"Ek sal en jy gaan." The interaction between the modal auxiliaries sal en gaan

Roné Wierenga

Virtual Institute for Afrikaans (VivA) & Afrikaans and Dutch, North-West University, South Africa
E-mail: rone.pink@gmail.com

Abstract

Traditionally *sal*, *wil*, *moet* and *kan* are regarded as the most common modal auxiliaries in Afrikaans. Recent literature has acknowledged that *gaan*, a lexical verb denoting 'movement away from a deictic centre', has acquired modal functions resembling those of *sal* in specific contexts. This article aims to explore the development of the grammatical functions of *gaan* and the interchangeability of *sal* and *gaan* as modal auxiliaries within the scope of grammaticalisation theory. A corpus study is undertaken to identify and analyse the nuances in the modal functions for which these auxiliaries compete when expressing modality in order to establish the extent to which *gaan* has grammaticalised. Similarly, the degree to which *gaan* and *sal* can be used as synonyms is explored in the formal register of the Taalkommissie's newspaper sub-corpus and the informal register of the News Commentary corpus.

Keywords: Modality; Afrikaans modal auxiliaries; grammaticalisation; corpus-based research; usage-based research.

1. Introduction[1]

Bybee, Perkins and Pagliuca (1994:176) define modality as "the grammaticalisation of speakers' (subjective) attitudes and opinions". Kiefer (1997:242) states that modality indicates whether or not a state or event is possible, thus considering modality as "the relativization of the validity of sentence meanings to a set of possible words". Within this scope, modality can be viewed as a semantic notion expressed by grammatical constructions (Nuyts 2016:1).

Much attention has been paid to modality as a theme in Afrikaans linguistic publications (see among others Breed 2012; Conradie 2016, 2018; Kirsten 2018; Erasmus 2019). Conradie (2018:261) states that the Afrikaans modal verbs are auxiliary verbs that can qualify and

---

[1] This article was submitted for publication in January of 2021 and is based on an assignment submitted as part of the course requirement for a corpus linguistics module completed in 2019 at the North-West University's Potchefstroom Campus. I would like to thank the faculty members, Prof Adri Breed and Dr Lande Botha for their patience, support and diligence which has led to the completion of this article. I would also like to thank my former classmate Petri Swanepoel who worked alongside me to complete the initial assignment and whose findings form part of the foundations upon which this article was developed.

relativise lexical verbs, other modal verbs, and even a proposition in its entirety. Traditionally, four main groups of modal auxiliary verbs are distinguished:

(i)   *durf* 'dare*', *kan* '*can', *mag* 'may', *moet* 'must', *sal* 'shall'[2], and *wil* 'will';
(ii)   *behoort* 'should' and *hoef* 'have to';
(iii)   *blyk* 'looks like*', *skyn* '*seems like*', *dien* 'serves to*', *meen* 'of the opinion that' and *wens* 'wish to'; and
(iv)   the modal particle *te*

(Ponelis 1979:246)

This traditional distinction does not recognise *gaan* 'go' as a modal auxiliary verb, even though *gaan* 'go' can function as a modal auxiliary in specific contexts in Afrikaans, as illustrated in example (1).

(1)   *Dit gaan Woensdag   reën.*
        it   go   Wednesday rain
        It is going to rain on Wednesday.

In this example, *gaan* 'go' does not function as the main verb, but rather as a modal verb expressing the speaker's certainty that it will rain. From example (1), it is clear that *gaan* 'go' can be used as a modal auxiliary in certain contexts in Afrikaans and should therefore be added to the traditional list of modal auxiliary verbs. The modal functions of *gaan* 'go' are recognised in more recent linguistic publications (see Breed 2012; Conradie 2018, 2020; Erasmus 2019:565, among others) but have not yet been fully explored. Relatively little has been published about *gaan* as a modal auxiliary. Of the available publications, Conradie most commonly accepts and includes *gaan* in research on modality in Afrikaans (see Conradie 2018, 2020).

Conradie (2018:262) describes *gaan* as a modal auxiliary often viewed as synonymous with *sal*, hence competing with *sal* for modal functions. This article places the focus on the verb phrases in which the modal auxiliaries *sal* and *gaan* are interchangeable in order to establish to what extent, and in which contexts, these modal auxiliaries compete for modal functions.

The rest of the article is structured as follows: Section 2 presents a brief outline of grammaticalisation theory as a means of studying the development of grammatical verbs. Section 3 outlines corpus research as a methodology for investigating the contexts and modal functions in which *sal* and *gaan* compete. The last section presents the findings of the corpus analysis and an interpretation of the data.

2.   Grammaticalisation theory and the modal auxiliaries *sal* and *gaan* in Afrikaans

This study is situated within the theoretical framework of grammaticalisation theory. This is a suitable framework for the analysis of *sal* and *gaan* as modal auxiliary verbs because it aims to describe the change that lexical verbs undergo to obtain grammatical functions, or that

---

[2] Directly translated, *sal* is likened to 'shall'; however, semantically *sal* correlates with the modal functions of 'will' (as discussed in section 2). For the purpose of analyzing the modal functions of *sal*, 'will' will be used as a translation throughout this article, because it supports a more thorough understanding of *sal* as a modal auxiliary than 'shall' does.

grammatical words undergo to acquire new grammatical functions (Bybee et al. 1994; Hopper & Traugott 2003; Lehman 2015).

