*-group verbs, and there appears to be greater similarity among the verbs within each group, i.e. based on the vocalism, than between verbs with similar consonantism between the groups. For example, **haltan* 'to hold' is a class VII, *a*-group verb, but it has a similar consonant structure to **helpan* 'to help,' which is a class III, *e*-group verb. If the consonant structure drives the analogical process, one would expect to see modern German *halten* take on the pattern of *helfen* – but it does not. Rather, it takes on the pattern of its fellow *a*-group class VI verbs, in that the present stem and participle have the same short vowel, and the preterite singular and plural have a long vowel.

This thesis is laid out in this manner: Chapter two reviews literature on the development of Germanic verbs of class VII, the loss of reduplication and the role that frequency plays in language change, and considers previous corpus-based historical studies. Chapter three describes the methods, including previous quantitative studies of strong verbs in Germanic languages as well as a discussion of the effects of style on language use. Chapter four describes the results and the effects of frequency on regularization, and it includes a case study on *hauen* in modern German. The fifth chapter concludes this thesis.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 Germanic verbs

One of the characteristic features of the Proto-Germanic verbal system is its assignment of verbs to one of two classes: a core group, smaller in number, which forms its preterite tense by means of vowel change (ablaut) and a larger group, usually morphologically derived, which forms its preterite by the addition of a dental suffix. Jakob Grimm was the first to employ the terminology "strong" and "weak," respectively, for these two categories The pattern of vowel alternation seen in the first type is a direct continuation of a feature of PIE, whereby the present tense (the source of the equivalent category in PGmc) was most commonly built to a root with **e*-vocalism, while the perfect tense (the source of the PGmc preterite) was built with **o*-vocalism of the root. In regard to the second class, the prerequisites for the PGmc dental preterite were to a certain extene already present in PIE, because of the fact that morphologically derived verbs could not form a perfect, but the employment of a dental formant as a solution to this problem is seen only in Germanic. The employment of ablaut, a process deeply rooted in the grammar of PIE, was continued in PGmc, where its use extended beyond verbal inflection into the realm of nominal and verbal derivation, as seen in pairs such as Old English (OE) *stelan* 'to steal' : *stalu* 'theft' and Gothic (Go) *ganisan* 'to be saved' (strong) : *nasjan* 'to save' (weak) (Fulk 2018:52-53). The dental preterite of weak verbs persists as *-te* in modern German and *-ed* in modern English. The strong verbs are typically called "irregular" and weak "regular" in

grammar books, but, as will become clear, the strong verbs follow a pattern, albeit one less obvious than adding the suffix *-ed* to form the preterite.

The strong verbs fall into two main subgroups, the *e-group* and the *a-group*. Strong verbs in the *e*-group derive from 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 past participle. In the preterite plural, classes IV and V show a lengthened e-grade (Prokosch 1939:164 ff.). The verbs in the e-group can be further divided into five classes based on the shape of their root, as shown in table 2.1.1. The preterite singular is formed in this group by changing the /e/ in the stem to an /a/, the preterite plural by dropping this vowel and inserting the vowel /u/ before a PIE sonorant, and the past participle by copying the vowel of the preterite plural, as in *singan : sang : sungun : sungan-* ('sing, sang, sung' class III) or *helpan : halp : hulpun : hulpan-* ('help, helped' class III). Classes IV and V show more complex relationships between the preterite singular and plural and the participle. The /u/ in the preterite plural and participle of class III is a result of anaptyxis before a PIE resonant or sonorant, and the origin of the /ĕ/ in classes IV and V will be discussed below.

**Table 2.1.1** Germanic strong verbs by class. Adapted from Salmons (2018:75-76), Prokosch 1939

| | *Class* | *Template* | *Present* | *Pret. Sg.* | *Pret. Pl.* | *Past part.* | *Meaning* |
|---|---|---|---|---|---|---|---|
| | I | CejC | dreiban | draib | dribun | driban- | 'to drive' |
| *e-* | II | CewC | leugan | laug | lugun | lugan- | 'to lie' |
| *group* | III | CeLC | helpan | halp | hulpun | hulpan- | 'to help' |
| | III | CeNC | drinkan | drank | drunkun | drunkan- | 'to drink' |
| | IV | CeL | stelan | stal | stēlun | stulan- | 'to steal' |
| | V | CeC | geban | gab | gēbun | geban- | 'to give' |
| *a-* | VI | CaC | faran | fōr | fōrun | faran- | 'to drive' |
| *group* | VII | | haitan | hehait | hehaitun | haitan- | 'to be called' |

The *a*-group of verbs comprises two classes. Class VI has an /a/ in the present and past participle and an /ō/ in the preterite, as in *faran : fōr(un) : faran-* ‘travel’ (Prokosch 1939:174). The origin of /ō/ is a matter of debate. The vowel could go back to PIE /ō/ or /ā/, after the low back merger combined /ā/ and /ō/ into /ō/ (Prokosch 1939:161), or it could be an extension of “the ablaut rule that yielded **ō* in the default past stems of vowel-initial verbs” (Ringe 2006:190), as was also seen with class V **ēt-* ‘eat’ (Ringe 2006:185, 189). In classes V and VI, the vowel of the preterite plural is the lengthened form of the vowel of the present (**etan : *ētun*, class V, Go *malan* : *mōlun* ‘to grind’). Because of the low back merger, a lengthened *a* would become an *ō.*

Class VII contains the reduplicating verbs, whose preterites retain the reduplication characteristic of the PIE perfect. It is assumed that the reduplicating forms, as seen in Gothic *haitan : haihait(un) : haitans* ‘to order, call, be called,’ cf. Modern German (MGm) *heißen*, are the older forms, and that the forms seen in Old High German (OHG), Old English (OE), and Old Norse (ON) are the result of contraction of reduplicated forms, i.e. haplology, in Northwest Germanic (NWGmc) (Prokosch 1939:176).

The origin of the ē in the preterite plurals of classes IV and V is a matter of debate. The consonant cluster /gb/ is not permitted in Germanic, and, faced with the zero-grade preterite plural form **gbun* for a Class V verb, a speaker of Proto-Germanic would have added a vowel. The /ē/ of the preterite plural is likely to have spread from the root stem /e/ and lengthened to /ē/ through analogical processes to the class VI verbs, where the preterite shows a lengthened vowel that is part of a marginally synchronically active alternation process (cf. Go *dags* ‘day’ : *fidurdogs* ‘four days’). For the Class IV verbs, like *neman* ‘take,’ the expected epenthetical vowel to repair a consonant cluster such as **nmun* is /u/, as seen in the class III verbs, e.g.

**hulpun*, which occurs before the PIE syllabic consonants /r l m n/. This strategy is functional in the class III verbs, and, because the Germanic class IV verbs have a PIE syllabic liquid in their nucleus, the expectation is that the class IV verbs would use the same strategy of repair as class IIIa. However, the class III verbs have a heavy coda (RC), and class IV does not (L). This makes them more structurally analogous to class V, which also has a light coda (C). The /ē/ could thus have spread via analogy from class V.

Class VII had several stem shapes, which were similar to classes I-III (van Coetsem 1994:124-127, compare table 2.1.1), and one hypothesis of their disposition suggests that the consonantism influenced the disposition of the *a*-group verbs (cf. van Coetsem 1990:124). The present study was able to test this, and the evidence does not bear this out (see 4.2). Modern German does not have reduplication as a strategy to form the past tense, and the class VII verbs in NWGmc developed an ablaut pattern in the past tense through reanalysis and restructuring (Jasanoff 2007:261). This pattern is similar to class VI, in that the vowels of the present and past participle are the same, and the preterite has one long vowel in both singular and plural.

**Table 2.1.2** Germanic verbs organized by stem shape. Adapted from Salmons (2018:78)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| I | ei + C | VII | a + RC | IV | e + LN | VI | a + L/N<br>a + C |
| II | eu + C | | | V | e + C | | |
| III | e + L/NC | | | | | | |

The weak verbs are an innovation in the Germanic family. Each verb class is morphologically distinct and has its own semantic range or grammatical function, which allows them to become productive grammatical categories. The first weak class is composed of

causatives or factitives, which are denominal, deadjectival, or deverbal, and end with *-jan* (Fulk 2018:294-295), as well as a few primary verbs such as Go *waurkjan* ‘to make, work.’ For example, Go *dauþjan* ‘to kill’ is derived from *dauþs* ‘dead,’ i.e. ‘to cause to be dead.’ The second weak class contains iteratives and denominatives (Fulk 2018:304) such as Go *salbōn* ‘to anoint,’ derived from *salba* ‘ointment, salve.’ The third weak class comprises duratives (Fulk 2018:309) such as OHG *habēn* ‘to have’, and the fourth, which is only semi-robustly attested in Gothic and once in ON, intransitive inchoatives and resultatives (Fulk 2018:315) such as Go *fullnan* ‘to fill up’ and ON *vakna* ‘to wake up.’ The origin of the dental preterite is still a matter of debate. The most widely accepted hypothesis of its origin is a combination of the PIE root *dhē-* ‘do’ with the root verb (Prokosch 1939:194), which then grammaticalized. It is hypothesized to have begun with the class ii verbs, the denominatives, with a meaning “to have done with (noun),” such as **salbo* ‘salve’ and **salbodeþ* ‘applied salve’ (Fulk 2018:292), and the long forms as seen in Gothic reduplicated plurals, e.g. *-dedun*, reflect the earliest PGmc forms, and the shorter forms seen in NWGmc, e.g. OHG *habētun* ‘they had,’ can be derived from these (Prokosch 1939:196 ff.). A recent counterhypothesis suggests that, because the derivations of the NWGmc forms require either *ad hoc* sound changes or restructuring, they are closer to the earlier form and the East Germanic (Gothic) forms are the innovations (Fulk 2018:292, Hill 2010). Hill (2010) suggests that the reduplicated forms in Gothic are incompatible with reconstructions of PGmc with evidence from NWGmc and the Gothic forms were innovated via the optative (448-451).

## 2.2 Reduplication in the Germanic verb

While six classes of strong Germanic verbs formed the preterite through vowel alternation (ablaut patterns) as discussed above, a small minority of strong verbs did so through reduplication. Proto-Germanic formed reduplicated preterites with a prefix of the first consonant of the root followed by an **e*, but if a root began with a cluster of **s* plus a consonant (e.g. **sl-*), the full cluster was reduplicated. If the first phoneme of the root was a vowel, the reduplication prefix was **e*. This reduplication process was inherited from PIE almost without change (Ringe 2006:248-249). However, because the reconstructed PIE perfect was almost exclusively reduplicated, Jasanoff argues that it is logical to say that "all Germanic strong verbs were once reduplicated" and the other classes lost their reduplication because the perfect was differentiated from the present through the vowel alternations, and reduplication was therefore unnecessary (Jasanoff 2007:242). He offers the example of **bîtan* 'bite' undergoing both reduplication and ablaut to give **bebait* and **bebitun*, in which the preterite is doubly marked, and over time, reduplication was lost, potentially beginning with verbs with multiple preverbal elements (Jasanoff 2007:242-243).

Of the Proto-Germanic daughter languages, reduplication is only attested in Gothic, although there are some relic forms in the Northwest Germanic languages. Gothic only has fourteen verbs with attested reduplicated preterites (Fulk 2018:269), which indicates how small this minority is. In the development of PGmc into the daughter languages, the reduplicating forms became less transparent because of sound changes, notably the low-back merger, the change from mobile stress to first syllable stress, and Verner alternations (van Coetsem 1990:85). In Gothic, the reduplication principle was made more transparent in three ways. First,

the effects of Verner's law were levelled (Jasanoff 2007:244, van Coetsem 1990:85). Verner's Law states that PIE /*p, *t, *k/ yield voiced fricatives in Germanic when following unstressed syllables, rather than voiceless fricatives as expected from Grimm's Law. The reduplicated syllable was not stressed, which means that the following syllable was stressed, as in *hai-háit* 'be called,' which Verner's Law would render **hai-gáit*. Evidence for this innovation is seen in Gothic *sai-so* 'sow', compared to ON *sera* < **sezo*, where the *r* is a result of rhotacization of *s* (Jasanoff 2007:244). One reduplicated preterite in Gothic appears spelled with both voiced and unvoiced consonants: *slepan* 'sleep' > *sai-slep* and *sai-zlep* (Fulk 2018:267), which could imply variation. Second, the expected differences between the singular and plural preterite stems were eliminated in class VII. Third, the vowel /*-ai-*/ [ɛ], seen typically in a breaking environment (before /r, h, hw/), was generalized as the reduplicating vowel in all cases (Jasanoff 2007:244).

The relic forms in Old Norse are seen only in the verba pura, as in *róa : rera* 'row,' which are "still transparently reduplicated," while in Old English, the only attestations which suggest reduplication are "poetic and/or Anglian" such as *heht* 'be called' (compare to West Saxon *hét*) (Jasanoff 2007:244-245). Old High German shows remnants in its "*r*-preterites" such as Bavarian and Allemannic *biru* (> *buan* 'dwell'), which is most similar to the Old Norse forms like *rera*, with a productive *r*-element (Jasanoff 2007:246). Although the attested verba pura in OHG are all weak (e.g. *sâen : sâta* 'sow'), the evidence from ON and Gothic points to their original inclusion in the reduplicating class (Jasanoff 2007:246-247).

The question thus arises as to why reduplication ceased to be productive in the Northwest Germanic languages. Because of the change to fixed stress in Germanic and Verner alternations, the reduplication process was not transparent. In Gothic, transparency was achieved through the systematic levelling of Verner alternations and the use of a constant vowel (see above). In

Northwest Germanic, reanalysis occurred. The preterites of strong verbs were mainly formed via ablaut, and, thus, the reduplicating class was under pressure from the ablaut system to conform and create its own ablaut pattern (see 2.3) (van Coetsem 1990:80).

