

## Frequency effects and structural change – the Afrikaans preterite

---

Johanita Kirsten

School of Languages, North-West University, Vaal Triangle Campus, South Africa  
E-mail: Johanita.Kirsten@nwu.ac.za

### Abstract

According to emergent grammar and exemplar theory in cognitive linguistics, the frequency of an item affects its behaviour in terms of structural change. In this article, I illustrate how high frequency items, such as preterital modal auxiliaries and copulas in Afrikaans, resist regularising with the rest of the Afrikaans verbal system. Items with a moderately high frequency can resist change for a time, but succumb to it eventually, such as *mog* (“might”) and *wis* (“knew”). While the course of change can also be affected by other factors, such as *het* (“have”) and *had* (“had”), and *dink* (“think”) and *gedink/dag/dog* (“thought”) show, the data in diachronic Afrikaans corpora from 1911 to 2010 confirm that high frequency items resist structural change to a large extent, while low frequency items do not. This links with the cognitive representation of language and language processing, and illustrates how the use of language shapes the structure of language.

**Keywords:** frequency, structural change, emergent grammar, exemplar model, preterite

### 1. Introduction

The preterite can be described as a synthetic past tense, where the past tense is indicated through inflection on the verb, such as *would* as the preterite of *will*, or *did* as the preterite of *do*. In many languages, there is a distinction between the present tense, the preterite and the perfect, such as in English with present tense *do*, preterite *did* and perfect *have done*. In more analytical languages, such as Afrikaans, which has only one present and one past tense, finer temporal distinctions often depend (at least partially) on other means than inflection, like adverbials, chronological order and context (Conradie 1998: 41).

In the development of Afrikaans from Dutch, deflection was widespread. This deflection caused the preterite to disappear almost completely, making way for the Dutch perfect as the new past tense – already in 1902 Du Toit (1902: 24) indicated that the “onfolmaak ferlede tyde” (literally “imperfect past tense”) had mostly disappeared – with the exception of a number of modals, the copula *wees* (“is”), and to a small extent the verbs *het* (“have”) and *weet* (“know”) (Ponelis 1979: 269; Conradie 1998: 37; Conradie 2006: 87). Conradie (1999: 20) gives an account of the remaining Afrikaans preterites with an indication of the extent to which it is still used, as shown in table 1:

**Table 1:** Remaining preterite forms in Afrikaans (Conradie 1999: 20)

Current PRET forms			Obsolescent in 20th century		
<b>Copula and auxiliary verb</b>			<b>Main and auxiliary verb</b>		
is	<i>was</i>	“was”	het	<i>had</i>	“had”
<b>Modal auxiliary</b>			<b>Modal auxiliary</b>		
sal	<i>sou</i>	“should”	mag	<i>mog</i>	“might”
moet	<i>moes</i>	“had to”	<b>Main verb</b>		
kan	<i>kon</i>	“could”	weet	<i>wis</i>	“knew”
wil	<i>wou</i>	“would”			
<b>Dubitative verb</b>					
dink	<i>dog/dag</i>	“thought”			

In the grammatical system of Afrikaans, the remainder of the preterite has been replaced by the Dutch perfect (used as a full-fledged past tense), and the historical present, which has been functionally extended to perform some of the earlier functions of the preterite (Conradie 1999: 21–2). The preterite is still fully functional in contemporary Dutch (Abraham 1999: 12) for example, but many other languages have been experiencing preterite loss, like Southern German and Yiddish in the West Germanic language family, as well as northern Italian, Hungarian, Polish, Czech, Russian, Ukrainian and Slavonic (Abraham 1999: 13). The catalyst for preterite loss is often that the perfect develops into a more general past tense (Abraham 1999: 14). This is exactly what happened in Afrikaans, and there are additional developments that aided preterite loss in Afrikaans to an even greater extent. These additional developments include regularisation of the Dutch verbs *hebben* (“have”) and *zijn* (“is”) to only *het* and *is*, the regularisation of the past participle to *ge-* (sometimes optional) + stem, and the functional extension of *het* to replace *is* as the past tense auxiliary used with mutative verbs (Conradie 1999: 22).

The process of preterite loss in the development of Afrikaans, and all the possible contributing factors, are explored in-depth in Conradie (1999). Conradie (1999) focuses primarily on the 18th and especially the 19th century, exploring the factors that contributed to early preterite loss in the formation of Afrikaans. After the initial development of Afrikaans, a fully standardised variety of Afrikaans was established by the 1930s, at which time it had also become an explicit marker of Afrikaner identity. Conradie (1999) gives only a few brief remarks on preterite use in the 20th century, reported in table 1. Other authors, such as De Villiers (1971) and Ponelis (1979), do not focus on historical developments when discussing Afrikaans preterites.

Two questions then arise: (1) what has happened to the last remnants of Afrikaans preterites since the formation of a standard variety a century ago? And (2) what can the recent history and current state of Afrikaans preterites reveal about why a finite set of preterites remained in a language where the rest of the verbal system has been almost completely regularised?

The answers to these questions will be sought in the exploration of diachronic corpora of standard written Afrikaans, and the interpretation of the relevant empirical findings within the theoretical framework of cognitive linguistics.

## 2. Theoretical framework

There are three major hypotheses that guide cognitive linguistic approaches to language (Croft and Cruse 2004: 1):

- language is not an autonomous cognitive faculty;
- grammar is conceptualisation; and
- knowledge of language emerges from language use.

In cognitive linguistics, language and language use is seen as based in the cognition of the individual (Bybee 2010b: 9), and the representation of linguistic knowledge is in principle the same as the representation of other conceptual structures (Croft and Cruse 2004: 2). Cognitive linguistics leave room for language to be influenced by extra-linguistic factors (Bybee and Hopper 2001: 19; Bybee 2010b: 193), and lead to a change from viewing language structure as a holistic, autonomous system to something more fluid and variable (Bybee and Hopper 2001: 2). Variability and variation is seen as a fundamental part of language (Bybee and Hopper 2001: 19; Croft 2001: 7), and not as an almost irrelevant manifestation of ‘performance’ in Chomskyan terms. Croft (2001: 8, 364) sees language as fundamentally dynamic and interactional, which not only accommodates variation but elevates it to the status quo, thereby setting aside the traditional focus on ‘competence’.

A further consequence of viewing change as a central component of language is that the clear distinction between synchronic and diachronic linguistics fades away. Bybee (2010a: 945; 2010b: 10) claims that linguists should seek explanations for current language structures in how these structures arose, because all synchronic states in language result from a long chain of diachronic developments (Bybee 2010a: 945). Consequently, a theory of language should focus on, or at least incorporate, the dynamic processes that continually create and recreate language (Bybee 2010b: 1).

The concept of ‘emergent grammar’ has been put forward in cognitive linguistics – the hypothesis that knowledge of language emerges from language use (Croft and Cruse 2004: 3). Of the three hypotheses of cognitive linguistics reported at the beginning of the section, this is the one most relevant to this study. Emergent grammar breaks with traditional ideas about grammar. It relativises language structure to speakers’ real experience with language, and sees structure as an ongoing reaction to the pressure of discourse rather than a pre-existing matrix (Bybee and Hopper 2001: 3). Grammar has no autonomous existence outside of mental representation and processing, and so it is continually adapted for use (Bybee and Hopper 2001: 2–3). A summary by Thompson and Hopper (2001: 48) is quite explanatory in this regard:

We could say, then, that what we think of as grammar is a complex of memories we have of how our speech community has resolved communicative problems. ‘Grammar’ is a name for the adaptive, complex, highly interrelated, and multiple categorized sets of recurrent regularities that arise from doing the communicative work humans do.