Grammaticalisation theory is a multifaceted theory that links well with cognitive grammar and is useful in functional usage-based approaches to language analysis (see Breed 2012; Conradie 2018; Kirsten 2018; Erasmus 2019). Grammaticalisation theory explains how lexical items evolve to acquire grammatical meanings (in the case of *gaan*, as will be shown) and how grammatical items take on more grammatical meanings (in the case of *sal*).

In short, grammaticalisation is a generalisation process that is accompanied by semantic bleaching and an increase in frequency as the item that undergoes grammaticalisation abandons its original lexical (or less grammatical) meaning in order to become useful in a wider variety of contexts (Hopper & Traugott 2003:1). The grammaticalisation process is unidirectional and multidimensional. This means that grammaticalisation takes place according to specific developmental routes by following a specific order. All constructions that undergo grammaticalisation do not necessarily go through or complete all the phases of grammaticalisation (Bybee et al. 1994:5). This is also the case with the grammaticalisation of *gaan* in Afrikaans, as demonstrated in examples (2) and (3).

(2)   *Ek gaan nou huis   toe.*
        I   go   now home to
        I am *going* home now.

(3)   *Ek gaan vanaand huiswerk   doen.*
        I   go   tonight   homework do
        I *will* do homework tonight.

Lexical *gaan* 'to go' is a main verb in Afrikaans that indicates movement, more specifically movement away from a deictic centre (as illustrated in example [2]). In example (3), the grammatical meaning of *gaan*, namely, to act as a future tense marker, is expressed. Hopper and Traugott (2003:3) state that verbs that express movement, obligation, or need often grammaticalise into future markers in Germanic languages. The lexical meaning of an item thus determines the grammaticalisation process as well as the grammatical functions that the grammaticalised item can assume. It cannot, however, be stated that all verbs of movement, obligation, or need will evolve into future markers, but only that these verbs have the potential for this development.

The grammaticalisation process may be accompanied by other phenomena, including phonological reduction, dependence on surrounding constructions, and competition. Competition arises when an item undergoing grammaticalisation assumes the same grammatical functions as an existing or established grammatical marker (Bybee et al. 1994:6). This is the case with *gaan* and *sal*. The increasing semantic shifts that take place during the grammaticalisation process give rise to the fact that *gaan* began assuming the grammatical functions of *sal,* resulting in the interchangeability of these two modal auxiliaries in certain contexts. This is illustrated in examples (4) and (5).

(4)   a. *Dit sal   hopelik   môre   reën.*
        it shall hopefully tomorrow   rain
        It *will* hopefully rain tomorrow.

       b. *Dit gaan   hopelik   môre   reën*
        it   go   hopefully tomorrow   rain
        It *will* hopefully rain tomorrow.

(5)   a. *Ja, ek sal   vanaand my huiswerk doen.*
        yes, I   shall tonight   my homework do
        Yes, I will do my homework tonight.

       b. *Ja,   ek gaan vanaand my huiswerk doen.*
        yes, I   go   tonight my homework do
        Yes, I *will* do my homework tonight.

In example (4a,b), *sal* and *gaan* are both future tense markers,[3] indicating that it will rain in the future. In the examples in (5a,b), both auxiliary verbs express epistemic modality, thus indicating the speaker's certainty of the proposition (that he/she will indeed do their homework that evening). In both cases, *sal* and *gaan* are syntactically interchangeable. The choice of auxiliary modal influences the semantic nuances of the sentence, as indicated in section 3.1.

Sal is, according to De Villiers (1971:87), one of the most functional modal auxiliary verbs in Afrikaans due to its usefulness in a variety of contexts including: i) In its impersonal form, *sal* can be used synonymously with *moet* 'must' (as in example [6]); and ii) *sal* contains the inherent core meanings "intention" or "insistence" which, in some contexts, make it interchangeable with *wil* 'will' (as in example [7]).

(6)   a. *Hy eis   dat sy moet gaan.*
        he demands that she must go
        He demands that she *must* go / He demands that she leave.

                                                                    [De Villiers 1971:88]

       b. *Hy eis   dat sy sal gaan.*
        he demands dat she shall go
        He demands that she *will* go / He demands that she leave.

(7)   a. *Ek wil dit doen.*
        I   will it do
        I *will* do it.

                                                                    [De Villiers 1971:88]

       b. *Ek sal   dit doen.*
        I   shall it   do
        I *will* do it.

Fischer (2007:182) presents the prototypical grammaticalisation path for the development of modal auxiliary verbs as shown in (8).

---

[3] Some approaches regard all future tense markers as markers of modality (Celle 2004; Salkie 2010).

(8)   ***Lexical verb > base modal > epistemic modal > subordinate/subjunctive***

Heine (1993:67) provides a summary of the most common typological findings on the grammaticalisation paths of auxiliary verbs. Only the remarks that apply to the grammaticalisation of the modal auxiliary verbs *sal* and *gaan* are presented here, together with the relevant typological sources from which Heine (1993) obtained this information.

  i) An agent-oriented modality marker (a deontic modality marker) can develop into a future time marker (Fleischman 1982; Bybee et al. 1994).

  ii) Possible aspectual markers can develop into future time markers (Dik 1987).

  iii) Future time markers tend to develop epistemic modality functions (Bybee & Pagliuca 1985; Bybee et al. 1994).

  iv) A future time marker also tends to assume speaker-oriented modality functions.

  v) A modality marker that reproduces ability tends to develop into a modal marker that expresses epistemic possibility.

From these findings the following adapted grammaticalisation path in (8') can be deduced.

(8')   *Agent-oriented modality > prospective > future time marker > epistemic / speaker-oriented modality*

This path can be simplified to the following in (8").