Van Coetsem's analysis considers the accent modification the most important factor (1990:80) because the Verner alternations were no longer conditioned. He considers the verbal prefixes as unbound, because they were unstressed, as evidenced in "infixations" of the reduplicating syllables as in *af-ai-auk* 'increase' (van Coetsem 1990:83). These cannot be termed "infixations" if the prefix is unbound, as van Coetsem argues, because an infix definitionally requires a bound morpheme. Jasanoff disputes the primacy of accent position in this reanalysis, arguing that because the shift to first syllable stress occurred in the Proto-Germanic period, Gothic reduplicated preterites were probably stressed on the first syllable, i.e. the reduplication prefix (2007:263). Gothic preserved the Proto-Germanic cluster rules in reduplication, and this type of onset preservation would not be likely in a language "where the stress was systematically located somewhere *other* than on the initial syllable" of the word (Jasanoff 2007:263).

Four main strategies for the reanalysis of the reduplicating class arose. First is the "*r*-preterite" as seen in ON *rera, sera, grera* (the verba pura) (van Coetsem 1990:72). Second is a reduction of the second (unstressed) syllable and contraction (van Coetsem 1990:73). Third is the development of a new ablaut pattern (van Coetsem 1990:73). The fourth strategy was simply to replace it with the weak preterite (van Coetsem 1990:75).

## 2.3 The loss of reduplication and development of a new ablaut pattern

Many theories have been proposed to explain the loss of reduplication, all involving word-internal prosody. The Germanic accent shift, from mobile to fixed stress, led to alternations in voicing (Verner's alternations or *grammatischer Wechsel*). Germanic words were stressed on the first syllable, but verbs with prefixes, including the reduplication prefix, had their stress on the root syllable, e.g. Go *us-ˋleiþan* 'to go out' and *ˋus-liþa* 'paralytic' (Prokosch 1939:118-119). The proposed theories often rely on the effects of this accent shift on transparency of the product of reduplication.

One theory, proposed by Fulk (1987), suggests contraction as the source of the loss of reduplication. This theory is based on the reduplicating verbs which begin with a vowel, such as Gothic *aikan*, which would reduplicate to *ai-aik(um)*, or with an /h/, such as *haitan*. This theory suggests that analogical processes drove the formation of a new pattern, as the reduplication sequence of these vowel-initial verbs contained an underlying sequence of reduplication vowel and root vowel. There is, however, not enough evidence to support an analogical argument: Only six reduplicating verbs in Gothic begin with a vowel, but few of them are well attested, and two (**aikan* and **alþan*) are only attested with a prefix (Jasanoff 2007:250). Fulk responds that because there are only fourteen reduplicating verbs attested in Gothic, six verbs could "exert outsize analogical influence" on the others (Fulk 2018:269).

Another theory, promoted by Bech (1969) and Vennemann (1994, 1997) which is not supported by any evidence, suggests that syncope of the root vowel and compensatory lengthening led to the loss of reduplication. The reduplicated forms were opaque, and they were reanalyzed as containing infixes. These infixes were extremely variable and were levelled

analogically to *-z-* (or *-R-*). Syncope of the root vowel turned **heRait* to **heRt* and then to **hēt*, with compensatory lengthening (cf. Jasanoff 2007:252 for a summary). Jasanoff rejects this theory for two main reasons: First, Vennemann's argument is predicated on what he knows about upcoming sound changes, which does not likely reflect the psychological reality of actual speakers of these languages. Second, the option of changing an opaque past tense form to the very productive weak class was a "simple recourse" (Jasanoff 2007:255), and, as will be discussed later, remains productive for infrequent strong verbs.

Jasanoff states that it is the opinion of the scholarly community that "some initial group of remodeled reduplicated forms was reinterpreted as having ablaut, and the rest of class VII took off from this nucleus" (2007:261). Van Coetsem's analysis suggests that the reduplication prefix was reinterpreted as the root, and that the original root was determined not to be part of the root, as in **le-laik*, where **le-* was interpreted as ablaut of **la-* (1990:75) The shift to first-syllable stress led to this reanalysis, which created the new ablaut pattern of *-a-* in the present and *-e-* in the preterite, bringing this unusual group of verbs in line with the way that the past tense is formed in the majority of the strong verbs, i.e. through ablaut (van Coetsem 1990:86), as well as making a monosyllabic root (van Coetsem 1990:76).

In his explanation of why there are no traces in Northwest Germanic of reduplicated preterites for verbs like **haldan* or **fallan* (CaRC), Jasanoff turns to compression. The preterite singular of **haldan* would have been **hegald*, and the preterite plural **hegaldun/*heguldun*. This "would have been subject to compression," which would have occurred via syncope of the second vowel to **hegldun* (Jasanoff 2007:268). The cluster *-gld-* is not permissible in Germanic languages, and, because speakers wanted to produce monosyllabic stems, they had two options to simplify this cluster: dropping the *l* to maintain transparency of reduplication while losing

transparency of the root (**hegdun*), or dropping the *g* to maintain transparency of the root while losing transparency of reduplication (**hēldun*, with compensatory lengthening). The reconstructed preterite plural, **heldun*, favors transparency of the root, and appears to follow an ablaut pattern from **hald-*. The preterite singular, however, was still transparently reduplicated, and a system which formed the preterite singular differently from the preterite plural "would naturally have been unstable" (Jasanoff 2007:269). Thus, speakers drew on analogy to the class VI verbs (**faran : *fōr : *fōrun*), where the preterite singular and plural had the same vowel, as did the infinitive and past participle, and changed entirely to ablaut, rendering the paradigm **haldan : *hēld : *hēldun* (Jasanoff 2007:269-270). Jasanoff considers this shift "the birth of class VII" as an ablauting class, because the remaining reduplicating forms were "isolat[ed] and marginaliz[ed]," and the new ablaut pattern **a : *ē* was extended to the verbs with roots in **-ai-, *-au-, *-ē-,* and **-ō-* (Jasanoff 2007:271).

The principal parts of selected class VII verbs in Germanic daughter languages are given in table 2.3.1.

**Table 2.3.1** Principal parts of Class VII verbs in Germanic languages.

| *Class VII* | *present* | *pret. sg.* | *pret. pl.* | *past participle* |
|---|---|---|---|---|
| *PGmc* | *haitan | *hehait | *hehaitun | *haitan- |
| | *lētan | *lelōt | *lelōtun | *lētan- |
| *Gothic* | haitan | haihait | haihaitun | haitans |
| | lētan | lailot | lailotun | lētans |
| *OHG* | heizan | hiaz | hiazun | giheizan |
| | lāzan | liaz | liazun | gilāzan |
| *OE (West Saxon)* | hātan | hēt | hēton | hāten |
| | lǣtan | lēt | lēton | lǣten |
| *ON* | heita | hét | hétu | heitinn |
| | láta | lét | létu | látinn |

## 2.4 Vocalism and *$\bar{e}_2$

Researchers from Brugmann to Vennemann have often conflated the problem of class VII with the origin of Proto-Germanic *$\bar{e}_2$ (Jasanoff 2007:261). It is logical to try to explain the origin of *$h\bar{e}_2t$ in class VII, but that is not the only vowel that needs to be explained, and any theory of class VII must account for all of these things (Jasanoff 2007:261). The previous section discussed theories of class VII, and this section discusses *$\bar{e}_2$.

Fulk (2018) summarizes four theories, starting with Grimm's (1822-37) hypothesis that a loss of the root-initial consonant led to loss of reduplication, as in (**he-hait* > **he-ait* > *hēt* or **ste-stald* > **ste-ald* > OE *steold* 'possessed'). Separately, Brugmann and Wood (both 1895) suggested that the reduplicating verbs had "heavy bases," as opposed to the "light bases" of classes I-V. Obstacles to this became clear with the acceptance of the laryngeal hypothesis, and it is currently out of favor. Third, van Coetsem hypothesized that *$\bar{e}_2$ resulted from the split of Proto-Indo-European /ei/ to /ee/ and /ii/, parallel to the split of /*eu*/ to /*eo*/ and /*iu*/ in OHG. This theory no longer has any proponents. The fourth hypothesis is the most recent, and it suggests that an *e* was inserted into the present stem before the root vowel, as in Northwest Germanic **r-e-ǣd* > OE *red,* or **x-e-ait* > *het* (Fulk 2018:267).

Van Coetsem proposed a theory based on contact assimilation of diphthongs. He raises the question of whether diphonematic diphthongs were subject to umlaut and states that there is direct proof that *-eu-* underwent umlaut. He argues that *$\bar{e}_2$ developed from /*ei*/ due to raising and lowering changes conditioned by umlaut and consonants (van Coetsem 1994:98). He argues that assimilation of a high vowel, e.g. **nisdo* > **nista* > **nesta* > *nest*, also occurred for diphonemic diphthongs, as in **steega* > **steiga* > OHG *stiega* 'to climb' (van Coetsem

1994:107 ff; see table 2.4.1). Jasanoff (2007) and Fulk (2018) argue that van Coetsem's theory lacks evidence and has largely been abandoned. According to Jasanoff, Proto-Indo-European **ei* cannot give both **ē₂* and **ī* in Germanic; there is no evidence, and it is not conditioned (2007:249).

**Table 2.4.1** Proposed paradigm of **steiga* > *stīgu*. Adapted from van Coetsem (1994:107 ff.)

| *era* | *present singular* | *present plural* | *past participle* |
|---|---|---|---|
| *Late PGmc* | *steegă̆ (ē₂)<br>*stiigi-<br>*stiigi | *steega- (ē₂)<br>*steege- (ē₂)<br>*steega- (ē₂) | *gistegan |
| *Later PGmc ("after e → i change had been disrupted")* | *steigă̆<br>*stiigi-<br>*stiigi- | *steiga-<br>*steige-<br>*steiga- | gistigan |
| *Later PGmc* | *stiigă̆ (ī)<br>*stiigi- (ī)<br>*stiigi (ī) | *stiiga- (ī)<br>*stiigi- (ī)<br>*stiiga- (ī) | gistigan |
| *OHG* | stīgu<br>stīgis<br>stīgit | stīgamēs<br>stīget<br>stīgant | gistigan |

Bammesberger suggests that, because Proto-Germanic **ē₁* was open, at least tending toward *a*, **ē₂* could be the product of contraction of *-e-ai-*, which was not recognized as **ē₁*, thus **ē₂* took its place (1986:66). Jasanoff's hypothesis is that a diphthong developed after Proto-

Indo-European **ei* changed to **ī*, "but early enough to have **ē₂* as its reflex in Northwest Germanic" (2007:274).

## 2.5 The role of type and token frequency in language change

Both type and token frequency play major roles in language change. Token frequency is the frequency with which a particular unit of language (which includes single words, a phrase, or even a particular letter) appears in a text (Bybee 2007:9). Type frequency counts patterns of language, as in morphological units or phonotactic sequences (Bybee 2007:9).

Token frequency influences how deeply a form is cognitively entrenched (Dammel et al. 2010:338). Irregular forms[2] are assumed to be more difficult to learn and therefore to be less likely to be retained as a language changes. Frequency of these irregular forms is a factor in whether the irregular form is maintained, as will be discussed in detail later. An irregular form that occurs frequently, such as *to be*, will be more readily available to the speaker's language memory than an irregular form that occurs infrequently. This cognitive entrenchment is the reason that particular strong or even suppletive verbs, such as *eat* or *go*, remain irregular, rather than becoming regular. Indeed, many of the core items of language "constitute most of the exceptions of a grammatical system" (Nübling 2010:139). Token frequency does not remain constant over time, however. Verbs can be replaced by new expressions, as happened in English with OE *cwedan* 'speak,' or the activity they refer to can become less common and the verb thus less used, such as verbs having to do with farming. The verbs that are most often affected are strong or irregular, and they become weak or regular verbs. On the other hand, as verbs

[2] Here, *irregular* includes both strong and suppletive verbs, as well as irregular verbs such as *have* (*has, had*).

grammaticalize, such as the German auxiliary *werden*, they increase in frequency and remain strong or irregular (Nübling 2010:152). As a verb becomes more frequent, it becomes more "resistant to regularizations" (Nowak 2010:372). These irregular verbs have to be learned by rote, rather than a more simple approach as with weak or regular verbs, where the past tense is formed by adding *-ed* to the end. This rote learning places a burden on the memory, but because these verbs are so frequently used, their representation in the memory is strengthened, and these rote-learned forms become more accessible (Nowak 2010:372).

Highly frequent items are not the only ones to resist regularization. Infrequently used fixed phrases, which Bybee calls "prefabs," can "show conservation of archaic language," such as *far be it from me* or *how goes it?*, which preserve the old verb-second Germanic order (Bybee 2007:16-17).

Frequency of use can influence the rate of sound change. High-frequency items undergo reduction "more quickly and radically than low-frequency items," (Bybee 2007:11). One type of sound change that occurs more quickly in high-frequency items is place assimilation. Nübling (2000 and 2010) investigates ten frequent verbs across ten Germanic languages. She finds that these verbs contract (dropping a root consonant as in Middle High German *haben* > *han* 'have') or assimilate (as in colloquial German *haben wir* > *habm wir* > *ham wir* > *hammir* > *hamma* 'we have'), and states that "obviously these assimilations are driven by high token frequency" (Nübling 2010:144-145).

Type frequency is less of a factor in the overall loss of strong verbs in the Germanic languages. Strong verbs account for approximately 2% of all verbs in modern German (Carroll et al. 2012:157 and see 2.5). They are, however, among the most frequently used. Irregular[3] verbs

[3] Bybee's terminology; includes strong and other irregular verbs.

comprise 22 of the 30 most frequent English verbs, and a study of child-directed speech in English found 292 irregular past-tense forms compared to 99 regular forms (Bybee 2007:101-102). Type frequency should not be discounted, however. The most frequent classes of strong verbs proved more resistant to change than the least frequent (Carroll et al. 2012:164), as will be further discussed in 2.6.