Recently, it has been established that the human brain has a much larger capacity for long-term memory than previously thought possible (Pierrehumbert 2001: 140; Bybee 2010b: 15), which solves the problem of storing large amounts of linguistic information previously anticipated by

many formal linguists. Moreover, it has been shown that human cognition has room for quite an extent of redundancy (Hare, Ford and Marslen-Wilson 2001: 196). This casts doubt on the necessity of abstract structures and systems for the acquisition of linguistic knowledge.

An important concept within emergent grammar is that of the ‘exemplar model’. Bybee (2010b: 19) explains that exemplars are built from a set of tokens the speaker has experienced, and regarded as similar in some ways. The categories in the speaker’s memory are regarded as large clouds of recalled tokens of specific categories, which is organised in a cognitive map where similar examples are closer and dissimilar examples further away from each other (Pierrehumbert 2001: 140). Token frequency is not overtly encoded in the exemplar model, but it plays an important role in the cognitive representation of a category (Pierrehumbert 2001: 143). The importance of use and frequency for the emergence and change of a language’s grammar is advocated by many linguists, among others Fenk-Oczlon (2001: 433), Bybee (2010a: 987) and Langacker (2010: 430). Humans’ cognitive apparatus are sensitive to frequency, and it tends to sort and represent events and elements according to context-relevant relative frequency without specific instruction or request (Fenk-Oczlon 2001: 433). High frequency leads to familiarity, which enables a speaker to recall tokens or constructions with more speed and ease, and to identify, recognise, anticipate and predict it more accurately (Fenk-Oczlon 2001: 432). High frequency strengthens mental representation, and familiarity eases and speeds up recall, causing resistance to structural change, while items with lower frequency are recognised and recalled with more effort, thus being more susceptible to change (Bybee 2010a: 962). On these grounds, Bybee and Hopper (2001: 10) claim that frequency affects linguistic behaviour in many ways (also see Hare et al. 2001: 181 and Deutscher 2005: 261 in this regard).

In such a usage-based theory, with a specific focus on frequency, among other things, quantitative studies become very important to understanding the breadth of language experience (Bybee 2010b: 12). To these ends, I used quantitative language data for this article, and I explain my methodology below.

### **3. Methodology**

As I have already hinted above, the use of quantitative, or at least quantifiable, data is particularly useful within cognitive linguistics and usage-based approaches (Bybee 2010b: 12), which stands in stark contrast to the generative framework (Gisborne and Hollmann 2014: 3). In the past 25 years or so, experimental and observational methods have amassed large amounts of data that calls into question a number of assumptions about language: a large amount of innate structure for language acquisition, that language is a highly modular system, and that the primary type of acceptability judgement data is reliable (Gries 2014: 16). It has been shown that seemingly impoverished language input during acquisition is rich with probabilistic structure, rendering many seemingly unlearnable things quite learnable (Gries 2014: 16). The concept of probability is important, and naturally involves frequency – the cumulative effect of usage frequency is a driver behind change, where high and low frequency have different effects on the behaviour and change of linguistic elements (Croft 2000: 3, 32; Deutscher 2005: 261; Leech, Hundt, Mair and Smith 2009: 90). Corpus linguistic methods are eminently suitable for investigating frequency, and other forms of quantifiable language data.

Corpus linguistics can be defined as the study or analysis of language through the use of electronic corpora (Leech et al. 2009: 24). In itself, corpora cannot reveal anything about

language, but as it is collections of electronic texts, it can be subjected to electronic analyses through computer software (Evison 2010: 122). A corpus does not necessarily contain any new information on (a) language, but the software offers a new perspective on the familiar (Biber, Conrad and Reppen 1998: 234; Evison 2010: 122).

One of the most important advantages of corpus linguistics is that large amounts of linguistic data can be analysed in terms of usage patterns – a lot more than would be possible to analyse manually – while contextual factors can be accounted for (Biber et al. 1998: 3–4; Reppen, Fitzmaurice and Biber 2002: viii; Conrad 2010: 234; Tognini Bonelli 2010: 20). Computers perform consistent, reliable analyses – they do not change their minds and do not become tired during analysis (Biber et al. 1998: 4). There is also the interactive component that allows the human analyst to make difficult linguistic judgements, while the computer takes care of record keeping (Biber et al. 1998: 4).

A diachronic corpus, a corpus consisting of the language use of a specific historical period, is a ‘snapshot’ of the language use of a specific period (Tognini Bonelli 2010: 22), which can be used to trace the course of a specific change. In contrast with certain aspects of the diachronic research tradition, Leech et al. (2009: 50) feel “that frequency evidence is far more important in tracing diachronic change than has generally been acknowledged in the past”. In corpus data, language change usually manifests through changes in frequency and contexts of use (Biber et al. 1998: 209). Of course, any change should always be described against the backdrop of the “always far greater and more comprehensive continuity in usage” (Mair 2006: 3).

A method that suits the investigation of ongoing language change particularly well, is comparative corpus linguistics, “or more specifically short-term diachronic comparable corpus linguistics” (Leech et al. 2009: 24). The corpora that I used are those used in Kirsten (2015). The empirical design of this article is loosely based on the model established by Mair (2006) and Leech et al. (2009). In this method, comparable corpora from different periods are used to investigate variation and change. Both studies used four corpora: Brown (written American English from 1961), LOB (British English from 1961), Frown (American English from 1992), and FLOB (British English from 1991). These corpora were then compared in terms of variety (American vs. British) and historical period (1961 vs. 1991/2). The most important advantage of this comparative corpus methodology is that it is firmly based on ‘observable’ differences between two or more corpora (Leech et al. 2009: 32).

An important concept in the above-mentioned methodology is that of comparability. Being comparable means that the composition of two or more corpora are the same in all aspects but one – that aspect in terms of which it is being compared (Leech et al. 2009: 28), in this case time periods. However, the Afrikaans corpora have deviated in certain ways from the model of Mair (2006) and Leech et al. (2009) for specific reasons.

Firstly, for practical reasons, the corpora are adaptations of the 30-year intervals that Mair (2006) used to investigate changes in English in the 20th century. Some of the source types included in the corpora are scarce and difficult to come by (see below), which is why it was not practically viable to include only sources from one year, every 30 years. This is why sources from every third decade was included (leaving 20-year gaps between every corpus). The decades covered by the corpora are: 1911–1920 (corpus #1), 1941–1950 (corpus #2), 1971–1980 (corpus #3) and 2001–2010 (corpus #4). There are more or less 261 000 words per decade

in comparable quantities for every genre (see below), with a maximum of around 2 000 words from a single text. In one case of extreme scarcity of texts (discussed further below), an absolute maximum of 5 000 words from a single text was maintained. The genres with the word count of each are given in table 2:

**Table 2:** Corpora details

Category	Genre	Word count
Fiction	Novels and short stories	± 60 000
Popular non-fiction	Biographical texts	± 20 000
	News reports	± 20 000
	Informative texts	± 60 000
	Religious texts	± 20 000
Academic	Texts from humanities	± 30 000
	Texts from the natural sciences	± 30 000
Manuscripts	Letters and diaries	± 21 000

Another way in which these corpora deviate from the model by Mair (2006) and Leech et al. (2009) is that they include some unpublished sources or manuscripts (letters and diaries). Even though these are comparatively small sections in the corpora, it involves some measure of unedited language use in the study, although it still primarily represents the written standard. All letters and diary entries obtained from private collections have been anonymised to protect the identities and personal information of the individuals mentioned in the texts.