(8")   *Agent-oriented modality > tense > epistemic / speaker-oriented modality*

In relation to these grammaticalisation routes, it is clear that *gaan* has grammaticalised from a main verb with lexical meaning, to a future tense marker, to a modal auxiliary with epistemic modal functions. Despite this development, *gaan* has not been completely dissociated from its lexical and time meanings, but rather occurs in all three forms in Afrikaans simultaneously.

In contrast, *sal* has no lexical meanings, and therefore evolved from an agent-oriented modality marker to a future tense marker and finally to a modal auxiliary expressing epistemic modality.

Due to the generalization in meaning that both *gaan* and *sal* have undergone – and are possibly still undergoing – in acquiring new grammatical meanings, an increase in the frequency of use of these words is expected in the linguistic data. Since semantic bleaching requires that a grammaticalised unit distances itself from its lexical meanings to become more versatile, grammaticalised items can be used in more contexts (Bybee et al. 1994:9). The expectation is therefore that the linguistic data will show that *sal* can be used as a modal auxiliary in more contexts than *gaan* because it does not have lexical meanings. It is also expected that *sal* will occur more frequently because it is not context-bound by lexical meaning like *gaan* is. Bybee et al. (1994:9) states that grammatical items are often used not only in contexts where their meaning is essential, but also in contexts where the speaker's need requires them. It is thus possible for two or more existing constructions to perform the same grammatical functions.

3.   Modalities

In analysing the competition between *sal* and *gaan* with regards to modal functions, it is necessary that modality is defined. There is no consensus amongst linguists as to the types of modal meanings that should be distinguished (compare, inter alia, Palmer 1977; Frawley 1992; Van der Auwera & Plungian 1998; Traugott & Dasher 2002; Erasmus 2019). It is not the focus of this article to re-evaluate the modality types or to identify and assess the reasonings behind various classifications of the types of modalities. Therefore, the three most commonly accepted types of modalities will be used as the classifications according to which the modal functions of *sal* and *gaan* will be analysed.

Valin (as quoted by Siewierska 1991:123) distinguishes three classes for modality, namely: inherent modality, objective modality, and epistemological modality. Wärnsby (2006:21) divides these three modality types between two overarching modal categories, namely: subject- and discourse-oriented modality. That is, some modalities apply to both the speaker as well as the person being addressed, or rather the discourse as a whole, while other modalities apply exclusively to the speaker. These approaches allow three modal types to be distinguished, namely dynamic modality, which is subject-oriented, and deontic and epistemic modality, which are discourse-oriented.

3.1   Epistemic modality

Epistemic modality (also known as propositional modality) refers to the "faith" of the speaker in what is being uttered, namely the proposition (Huddleston & Pullum 2002:178). The speaker assesses and positions him-/herself in terms of their belief in the "truth" or "authenticity" of the proposition presented. The extent to which the speaker believes in the credibility of the utterance/proposition is therefore epistemic in nature. Despite the fact that epistemic modality does take into account the attitude of the speaker, it is primarily a question of whether the proposition expressed by the utterance is true or false. The external conditions are of secondary importance to the speaker (Traugott 2011:384). Consider examples (9-11).

(9)   *Ek gaan moontlik   more   klas   toe kom.*
         I   go   possibly tomorrow class to   come
         I am possibly going to come to class tomorrow.

(10)   *Waarskynlik sal   hy nie sterf   aan 'n verkoue nie.*
         probably   shall he not die   of   a cold   not
         He will probably not die of a cold.

(11)   *Hy sal   verseker nie met 'n gebreekte been kan dans   nie.*
         he shall certainly not with a broken   leg   can dance not
         He *will* certainly not be able to dance with a broken leg.

Epistemic modality relies on the speaker's knowledge of the proposition or the broader context in which the proposition is uttered. Traugott (2011:384) identifies the following meaning parameters for epistemic modality: i) possibility, ii) probability and iii) derived certainty. These parameters can be divided into five categories based on the following question: Is it impossible,

possible, hypothetical, probable or assured (provable)? The following questions are constructed to test for each of these five categories:

- Is this proposition possible?
- Is there a degree of doubt in the proposition?
- Can the proposition be substantiated by provable facts?
- Can the proposition be guaranteed?
- Does the proposition express a wish or desire?
- Is a prediction made?
- Is the proposition hypothetical?

3.2   Deontic modality

Deontic modality, like epistemic modality, is discourse-oriented. Deontic modality emphasizes the degree to which a force is applied to the subject of a sentence. This force aims to cause the subject to perform a certain action. Most commonly, the power of the proposition is expressed by the speaker. In the case of laws or rules, however, this power can also be exercised by an unidentified third party. The speaker is thus the means through which the power emerges, but the person, institution, or authority from which the command comes is the deontic source (Huddleston & Pullum 2002:178).

Nuyts (2016:36) emphasizes, in particular, the moral desirability of the proposition in relation to the speaker. 'Morality' is loosely used in this context to include broader codes of conduct such as social norms, as well as the personal ethical and religious criteria of the speaker. The speaker serves as a deontic assessor, allowing participants to express different moral assessments of the same situation. Thus, a sliding scale can be used to express deontic orientation. This sliding scale ranges from absolutely morally essential, through varying degrees of desirable morality, to absolutely morally unacceptability (Nuyts 2016:36).