## 2.6 Language change and strong verbs

The Germanic languages exhibit a tendency toward frequency-based patterns (see 2.5), and these patterns are a continuation of changes to the strong verb system that were attested earlier in the language. A notable pattern is that strong verbs tend to become weak over time, but the least frequent strong verbs are more likely to become weak over time. The strong verbs represent a minority of all verbs in the Germanic languages, but some classes of strong verbs have a high enough type frequency to remain in the strong class[4] - although this does not guarantee that there will be no changes to the paradigm. A less frequent strong verb could become weak, or it could move to another strong class, or its paradigm could be levelled to create a single characteristic vowel in the past.

Counts of strong verbs in modern Germanic languages range from around 200 in German (Carroll et al. 2012:157) to 210 in Dutch (Knooihuizen & Strik 2014:178), compared to the 500 strong verbs Seebold (1970) lists for Proto-Germanic. It is important to note that not all verbs in Seebold (1970) are attested in all the daughter languages, thus it cannot be assumed that Dutch or German has lost around 300 strong verbs each (Knooihuizen & Strik 2014:178 note 6). However,

[4] This is not tested in the present study.

data clearly points to a trend of movement from strong to weak (Lieberman et al. 2007, Dammel et al. 2010, Carroll et al. 2012, Knooihuizen & Strik 2014, Strik 2014). Carroll et al. find that the rate of weakening of strong verbs is inversely proportional to their frequency (2012:155-156).

Synchronic studies of regularization and levelling in Germanic languages show that change in the strong verb system of the Germanic languages is ongoing, and their data represents the continuation of the process described in the previous sections.

In their 2010 study, Dammel et al. investigate levelling in English, German, Dutch, and Swedish paradigms. Where historically all four languages had a four-part ablaut system, with preterite singular and preterite plural having different vowels, the modern languages have a unified preterite.[5] Swedish levelled to the preterite singular and Dutch to the preterite plural, while in English and German, the direction of levelling depended on the ablaut class (Dammel et al. 2010:338-339). See table 2.6.1. No matter which direction the levelling ultimately took, the process occurred analogically, i.e. built on existing majority patterns in the strong verb system.

The three-part verbal system in the modern Germanic languages can be described based on the vowels present in their principal parts. A verb like *give : gave : given* would be classified as ABA, whereas *sing : sang : sung* would be ABC. In Swedish, levelling toward the preterite singular strengthened the ABA ablaut pattern, seen in classes I, II, and V (Dammel et al. 2010:339), which accounts for 71% of strong verbs in Swedish (e.g. *skriva : skrev : skrivit* 'write') (Dammel et al. 2010:341). Dutch (e.g. *binden : bond : gebonden* 'bind'), German (e.g. *bieten : bot : geboten* 'bid'), and English (e.g. *bear : bore : born*) increased the ABB pattern, which accounts for 84%, 51%, and 60% of ablauting strong verbs in these languages, respectively (Dammel et al. 2010:341).

[5] Only high-frequency verbs in Dutch (e.g. *spreken* 'speak,' *nemen* 'take') maintain the four-vowel ablaut pattern (Nowak 2010:371).

Category, or type, frequency also plays a role in the levelling process. In the Germanic languages, the perfect was originally resultative and evolved a past-time meaning over time. In Dutch and German, it became a past tense form, while in Swedish and English, it retained the aspectual function (Dammel et al. 2010:346-347). Prior to this innovation, which was ongoing in the 9th century, the only past tense form was the preterite, which is aspectually complete for telic verbs and otherwise simply the past tense (Rittenhouse 2014:17-19).

The perfect is formed in Germanic languages with an auxilliary and the past participle, and the preterite of strong verbs is formed through a system of vowel changes, as discussed above. The vowel of the past participle can match any of the other vowels in the paradigm or be different from both (compare table 2.1.1). When the preterite singular and plural levelled, the past participle did not change, but the preterite vowel often levelled to the vowel of the participle, as discussed above. As the perfect grammaticalized in German and Dutch, it became more frequent than the preterite, which could have an effect on the cognitive entrenchment of the strong preterite forms. In spoken modern German, 79% of past tense forms (not counting auxiliaries) are perfect, and in Dutch 54%, while in English, 68% of past tense forms are preterite, and in Swedish 57% (Dammel et al. 2010:352). The extension of perfect to the preterite occurred in the Middle High German (MHG) to Early New High German (ENHG) period, during which ablaut levelling also occurred. This increased frequency of use influenced the strengthening of the ABB pattern, i.e. levelling of the preterite forms to the vowel that matched the past participle (Dammel et al. 2010:353). This increases the opportunity for analogical formations.

**Table 2.6.1** The principal parts of e-group verbs in four NW Germanic languages, showing which vowel was generalized to the unified preterite. Verbs are shown in historical forms. Adapted from Dammel et al. 2010.

| *Language* | *Verb class* | *Present* | *Pret. sg.* | *Pret. pl.* | *Past participle* | *Pattern* |
|---|---|---|---|---|---|---|
| ***Swedish*** | 1 | skriva | skrev | (skrivo) | skrivit | ABA |
| | 2 | bjuda | böd | (budo) | budit | ABA |
| | 3 | finna | fann | (funno) | funnit | ABC |
| | 4 | bära | bar | (båro) | burit | ABC |
| | 5 | giva | gav | (gåvo) | givit | ABA |
| ***Dutch*** | 1 | rîden | (rêd)[6] | rêden | gerêden | **ABB** |
| | 2 | bieden | (bôd) | bôden | gebôden | **ABB** |
| | 3 | vinden | (vand) | vonden | gevonden | **ABB** |
| | 4 | stelen | stal | stâlen | gestôlen | ABB'C |
| | 5 | geven | gaf | gâven | gegeven | ABB'A |
| ***German*** | 1 | rîten | (reit) | riten | geritten | **ABB** |
| | 2 | bieten | bot | (buten) | geboten | **ABB** |
| | 3 | finden | fand | (funden) | gefunden | ABC |
| | 4 | stehlen | (stal) | stâlen | gestolen | ABC |
| | 5 | geben | (gab) | gâben | gegeben | ABA |
| ***English*** | 1 | wrîtan | wrât | (writon) | writen | ABC |
| | 1 | bîtan | (bât) | biton | biten | **ABB** |
| | 2 | clêofan | (clêaf) | (clûfon) | clofen | **ABB** |
| | 3 | singan | sang | (sungon) | sungen | ABC |
| | 3 | spinnan | (span) | spunnon | spunnen | **ABB** |
| | 4 | beran | (bær) | (bæron) | boren | **ABB** |
| | 5 | gifan | (geaf) | gêafon | gifen | ABA |

The ABB pattern in German and Dutch is so strong that Nowak (2010) suggests that an *x-o-o* system represents the formation of a new eighth ablaut class. This class developed via analogy "through the convergence of several strong verbs which all share the pattern *x-o-o*" (Nowak 2010:362). In Dutch, this pattern evolved from the levelling in the preterite of class III (Nowak 2010:364) and, because of its high type frequency, it stabilized the whole class of strong verbs, because less frequent strong verbs can remain strong by shifting class to this *x-o-o* pattern

[6] Parentheses are given as in the source. The authors seem to have assumed that the preterite plural was selected even when both preterite vowels were the same.

(Nowak 2010:373). In German, the *x-o-o* pattern is seen in less frequent strong verbs. The *x-o-o* verbs in German are infrequent, and many of them show evidence of weak variants in ENHG (Nowak 2010:366). The pattern may even have been a stopping point on the path to weakening for some verbs, like *bellen* 'bark' (Nowak 2010:367). The pattern is still productive today: Infrequent strong verbs, such as *schwimmen* 'swim' or *spinnen* 'spin,' both of class IV, are occasionally found in colloquial language with an *o* in the preterite, rather than the expected *a* (*schwimmen : schwamm/schwomm : geschwommen*) (Nowak 2010:368).

The tendency toward weakening over time is very strong in Frisian. Strik (2014) tabulates the change of inflection from Old to Early Modern Frisian and finds that 86 verbs moved from strong to weak, 12 from one strong class to another, and 9 from weak to strong (Strik 2014:458). In class VII, one verb moved into it from another strong class, two verbs moved from class VII into a different strong class, and 17 became weak (Strik 2014:458).

## 2.7 Previous corpus-based historical studies

Several studies have been done on the regularization of strong verbs in modern Germanic languages. Lieberman et al. (2007) compiled a list of strong verbs in Old, Middle (ME), and Modern English (ModE) from grammar textbooks, which was then annotated by hand. They excluded any verbs from OE which were not present in ModE, leaving them with 177 verbs (Lieberman et al. 2007:713). They then searched the CELEX2 database to determine the frequency of each verb in each period of the language and grouped the verbs by logarithms of frequency, which ranged from 1 to $10^{-6}$ (Lieberman et al. 2007:713). The 13 most frequent verbs remained strong from OE to ModE, while of the 12 least frequent, only one remains strong

(Lieberman et al. 2007:713). See table 2.7.1. They then calculated that for English, strong verbs "regularize at a rate that is inversely proportional to the square root of their usage frequency" (Lieberman et al. 2007:714), that is, infrequently used strong verbs are more likely to become weak than frequently used ones, and there is a quantifiable rate at which this occurs, depending on the frequency of use. They also state that they can predict the future of strong verbs, suggesting that "only 83 of the 177 verbs studied will be irregular in 2500" (Lieberman et al. 2007:715). 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. Language, as a tool used by speakers, has variables in its usage both over time and synchronically, such as frequency and style, which affect the composition of the grammar.

**Table 2.7.1** Regularization rate of verbs from OE to ModE. Adapted from Lieberman et al. 2007:713-714.

| *Frequency* | *Number of strong verbs (OE)* | *Number of strong verbs (ModE)* | *Regularization rate (%)* |
|---|---|---|---|
| $10^{-1}$-1 | 2 | 2 | 0 |
| $10^{-2}$-$10^{-1}$ | 11 | 11 | 0 |
| $10^{-3}$-$10^{-2}$ | 37 | 33 | 10 |
| $10^{-4}$-$10^{-3}$ | 65 | 37 | 43 |
| $10^{-5}$-$10^{-4}$ | 50 | 14 | 72 |
| $10^{-6}$-$10^{-5}$ | 12 | 1 | 91 |

**Table 2.7.2** Regularization rate of verbs from OHG to MGm. Adapted from Carroll et al. 2012:159.

| *Frequency* | *Number of strong verbs (OHG)* | *Number of strong verbs (MGm)* | *Regularization rate (%)* |
|---|---|---|---|
| $10^{-1}$-1 | 1 | 1 | 0 |
| $10^{-2}$-$10^{-1}$ | 8 | 8 | 0 |
| $10^{-3}$-$10^{-2}$ | 48 | 47 | 2.1 |
| $10^{-4}$-$10^{-3}$ | 61 | 56 | 8.2 |
| $10^{-5}$-$10^{-4}$ | 40 | 25 | 37.5 |
| $10^{-6}$-$10^{-5}$ | 6 | 1 | 83.3 |

Carroll et al. (2012) sought to replicate Lieberman et al.'s study for German. As Lieberman et al. did, Carroll et al. used reference books to find strong verbs in the four traditional periods: OHG, MHG, ENHG, and NHG (2012:157). They excluded any verbs which were not attested in OHG but appeared later, as well as those which had both strong and weak forms attested in any single period (Carroll et al. 2012:158). Frequency data was collected from the CELEX2 database and the *Frühneuhochdeutschkorpus* (ENHG), and verbs were categorized by log frequency (Carroll et al. 2012:158-159). They found a similar result: Less frequent verbs were more likely to weaken than the most frequent (2012:159). See table 2.7.2. However, they did not find a constant rate of change, as Lieberman et al. found in English. Their data suggest that weakening does not occur at a constant rate in German, and that English verbs weaken 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).

Carroll et al. (2012) also investigated whether the frequency of each verb class had an effect on rate of weakening and found that the verbs in classes I-V were more frequent and less likely to weaken than the verbs in classes VI and VII. The rate of weakening of classes I-V, i.e.

the *e*-group, from OHG into NHG ranged from 6.3% (V) to 10.5% (III), with class IV as an outlier in that group at 20%. Within the *a*-group, class VI weakened at a rate of 26.7%, while class VII weakened at a rate of 41.7% (Carroll et al. 2012:164). See table 2.7.3. Regularization or weakening is an expected process 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*[7].

Change does not only occur from strong to weak. Weak verbs can occasionally become strong. Newberry et al. (2017) investigated language change using techniques developed in evolutionary biology. They counted verbs with multiple preterite forms which occurred at least 50 times in the Corpus of Historical American English (1810-2009) (Newberry et al. 2017:1). They 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*[8] 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. 2017:224). In this case, phonological form affects morphological categorization, mimicking the PIE verb classes, and, rarely, weak verbs can be drawn into the strong classes. 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.

[7] However, this verb does not conform to any existing ablaut patterns and would not be likely to follow a strong inflection. A verb like *ping(en)* (what a computer does to contact another computer), which does, in fact, conform to a strong pattern (*sing(en)*) and inflects weakly (*pinged, pingte*), is a more compelling example.
[8] *Dive* may have had both strong and weak variants in competition. (Jared Klein, p.c.).