One category in one of the corpora is incomplete, where there is an almost complete lack of available texts (mentioned above): the *Natural sciences* category in corpus #1 consists of only 9 260 words, as there are very few Afrikaans texts in the natural sciences available from that period. Natural scientists in South Africa still wrote predominantly in either Dutch or English during that time, causing very few Afrikaans texts from that specific period to have been preserved. In all cases where it was applicable, the numbers from corpus #1 have been normalised (i.e. increased proportionally as if consisting of the same number of words than the other corpora) for the sake of comparability.

In the initial analyses of the data, two analysis tools were used. Firstly, word lists, including frequency lists, and concordances of the relevant words were compiled and analysed using WordSmith Tools 6.0. Secondly, tables and graphs summarising the results from the analyses were compiled in Microsoft® Excel 2010. Furthermore, when I noticed a change in frequency, I performed a statistical test of significance (Biber et al. 1998: 275). I performed log likelihood tests with the log likelihood calculator of Rayson (2015), which was developed specifically for corpus data. This test indicates whether the difference in frequency between two data sets, or between several consecutive data sets, can be attributed to coincidence, or whether it is significant. A result of less than 3.84 is regarded as insignificant (i.e.  $p > 0.05$ ), between 3.84 and 6.63 indicates a low level of significance (when the  $p$  value is between 0.01 and 0.05), and more than 6.63 is regarded as significant (i.e.  $p < 0.01$ ) (Rayson 2015).

#### 4. A century of preterite use in written Standard Afrikaans

##### 4.1 Obsolescent preterites

Conradie (1999: 20) indicates three obsolescent preterite forms: *had* (“had”), the preterite of auxiliary and main verb *het* (“have”); *mog* (“might”), the preterite of the modal auxiliary *mag* (“may”); and *wis* (“knew”), the preterite of the main verb *weet* (“know”).

The first of these to be addressed is *had*. Early in the twentieth century the use of *had* was already quite rare, and declining even further (Kirsten 2013: 69–70). However, even in the fairly small corpora of this study, there are examples of use throughout the century. The frequency of use is indicated in table 3.

**Table 3:** Frequencies of *had* 1911–2010

	1911–1920	1941–1950	1971–1980	2001–2010
<b>main verb</b>	125	15	6	2
<b>auxiliary</b>	71	2	–	–

Compared to the form *het* (“have”), which is consistently used between four and a half and five thousand times in each corpus, *had* and the related form *hadden* is already quite rare in corpus #1, contributing a mere 4% to the total uses of *het*-forms; in the corpora of the following decades it decreases even further over time. This corresponds to the pattern of items with low frequency (relatively) being particularly susceptible to change, in this case to a further decrease in usage frequency. In contrast with *het*, *had* is used more frequently as a main verb than as an auxiliary in corpus #1 and #2, and in corpus #3 and #4 it is only used as a main verb, but its occurrence is very rare.

The next preterite on the list is the modal auxiliary *mog* (“might”), which is the preterite of *mag* (“may”). The frequencies of *mog* and related forms, as well as *mag* and related forms, are indicated in table 4.

**Table 4:** Frequencies of *mag* and *mog* and related forms 1911–2010

	1911–1920	1941–1950	1971–1980	2001–2010
<b>present</b>	147	170	99	100
<b>preterite</b>	24	6	–	2

The one important pattern in the table above is the radical decline in the preterite form(s), which is almost or completely absent in all but corpus #1. However, unlike *had*, which has a regular past tense *het* and *gehad*, *mog* does not. In order to determine the use of *mag* in the past tense, table 5 distinguishes between its uses in different tenses.

**Table 5:** Frequency of *mag* in different tenses 1911–2010

	1911–1920	1941–1950	1971–1980	2001–2010
<b>past</b>	2	6	2	6
<b>present</b>	131	164	97	94
<b>future</b>	1	–	–	–

Compared to the present tense, *mag* is not used very frequently in the past tense; however, only in corpus #1 is *mog* used more frequently in the past tense than *mag*. This indicates that *mag* took over the functional load of *mog* in its absence, once again showing how a lower frequency can contribute to structural change.

The last obsolescent preterite is *wis* (“knew”), the preterite of *weet* (“know”). The regular past tense *geweet* has replaced the preterite to a large extent, the details of which is given in table 6.

**Table 6:** Frequencies of *weet*, *wis* and *geweet* 1911–2010

	1911–1920	1941–1950	1971–1980	2001–2010
<b>weet</b>	281	239	277	298
<b>wis</b>	16	7	–	2
<b>geweet</b>	27	42	41	55

From table 6 it is clear that *geweet* has taken over the functional load of *wis* almost completely, and is the more frequent of the two in corpus #1 already. I found no uses of *gewis* as the past tense of *weet* in the corpora.

All three of the obsolescent preterites show the same pattern: they are already somewhat low in frequency compared to other Afrikaans preterites (see section 4.2 and 4.3), and in the course of the century their frequencies drop even further. Next up are the much more frequently used modal auxiliaries with fully productive preterites.

## 4.2 Modal auxiliaries

Before I investigate the use of the preterite of Afrikaans modal auxiliaries, I would like to give a quick overview of how I conceptualise modality in this article. As the focus of this article is not on modality, but on the Afrikaans preterite, which happens to involve modal auxiliaries, I will not go into modality in any depth.

Van der Auwera and Plungian (1998: 80) define modality as those semantic domains that concern possibility and necessity. Within this framework there are four domains: (1) participant-internal modality, which refers to ability and necessity internal to the participant; (2) participant-external modality, which refers to circumstances outside the participant, concerning possibility and necessity; (3) deontic modality, a subdomain of participant-external modality, which refers to permission or obligation outside of the participant; and (4) epistemic modality, which refers to the speaker’s judgements on the likelihood of something (Van der Auwera and Plungian 1998: 80–1). Two additional concepts that are often regarded as within the purview of ‘modality’ are volitive modality and evidential modality (Van der Auwera and Plungian 1998: 84). While Van der Auwera and Plungian (1998: 86) do not accept these concepts within their deliberately restricted definition of modality, they admit that they are at least closely linked to the more central domains of modality above. For the purpose of this article, I will use ‘modality’ as an umbrella term including also more peripheral types of (possible) modality.

There is a link between modality, tense, and aspect, which can be seen in, for instance, Heine’s (2003: 594) grammaticalisation paths, and Patard (2014: 69) notes that the past tense can convey modal meaning in many languages. While imperfectivity would more often be



associated with modality, the past tenses in Germanic languages which are used to convey modality are usually aspectually neutral (Patard 2014: 70, 72). Patard (2014: 87) links the modal uses of past tenses with pragmatic inferences, which become conventionalised to some extent, and she summarises it as follows:

Adopting a dynamic perspective, one may say that the modal uses of past tenses reflect a semantic evolution affecting these forms within particular environments (or constructions): past tenses have been increasingly used to convey modal effects (through inferential processes) so that, in certain contexts, the interpretation in terms of their source meaning is progressively backgrounded (or even overruled) while the target modal meaning is increasingly focused upon and put to the foreground (Patard 2014: 88).