The meaning parameters of deontic modality are: i) obligation, ii) moral expectation and iii) consent (Traugott 2011:383; Palmer 2001). Deontic modality is mainly performative, because the addressee's behaviour is morally assessed and changed by the proposition (Wärnsby 2006:19). Note that the modification does not necessarily have to be performed for the utterance to be termed deontic as illustrated in examples (12-14).

(12)   *Moenie in die huis   hardloop nie!*
      do not   in the house run   not
      Do not run in the house!

(13)   *Daar mag nie in die dam geswem word nie.*
      there may not in die dam swam   not
      Swimming in the dam is not allowed.

(14)   *Kan ek asseblief by Lenie gaan kuier?*
      can I   please   at Lenie go   visit
      May I please go visit Lenie?

The difference between epistemic and deontic modality is that epistemic modality emphasises the speaker's commitment or confidence in the proposition, while deontic modality focuses on the consent or permission, which the speaker has in order to execute the proposition.

The subcategories for deontic modality are defined by the question: Is it obligatory, recommended, permissible or prohibited? The following questions are constructed to test for these categories:

- Is it compulsory?
- Is it enforceable?
- Is it recommended?
- Is it discouraged?
- Is it permissible?
- Must permission be obtained?
- Is it prohibited?

3.3   Dynamic modality

Dynamic modality, also known as inherent modality, is subject-oriented. In contrast to epistemic and deontic modality, dynamic modality deals exclusively with the physical abilities of the subject. Dynamic modality emphasizes the subject's ability or lack thereof to execute the action in the proposition. This ability can be internal or external to the subject, or even simply a wish or desire (Palmer 2001:10). Consider the examples in (15-17).

(15)   *Ek kan vinniger swem as   wat   my tyd   vandag was.*
        I   can faster   swim than what my time today   was
        I can swim faster than what my time today indicated.

(16)   *Hy wil volgende jaar weer   deelneem   aan die sportliga.*
        he will next   year again participate in   the sport league
        He wants to participate in the sport league again next year.

(17)   *Ek kan nie in die reën   swem nie.*
        I   cannot in the rain   swim not
        I cannot swim in the rain.

Dynamic modality encodes the intrinsic abilities of the subject, whether physical or mental, to perform an action. However, the agent's ability to perform the action may be limited by external circumstances (Nuyts 2016:34). This limitation is demonstrated in example (17). Although the agent in this scenario has the physical ability to swim, this ability is limited by the presence of rain which makes the agent reluctant to swim.

The meaning parameters of dynamic modality are: i) ability and ii) volition (or desire). Unlike epistemic and deontic modality, dynamic modality has no subclassifications beyond its parameters. The following questions are constructed to test for these parameters:

- Is it viable?
- Is there the capacity to do so?
- Does the speaker have the will to perform the action?
- Does the speaker have the intention of performing the action?
- Is a desire expressed?

4.   Corpus and data analysis

A thorough analysis of the grammatical functions of *sal* and *gaan* is underpinned by the parameters of a corpus analysis, as outlined below.

i) The frequencies of use of *sal* and *gaan* in the corpora are an indication of grammaticalisation. (Since *sal* has no lexical or inchoative uses, only the grammatical uses of *gaan* that express tense or modality will be compared with the usage frequencies of *sal*.)

ii) The relationship between the frequency of use of grammatical *gaan* versus lexical *gaan* can provide an indication of the usability of *gaan* as a grammatical marker in Afrikaans. The level of semantic bleaching that *gaan* has undergone can thus be established.

iii) Collostructional analyses[4], like a collexeme analysis and distinctive collexeme analysis, can provide insight into the types of main verbs that commonly occur with the modal auxiliaries *sal* and *gaan.*

iv) Finally, the modal meanings that *sal* and *gaan* express will be examined to determine which of these modal auxiliaries most commonly performs each function.

For the purposes of this study, two corpora are used, namely the *Taalkommissie* or Language Commission Corpus (TK-corpus) and the *NWU/Kommentaarkorpus 1.4* or News Commentary Corpus.[5] The TK-corpus is a stratified corpus that consists of a variety of genres. For the purpose of this study, the newspaper sub-corpus of the TK-corpus will serve as a source of formal Standard Afrikaans language examples. In contrast, the Commentary corpus is a spontaneous corpus that consists of the comments that newspaper readers make on two popular online newspapers' websites. Only the data from one of these online newspaper websites will be used in this study in order to ensure that the corpus sizes are comparable. The comparability of the corpora is set out in Table 1.

---

[4] Gries and Stefanowitsch (2004a; 2004b) developed the collostructional analyses as a statistical method of calculating the associations between constructions and the item that precedes or follows it. One of the techniques of a collostructional analysis is a collexeme analysis. Collexeme analyses are used to gain insight into a specific slot in a construction. These analyses are discussed in section 3.4.
[5] Corpora were obtained using the Virtual Institute for Afrikaans' (VivA) online corpus portal (VivA 2020).

Table 1. Comparability of the corpora

	TK-newspaper sub-corpus[6]	**Commentary sub-corpus**[7]
Corpus size (words)	16 537 677	14 748 603
Register	Formal	Informal
Text types	Newspaper articles Edited and censored	Commentary Unedited and uncensored
Corpus build	Stratified	Spontaneous

The TK-newspaper sub-corpus is significantly larger than the Commentary sub-corpus. It is therefore necessary that the results from these corpora are normalized to ensure that the findings are statistically comparable. Since the use of *gaan* as a search string will recall all the instances of *gaan* in the corpora (including the lexical and inchoative uses of *gaan),* provisions will be made during the normalization process to eliminate all instances wherein *gaan* does not express future tense or modality. This will ensure that the results for *gaan* are comparable with the results for *sal,* and that the findings are accurate.