**Table 2.7.3** Regularization rate of verbs by class from OHG to MGm. Adapted from Carroll et al. 2012:164.

| *Verb class* | *Number of verbs in OHG* | *Regularization rate (%)* |
|---|---|---|
| I | 37 | 9.1 |
| II | 28 | 8.7 |
| III | 46 | 10.5 |
| IV | 13 | 20 |
| V | 15 | 6.3 |
| VI | 20 | 26.7 |
| VII | 15 | 41.7 |

Computer-assisted studies of corpora can also allow researchers to elucidate the process through which a form changed. Somers (2011) investigates the generally accepted theory that the second person singular *-st* ending in OHG formed as a result of cliticization of the subject pronoun with the original *-s* ending and reanalysis. Because German is a V2 language and speakers often start with a non-subject element, the subject pronoun often follows the verb in an unstressed position, which allows for cliticization.

| | | | | |
|---|---|---|---|---|
| (1) Tatian | (150, 7) | | | |
| | thu | grim | man | bi**st**, |
| | you | fierce | man | are |
| | nimi**st** | thaz | thu | ni sázto**s** |
| | take | that-DEM | you | NEG establish-PRET.IND |
| | ‘You are a fierce man. You take that which you did not establish.’ | | | |
| (2) Otfrid | (II 22, 33; 36) | | | |
| a) | Ouh | gibórge**s** | thu | thés |
| | also | be wary-PRES.IND | you | this-GEN |
| | bítit | er | thih | fisges |
| | asks | he | you-ACC | fish-GEN.SG |
| | ‘Also be wary of this, he asks you for fish.’ | | | |
| b) | harto | bórge**st** | thu | thés |
| | very | be wary-PRES.IND | you | this-GEN |
| | bítit | er | thih | éies |
| | asks | he | you-ACC | egg-GEN-SG |
| | ‘Be very wary of this, he asks you for egg.’ | | | |

(3a) Tatian (62, 12)

| | | | |
|---|---|---|---|
| Fon thinen | uuorton | uuirdis**tu** | girehfestigot |
| through your | words-DAT.PL | become-you | justified |

'Through your words, you become justified (the) justified (one)'

(3b) Otfrid (V 7, 19)

| | | |
|---|---|---|
| Wenan | súachis**tu** | sar? |
| who-ACC | seek-you | right now? |

'Whom do you seek right now?'

There are enough examples of second person singular forms being written in this cliticized manner to allow historical linguists to conclude that this was the way in which the *-st* ending developed. Somers describes an alternative theory, proposed by Sihler in 1986, that the *-st* ending resulted from analogy to the preterite present verbs (Somers 2011:143). Somers uses corpus data derived from Otfrid's *Evangelienbuch* and Tatian to argue that both cliticization and analogy were essential to the spread of the *-st* ending, but cliticization was primary (Somers 2011:144). She finds that the shape of the verbal ending is predictive of the presence of *-st* versus *-s*. The second person singular form of 'to be,' *bist*, is the origin of the form. Because forms of 'to be' are typically the most frequent verbs in a corpus, there is a link here between frequency and language change. A frequently used irregular form can create a pattern for speakers to build an analogy from or to reinforce a pattern that isn't supported by type frequency. In support of the analogy argument, she found that other verbal inflection patterns that ended in *-is,* that is, the present indicative of strong and class i weak verbs, were more likely to show the *-st* ending, while the classes ii and iii weak verbs, whose stems end in ô and ê respectively, were more likely to show the *-s* ending (Somers 2011:168-169; see table 2.7.4). In support of the cliticization argument, she found that position of the verb, whether it was V1/V2 or Vfinal, had an effect on the presence of the *-st* ending. If a verb was in final position, it was more likely to have the original ending, and verbs in first or second positions were more likely to have the innovative

ending (Somers 2011:168, 171-172; see table 2.7.5). Verbs in final position are inherently unable to host clitics.

**Table 2.7.4** Shape of ending and number of tokens by type of ending. Clitic group indicates a formation such as *nimistu*. Adapted from Somers 2011:169.

| | *-st ending* | *clitic group* | *-s ending* |
|---|---|---|---|
| *-is* | 66 (60%) | 14 (12.73%) | 30 (27.27%) |
| *-ôs/-ês* | 8 (34.78%) | 1 (4.35%) | 14 (60.87%) |
| *total* | 74 | 15 | 44 |

**Table 2.7.5** Syntactic position and number of tokens by ending. Clitic group indicates a formation such as *nimistu*. Adapted from Somers 2011:168.

| | *-st ending* | *clitic group* | *-s ending* |
|---|---|---|---|
| *V1* | 6 | 2 | 4 |
| *V2* | 47 | 13 | 15 |
| *Vfinal* | 21 | -- | 25 |
| *total* | 74 | 15 | 44 |

## 2.8 Summary

Previous studies have found a trend toward weakening in Germanic verbs, as well as a strong tendency toward frequency-based analogical patterns. As the verbs of class VII moved from forming a past tense via reduplication, they developed a new ablaut pattern. This is not a unique process, because Nowak (2010) suggests that an eighth ablaut class has been created in the last few centuries as a result of levelling across paradigms. Strong and irregular verbs are vastly outnumbered by weak verbs, yet some of the most frequent verbs are irregular and have remained so over centuries. The inflectional endings of a frequently used verb can also lead to analogical change, as Somers (2011) shows. Frequency of use has implications for language change, including weakening.

# CHAPTER 3

# METHODS

This section focuses on the methodology employed for linguistic analysis of the changes to the class VII verbs in German. The section is structured as follows: in 3.1, previous quantitative studies which investigated weakening of strong verbs in other Germanic languages will be discussed. In 3.2, the effects of audience design on style will be discussed. In 3.3, the methods and hypotheses of this study will be introduced. Section 3.3 is divided further into 3.3.1, in which the method of data collection for frequency of the Germanic verbs will be described, and 3.3.2, in which the method for the variationist analysis will be described.

## 3.1 Previous quantitative studies

In the last decade, several quantitative studies on diachronic language change have been published. The advent of computers has allowed researchers to perform analyses on large corpora that would have been unthinkable fifty years ago. Modern analytical software packages allow linguists to empirically test previous theories of language change, as well as to quantify variation within a historical corpus, for example. It is also possible to quantify the frequency of items within a corpus. This allows modern linguists to get a bigger picture of diachronic language change than was possible in the era before computer-aided analyses. This section discusses several of the quantitative diachronic studies from the last decade.

The Germanic languages show a strong tendency toward analogy based on extant patterns, especially in the verbal inflections. Ablaut levelling can lead to the strengthening of particular patterns (see 2.6). In their 2010 study, Dammel et al. investigate the ways in which the preterite was levelled in strong verb paradigms in English, German, Dutch, and Swedish. They sort the strong verbs in these languages by historical verb class and vowel alternations in the modern language. They additionally look at the frequency of use of the periphrastic perfect versus the preterite in spoken language in each of these languages, because some verbs have mixed paradigms (i.e. strong preterite and weak participle or vice versa). As discussed in 2.1, Germanic verbs originally had a four-part system, where the preterite singular and plural had vowels that differed with respect to quality. In Swedish, this was uniformly levelled to the vowel of the singular, and in Dutch, it was uniformly levelled to the vowel of the plural (Dammel et al. 2010:338). These levelling processes led to the reorganization of ablaut patterns. In Swedish, the ABA and ABC patterns (compare table 2.6.1) maintain a distinction between preterite and past participle (Dammel et al. 2010:339), and the ABA pattern accounts for 71% of verbs (Dammel et al. 2010:342). In Dutch, the ABB pattern was preferred, leading to a uniform vowel in the past tense (Dammel et al. 2010:340). Indeed, 84% of strong verbs have this ABB pattern (Dammel et al. 2010:342).

Neither German nor English levelled in a uniform manner. The selection of the vowel depended on the ablaut class, although in English the preterite was occasionally remodeled based on the past participle (Dammel et al. 2010:339). German shows ABA (class V), ABB (classes I and II), and ABC (classes III and IV) (Dammel et al. 2010:340). The distribution in German is less stark than in Dutch, with 51% following ABB, 20% ABA, and 29% ABC (Dammel et al. 2010:342). English has a more complicated process of levelling, where several classes went in

multiple routes (e.g. class I shows both ABB and ABC in modern English) (Dammel et al. 2010:341). The ABB pattern is nonetheless the most common, with 60% of verbs, while 18% of verbs follow the ABA pattern and 22% ABC (Dammel et al. 2010:342).

The above study also investigated the direction of strong verb weakening in these languages. They examined whether oscillations occurred in the preterite or in the participle, such as German *backen : buk/backte : gebacken* 'bake' or English *weave : weaved/?wove : woven/weaved*. Verbs with mixed paradigms, i.e. a strong preterite and weak participle or vice versa, suggest that verbs do not change class in one step, but rather that they go step by step (Dammel et al. 2010:343). Because German, Dutch, and English show mixed paradigms of the type weak preterite – strong participle, this suggests that the past participle is the more stable form (Dammel et al. 2010:345), which they posit is because it is more cognitively entrenched, because the past participle is the more frequently used form (352). As Bybee's and Nübling's research indicates (see 2.5), more frequently used forms are more likely to persist, even if they are irregular. Dammel et al. discuss the grammaticalization of the perfect in these four languages, which has not happened to a uniform degree across them. The perfect is the least grammaticalized in English and the most grammaticalized in German. English uses the perfect primarily for events in the "extended now," e.g. *I have lived in Athens for three years*, and Swedish does as well, with the addition of events with "current relevance," Dutch adds a past tense reading to the usages of perfect, and German additionally uses the perfect as a general past tense (primarily in spoken language, where written language uses the preterite) (Dammel et al. 2010:348). The authors argue that the increased frequency of use of the perfect in German allows the past participle to be more easily recalled by speakers, whereas the preterite form, being less frequently used, shows variation.

Reorganization of ablaut patterns in the levelling process can potentially even lead to the formation of a new ablaut pattern. Nowak (2010) investigates the extent of analogical spread of the pattern *x-o-o* (an ABB pattern) in Early New High German (late 14th-18th c) and Middle Dutch (14th-17th c) by tabulating verbs in German and Dutch by their vowel alternations across the paradigm. She additionally looks at the effects of frequency on the changes of vowel alternations over time. She suggests that an eighth ablaut pattern, *x-o-o*, has formed. The class II paradigm, seen in German *bieten : bot : geboten* 'offer' and Dutch *bieden : bood : geboden*[9] 'offer', was extended by analogy to verbs from other classes, such as German class V *weben : wob : gewoben* 'weave' (where *weben : wab : geweben* is expected) and Dutch class V *trekken : trok : getrokken* 'tear' (where *trekken : trak : getrekken* is expected) (Nowak 2010:361). This occurred more frequently in Dutch than in German (op. cit.) and is a large factor in the prevalence of the ABB pattern as discussed above. The class III verbs form the majority of verbs affected by this process, and moreso in Dutch (35/46 *x-o-o* verbs) than in German (12/24) (Nowak 2010:363). This is a result of the different direction of ablaut levelling in the preterite between the two languages, as discussed above. Dutch levelled the paradigm to the vowel of the preterite plural, and in class III, the preterite plural vowel was /o/. This led to the extreme prevalence of the *x-o-o* pattern in Dutch, because "all verbs of class III joined the eighth ablaut class" (Nowak 2010:365). In Early New High German, the class IIIa verbs, which have a CVNC stem, underwent a lowering of the /u/ in the past participle to /o/ before double nasals only (Nowak 2010:365). The class IIIb verbs, which have a CVLC stem, uniformly lowered the past participle to /o/ in Middle High German (op. cit.). Variation was introduced into a previously unambiguous paradigm as the past participles of double nasal verbs lowered from /u/ to /o/,

[9] In both *bood* and *geboden*, the vowel is /ō/.

which left the preterite plural without paradigmatic support, and it levelled to the /a/ of the preterite singular (Nowak 2010:365). The verbs with a nasal + consonant in the stem, however, levelled to the /u/ of the preterite plural, because that was supported by the past participle (Nowak 2010:365, and compare Dammel et al. 2010). Compare *sinnen : sann : (sunnen) : gesonnen* 'to think' and *singen : sang : (sungen) : gesungen* (both IIIa) to *werden : (ward) : wurden : geworden* 'to become,' IIIb.

While in Dutch, all class III verbs moved to the *x-o-o* pattern, it was only the infrequent verbs from class III in German that did so (Nowak 2010:365). The infrequent verbs, such as *weben*, *fechten* 'fight, fence', and *bellen* 'bark', show both strong and weak preterite variants in ENHG (Nowak 2010:366). Because many of these verbs, such as *bellen*, are now weak and because several verbs that remain strong today, such as *weben*, show weak variants, it is suggestive that the *x-o-o* pattern is a stopping point on the way to a weak conjugation (Nowak 2010:367).

Productivity of strong verb inflection patterns can also be investigated through quantitative analysis. Knooihuizen and Strik (2014) studied the productivity of strong verb patterns using nonce words built around the strong verb patterns in Dutch. They asked native speakers of Dutch to produce inflections of these nonce verbs, to rate inflections of nonce verbs, or to produce strong preterite forms and past participles of weak verbs (Knooihuizen & Strik 2014:183) in order to "quantify the relative productivity of the default weak inflection compared to strong" (Knooihuizen & Strik 2014:182). They expected the weak inflection to be far more productive than the strong, and that the strong verb classes with a high pattern frequency (e.g. I, IIb, IIIa) would be more productive than the other classes. They found, contrary to their expectation, that a third of the forms produced for the nonce verbs were strong (Knooihuizen &

Strik 2014:184). They also found that, as expected, the nonce verbs patterned after the more type-frequent strong verb classes showed higher rates of strong inflection, while the nonce verbs patterned after the root shape of the least type frequent classes, particularly class VII, showed very little strong inflection (Knooihuizen & Strik 2014:184-185). Consonantism of the stem played a role in the productivity of strong inflections, in that plosives and fricatives were more likely to produce strong inflections than nasals and liquids, which reflects the consonantism of existing verbs in Dutch (Knooihuizen & Strik 2014:186). Speakers were also more likely to judge the strong inflections of nonce verbs in the more frequent classes as acceptable than those of the less frequent, with class I being most acceptable and VI the least (3.87 on a 7-point Likert scale vs. 2.21) (Knooihuizen & Strik 2014:188-189). An interesting point in light of Nowak's (2010) study is that they found that /o/ and /ɔ/ were popular in both the elicitation of inflections of nonce verbs and in the forced elicitation of strong inflections of weak verbs (Knooihuizen & Strik 2014:191). Speakers of modern Dutch show preferences for particular inflectional patterns, and "we must assume" that such preferences existed among language users in the past. This can help to explain why inflectional change is not always weakening, but is sometimes seen as movement into another strong class (Knooihuizen & Strik 2014:195). In modern Dutch, the vowel /o/, which originated in classes II, III, and IV, has spread into the *a*-group class VI, and this suggests that this vowel has gained "iconicity" as a past tense marker (Knooihuizen & Strik 2014:194).