In many cases the original temporal and aspectual reference is still possible during the transitional phase, from temporal to modal meanings (Patard 2014: 88). In Afrikaans, for instance, *sou* (“should”) is in a transitional phase, as it is still occasionally used to indicate past future tense, although it is more often used with purely modal meaning in both the past and present tense in the data (see table 7), possibly moving in a similar direction as English *should* (see Rossouw and Van Rooy 2012: 8), from a temporal to a modal meaning.

#### 4.2.1 “*Sou*”

The modal auxiliary *sal* (“shall”) indicates intention or prediction, usually with a future tense implicature, and its preterite is *sou* (“should”). *Sou* is used in two main categories: one is to indicate the past future tense (1), and the other is to indicate modality (2). The modal meanings of *sou* include hypotheticals, and epistemic possibility, and it can also indicate intention in the past tense.

- (1) Hij sou nie die laaste wees deur die Hollands-Afrikaanse volk gehou, in sijn bestaan bedreïgd deur die magtige Albion. (corpus #1, Religious)  
[He would not be the last one held by the Hollands-Afrikaans nation, in his existence threatened by the mighty Albion.]
- (2) Dit sou beter wees as hy hier sou gebly het om vir ons te help. (corpus #2, Fiction)  
[It would have been better if he would have stayed here to help us.]

The inflected Dutch forms *zou* and *zouden* are still present in corpus #1 (37 and 4 times, respectively); however, I regard them as equivalent for this purpose, as they are after all preterites. Table 7 gives the frequencies of the two main categories of *sou*.

**Table 7:** Temporal and modal frequencies of *sou* 1911–2010

	1911–1920		1941–1950		1971–1980		2001–2010	
<b>past future tense</b>	51	12,3%	82	18,7%	50	15,7%	53	19,9%
<b>modal</b>	362	87,7%	356	81,3%	269	84,3%	213	80,1%
<b>Total</b>	413		438		319		266	

The only clear change that can be seen from the table is the noticeable decline in total frequency from the corpus #2 to #4. A statistical test of significance shows that this decrease is indeed significant – while the difference between corpus #1 and #2 is not significant, the decrease from

corpus #2 onward is<sup>1</sup>. While the modal uses are up to four times as frequent as the temporal uses, the ratio between the two does not change in the course of the century, despite the decline in overall frequency.

The most prominent modal use of *sou* is to indicate hypotheticals or the irrealis, and while the modality itself is not bound by tense, the assertion or event can be presented in terms of the past (3) or the present (4).

- (3) 'n Paar jaar gelede sou niemand dit gewaag het nie. (corpus #3, Biographical)  
[A few years ago no-one would have dared.]
- (4) As jy regtig verder met hom wil werk (berading) sou ek voorstel dat jy hier begin. (corpus #4, Manuscript)  
[If you really want to work with him still (counselling) I would propose that you start here.]

Table 8 gives further details in this regard.

**Table 8:** Modal *sou* in the past and present tense 1911–2010

	1911–1920		1941–1950		1971–1980		2001–2010	
<b>past tense</b>	160	44,2%	201	56,5%	175	65,1%	137	64,3%
<b>present tense</b>	202	55,8%	155	43,5%	94	34,9%	76	35,7%
<b>Total</b>	362		356		269		213	

In the table above, a clear trend presents itself: in corpus #1 *sou* is used in the present tense more frequently than in the past tense, but the two tenses swop places in corpus #2. In corpus #3 the past tense uses increase even more proportionally, if not in raw frequency. While there is still a substantial use of modal *sou* in both tenses, the raw frequency decreases significantly<sup>2</sup>.

Something that complicates matters slightly is the frequent use of *sou* in combination with other modal auxiliaries (predominantly *kan* (“can”) and *kon* (“could”), but also *wil* (“will”), *wou* (“would”), *moet* (“must”) and *moes* (“must” PRET)). The question arises to which extent either the present tense or the preterites of the additional modal auxiliaries are combined with *sou*. In principle there are four options:

- i) *sou* and the preterite in the past tense;
- ii) *sou* and the present tense form in the past tense;
- iii) *sou* and the preterite in the present tense; and
- iv) *sou* and the present tense form in the present tense.

An example sentence of each, in this order, is given in examples 5 to 8.

- (5) Ander was bang om die reis te onderneem, omdat hulle dan deur die land van hulle vyande, met wie hulle gedurig in oorlog verkeer het, sou moes gaan. (corpus #2, Informative)  
[Others were scared to undertake the journey, because they would have had to travel through the land of their enemies, with whom they were continually at war.]

<sup>1</sup> The results of log-likelihood tests are: all four corpora consecutively 53.99; corpus #1 to #2: 0.80 (insignificant); corpus #2 to #3: 18.13; corpus #3 to #4: 4.99 (low level of significance).

<sup>2</sup> Log-likelihood for all four corpora consecutively: 74.92.

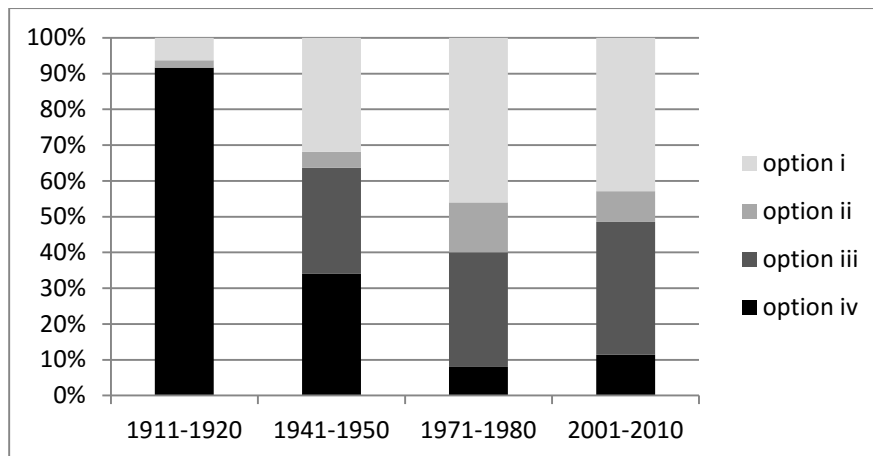
- (6) Maar hy was te lomp en het die mense afgeskrik deur opelik te verklaar, dat die Raad miskien sou kan ontwikkel tot iets nog groters. (corpus #1, Humanities)  
[But he was too clumsy and he scared the people away through declaring openly that the Council possibly could develop into something even bigger.]
- (7) Kearns & Paddison (2000) beweer dat stedelike bestuur ondernemender moet word, wat die keuse vir welvaartstaatmodel sou kon teenspreek. (corpus #4, Natural Science)  
[Kearns & Paddison (2000) claims that city management should become more enterprising, which could contradict the choice for the prosperity model.]
- (8) In 'n eenrigting-praatjie of lesing sou televisie wel die spreker se ooredingsvermoë sterk kan projekteer. (corpus #3, Informative)  
[In a one-way talk or lecture television would indeed be able to project the speaker's persuasiveness.]

The examples show that all the options are possible, which leads to the question of preference. More details in this regard are given in table 9.

**Table 9:** *Sou* in combination with other modal auxiliaries 1911–2010

	1911–1920	1941–1950	1971–1980	2001–2010
<b>option i</b>	3	14	23	15
<b>option ii</b>	1	2	7	3
<b>option iii</b>	–	13	16	13
<b>option iv</b>	44	15	4	4

The frequencies are particularly low, but one deduction can be made – corpus #1 shows a preference for the present tense form of the additional modal auxiliaries when used in the present tense, but that preference does not continue as time goes by (see figure 1), illustrating what Conradie (1999: 28) calls “the relatively modern phenomenon of preterite agreement on concatenated modal verbs”. The frequencies suggest an increasing preference for the use of the preterite all around, although the frequencies are too low to confirm clear trends with certainty.