As already mentioned, grammaticalised items occur more frequently in language. As such, it is expected that *sal* and *gaan* will occur frequently in the corpora. Due to time constraints, it was not viable to check each of these instances manually in order to classify the modal functions expressed. For this reason, the population proportion test from Select Statistic Services (2019) was used to determine the sample size required to ensure that the findings are representative of the entire population. This population proportion test determines how many instances of *sal* and *gaan*, respectively, have to be manually classified in order for the findings to be representative of the total number of instances of each auxiliary modal in each of the corpora.

The statistical calculations on Select Statistic Services (2019) determine the population proportions for *sal* and *gaan* in both corpora with 99% accuracy. According to these calculations, the sample size for each modal auxiliary in each corpus is 660 instances.

Wordsmith Tools' sampling function was used to extract the 660 appropriate (in other words grammatical) hits of *sal* and *gaan* in each of the corpora. Wordsmith's sampling function is programmed to randomly reduce the hits. The hits in both corpora were reduced twice to ensure that the samples were indeed random and representative of the corpora as a whole.

4.1   Overview of the corpus results

The 660 instances of *sal* and *gaan* in both corpora are controlled as follows: i) From the sample (660 hits) of *gaan* in the TK-corpus and the Commentary corpus respectively, the grammatical uses of *gaan* are separated from the lexical and inchoative uses of *gaan*; ii) The sample (660 hits) of *gaan* will be checked in both corpora, and inaccurate non-grammatical uses of *gaan* will be eliminated. These results are visually depicted in Figure 1.

---

[6] Forthwith referred to as the TK-corpus.
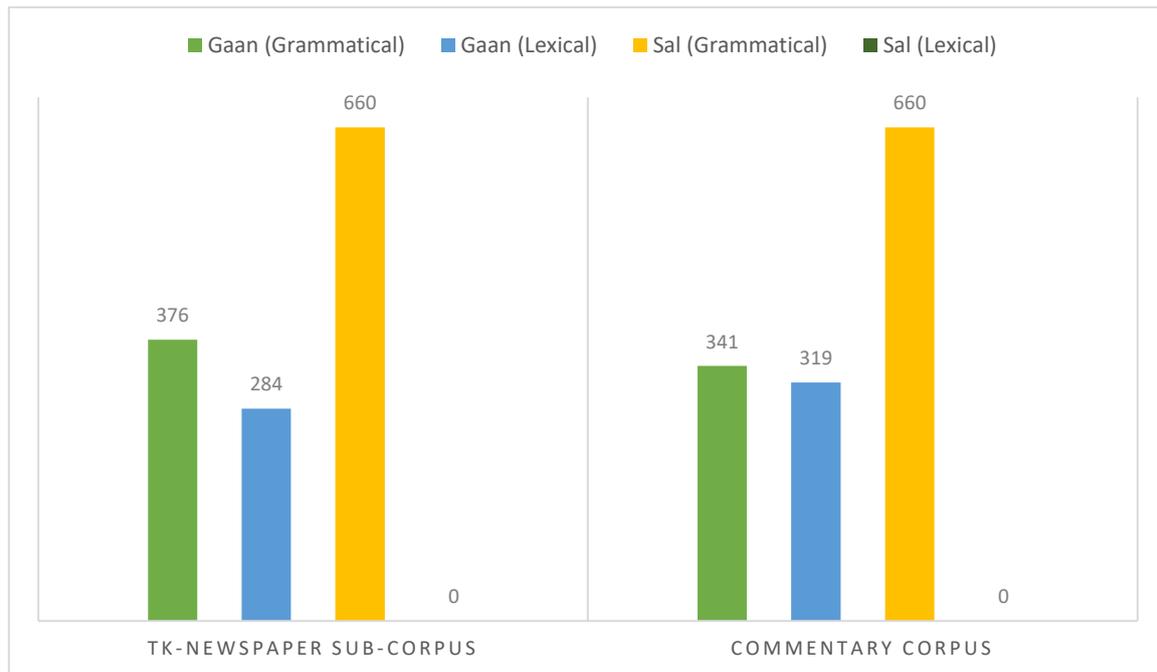[7] Forthwith referred to as the Commentary corpus.

Figure 1. Raw frequencies of *sal* and *gaan* in the samples

These findings will be discussed in the sections that follow.

The following ratios for lexical *gaan* versus grammatical *gaan* were obtained from the samples (660) for both corpora: i) TK-corpus 284:376, ii) Commentary corpus 341:319.

In the TK-corpus sample, 43% of the uses of *gaan* were lexical instances, compared to 52% lexical instances in the Commentary corpus. However, these results are influenced by the elimination of the inchoative instances of *gaan*. Since this study is not concerned with the inchoative functions of *gaan*, these instances were eliminated from the corpus findings. However, the expectation is that *gaan* will have more instances of grammatical functions than instances of its lexical use, should the inchoative instances be included.

Henceforth only the instances wherein *gaan* expresses modal meanings are included in the discussion.

The analysis of the corpus data for *sal* and *gaan* provided some difficulties. These trials influence the results presented in the following sections and are therefore outlined in this section. Therefore a few general remarks regarding the data analysis are in order: Some instances encountered in both corpora were ambiguous. These instances include i) cases where the modal auxiliary was used to pose a question, ii) cases where figurative language was used, and iii) cases where the intention of the speaker is unclear. These cases are illustrated in (18-23).