A follow-up study investigated the effects of Dutch influence on strong inflection productivity in Frisian (Knooihuizen et al. 2018), where the seven strong verb classes split into 16 different inflectional patterns, each of which contains a small number of verbs. This leads to a low type frequency for the majority of strong verb inflectional patterns and no basis for

analogical formations (Knooihuizen et al. 2018:64). The /o/ vowel does not occur as often in Frisian inflectional patterns as in Dutch, and the authors did not expect it to occur as widely as in Dutch, even though these speakers are bilingual in Dutch and Frisian and could have made analogous formations to the typologically similar Dutch past tense. Similar to the Dutch results, Frisian speakers produced strong inflections of nonce verbs about a third of the time, and the more frequent verb patterns were more productive (Knooihuizen et al. 2018:68). Unlike the Dutch data, and as the authors expected, there were not many past tense forms with an /o/, because this vowel does not occur as frequently in verbal inflection as it does in Dutch (Knooihuizen et al. 2018:68).

These studies highlight the importance of analogy in verbal inflection patterns for speakers of Germanic languages. Analogical processes are so strong that they are noticed and commented on by the general public. The popular German language columnist Bastian Sick in his column "Zwiebelfisch" 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 gedrucken [sic][10]" (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 geklungen. Ich bringe, ich brang, ich habe … Nun ja, mit den Analogien ist das so eine Sache[11]" (Sick 2008). Knooihuizen and Strik cite a tweet from a language society, asking whether Queen Beatrix "woof" or "wuifde" ('waved') at the crowds, where the former is the strong preterite and the latter the weak (2014:173).

[10] he "had milked (weak participle for strong verb) the cow. Luckily the newspaper didn't print (strong participle for weak verb) it."
[11] "I sing, I sang, I have sung. I ring, I rang, I have rung. I bring, I brang, I have … Well, analogies are like that."

## 3.2 Audience design and style

Style plays a major role in synchronic variation. Speakers have a variety of styles to choose from in their linguistic repertoires, which are dependent upon the domain of use. That is, in a casual conversation between friends, a speaker may choose one register, and in a formal setting, such as a job interview, they may choose a different one. The more formal styles are typically more conservative and show change slowly, if at all (cf. 2.5 and Bybee's [2007] "prefabs"), while ongoing change can be apparent in informal styles (Weerman et al. 2013).

Allan Bell describes the concept of "audience design" to explain shifts in style by a single individual in different contexts: "Style is what an individual speaker does with language in relation to other people" (Bell 2011:141) and it "derives its meaning from the association of linguistic features with particular social groups" (142). According to this model, both speakers *and hearers* are crucial to the use of language (Bell 2011:144), because the speaker will alter their produced speech depending on who the listener is, in order to signal membership in a particular group.

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). Informal written language is not necessarily equivalent to spoken language, 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" (Weerman et al. 2013:356). Weerman et al. investigated synchronic variation in case marking in 17$^{th}$-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, with the knowledge that the genitive and dative cases were in the process of being eliminated from the language, as they are absent in modern Dutch. 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).

Informal texts are more likely to be reflective of spoken language than formal texts are, because formal language is often more archaic. Additionally, formal texts are often edited, and, as such, they pass through the hands of multiple editors, where variation in style is eliminated in favor of an existing, prescriptive norm. Not all edited texts have variation eliminated from them, however, and the division between formal and informal is not inherently based on whether the text has been edited. Works of fiction may depict casual styles in dialogue or in the narrator's voice, and newspapers may include transcriptions of interviews, which may use casual styles. Works of fiction have a target audience for which the language is crafted, such as a novel aimed at teenagers using youth slang. Some genres of fiction, notably literary fiction, have a target audience that expects a more archaic style, and, as such, these texts hew to that norm. In sum, informal registers more accurately reflect ongoing change than formal ones, which may not show ongoing change at all.

## 3.3 This study

This corpus-based diachronic study investigates the factors that influenced the disposition of the class VII verbs from Old High German into Modern German. In Modern German, the formerly reduplicating verbs are primarily strong and follow an ABA ablaut pattern (see table 2.6.1). Based on data from previous studies (Lieberman et al. 2007, Carroll et al. 2012, Dammel et al. 2010, Nowak 2010, Knooihuizen & Strik 2014, Knooihuizen et al. 2018), token frequencies are investigated for an effect on the rate of weakening. In this study, the variable of type frequency is eliminated by looking only at class VII verbs.

Hypothesis 1: Less frequent verbs will become weak at a higher rate than more frequent verbs. If a verb becomes less frequent over time, it will be more likely to become weak.

Hypothesis 2: The register of a text will have an effect on the rate of usage of prescribed strong forms vs. weak forms. The more formal the text, the more likely it is that the text will evince conservative traits, even when change is underway.

### 3.3.1 The verbs under consideration

This study is a corpus study from the year 800 CE to 2018 CE in order to test whether low token frequency over time is a predictive factor in the weakening of historically strong Gmc class VII verbs. A list of class VII verbs was collected from grammars of Old English and Gothic, and the PGmc forms were found using Seebold 1970. The attested OHG and Old Norse versions of these verbs were also tabulated. Verbs that were attested in both OHG and OE are given in table 3.3.1.1. Because *hang* has both weak and strong homonyms (i.e. transitive and

intransitive, respectively), it was excluded from analysis. The verbs *sow* and *salt* had too few attestations in OHG and were excluded for that reason.

German has a wide array of verbs with prefixes that alter the meaning, e.g. *fangen* 'catch' vs. *empfangen* 'receive.' Because these prefixed verbs contribute to the overall frequency of a verbal root, a non-exhaustive list of prefixed verbs built on these roots was generated for modern German and tabulated as well (Appendix A).

**Table 3.3.1.1** The attested class VII verbs in Old High German and Old English, given as infinitive and preterite singular, with their modern equivalents. Verbs that have become weak, or show variation, are in *italics*.

| ***PGmc*** | ***OHG*** | ***NHG*** | ***OE*** | ***NE*** |
|---|---|---|---|---|
| *fallan | fallan : fell | fallen : fiel | feallan : fēoll | fall : fell |
| **falþan* | *faldan : fial* | *falten : faltete* | *fealdan : fēold* | *fold : folded* |
| *fāhan | fāhan : fiang | fangen : fing | fōn : feng | none (catch, seize) |
| *haitan | heizan : hiaz | heißen : hieß | hātan : hēt | none (to be called) |
| *haldan | haltan : hialt | halten : hielt | healdan : hēold | hold : held |
| *hāhan | hāhan : hiang | hängen : hing | hōn : heng | hang : hung |
| **hawwan* | *houuan : hio* | *hauen : hieb/haute* | *hēawan : hēow* | *hew : hewed* |
| *hropan | ruofan : riof | rufen : rief | hrōpan : hrēop | none (to call, cry out) |
| *hlaupan | loufan : liof | laufen : lief | hlēapan : hlēop | leap : leapt |
| *lætan | lāzan : liaz | lassen : ließ | lǣtan : lēt | let : let |
| *rædan | rātan : riat | raten : riet | rǣdan : rēd | read : read |
| **sœan* | *sāian : sāte* | *säen : säte* | *sāwan : sēow* | *sow : sowed* |
| **saltan* | *salzan : sialz* | *salzen : salzte* | *sealtan : sēolt* | *salt : salted* |
| *skaidan | skeidan : skied | scheiden : schied | scēadan : scēd | shed : shed |
| *slepan | slāfan : sliof | schlafen : schlief | slǣpan : slep | sleep : slept |
| **waldan* | *waltan : wiald* | *walten : waltete* | *wealdan : wēold* | *wield : wielded* |

### 3.3.2 German frequency analysis

German frequency data was gathered from three sources. Old High German data was gathered using the public installation of ANNIS software at Humboldt Universität and the Old

High German corpora from the Deutsch Diachron Digital project. These combined corpora contain 614,462 tokens. A list of the texts is included as Appendix B. The total number of tokens for each lemma was noted. ANNIS allows the user to search by inflectional form, e.g. 3.SG.PAST, and each lemma was searched for all persons and numbers in the past tense.

Middle High German data was gathered from the Mittelhochdeutsche Begriffsdatenbank (MHDBDB). The MHDBDB is not a parsed corpus, nor is it intended for the use of linguists. The database contains 10,656,235 tokens from 666 documents. Users are able to search by lemma and count tokens, and the results are listed as they appear in the source documents, including all spelling variation. A weak preterite was considered as one that had a /t/ following the root (as in *falten : faltete* 'fold(ed)'). It is possible that the strong preterite counts contain some subjunctives, because the German subjunctive is based on the preterite form, and that some were missed because of spelling variation. A test study of the ambiguous form *liessen* (1/3.PL.PRET.IND or 1/3.PL.PRES.SUBJ) showed that 124/124 tokens were in the preterite indicative form, so this may not be a major concern.

Early Modern and Modern German data was gathered at the *Digitales Wörterbuch der deutschen Sprache* (Digital Dictionary of the German Language, DWDS), using the *Referenz- und Zeitungskorpora* (aggregated reference and newspaper corpora), which includes the German Text Archive (1473-1927), the DWDS-Kernkorpus (core corpus) (1900-1999), the DWDS-Kernkorpus 21 (2000-2010), *Der Tagesspiegel* (The Daily Mirror, a newspaper, 1996-2005), and *Die ZEIT* (Time, a newspaper, 1946-2018). The corpus contains 1,308,074,980 tokens from four genres of text: *Wissenschaft* (scholarly), *Belletristik* (literature), *Gebrauchsliteratur* (nonfiction with a particular purpose), and *Zeitung* (newspaper). The search function at DWDS allows users to search by lemma and exact form, and total tokens were counted via the lemma search, and the

preterite forms were obtained using the exact form search. All inflectional forms of the preterite were searched, e.g. *lief - liefst - liefen - lieft* 'ran 1/3.sg, 2.sg, 1/3.pl, 2.pl.' Spelling variation was found through the lemma search. The earliest tokens showed spelling variation, and exact form searches were created based on this, e.g. *lieff – lieffst, etc.* and the weak variant *laufft*e *– laufftest, etc* alongside the expected *laufte – lauftest, etc.* Because the resources at DWDS limit the download to the first 5,000 hits, when the number of hits was below 5,000, the results were downloaded as a spreadsheet. The weak form of *hauen* 'hew' had many false hits because of French-language text in 19th-century novels, as well as French borrowings, e.g. *haute cuisine*. These were cut from the data under consideration. The preterite of the weak verb *abschlaffen* 'sag, wilt' was picked up through the exact-spelling search for *schlaffte*, and these tokens were excluded.

Prefixed verbs contribute to the overall frequency of the root verb, and token counts for the lemmas of prefixed verbs were taken for modern German from the aggregated corpus at DWDS, and net frequency of the root verb was calculated.

Frequency of *hauen* at specific time points was taken from the *Wortverlaufskurve* (word-trend curve) of the aggregated corpus at DWDS.

### 3.3.3 Variationist analysis

Because *hauen* is the only verb under consideration that shows variation in contemporary German, a variationist case study was able to be performed. This synchronic variation makes it an ideal case study for the roles which genre or formality play. A search for *haute* and *hieb* was performed in the individual corpora at DWDS for the newspapers *Berliner Zeitung* (1994-2005),

which is a member of the BILD-family of tabloids; *Die Zeit* (1946-2018), a respected center-left newspaper based in Hamburg which is often considered highbrow (Warwick); and *Der Tagesspiegel* (1996-2005), a Berlin-based news daily with "classic liberal" leanings (Warwick). Two additional sources that are available through DWDS are blogs and film and television subtitles. The blog corpus is part of a reference corpus for internet-based communication and is regularly updated. It has 102,500,000 tokens. The subtitles corpus was compiled in 2013-2014 and is based on data available from opensubtitles.org and contains 75,600,000 tokens (DWDS Special Corpora). A search was performed for the 1/3.SG form of both the strong and weak variants (i.e. *haute* and *hieb*) in each individual corpus, and the results were exported for download. Tokens were coded by preterite type (strong/weak), source, and time period (decade), then a multiple linear regression analysis was performed in Rbrul.

These newspapers represent a variety of sociocultural statuses to which usage of preterite forms can be indexed. A respected, highbrow paper like *Die Zeit* would be expected to use the standard (strong) preterite form most frequently, while the tabloid newspaper (*BZ*) would be expected to use the nonstandard (weak) form. *Der Tagesspiegel* is assumed to fall between the two others. Because subtitles are intended to replicate spoken language, it is assumed that these will favor the informal *haute* over more formal *hieb* because many movies and TV shows feature casual speech. Blogs were chosen to represent casual speech as well, as they are often written by people for a familiar audience (compare Bell 2011), and people can choose to represent themselves in the way they usually speak. However, bloggers can also choose to represent themselves more formally, for an audience of strangers, so this could affect the relative rates of strong and weak preterite forms of *hauen*.

In sum: *Die Zeit* and, to a lesser extent, *Der Tagesspiegel* are expected to favor the standard strong preterite *hieb*, while *Berliner Zeitung,* subtitles, and blogs are expected to favor the colloquial weak preterite *haute*.