**Figure 1:** *Sou* in combination with other modal auxiliaries

The figure confirms that the use of *sou* seems to usually trigger the use of the preterite of the other modal auxiliary it is used with in corpus #3 and #4, regardless of whether the rest of the

sentence is in the past or the present tense. Finally, the use of *sal* and the preterite *sou* is summarised in table 10.

**Table 10:** Frequencies of *sal* and *sou* 1911–2010

	1911–1920		1941–1950		1971–1980		2001–2010	
<b>sal</b>	1 083	72,4%	952	68,5%	915	74,1%	877	76,7%
<b>sou</b>	413	27,6%	438	31,5%	319	25,9%	266	23,3%
<b>Total</b>	1 496		1 390		1 234		1 143	

There seems to be a general decline of both *sal*<sup>3</sup> and *sou*<sup>4</sup>, and the ratio between the two varies but does not signal definite change. However, as *sal* is used primarily with regard to future reference (an implicature in its capacity as modal auxiliary), the cause and implications of this change fall outside the scope of this article.

#### 4.2.2 “Wou”

The modal auxiliary *wil* (“will”), which indicates intention or desire, takes the preterite form *wou* (“would”). In contrast with *sou*, as well as English *will* and *would*, both *wil* and *wou* are only ever used with modal meaning, and not temporal reference. Both *wil* and *wou* are used for participant-internal modality, *wou* is seldom used with other modal auxiliaries apart from *sou*, and it is only used in the past tense, illustrated in example (9).

- (9) Ek wou hom die laaste ent uittrek, maar hy het byna histeries geskree: “Moenie aan my vat nie!” (corpus #4, Fiction)  
 [I wanted to pull him out the last stretch, but he shouted almost hysterically: “Don’t touch me!”]

The frequencies of *wil* and *wou* are shown in table 11.

**Table 11:** Frequencies of *wil* and *wou* 1911–2010

	1911–1920		1941–1950		1971–1980		2001–2010	
<b>wil</b>	426	79,3%	344	75,3%	393	78,4%	391	82%
<b>wou</b>	111	20,7%	113	24,7%	108	21,6%	86	18%
<b>Total</b>	537		457		501		477	

There is a slight decrease in frequency of *wou* from corpus #3 to #4, but it is statistically insignificant<sup>5</sup>. If I include *wil*, the total frequencies and proportions show variability but does not suggest change. Thus, there seems to be no changes occurring with regard to *wou*.

<sup>3</sup> Log-likelihood for all four corpora consecutively: 23.0; corpus #1 to #2: 8.50; corpus #2 to #3: 0.58 (insignificant); corpus #3 to #4: 0.81 (insignificant).

<sup>4</sup> Log-likelihood for all four corpora consecutively: 53.99; corpus #1 to #2: 0.80 (insignificant); corpus #2 to #3: 18.13; corpus #3 to #4: 4.99 (low level of significance).

<sup>5</sup> Log-likelihood for corpus #3 to #4: 2.57 (insignificant).

### 4.2.3 “Kon”

The modal auxiliary *kan* (“can”) conveys both participant-internal ability and participant-external possibility. The preterite *kon* (“could”) places the ability or possibility in the past (10), except when a wish or desire is being expressed (11), or if it is used in combination with *sou* (12).

- (10) Hy kon ook ure aaneen die ou Joodse geskifte lees en met Josef daaroor gesels. (corpus #4, Religious)  
[He could read old Jewish texts for hours on end and discuss it with Josef.]
- (11) Ek wens jy kon bietjie in my kop inklim en kyk wat gaan aan. (corpus #4, Manuscript)  
[I wish you could climb into my head for a bit and see what’s going on.]
- (12) En veral met John *sou* ek kon uitgaan – hy is immers verlief op jou! (corpus #2, Fiction)  
[And especially with John I could go out – he is after all in love with you!]

If example (11) would be reformulated with *kan*, the difference in meaning becomes clear where *kon* indicates a wish for an unrealistic or impossible matter, while *kan* indicates a wish for a real possibility. The original association of temporal distance between *kan* and *kon* is extended here to epistemic distance.

De Villiers (1971: 29) claims that when *kan* is used with epistemic modality in the past tense in formal texts, it would sometimes not undergo preterite assimilation. However, in a sample of 400 uses of *kan* from each corpus, I did not find any examples of this occurring. This does not necessarily mean that the possibility does not exist, rather that it is just too rare to surface in my data.

Details regarding frequencies of *kon* and *kan* are given in table 12.

**Table 12:** Frequencies of *kan* and *kon* 1911–2010

	1911–1920		1941–1950		1971–1980		2001–2010	
<b>kan</b>	959	74,6%	1 053	75,7%	1 290	81,7%	1 357	80,5%
<b>kon</b>	326	25,4%	344	24,6%	289	18,3%	329	19,5%
<b>Total</b>	1 285		1 397		1 579		1 686	

It seems that *kon* also shows no definite change – the total frequencies and the ratio with *kan* shows variability, but it is not statistically significant<sup>6</sup>, and does not show a strong direction of change. However, the increase of *kan* is significant<sup>7</sup>.

### 4.2.4 “Moes”

The modal auxiliary *moet* (“must”) indicates both deontic and epistemic modality, with the preterite *moes*, which is used solely in the past tense. While *moet* merges with *nie* in the negative

<sup>6</sup> While the decrease from corpus #2 to #3 shows a low level of significance (with a log-likelihood result of 4.49), the increase to corpus #4 cancels this out.

<sup>7</sup> Log-likelihood for all four corpora consecutively: 96.23; corpus #1 to #2: 4.64 (low level of significance); corpus #2 to #3: 25.34; corpus #3 to #4: 1.48 (insignificant).

to form *moenie*, *moes* remains separate as *moes nie* in the negative. The use of *moes* is illustrated in (13).

- (13) Die Transvaal wil knoei: hulle sê ons moes die Engelse nie gehelp *het* nie. (corpus #1, Biographical)

[*The Transvaal wants to tamper: they say we should not have helped the English.*]

The frequencies of *moes* and *moet* are given in table 13.

**Table 13:** Frequencies of *moet* and *moes* 1911–2010

	1911–1920		1941–1950		1971–1980		2001–2010	
<b>moet</b>	807	72,6%	1 018	78,2%	925	83,3%	785	77,4%
<b>moes</b>	304	27,4%	284	21,8%	185	16,7%	229	22,6%
<b>Total</b>	1 111		1 302		1 110		1 014	

The frequencies of *moes* are variable but do not suggest continued change<sup>8</sup>. The present tense *moet* seems to decline somewhat significantly in frequency from corpus #2 onwards<sup>9</sup>.

#### 4.2.5 Cursory remarks: modal auxiliaries

The only modal auxiliary that increases significantly is *kan* – the overall increase is statistically significant, and the differences between the consecutive corpora are all significant, except for the difference between corpus #3 and #4. The other modal auxiliaries, including *kon*, fluctuate or decrease. The unmarked forms are used more frequently than the preterites throughout, even *wil* and *wou*, which have the lowest frequency of the modal auxiliaries. The only decline that persists throughout is *sal*, but as that involves future reference, it is not the subject of this article.

Next up is the third category of preterite use – the copula *wees* (“be”) with the preterite *was* (“was”).