(18)   *Daar gaan dadels van kom.*
          there go   dates   of   come
          Nothing will come of it.

(19)   *Ek stem dit sal   bloedgrond   word.*
        I   agree it   shall bloodbath (bloodground) become
        It will become a bloodbath / tainted ground.

(20)   *Wie gaan dit keer?*
        who go   it   stop
        Who's going to stop it?

(21)   *Hoekom sal   jy   dan 'n oornagtassie   pak?*
        why   shall you then an overnight bag pack
        Why would you then pack an overnight bag?

(22)   *Ons sal   veel beter moet speel.*
        we shall a lot better must play
        We will have to play a lot better.

(23)   *Ek gaan stem.*
        I   go   vote
        I will vote. / I am going to vote.

Examples (18) and (19) demonstrate the use of the auxiliary modals in figurative language. Example (18) is an expression which directly translates to 'dates will come of this'. Figurative uses like these are troubling because the modal auxiliary expresses epistemic surety, yet the proposition (that an unspecified situation will result in dates being produced) is impossible, therefore contradicting the meaning expressed by the modal auxiliary. In examples (20) and (21), questions containing the modal auxiliaries are posed. The way in which these questions are phrased makes the meanings expressed by the modal auxiliaries ambiguous. Depending on the interpretation of these questions, *gaan* and *sal* can both be understood as future tense markers or modal auxiliaries in these examples. Lastly, *sal* in example (22) expresses modal meaning, however it is unclear which modal meaning is implied. Cases like these are difficult to classify using the test questions provided in section 2. Similarly, cases containing contextual references to places wherein *gaan* is the main verb (as is the case in example [23]), simultaneously express the speaker's epistemic commitment as well as the lexical meaning 'going'. In these cases, *gaan* can be grammatical or lexical depending on the pragmatic context.

These ambiguous cases cannot be undisputedly classified according to the parameters of this study. Therefore, these instances are accounted for in the data tables presented in the following sections. However, they are not included in the interpretation of the data.

4.2   TK-corpus

49 424 instances of *gaan* and 73 220 instance of *sal* were recorded in the TK-corpus. *Sal* occurred nearly twice as often as *gaan* in this corpus. Although noteworthy, this difference is not statistically significant. It is, however, interesting to note that, whilst *gaan* occurred much less frequently than *sal,* more than half of the instances of *gaan* included in the samples were cases wherein *gaan* functioned as a tense marker or modal auxiliary. Samples of each of these modal auxiliaries were made and the modal functions for *sal* and *gaan* are presented in Table 2.

Table 2. Frequencies of *sal* and *gaan* in the TK-corpus

Types of modality	Instances of *gaan* from sample (376)	Relative frequency of *gaan* (49 424)	Instances of *sal* from sample (660)	Relative frequency of *sal* (73 220)
Epistemic	305	81.11%	540	81.81%
Deontic	5	1.33%	43	6.52%
Dynamic	53	14.09%	66	10.00%
Questions	4	1.06%	3	0.45%
Unclear	7	1.86%	5	0.76%
Figurative	2	0.53%	3	0.45%

From Table 2, the following conclusions can be drawn. Firstly, epistemic modal uses are the most frequent form of modality in the TK-corpus, since the TK-corpus contains many publications that have been written in the formal register and reviewed by language editors it is possible that this preference for epistemic modality is related to the nature of the texts included in this corpus. Dynamic modality is the second most frequent modality found in the TK-corpus and deontic modality the least frequent modal function. The findings for each sub-class of the modal functions are presented in Table 3.

Table 3. Comparison of the modalities in the TK-corpus

Modalities	*Gaan* (363)	*Sal* (649)
Epistemic:	**305**	**540**
Impossible	30	31
Possible	63	85
Hypothetical	46	91
Probable	24	43
Certain	142	290
Deontic:	**5**	**43**
Obligatory	2	34
Recommended	2	5
Permissible	1	3
Prohibited	0	1
Dynamic:	**53**	**66**
Ability	5	11
Volition	48	55

The epistemic modal functions of *sal* (81.81%) are slightly more frequent than those of *gaan* (81.11%) in the TK-corpus. *Gaan* (14.09%) occurs more frequently in dynamic modal functions than *sal* (10.00%). For deontic modal functions *sal* (6.52%) seems to be preferred to *gaan* (1.33%). The preference for *sal* in deontic and epistemic modal functions is expected because the TK-corpus contains mostly edited and censored formal written language such as can be found in newspaper articles. Consider examples (24-29).

(24)   *En dit gaan aaklig   wees as die mense dan teruggaan op hul woord.*
          and it   go   horrible be   if the people then go back   on their word
          And it will be horrible if the people are going back on their word.   (Epistemic)

(25)   *Hulle mag maar gaan waar hulle wil.*
     they   may but   go   where they want
     They may go wherever they please.   (Deontic)

(26)   *Die jong   man gaan volgende jaar ingenieurswese studeer.*
     the young man go   next   year engineering   study
     The young man is going to study engineering next year.   (Dynamic)

(27)   *Die reguleerder sal   dus appélleer teen   die beslissing.*
     the regulator   shall thus appeal   against the blessings
     The regulator will appeal the judgement.   (Epistemic)

(28)   *Jy   sal   nie teruggaan nie.*
     you shall not go back   not
     You will not go back.   (Deontic)

(29)   *Ek sal   dit wil doen om 'n verskil   in mense se lewens te kan maak.*
     I   shall it will do   to a difference in people's lives   can make
     I will want to do it in order to make a difference in other people's lives.   (Dynamic)

It can, therefore, be deduced from these findings that *sal* occurs more frequently than *gaan* and expresses the majority of the modal functions in the TK-corpus. Although a distinction can be made between the semantic values of *sal* and *gaan,* allowing *gaan* to be preferred for dynamic modalities and *sal* for deontic modalities, there is no preference between these two auxiliaries for epistemic modal functions since the difference in relative frequency is very small (0.70%).