# CHAPTER 4

# RESULTS

## 4.1 Weakening over time

This section presents the counts of strong and weak preterite tokens in the three periods of German. In OHG, all thirteen verbs are strong and show no variation in the preterite form. See appendix C. In MHG, *falten* appears to be in the process of becoming weak, as the weak preterites outnumber the strong (58 vs 42), and *rufen* shows some variation in the preterite (150 weak vs 1414 strong); thus, twelve of thirteen verbs can be considered strong. In NHG, both *falten* and *walten* have become weak, and *hauen* shows considerable variation in its preterite form (31% weak preterites). *Rufen* continues to show some oscillations in the preterite form (mainly prior to 1800), but the strong form predominates by far. In NHG, therefore, ten of thirteen class VII verbs continue to be strong, while two have become weak and one is in the process of weakening, which is a 23% regularization rate from OHG. The verbs which have become weak are the least frequently used verbs in their category, with token counts below 2 per 100,000, as will be discussed in 4.2.

## 4.2 Effects of frequency on regularization

As expected, the least frequent verbs showed the most variation and were more likely to become weak. This change is observable over time. In OHG, the most frequent verb was *heißen*,

at 93.09 tokens per 100,000, followed by *lassen*, at 33.04. The least frequent verb was *falten*, at 0.33 tokens per 100,000. In Middle High German, variation was seen in the preterites of *falten* (58% weak, 1.20 tokens/100,000), *hauen* (3.5% weak, 7.80 tokens/100,000), and *rufen* (9.6% weak, 22.51 tokens/100,000). In NHG, variation is seen for all verbs, but only *hauen* shows more than 5% variation. See table 4.2.1. The verb *heißen* also meant 'to order, to command' in the Old and Middle High German periods, which can account for the greater frequency of its use in MHG than NHG. A table of all token counts and frequencies is included as Appendix C.

**Table 4.2.1** Frequencies of class VII verbs in German, ordered by frequency in NHG. Frequency is per 100,000 tokens. Variation is calculated as the number of tokens of the weak preterite divided by the total number of preterite tokens. No variation was seen in OHG. Verbs given in *italics* are weak in modern German. Entries in **brown** denote that the verb is weak at this point. The double orange line separates the strong from weak verbs in MGm. The box highlights verbs which increase noticeably in frequency when their prefixed versions are considered.

| *Verb* | *OHG frequency* | *MHG frequency* | *variation (%) MHG* | *NHG frequency* | *variation (%) NHG* |
|---|---|---|---|---|---|
| *falten* | 0.33 | **1.2** | **58** | **0.34** | **n/a** |
| *walten* | 9.11 | 10.82 | 0 | **0.58** | **n/a** |
| hauen | 1.14 | 7.8 | 3.55 | **1.15** | **30.88** |
| raten | 7 | 13.63 | 0 | 2.56 | 0 |
| scheiden | 11.39 | 31.21 | 0.69 | 2.56 | 0.4 |
| fangen | 9.93 | 44.42 | 0 | 3.94 | 0.01 |
| schlafen | 8.3 | 12.4 | 0 | 3.94 | 0.03 |
| rufen | 9.93 | 22.51 | 9.59 | 5.89 | 0.64 |
| fallen | 11.23 | 16.8 | 0.15 | 17.51 | 0.01 |
| laufen | 6.02 | 24.55 | 0.07 | 21.91 | 0.02 |
| heißen | 93.09 | 121.46 | 0 | 44.32 | 0.01 |
| halten | 11.39 | 19.41 | 0 | 59.72 | 0 |
| lassen | 33.04 | 184.67 | 0 | 114.15 | 0 |

For *scheiden* and *rufen*, the majority of variation in the preterite was seen prior to 1800. With *rufen*, the majority of the 266 weak preterite tokens from 1600-99 were from the same text, and the case is similar with the 246 tokens from 1700-99 – which suggests that this variation may not have been widespread, but rather may have been regional.

When prefixes were taken into account, the frequencies of *raten, scheiden, fangen,* and *schlafen*, inside the box in table 4.2.1, all increased to over 5 per 100,000. *Hauen* only increased from 1.15 to 1.9 per 100,000. See table 4.2.2. This suggests that there is a threshold value for frequency at which a strong verb is likely to become weak, and it is above 2 and below 5 tokens per 100,000. This is comparable to the relationship between frequency and weakening seen by Carroll et al. (2012), as discussed below.

**Table 4.2.2** Total frequency (per 100,000 tokens) of class VII verbs in modern German with percent variation (i.e. percent of weak tokens for expected strong). Compare the frequencies of the verbs in the box to those in table 4.2.1.

| *Root verb* | *Net frequency* | *Variation (%)* |
|---|---|---|
| *hauen* | 1.49 | 30.88 |
| *walten* | 1.9 | (weak) |
| *schlafen* | 5.15 | 0.03 |
| *raten* | 11.39 | 0 |
| *fangen* | 17.73 | 0.01 |
| *rufen* | 22.61 | 0.64 |
| *scheiden* | 23.68 | 0.69 |
| *laufen* | 27.21 | 0.02 |
| *fallen* | 30.31 | 0.01 |
| *heißen* | 44.32 | 0.01 |
| *halten* | 84.74 | 0 |
| *lassen* | 131.5 | 0 |

As was expected, the most frequent of these verbs have remained strong in modern German. This result is comparable to Carroll et al.'s (2012) results. The verbs which were found to remain strong in this study are the most frequent in their class, and they were in the three highest frequency bins (normalized to greater than 100 tokens per 100,000) in Carroll et al.'s (2012) study: *halten, fallen, fangen, lassen, laufen, raten, rufen,* plus *heißen*, which was not in their study, and *schlafen*, which was in the fourth highest frequency bin (greater than 10 tokens per 100,000). Those which became weak were in the fifth and sixth frequency bins (less than 10

tokens per 100,000): *walten, falten*. They did not study *hauen* because it showed variation (Joseph Salmons, p.c.), but its frequency in OHG was comparable to *falten*, higher in MHG, and is comparable to both *walten* and *falten* in NHG. Because the body of text available from the MHG period includes epic poetry, wherein swords are drawn and bodies are hewn, and heads hewn from said bodies, the greater frequency during this period is not unexpected. Many of the most recent tokens of *hauen* regard soccer, for example, "er haute den Ball ins Tor" ('he kicked the ball into the net'). The first attestation of this in the DWDS aggregated corpus is from 2002: "Müller haute den Ball ins Tor" (Der Tagesspiegel, 2002-08-10). There is one attestation, from 1966 in *Die Zeit*, of this usage with the strong preterite: "Dort kam der Mittelstürmer herangebraust […] und hieb den Ball aus dieser Entfernung in das Tor.[12]" *Hauen* will be discussed in further detail in 4.3.

This data supports the first hypothesis of this study, that the least frequent verbs are more likely to become weak over time. Because this study focused narrowly on one class and did not group all 150+ strong verbs into frequency bins, the finer grained clarity of token frequency, independent of type frequency thus provided allowed a potential frequency threshold for weakening to be seen. Additionally, the verbs that begin with a moderately high frequency and later become less frequent (see *walten* and *hauen*), they are more likely to become weak. See figure 4.2.1.

[12] "Then the center forward came rushing up and heaved (hewed) the ball from that distance into the net."

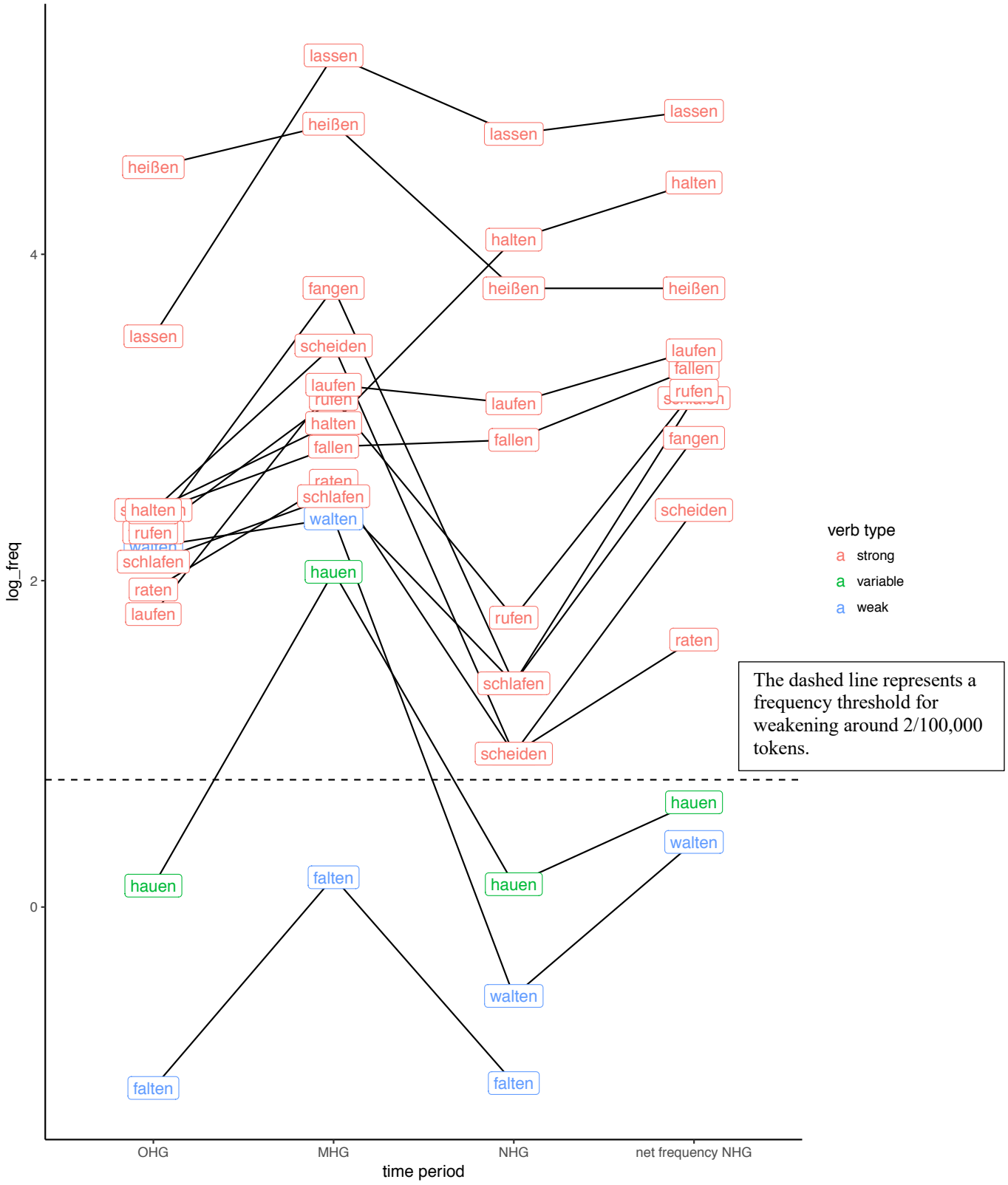


**Figure 4.2.1** Frequency of the 13 verbs under study over time, including net frequency in NHG. Colors indicate inflectional category in modern German.

## 4.3 Case study: *hauen*

*Hauen*, as the only one of these verbs which shows variation in modern German, was further investigated in NHG. Token counts of strong and weak preterite forms, as well as of the lemma, were found for the time periods 1473-1599, 1600-1699, 1700-1799, 1800-1899, 1900-1949, 1950-1999, and 2000-2018. In the earliest time periods, the strong form predominated. In the period 1900-1949, the weak form increased, and in 1950-1999, it overtook the strong form. See table 4.3.1 and figure 4.3.1. The overall frequency of *hauen* decreased significantly beginning in 1670-1679, from a peak of 4.85 tokens per 100,000, to 1.97 per 100,000 in 1740-1749. Even though the token count of *hauen* increases in the 20th century, its frequency is steady at around 1 per 100,000. See figure 4.3.2.

**Table 4.3.1** Token count of preterite forms of *hauen* over time. Data taken from the aggregated reference corpus at DWDS.

| *hauen* | *1473-1599* | *1600-99* | *1700-99* | *1800-99* | *1900-49* | *1950-99* | *2000-18* |
|---|---|---|---|---|---|---|---|
| *strong* | 7 | 238 | 164 | 382 | 174 | 191 | 125 |
| *weak* | 0 | 5 | 15 | 11 | 35 | 213 | 289 |
| *lemma* | 111 | 1684 | 1294 | 1323 | 951 | 4590 | 5092 |