#### 4.3. The copula “wees”

The copula *wees* (“be”) is realised in several different ways in Afrikaans – the infinitive *wees*, the most frequent form *is* in the present tense, the preterite *was*, and the regular past tense *gewees*. Both *is* and *was* can combine with *gewees*, even though *gewees* can also be used with modal auxiliaries and *het*. Examples 14–17 illustrate the different realisations of *wees* in the past tense: in (14) the preterite alone is used in the past tense, in (15) the preterite of a modal auxiliary leads to the regular past tense *gewees* with *het*, in (16) the preterite and regular past tense are used in combination, and in (17) the present tense *is* is used with the regular past tense.

- (14) Vir wolwe was hy uiters bedug! (corpus #4, Biographical)

[*Of wolves he was very careful!*]

<sup>8</sup> While the decrease from corpus #1 to #3 shows a low level of significance (with a log-likelihood result of 32.19), the increase to corpus #4 cancels this out to some extent.

<sup>9</sup> Log-likelihood for corpus #1 to #2: 25.0 (increase); corpus #2 to #3: 3.95 (decrease, low level of significance); corpus #3 to #4: 11.95 (decrease).

- (15) Dat hy só vermetel kon gewees *het*! (corpus #3, Fiction)  
[*That he could have been so audacious!*]
- (16) Dit was ook wel die geval in L gewees. (corpus #2, Manuscript)  
[*It was indeed also the case in L.*]
- (17) Ons poginge is dan ook nie vrugteloo gewees nie. (corpus #1, News report)  
[*Our attempts were not fruitless then.*]

According to Steyn (1976: 44) the combination “*was gewees*” is redundant, just like double negation, and while it is not ungrammatical, it does not indicate any semantic distinction from bare *was* (see also Conradie 1977: 61). Steyn (1976: 44) attributes the redundant *gewees* to Dutch influence, and claims that the combination occurs primarily in spoken language, while bare *was* is preferred in writing. He further notes the combination *is gewees*, although he believes it to be less frequent than *was gewees* (Steyn 1976: 44). Ponelis (1979: 266) indicates another equivalent of *was*, namely *het gewees* without a modal auxiliary, as in “Dit *het* so *gewees*” [“It has been so”].

When considering the lexical item *was* in unannotated corpora, such as those used in this article, it should be kept in mind that it represents several homonyms. Apart from being used as the preterite of the copula *wees*, it can also be an auxiliary in the passive voice (18), a main verb which means “to wash” (19), and a noun meaning “wax” (20).

- (18) Dit het gelei tot die ontstaan van die sogenaamde “Hyphenated techniques” waar chromatografie gekoppel word aan analise tegnieke wat tradisioneel op hul eie gebruik was. (corpus #4, Natural Science)  
[*That led to the emergence of the so-called “Hyphenated techniques” where chromatography is linked with analysis techniques that were traditionally used on their own.*]
- (19) Ek was die kleiner kinders se klere en hang dit buite. (corpus #4, Fiction)  
[*I wash the smaller children’s clothes and hang it outside.*]
- (20) Die bome het blare soos vlerke van was; die son smelt die vlerke van die bome. (corpus #3, Fiction)  
[*The trees have leaves like wings of wax; the sun melts the wings of the trees.*]

Also, in corpus #1 it still occurs as a past tense auxiliary with mutative verbs, as *het* had not yet replaced *is* and *was* completely in those contexts. Table 14 summarises the frequencies of the different verbal uses of *was*.

**Table 14:** Frequencies of the different verbal uses of *was* 1911–2010

	1911–1920	1941–1950	1971–1980	2001–2010
<b>main verb</b>	10	9	17	8
<b>past auxiliary</b>	27	–	–	–
<b>passive auxiliary</b>	156	92	65	42
<b>copula</b>	2 035	1 442	1 370	1 397
<b>Total</b>	2 228	1 543	1 452	1 447

Table 14 shows that the main use of *was* is as preterital copula, and that it does not occur as a past tense auxiliary after corpus #1. There is a significant drop in total frequency<sup>10</sup>, and that of the copula<sup>11</sup>, although the significance can be attributed to the difference between corpus #1 and #2.

Because *gewees* does not always occur in combination with *was*, but also in combination with modal auxiliaries<sup>12</sup>, it is possible that increase in frequency of *gewees* could contribute to the decrease of *was*. I explore this possibility in table 15.

**Table 15:** Frequencies of *gewees*

	1911–1920	1941–1950	1971–1980	2001–2010
+ modal aux.	58	31	35	22
+ <i>is</i>	35	3	–	–
+ <i>was</i>	23	9	3	1
<b>Total</b>	116	43	38	23

However, table 15 shows that *gewees* does not increase in frequency – on the contrary, it decreases significantly<sup>13</sup> – so it cannot be responsible for the decrease in frequency of *was*. Furthermore, the use of *gewees* in combination with *was* cannot be counted as competition, because *was* is used anyway, but that does not change anything in this regard.

It is also clear from the table that *is* in combination with *gewees* is absent in corpus #3 and #4, which does not necessarily mean that this usage has disappeared completely from the language. It may simply indicate that it is too rare to occur in the decades from 1971 onwards in the corpora used in this study, in stark contrast with corpus #1. There is also a decrease in frequency of uses together with *was*, which suggests that it might be falling out of use, at least with regard to written Standard Afrikaans.

As an explanation for the decline of *was* cannot be found in the use of *gewees*, I turn to the frequency of the present tense *is* and infinitive *wees* to determine whether there might be a shift towards the present tense.

**Table 16:** Frequencies of the different realisations of *wees* 1911–2010

	1911–1920		1941–1950		1971–1980		2001–2010	
<b>was</b>	2 035	31,4%	1 442	25,8%	1 370	24,2%	1 397	23,6%
<b>is</b>	3 910	60,4%	3 602	64,5%	3 825	67,5%	4 036	68,1%
<b>wees</b>	532	8,2%	542	9,7%	473	8,3%	493	8,3%
<b>Total</b>	6 477		5 586		5 668		5 926	

<sup>10</sup> Log-likelihood for all four corpora consecutively: 233.26; corpus #1 to #2: 123.34; corpus #2 to #3: 2.29 (insignificant); corpus #3 to #4: 0.03 (insignificant).

<sup>11</sup> Log-likelihood for all four corpora consecutively: 177.57; corpus #1 to #2: 100.09; corpus #2 to #3: 1.47 (insignificant); corpus #3 to #4: 0.18 (insignificant).

<sup>12</sup> There are no examples of *gewees* with *het* without modal auxiliaries, contrary to what Ponelis (1979: 266) suggests.

<sup>13</sup> Log-likelihood for all four corpora consecutively: 83.12; corpus #1 to #2: 34.62; corpus #2 to #3: 0.28 (insignificant); corpus #3 to #4: 3.78 (low level of significance).



The frequencies of *is* and *wees* both vary throughout. There is a small proportional decrease of *was* compared to *is* until corpus #3, but not sufficient to explain the decrease of *was* – especially considering the variance in total frequency of the different realisations of *wees*. The only explanation left is that extra-linguistic factors, like standardisation or socio-political context, caused the decrease from corpus #1 to #2.