This observation is significant because epistemic modality is regarded as the final phase of the general grammaticalisation path for modal auxiliaries (refer to section 2). Both *sal* and *gaan* has thus completed the final phase of the grammaticalisation path and are equally grammaticalised in their modal uses. Yet, *gaan* is more grammatically constrained because it still retains lexical meaning, whereas *sal* has no lexical meanings and can, therefore, be used in a wider variety of contexts.

4.3   Commentary-corpus

A total of 88 220 instances of *gaan* and 73 230 instances of *sal* were identified in the Commentary corpus. It is striking that *gaan* has occurred considerably more frequently in the informal social Afrikaans that appears in the Commentary Corpus than in the formal Standard Afrikaans of the TK-corpus. This high frequency of use of *gaan* is all the more striking when one considers that significantly fewer grammatical uses of *gaan* were found in this sample than in the sample from the TK-corpus (341/660 versus 376/660). These results are summarised in Tables 4 and 5.

Table 4. Frequencies of *sal* and *gaan* in the Commentary-corpus

Types of modality	Instances of *gaan* from sample (341)	Relative frequency of *gaan* (88 220)	Instances of *sal* from sample (660)	Relative frequency of *sal* (73 230)
Epistemic	289	84.75%	542	82.12%
Deontic	3	0.88%	13	1.97%
Dynamic	9	2.64%	52	7.88%
Questions	32	9.38%	23	3.48%
Unclear	6	1.76%	15	2.27%
Figurative	2	0.59%	18	2.73%

Table 5. Comparison of the modalities in the Commentary-corpus

Modalities	*Gaan* (out of 301)	*Sal* (out of 604)
Epistemic:	**289**	**539**
Impossible	47	52
Possible	72	65
Hypothetical	45	150
Probable	10	52
Certain	115	220
Deontic:	**3**	**13**
Obligatory	2	5
Recommended	0	6
Permissible	0	0
Prohibited	1	2
Dynamic:	**9**	**52**
Ability	6	8
Volition	3	44

Sal is also used less frequently (89.24%) to express epistemic modality than *gaan* (96.01%) in the Commentary corpus. From this it can be deduced that epistemic modality is the most frequently occurring modal function in informal registers of Afrikaans and that the preferred modal auxiliary verb for epistemic modality in informal Afrikaans registers is *gaan*. *Sal* (2.64%) occurs more frequently in dynamic modal functions than *gaan* (7.88%). These findings contrast with those of the TK-corpus where *gaan* is more commonly used in dynamic modal functions.

(30)   *As   dit so onaangenaam gaan wees, sal   sy   nie weer   kom   nie.*
    if it this   unpleasant   go   be   shall she not again come not
    If it is going to be unpleasant, she will not come again.   (Epistemic: Informal)

It is interesting that *sal* seems to be used more commonly in informal registers of Afrikaans to express dynamic modality than *gaan* is, more so when taking into consideration that *gaan* is most frequently used to express volition in the TK-corpus. Therefore, *sal* seems to be the preferred modal auxiliary to express needs and desires in informal registers while *gaan* seems to be preferred in formal registers.

(31)   *Ons gaan dit in ons tweede uitgawe van 2007 publiseer.*
         we   go   it in our second edition   of   2007 publish
         We are going to publish it in our second edition of 2007.   (Dynamic: TK-korpus)

(32)   *Ons sal   onmiddellik daarmee begin.*
         we   shall immediately with it   begin
         We will begin immediately.   (Dynamic: Commantary corpus)

4.4   Collexeme and distinctive collexeme analysis

A collexeme analysis is a statistical analysis that measures the degree of the forces of attraction and repulsion between a lemma and a linguistic item that occurs in a particular construction. Collexeme analyses form part of the collostructional methods of language analysis developed by Gries and Stefanowitsch (2016). The purpose of the present collexeme analysis is to determine which verbs most commonly occur with *sal* and *gaan* in contexts wherein these auxiliary verbs perform modal functions. Knowledge of the main verbs that tend to collocate with each of these modal auxiliary verbs provides an indication of the syntactic and contextual semantic nuances in Afrikaans that result in one of these modal verbs being preferred over the other.

Gries and Stefanowitsch's (2016) collexeme analysis software was used to perform this calculation. The fifty strongest collocations of *sal* and *gaan* are highlighted and interpreted respectively. For this purpose, the results from the two corpora are combined. Only collocation strengths greater than 1.31 are considered significant, because these collocation strengths have a p-value less than 0.05 and are therefore statistically significant.

Certain tendencies for the collocations of *sal* and *gaan* can be deduced from an analysis of the top fifty strongest collocations of each modal verb. The following verb groups tended to collocate with *gaan:* i) main verbs that denote 'movement' as an inherent meaning, ii) verbs of negative experience or actions, iii) choice verbs, iv) verbs that denote moral aversion and v) verbs of creation.