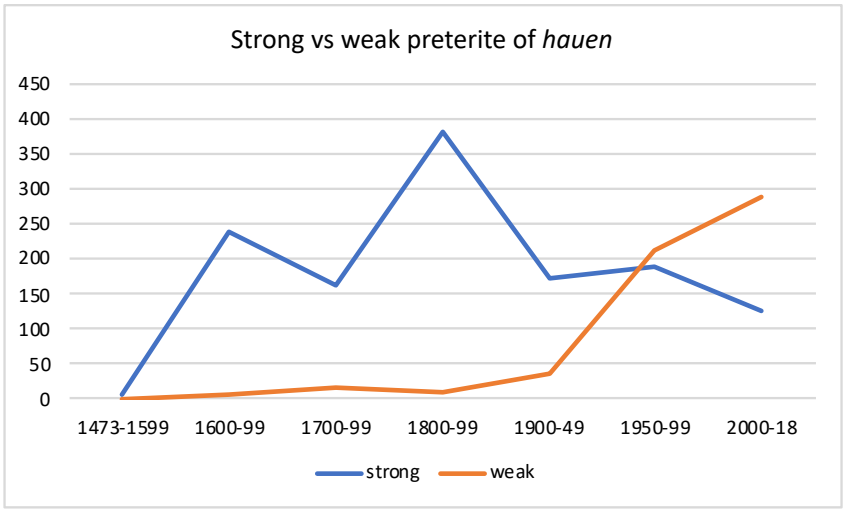


**Figure 4.3.1** Graphical presentation of table 4.3.1.

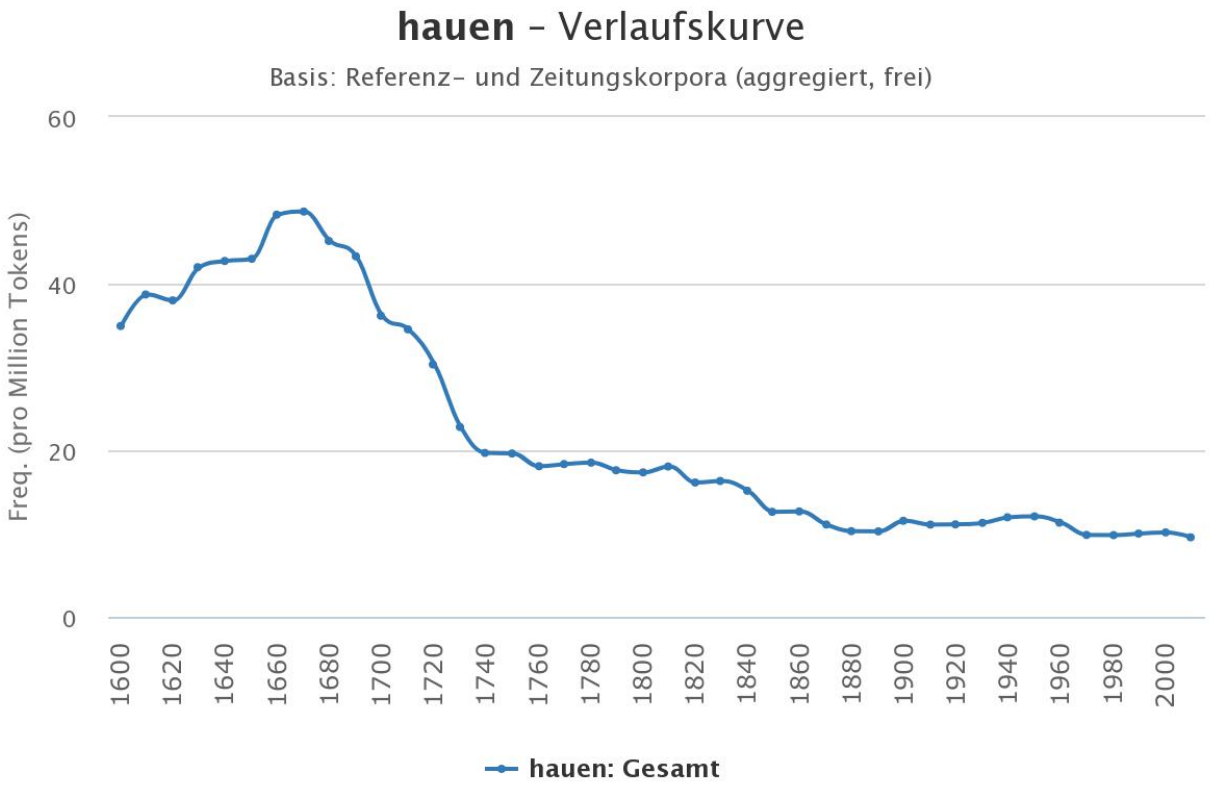


**Figure 4.3.2** Frequency of *hauen* since 1600, by decade. (DWDS Word Information)

A linear regression analysis was performed in Rbrul in order to evaluate potential factors involved in the variation seen in contemporary German. Model data is presented in tables 4.3.2 and 4.3.3. The data generated from DWDS includes the genre of text (see 3.3.2), bibliographic information (author and title), and year. Tokens were coded by preterite type (strong or weak) and time period (1-7, according to the date ranges in table 4.3.1, i.e. century until 1900, then 50-year period; treated as both categorical and continuous). Preterite type was the response variable.

The first model treated time period as a categorical variable. Genre of text had a moderate effect, with Wissenschaft favoring the strong preterite (factor weight 0.669) and Gebrauchsliteratur the weak (factor weight 0.362), although the p value for this effect was 0.065, which does not reach statistical significance. The time periods beginning in 1950 nearly categorically favor the weak preterite (factor weights 0.0262 for 1950-99 and 0.0129 for 2000-18). The p value of the effect of time period is $3x10^{-51}$, which is highly statistically significant. A second model was created using time period as a continuous variable. The effects of genre are different in this model, nearly the reverse of the previous one, with Wissenschaft favoring the weak preterite (factor weight 0.155) and Zeitung the strong (factor weight 0.766). With every increase in time period, the log odds of using the strong preterite decreased by 0.724. The interaction between genre and time period was calculated, and as the time period increased, the log odds of favoring the strong preterite decreased by 0.228 for Belletristik and 0.3474 for Zeitung. The p value of the interaction is 0.0601, which approaches statistical significance.

**Table 4.3.2** Rbrul model 1. A factor weight closer to 1 favors the strong preterite, and closer to 0 favors the weak. Numbers in [brackets] indicate non-statistically significant results.

| hauen: strong vs weak | input probability 0.961 | $R^2$: 0.566 | deviance: 1501.637 | AIC: 1521.637 |
|---|---|---|---|---|
| | | | | |
| **Genre** | **factor weight** | | **Time period** | **factor weight** |
| Wissenschaft | [0.669] | | 1 | >0.999 |
| Belletristik | [0.511] | | 2 | 0.537 |
| Zeitung | [0.455] | | 3 | 0.453 |
| Gebrauchsliteratur | [0.362] | | 4 | 0.184 |
| | | | 5 | 0.121 |
| | | | 6 | 0.0262 |
| | | | 7 | 0.0129 |

**Table 4.3.3** Rbrul model 2. A factor weight closer to 1 favors the strong preterite, and closer to 0 favors the weak. Numbers in [brackets] indicate non-statistically significant results.

| hauen: strong vs weak | input probability 0.995 | $R^2$: 0.504 | deviance: 1535.816 | AIC: 1551.816 |
|---|---|---|---|---|
| | | | | |
| **Genre** | **factor weight** | | **Genre/time interaction** | **log odds** |
| Zeitung | [0.766] | | Wissenschaft +1 | 0.587 |
| Belletristik | [0.749] | | Gebrauchsliteratur +1 | -0.0116 |
| Gebrauchsliteratur | [0.333] | | Belletristik +1 | -0.228 |
| Wissenschaft | [0.17] | | Zeitung +1 | -0.3474 |
| | | | | |
| **time period** | **log odds** | | | |
| +1 | -0.724 | | | |

Both of these linear regression models show that time is a significant factor in the variation of the preterite of *hauen* seen in the NHG period, and that as time increases, use of the weak preterite increases. It has already been shown that token frequency of *hauen* decreases over time. Decreasing token frequency correlates positively with increased realization of *hauen* as a weak preterite. Taken together with the statistical significance of the linear regression model, the data suggest a causal relationship between the decreased frequency of use and the shift from

strong to weak verb. Both of these models suggest that genre of text is a factor, although neither reaches significance, but they do not agree on which genre of text favors which preterite.

In order to refine the results of the above, time period was broken down by decade beginning in 1900. The strong preterite predominates until 1970-79, when the weak preterite begins to overtake it (28 vs. 34, respectively). From 1980-89 the strong preterite is, again, slightly lower than the weak (31 vs 36, respectively). During the 1990s, the weak preterite becomes predominant, with twice as many tokens as strong (121 vs 63). The 2000s continue this trend, with 211 weak vs 103 strong. The current decade does not have as many tokens of either, although the data only covers nine years rather than ten. Regardless, the weak preterite outnumbers the strong 78 to 21. See table 4.3.4 and figure 4.3.3.

**Table 4.3.4** Token counts of preterite forms of *hauen* since 1900. (DWDS Corpora)

| | *1900-09* | *10-19* | *20-29* | *30-39* | *40-49* | *50-59* | *60-69* | *70-79* | *80-89* | *90-99* | *2000-09* | *2010-18* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Weak* | 1 | 2 | 10 | 19 | 3 | 7 | 16 | 34 | 36 | 121 | 211 | 78 |
| *Strong* | 21 | 73 | 30 | 26 | 23 | 29 | 39 | 28 | 31 | 63 | 103 | 21 |

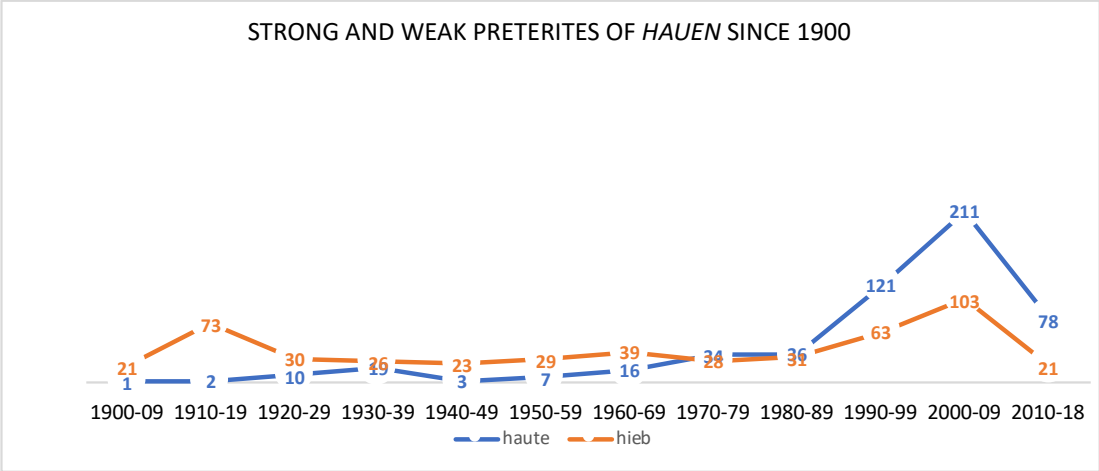


**Figure 4.3.3** Graphical presentation of table 4.3.4.

Data was grouped into seven 20-year periods and was considered as a categorical value, with genre included as a second categorical variable. The periods beginning in 1960 strongly favor the weak preterite (factor weights from 0.388 to 0.171; see table 4.3.5), which was statistically significant ($p=5.39x10^{-47}$). Genre was not a significant factor (p=0.334). It is possible that the genres under consideration are too broad to be of use in this situation; therefore, a deeper look into style was warranted.

**Table 4.3.5** Rbrul model of *hauen* by decade since 1900. A factor weight closer to 1 favors the strong preterite. A number given in [brackets] indicates a lack of statistical significance.

| hauen: strong vs weak | input probability 0.475 | $R^2$: 0.364 | deviance: 1164.981 | AIC: 1194.981 |
|---|---|---|---|---|
| **Years** | **Factor weight** | | **Genre** | **Factor weight** |
| 1900-19 | 0.928 | | Wissenschaft | [0.733] |
| 1940-59 | 0.703 | | Belletristik | [0.479] |
| 1920-39 | 0.465 | | Zeitung | [0.425] |
| 1960-79 | 0.388 | | Gebrauchsliteratur | [0.349] |
| 1980-99 | 0.224 | | | p=0.334 |
| 2000-18 | 0.171 | | | |
| | $p=5.39x10^{-29}$ | | | |

4.4 The effect of style on the preterite of *hauen*

All textual sources had a higher number of weak preterite tokens than strong. *BZ* had 78 weak to 64 strong, *Die Zeit* 205 vs 139, *Der Tagesspiegel* 68 vs 39, subtitles 97 vs 3, and blogs 31 vs 19 (see table 4.4.1). A Fisher's Exact test was performed in R, with a p-value of $1.69x10^{-15}$. This supports the hypothesis that difference in the use of the strong or weak preterite based on style is not due to chance.

**Table 4.4.1** Token counts of strong and weak preterite forms of *hauen* in five sources.

| *hauen* | blogs | subtitles | *Berliner Zeitung* | *Der Tagesspiegel* | *Die ZEIT* |
|---|---|---|---|---|---|
| *weak* | 31 | 97 | 78 | 68 | 205 |
| *strong* | 19 | 3 | 64 | 39 | 139 |

A model was built in Rbrul using source as a predictive factor for variation, and it was statistically significant ($p=4.02x10^{-15}$). All sources except subtitles favored the strong preterite slightly (factor weights 0.695 to 0.614), and, unexpectedly, *BZ* favored the strong preterite most strongly. Time period (decade) was added as a continuous variable, and its log odds of favoring the strong preterite decreased by 0.253 for every decade. This addition changed the factor weights of the sources, with blogs nearly categorically favoring the strong preterite, *Die Zeit* favoring the strong preterite (factor weight 0.736), *Tagesspiegel* favoring neither (factor weight 0.554), and *BZ* and subtitles categorically favoring the weak (factor weights 0.00646 and 0.00105). See table 4.4.2.

The data support the second hypothesis, that register affects usage of the weak preterite form. As expected, the prestigious newspaper *Die Zeit* favors the standard preterite form, *Tagesspiegel* marginally favors the strong, and both *BZ* and the subtitles favor the weak. Unexpectedly, blogs favor – categorically – the strong form. The principle of audience design suggests that a blogger would tailor their message to the expected audience (see Bell 2001 and 3.2). Some topics may have a more informal expected audience, e.g. popular culture, while others may have a more formal expected audience, e.g. politics or religion. Without further information about the specific blogs in question, conclusions regarding blogs as a unified genre cannot be drawn.

**Table 4.4.2** Rbrul model of preterite forms of *hauen* using sources as a marker for style, after the addition of time period.

| Hauen: strong vs weak | input probability 0.355 | $R^2$: 0.324 | deviance: 842.865 | AIC: 862.865 |
|---|---|---|---|---|
| **Source** | **factor weight** | | | |
| blogs | >0.999 | | | |
| Die Zeit | 0.736 | | | |
| Der Tagesspiegel | 0.554 | | | |
| Berliner Zeitung | 0.00646 | | | |
| subtitles | 0.00105 | | | |
| | | | | |
| **time period +1** | **log odds -0.253** | | p=0.0065 | |

# CHAPTER 5

# CONCLUSION

This thesis has investigated the effects of frequency and style on the weakening of class VII verbs from OHG to NHG. It builds on work previously done by Lieberman et al. (2007) and Carroll et al. (2012). It takes into consideration token frequency, which is a dynamic variable, rather than an inherent property of the grammar, and this usage-based approach provides a more accurate picture of the reasons why strong verbs weaken over time than looking solely at phonological form (i.e. stem shape) or membership in a verb class. By looking only at one verb class, this study also eliminates the variable of type frequency, which can play a confounding role in the analysis of verb weakening. It offers insight into reasons why some class VII verbs remain strong and others become weak. This study identifies a threshold of weakening for strong verbs based on frequency, around 2/100,000 tokens.