#### 4.4. The outlier: “*dag/dog*”

For the sake of thoroughness, I will lastly investigate a bit of an outlier: the main verb *dink* (“think”) with two possible preterite realisations, *dag* and *dog*. The contemporary Afrikaans verb form *dink* developed from the Dutch form *denken* (and *denk* and *denkt*) that take the preterite *dachten* (and *dacht*) – the stem vowel of the present tense thus changed in the development from Dutch. Of the two preterites of *dink*, then, *dag*’s vowel corresponds to that of the Dutch preterite, while *dog* does not. However, these two contemporary forms are complete equivalents. Also, there is a regular past tense form of *dink*, namely *gedink*, which is merely the past tense of *dink*, while *dag/dog* underwent semantic change to mean more or less “wrongly assume” (Conradie 2006: 88). Regular past tense forms *gedag* and *gedog* are also attested. With these complications in mind, I give the frequencies of *dink* and all its related forms in table 17.

**Table 17:** Frequencies of *dink* and related forms 1911–2010

	1911–1920	1941–1950	1971–1980	2001–2010
<b>dink</b>	52	156	180	175
<b>denk/t/en</b>	102	–	–	–
<b>dag</b>	13	2	3	1
<b>dacht</b>	1	–	–	–
<b>dog</b>	4	2	2	2
<b>gedink</b>	10	31	44	25
<b>gedenk*</b>	15	–	–	–
<b>gedag</b>	1	1	–	–
<b>gedacht</b>	2	–	–	–
<b>gedog</b>	1	1	2	1

\* only when synonymous with *gedink*

The frequencies are in many cases too low to claim anything with certainty, but a few remarks are in order. Prominent variation still occurs in corpus #1 – the original Dutch forms and the new regular past participles *gedenk* and *gedink* occur, with the older stem vowel still more frequent than the new. All of the old forms are absent from corpus #2 onwards; only the word *gedenk* remained, but with a modified meaning of commemoration, similar to *herdenk* (which also preserved the original stem vowel). The original vowel of the preterite, in *dag* and *gedag*, is also more frequent than the new one in corpus #1, although the frequencies thereafter are too low to say anything more.

I have two reasons for calling *dink* and its related forms an outlier – (1) it is not nearly as frequent as even the least frequent modal auxiliary that is not obsolescent, but (2) no sources regard these forms as obsolescent; in fact, De Villiers (1971: 24) describes it as frequently used in spoken language. Also, *dag* and *dog* occur either in informal contexts or in reported speech

in all but corpus #1. All of this indicates that it should be regarded as informal, and that it is probably more prevalent in the spoken language than in the written standard. For these reasons, I refrain from labelling it as obsolescent, and can only say that it is not particularly frequent in the written Standard Afrikaans in the corpora used in this study.

## 5. Conclusion

The corpus data I used for this article confirms that the preterites that Conradie (1999: 20) labels as obsolescent (*had, mog, wis*) are indeed so, and are all increasingly rare from corpus #2 onwards. While I cannot say that they are completely absent from the standard variety (because of the limitations of the corpus data), other options have almost entirely replaced these preterites – *het* and *het gehad* for *had*, *mag* for *mog* and *geweet* for *wis*. This answers one part of the first research question regarding the recent history of Afrikaans preterites. The use of *sou* to indicate the past future tense also decreases in frequency in the consecutive corpora used for this study, together with a more general decline in preterite use. Table 18 summarises the use of preterites without alternative past tense forms (thus excluding the obsolescent forms and *dag/dog*) compared to their present tense equivalents, answering the remaining part of the first research question.

**Table 18:** Total frequencies of productive preterite use compared to present tense use

	1911–1920		1941–1950		1971–1980		2001–2010	
<b>preterite</b>	3 189	30,7%	2 621	27,3%	2 271	23,6%	2 307	23,7%
<b>present</b>	7 185	69,3%	6 969	72,7%	7 348	76,4%	7 446	76,3%
<b>Total</b>	10 374		9 590		9 619		9 753	

Two preterites, *moes* and *was*, reflect the same frequency pattern as the total preterite use, while *sou* and *wou* increase from corpus #1 to #2 and then decreases further on, and *kon* is variable but does not change significantly. The present tense form *kan* increases, on the other hand, which causes *kon* to decrease proportionally, while *sal* decreases with *sou*, *wil* remains more or less stable and *moet* behaves similarly to *sou* and *wou*. In general, then, it would seem like there is a subtly growing preference for the present tense rather than the preterite, although it is not nearly clear and definitive enough for me to make any final claims on the matter. This can be amended by using additional corpora in future research.

I now return to the theoretical issue regarding the relationship between frequency and language change to address the second research question. Like I explained in section 2: the higher the frequency of a construction is, the more it resists changes like regularisation and analogy, while lower frequency items are more susceptible to systematic change, tying in with the tenets of emergent grammar. For instance, the frequency of *mag* is the lowest of all the modal auxiliaries, and following the more general pattern, the past tense uses are even less frequent. This caused *mog* to fall increasingly into disuse, to the extent that it can be labelled as almost obsolete in contemporary Standard Afrikaans. The same is true for *weet* and *wis* – while *weet* is rather frequent for a lexical verb compared to most other lexical verbs in the data, the past tense uses seem too infrequent to resist being assimilated into the regularised verbal paradigm. It would seem that both *mag/mog* and *weet/wis* have been border cases for a time, being almost frequent enough to resist regularisation, but just not quite.

In contrast, the frequency of both the present tense and the preterite of all the other modal auxiliaries are particularly high (the preterites are all in the top 270 and the present tense forms in the top 75 most frequent tokens of over 22 000 in each of the corpora), resisting being regularised with the rest of the Afrikaans verbal system. The same, and more, is true for the single most frequent verb in Afrikaans, *wees* and its inflected forms – not only did it retain the preterite *was*, but also the infinitive *wees* (with a regular past tense *gewees*) together with the most frequently used present tense *is*. Infinitive forms that differ from the present tense are otherwise largely absent from the Afrikaans verbal system.

An anomaly in this regard is the verb *het*: as main verb it has the infinitive *hê* and the partially regularised past tense *gehad*, while the preterite *had* as main and auxiliary verb has become obsolescent. As the second most frequent verb in Afrikaans, would it not exactly have retained the preterite, like the case is with *wees*, if my argument above was true? A possible explanation for its preterite loss despite the odds can be found in the course according to which *het* developed from Dutch to Afrikaans: The Afrikaans form *het* originated from dialectal Dutch, particularly Hollands, which employed *het* as singular and *hewwe* as plural, rather than the more extended Standard Dutch paradigm with *heb*, *hebben*, *hebt* and *heeft* (Ponelis 1993: 386; Conradie 2006: 89). The singular *het* then persisted into Afrikaans, while *hewwe* developed into the infinitive *hê* (Ponelis 1993: 386; Conradie 2006: 89). This means that *het* would have already started to regularise during the early development of Afrikaans, gaining particular momentum during the 19th century (Conradie 2006: 89–90). In this process the preterite *had* lost its foothold during final regularisation of the verbal system, despite the high frequency of *het*.

The data set of this article confirms and illustrates the role of usage frequency in morphosyntactic change – the higher the frequency, the more resistant to change, where lower frequency items adapt more easily to a systematic structural change like regularisation. This links back to the role of cognitive grammatical representation, where higher frequency leads to more familiarity, and easier recall of an item, which is why changes in the system do not as easily affect these items. All of this favours emergent grammar and specifically the exemplar model, rather than viewing language as an abstract, rule-governed system.

### Acknowledgements

I would like to thank the North-West University for granting me six months research leave, during which I completed my PhD thesis and submitted this article. This article is based on a section of my PhD thesis, and I thank my promoter Bertus van Rooy for valuable feedback on my thesis. I also thank the anonymous reviewers who gave very valuable feedback on an earlier version of this article.