This study suggests that the effects of frequency are potentially greater than seen in previous studies. Through the investigation of only one class of verbs, confounding factors such as type frequency were controlled for, and the individual changes in frequency over time of each of these verbs were able to be tracked, rather than as a bulk "frequency bin" as seen in Lieberman et al. (2007) and Carroll et al. (2012). This allowed for a detailed analysis of the effects of frequency over time and the suggestion of a potential threshold for weakening.

While strong verbs in the Germanic languages have a tendency to become weak over time, the most frequent ones have remained strong. A usage-based acquisitional model can be suggested as an explanation for this, because frequently used irregular forms are more easily able

to be recalled than infrequent ones. The data from this study suggests that a threshold frequency exists, below which a strong verb is more likely to become weak. Further investigations of this in other verb classes are warranted.

This study also points to the cohesiveness of the *a*-group verbs as a group, which suggests there is some factor involving the vocalism that drives patterning in the Germanic languages. Four strategies have been suggested to have been involved in the loss of reduplication through reanalysis (cf. 2.2 and 2.3): the so-called *r*-preterite in the verba pura, reduction and contraction of the unstressed second syllable, development of a new ablaut pattern, and replacement with the weak preterite. This study has built on Lieberman et al. and Carroll et al. to develop a fifth causal element in the loss of reduplication of class VII and, more broadly, the weakening of the Germanic verb. The class VII verbs in German followed a short-long-short vowel alternation pattern which is very similar to that seen in class VI, the other member of the *a*-group, rather than remodeling themselves based on the consonant structures seen in the *e*-group. This study, by quantifying the weakening of the formerly reduplicating verbs, could shed light on the developmental path of PIE reduplicating verbs into (Proto-)Germanic.

**References**

Bammesberger, Alfred. 1986. *Der Aufbau des germanischen Verbalsystems* (Indogermanische Bibliothek. Erste Reihe, Lehr- und Handbücher). Heidelberg: C. Winter.

Bech, Gunnar. 1969. *Das germanische reduplizierte Präteritum.* Kongelige Danske Videnskabernes Selskab, Historisk-filosofisk Meddeleser 44. 1. Copenhagen: Munksgaard.

Bell, Allan, 2001. Back in style: reworking audience design. In: P. Eckert, J.R. Rickford (eds.), *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press. 139–169.

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Carroll, Ryan, Ragnar Svare & Joseph Salmons. 2012. Quantifying the evolutionary dynamics of German verbs. *Journal of Historical Linguistics* 2(2). 153-172.

Dammel, Antje, Jessica Nowak & Mirjam Schmuck. 2010. Strong-verb paradigm leveling in four Germanic languages: A category frequency approach. *Journal of Germanic Linguistics* 22(04). 337–359.

Deutsch Diachron Digital. www.deutschdiachrondigital.de. (February 23, 2019.)

DWDS. Digitales Wörterbuch der Deutschen Sprache. www.dwds.de. (February 23, 2019.)

DWDS Corpora. www.dwds.de/r. (February 23, 2019.)

DWDS Special Corpora. www.dwds.de/d/k-spezial. (February 23, 2019.)

DWDS Word Information. www.dwds.de/stats. (February 23, 2019.)

Fulk, R.D. 1987. Reduplicating verbs and their development in Northwest Germanic. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 109. 159-179.

---. 2018. *A comparative grammar of the early Germanic languages.* (Studies in Germanic Linguistics.) Amsterdam: Benjamins.

Hill, Eugen. 2010. A case in grammaticalized inflectional morphology: Origin and development of the Germanic weak preterite. *Diachronica* 27(3). 411-458.

Jasanoff, Jay H. 2008. From reduplication to ablaut: The class VII strong verbs of Northwest Germanic. *Historische Sprachforschung* 120. 241–284.

Johnson, Daniel Ezra. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1). 359-383.

Knooihuizen, Remco, Odile A. O. Strik & Gerbrich de Jong. 2018. Frisian strong and weak verbs in the face of Dutch influence: a synchronic and experimental approach. *The Journal of Comparative Germanic Linguistics* 21(1). 57–74.

Knooihuizen, Remco & Oscar Strik. 2014. Relative productivity potentials of Dutch verbal inflection patterns. *Folia Linguistica* 35(1). 173-200.

Lieberman, Erez, Jean-Baptiste Michel, Joe Jackson, Tina Tang & Martin Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature* 449. 713-716.

MHDBDB. Mittelhochdeutsche Begriffsdatenbank. mhdbdb.sbg.ac.at/. (February 23, 2019.)

Newberry, Mitchell G., Christopher A. Ahern, Robin Clark, & Joshua B. Plotkin. Detecting evolutionary forces in language change. *Nature* 551. 223-226.

Nowak, Jessica. 2010. On the emergence of an eighth ablaut class in German and Dutch. *Journal of Germanic Linguistics* 22(04). 361–380.

Nübling, Damaris. 2000. *Prinzipien der Irregularisierung: eine kontrastive Analyse von zehn Verben in zehn germanischen Sprachen* (Linguistische Arbeiten 415). Tübingen: Niemeyer.

Nübling, Damaris. 2011. How do exceptions arise? On different paths to morphological irregularity. In Horst J. Simon & Heike Wiese (eds.), *Expecting the Unexpected: Exceptions in Grammar*. Berlin: De Gruyter Mouton.

Prokosch, Eduard. 1939. *A Comparative Germanic Grammar.* Baltimore: Linguistic Society of America.

Ringe, Don. 2006. *From Proto-Indo-European to Proto-Germanic: A linguistic history of English, Volume 1.* Oxford: Oxford University Press.

Rittenhouse, Rose. 2014. Verbal periphrasis in two early Germanic languages: A comparative study of the passive and perfect in the Old High German *Evangelienbuch* and the Old Saxon *Hêliand*. Madison, WI: University of Wisconsin dissertation.

Salmons, Joseph. 2018. *A History of German.* Oxford: Oxford University Press.

Somers, Katerina. 2011. The introduction and extension of the *-st* ending in Old High German. *Journal of Germanic Linguistics* 23(2). 141-181.

Strik, Oscar. 2014. Stability and change in strong verb inflection between Old and Early Modern Frisian. *Amsterdamer Beiträge zur älteren Germanistik* 73(1). 449–496.

Van Coetsem, Frans. 1990. *Ablaut and reduplication in the Germanic verb* (Indogermanische Bibliothek. Erste Reihe, Lehr- Und Handbücher). Heidelberg: C. Winter.

Van Coetsem, Frans. 1994. *The vocalism of the Germanic parent language: systemic evolution and sociohistorical context* (Untersuchungen Zur Vergleichenden Grammatik Der Germanischen Sprachen; 4. Bd). Heidelberg: C. Winter.

Vennemann, Theo. 1994. Zur Entwicklung der reduplizierenden Verben im Germanischen. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 116. 167-221.

---. 1997. The development of reduplicating verbs in Germanic. In Irmengard Rauch and Gerald F. Carr, eds., *Insights in Germanic linguistics II: Classic and contemporary*. Bloomington: Indiana University Press. 297-336.

Warwick University. 2010. Current news. warwick.ac.uk/fac/arts/modernlanguages/german-studies/resources/wgsw/for-school-students/current-news/germany/. (February 23, 2019.)

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.

**Appendix A: List of texts included in OHG corpus**

Deutsch Diachron Digital

Benediktinerregel

Isidor

Kleinere Althochdeutsche Sprachdenkmäler

Monsee Fragments

Murbacher Hymnen

Otfrid: Evangelienbuch

Physiologus

Tatian

Notker: Martianus Capella, Psalmen, Boethius Categoriae, Boethius De Interpretatione, Kleinere Ars Rhetorica, Kleinere De Musica, Kleinere De Partibus logicae, Kleinere syllogismus

**Appendix B: Token count and frequency of prefixed verbs**

| ***verb*** | ***n*** | ***frequency*** |
|---|---|---|
| *auffallen* | 22302 | 1.704948 |
| *einfallen* | 20904 | 1.598074 |
| *gefallen* | 111321 | 8.510292 |
| *zerfallen* | 12982 | 0.992451 |
| *fallen* | 228986 | 17.50557 |
| ***total *fallen*** | **396495** | **30.31134** |
| *abfangen* | 4253 | 0.325134 |
| *anfangen* | 78229 | 5.980468 |
| *befangen* | 3625 | 0.277125 |
| *empfangen* | 65229 | 4.986641 |
| *verfangen* | 3548 | 0.271238 |
| *fangen* | 77028 | 5.888653 |
| ***total *fangen*** | **231912** | **17.72926** |
| *anhalten* | 22052 | 1.685836 |
| *aufhalten* | 37856 | 2.894024 |
| *behalten* | 75087 | 5.740267 |
| *enthalten* | 134588 | 10.28901 |
| *verhalten* | 57654 | 4.407546 |
| *halten* | 781240 | 59.72441 |
| ***total *halten*** | **1108477** | **84.74109** |
| *abhauen* | 3306 | 0.252738 |
| *einhauen* | 491 | 0.037536 |
| *zerhauen* | 639 | 0.04885 |
| *hauen* | 15045 | 1.150163 |
| ***total *hauen*** | **19481** | **1.489288** |
| *ablaufen* | 15704 | 1.200543 |
| *anlaufen* | 7733 | 0.591174 |
| *entlaufen* | 1833 | 0.14013 |
| *mitlaufen* | 1143 | 0.08738 |
| *verlaufen* | 42911 | 3.280469 |
| *laufen* | 286602 | 21.91021 |
| ***total *laufen*** | **355926** | **27.20991** |
| | | |
| *ablassen* | 4169 | 0.318713 |
| *anlassen* | 1336 | 0.102135 |
| *belassen* | 8206 | 0.627334 |
| *entlassen* | 40461 | 3.093171 |
| *verlassen* | 172118 | 13.15811 |
| *zerlassen* | 562 | 0.042964 |
| *lassen* | 1493218 | 114.1539 |
| ***total *lassen*** | **1720070** | **131.4963** |
| *anrufen* | 23172 | 1.771458 |
| *aufrufen* | 31471 | 2.405902 |
| *berufen* | 55612 | 4.251438 |
| *verrufen* | 1027 | 0.078512 |
| *rufen* | 184499 | 14.10462 |
| ***total *rufen*** | **295781** | **22.61193** |
| *bescheiden* | 46943 | 3.588709 |
| *entscheiden* | 229336 | 17.53233 |
| *scheiden* | 33475 | 2.559104 |
| ***total *scheiden*** | **309754** | **23.68014** |
| *abraten* | 3280 | 0.25075 |
| *anraten* | 2074 | 0.158554 |
| *beraten* | 53261 | 4.071708 |
| *erraten* | 5368 | 0.410374 |
| *verraten* | 49644 | 3.795195 |
| *raten* | 35393 | 2.705732 |
| ***total *raten*** | **149020** | **11.39231** |
| *ausschlafen* | 1538 | 0.117577 |
| *einschlafen* | 6814 | 0.520918 |
| *entschlafen* | 1699 | 0.129886 |
| *verschlafen* | 5688 | 0.434837 |
| *schlafen* | 51582 | 3.943352 |
| ***total *schlafen*** | **67321** | **5.14657** |
| *bewalten* | 2 | 0.000153 |
| *verwalten* | 17236 | 1.317661 |
| *walten* | 7589 | 0.580166 |
| ***total *walten*** | **24827** | **1.89798** |

**Appendix C: Token counts of verbs in time periods**

| *Verb* | *Strong pret OHG* | *Weak pret OHG* | *Lemma count OHG* | *Strong pret MHG* | *Weak pret MHG* | *Lemma count MHG* | *Strong pret NHG* | *Weak pret NHG* | *Lemma count NHG* |
|---|---|---|---|---|---|---|---|---|---|
| *fallen* | 26 | 0 | 69 | 684 | 1 | 1790 | 88011 | 6 | 228986 |
| *falten* | 0 | 0 | 2 | 42 | 58 | 128 | 0 | 857 | 4402 |
| *fangen* | 14 | 0 | 61 | 1062 | 0 | 4734 | 22652 | 2 | 77028 |
| *halten* | 11 | 0 | 70 | 1125 | 0 | 2069 | 146172 | 1 | 781240 |
| *hauen* | 2 | 0 | 7 | 136 | 5 | 831 | 1276 | 570 | 15045 |
| *heißen* | 91 | 0 | 572 | 9108 | 0 | 12943 | 153775 | 9 | 579712 |
| *laufen* | 15 | 0 | 37 | 1441 | 1 | 2616 | 63344 | 11 | 286692 |
| *lassen* | 37 | 0 | 203 | 3681 | 0 | 19679 | 297264 | 1 | 1493218 |
| *raten* | 28 | 0 | 43 | 434 | 0 | 1452 | 7011 | 0 | 35393 |
| *rufen* | 30 | 0 | 61 | 1414 | 150 | 2399 | 87572 | 563 | 184499 |
| *scheiden* | 3 | 0 | 70 | 2638 | 1 | 3326 | 9042 | 63 | 33475 |
| *schlafen* | 8 | 0 | 51 | 316 | 0 | 1321 | 11565 | 4 | 51582 |
| *walten* | 10 | 0 | 56 | 477 | 0 | 1153 | 0 | 652 | 7589 |