Corpus acknowledgements:

- I would like to thank the following individuals and institutions for providing sources for my corpus compilation:
  - CText (Centre for Text Technology, NWU, Potchefstroom Campus) for use of the PUK-Protea Korpus and the Taalkommissie-korpus.
  - Ana Deumert for use of the Corpus of Cape Dutch Correspondence (Deumert 2004).
  - National Archives repository in Pretoria for making available letters from 1911–1920, 1941–1950 and 1971–1980.
  - North-West University Archives for providing me with letters from 1941–1950.

- Epog Archive for providing me with letters from 2001–2010.
- All the anonymous private individuals who provided me with letters and diaries from private collections.
- My temporary research assistant, Chantelle Kruger, for hours of scanning, typing and quality control.
- My promoter, Prof. Bertus van Rooy, for guidance and advice (although any errors in compilation or judgement are completely my own).

## References

- Abraham, W. 1999. Preterite Decay as a European Areal Phenomenon. *Folia Linguistica* XXXIII(I): 11–18.
- Biber, D., S. Conrad and R. Reppen. 1998. *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Bybee, J. 2010a. Diachronic Linguistics. In D. Geeraerts and H. Cuyckens (Eds.) *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press. pp. 945–987.
- Bybee, J. 2010b. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Bybee, J. and P.J. Hopper. 2001. Introduction to frequency and the emergence of linguistic structure. In J. Bybee and P.J. Hopper (Eds.) *Frequency and the emergence of linguistic structure*. Amsterdam/Philadelphia: John Benjamins. pp. 1–24.
- Conrad, S. 2010. What can a corpus tell us about grammar? In M. McCarthy and A. O’Keeffe (Eds.) *The Routledge handbook of corpus linguistics*. London/New York: Routledge. pp. 227–240.
- Conradie, C.J. 1977. Die hulpwerkwoord het binne die Afrikaanse werkwoordgroep. *Taalfasette* 22(1): 56–72.
- Conradie, C.J. 1998. Tempusgebruik in Afrikaanse narratiewe. *South African Journal of Linguistics* 16(2): 37–43.
- Conradie, J. 1999. Preterite Loss in Early Afrikaans. *Folia Linguistica* XXXIII(I):19–37.
- Conradie, J.C. 2006. Defleksie en grammatikalisasie: die Afrikaanse hulpwerkwoord *het*. *Tydskrif vir Taalonderrig* 40(1): 86–98.
- Croft, W. 2000. *Explaining Language Change: An Evolutionary Approach*. Harlow: Pearson Education.
- Croft, W. 2001. *Radical Construction Grammar. Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Croft, W. and D.A. Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

- Deutscher, G. 2005. *The Unfolding of Language: An Evolutionary Tour of Mankind's Greatest Invention*. New York: Henry Holt and Company.
- De Villiers, M. 1971. *Die grammatika van tyd en modaliteit*. 2de druk. Kaapstad: Balkema.
- Du Toit, S.J. 1902. *Fergelykende Taalkunde fan Afrikaans en Engels*. 2de uitgawe. Paarl: D.F. du Toit & Co.
- Evison, J. 2010. What are the basics of analysing a corpus? In M. McCarthy and A. O'Keeffe (Eds.) *The Routledge handbook of corpus linguistics*. London/New York: Routledge. pp. 122–135.
- Fenk-Oczlon, G. 2001. Familiarity, information flow, and linguistic form. In J. Bybee and P.J. Hopper (Eds.) *Frequency and the emergence of linguistic structure*. Amsterdam/Philadelphia: John Benjamins. pp. 431–448.
- Gisborne, N. and W.B. Hollmann. 2014. Theory and data in cognitive linguistics. In N. Gisborne and W.B. Hollmann (Eds.) *Theory and data in cognitive linguistics*. Amsterdam/Philadelphia: John Benjamins. pp. 1–14.
- Gries, S.T. 2014. Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics. Some necessary clarifications. In N. Gisborne and W.B. Hollmann (Eds.) *Theory and data in cognitive linguistics*. Amsterdam/Philadelphia: John Benjamins. pp. 15–48.
- Hare, M.L., M. Ford and W.D. Marslen-Wilson. 2001. Ambiguity and frequency effects in regular verb inflection. In J. Bybee and P.J. Hopper (Eds.) *Frequency and the emergence of linguistic structure*. Amsterdam/Philadelphia: John Benjamins. pp. 181–200.
- Heine, B. 2003. Grammaticalization. In B.D. Joseph and R.D. Janda (Eds.) *The handbook of historical linguistics*. Malden/Oxford/Melbourne/Berlin: Blackwell. pp. 575–601.
- Kirsten, J. 2013. Totius en die standaardisering van vroeg-moderne Afrikaanse werkwoorde. *Tydskrif vir Geesteswetenskappe* 53(1): 60–75.
- Kirsten, J. 2015. The use of was in Afrikaans passive constructions: a diachronic corpus study. *Southern African Linguistics and Applied Language Studies* 33(2): 159–170.
- Langacker, R.W. 2010. Cognitive Grammar. In D. Geeraerts and H. Cuyckens (Eds.) *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press. pp. 421–462.
- Leech, G., M. Hundt, C. Mair and N. Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University press.
- Mair, C. 2006. *Twentieth-century English: History, variation and standardization*. Cambridge: Cambridge University Press.

- Patard, A. 2014. When tense and aspect convey modality. Reflections on the modal uses of past tenses in Romance and Germanic languages. *Journal of Pragmatics* 71: 69–97.
- Pierrehumbert, J.B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee and P.J. Hopper (Eds.) *Frequency and the emergence of linguistic structure*. Amsterdam/Philadelphia: John Benjamins. pp. 137–157.
- Ponelis, F.A. 1979. *Afrikaanse Sintaksis*. Pretoria: Van Schaik.
- Ponelis, F.A. 1993. *The Development of Afrikaans*. Frankfurt am Main/Berlin/Bern/New York/Paris/Wien: Peter Lang.
- Rayson, P. 2015. *Log-likelihood and effect size calculator*. Available online: [ucrel.lancs.ac.uk/llwizard.html](http://ucrel.lancs.ac.uk/llwizard.html) (Accessed 12 August 2015).
- Reppen, R., S.M. Fitzmaurice and D. Biber. 2002. *Using Corpora to Explore Linguistic Variation*. Amsterdam/Philadelphia: John Benjamins.
- Rossouw, R. and B. Van Rooy. 2012. Diachronic changes in modality in South African English. *English World-Wide: A Journal of Varieties of English* 33(1): 1–26.
- Steyn, J.C. 1976. 'n Paar sintaktiese onvasthede in Afrikaans. *Taalfasette* Jun/Jul: 25–53.
- Thompson, S.A. and P.J. Hopper. 2001. Transitivity, clause structure, and argument structure: Evidence from conversation. In J. Bybee and P.J. Hopper (Eds.) *Frequency and the emergence of linguistic structure*. Amsterdam/Philadelphia: John Benjamins. pp. 27–60.
- Tognini Bonelli, E. 2010. Theoretical overview of the evolution of corpus linguistics. In M. McCarthy and A. O'Keeffe (Eds.) *The Routledge handbook of corpus linguistics*. London/New York: Routledge. pp. 14–27.
- Van der Auwera, J. and V.A. Plungian. 1998. Modality's semantic map. *Linguistic Typology* 2: 79–124.