At least half of the fifty strongest collocations with *gaan* were verbs that contain an inherent meaning that expresses movement. Examples of these verbs include: *reis* 'to travel', *aangee* 'to give/pass (sth)', and *wegkom* 'to get away'. In addition, 7 out of 50 words expressed negative experiences or actions, such as *doodskiet* 'to kill by shooting', *wegkom* 'to get away', *vang* 'to catch', *seermaak* 'to hurt' and *doodgaan* 'to die'. Some of these words, such as *seermaak* 'to hurt' and *doodskiet* 'to kill by shooting', can also be considered threats. Only three of the top fifty collocations were choice verbs such as *besluit* 'to decide', *kies* 'to choose' and *uitsoek* 'to select'. Likewise, only five of the top verb collocations are verbs of moral aversion, such as *rook* 'to smoke' and *dobbel* 'to gamble'. Finally, six of the fifty verb collocations are verbs of creation. These verbs include, among others, *skep* 'to create', *bou* 'to build' and *skryf* 'to write'.

Similar trends can be detected in the fifty strongest collocations of *sal*. Five verb groups are identified, including negative action verbs and verbs with the prefix *er-*. Thirteen of the fifty verbs that *sal* was found to have the strongest collocations with are verbs that can be considered negative actions. These words include *afbrand* 'to burn down', *klap* 'to slap', *oorspoel* 'to flood' and *benadeel* 'to harm'. In addition, five verbs containing the prefix *ver-* can be

identified. These verbs are *verwerp* 'to reject', *verruil* 'to exchange', *vermag* 'to accomplish', *verneem* 'to hear' and *vervang* 'to replace'.

In the collexeme analysis of *sal*, it is noticeable that this modal verb often strongly coincides with particle verbs, or prepositional verbs like **uit**loop 'to flow out', **terug**kyk 'to look back' and **af**lewer 'to deliver/drop off'. This tendency is attributed to the relationship between the modal auxiliary verb and the particle verb, namely that the use of the modal auxiliary verb *sal* requires the subsequent particle verb to occur in its combined form (Ponelis 1979:235). See examples (33) and (34) for illustration:

(33)   *Hy sal   die boeke aflewer.*
        he shall the books deliver
        He *will* deliver the books.

(34)   *Die maatskappy lewer   die boeke af.*
        the company   deliver the books off
        The company delivers the books.

In example (33) the preposition and verb combine to form a single word unit following the modal auxiliary verb *sal.* In contrast, the absence of the modal auxiliary verb in example (34) allows for a sentence structure wherein the preposition and the verb occur separately. In cases like these where the preposition and the verb occur separately in the sentence, they are still considered to function as a unit. The modal verb therefore requires the unit form of the prepositional verb. This is confirmed by the absence of the verb form without the preposition in the collocation lists for *sal*.

The collexeme analysis also indicated strong repulsive forces between some main verbs and *sal* and *gaan*, specifically with main verbs with the past tense prefix *ge-*. Some examples include *geslaap* 'slept', *gebad* 'bathed' and *gekneus* 'bruised'. These results are indicative of the fact that modal verbs capable of functioning as future tense markers cannot occur in constructions denoting the past tense in their future tense forms. It is therefore syntactically marked to state: *Ek sal gister gegaan het* 'I will have gone yesterday'. In such cases, the past tense form of the modal verb is required, as example (35) illustrates.

(35)   *Ek sou   gister   gegaan het.*
        I   would have   yesterday gone   had
        I would have gone yesterday.

In addition to the collexeme analysis, a distinctive collexeme analysis was also performed. A distinctive collexeme analysis is a statistical calculation that forms part of the collostructional methods of analysis and indicates whether a verb is more likely to occur with *sal* than with *gaan,* and vice a versa. A distinctive collexeme analysis, unlike a collexeme analysis, does not indicate the words a construction is likely to occur with, but rather with which modal auxiliary a verb is most likely to occur. This analysis was also performed using Gries and Stefanowitsch's (2016) calculators. The thirty strongest collocations for *sal* and *gaan* are presented in Table 6.

Table 2. Main verb collocations for sal and gaan in both corpora

Gaan		Sal	
Verb	**Collocation strength**	**Verb**	**Collocation strength**
SE Sê?	295.63	WEET	273.63
HAAL	216.23	KAN	238.61
HE Hê?	172.65	WEES	194.09
KOM	169.4	SIEN	159.88
GESE GESê?	151.07	GAAN	56.26
GEBEUR	149.44	HELP	54.08
LE Lê?	132.34	STEM	50.54
SOEK	272.01	BID	46.59
SPEEL	239.84	BESKERM	44.18
KYK	484.17	DINK	41.25
KUIER	109.54	OPHOU	37.02
STUDEER	111.57	BLY	35.14
GEE	76.15	AANHOU	32.74
DRUK	132.34	HOOR	31.85
GLO	66.78	HERSTEL	31.36
LYK	51.71	DRA	30.17
STYG	49.27	DOEN	27.45
SIT	320.06	LEER	26.8
OPTREE	46.83	GENEES	26.38
LEWE	40.31	SAAMSTAAN	23.95
BESLUIT	37.46	VERGEET	22.3
GROEI	37.46	KRY	19.99
KLA	35.02	AGTERKOM	19.42
SLAAP	35.02	SKRIK	18.59
KOOP	97.32	VOORSIEN	18.2
KON	177.13	VAT	16.6
UITVIND	29.73	OOPGAAN	14.98
GESELS	29.32	VERGEWE	14.87

The results from the distinctive collexeme analysis indicate that verbs denoting stativity or states, like *weet* 'to know', *wees* 'to be' and *genees* 'to heal', prefer to occur with *sal* as opposed to *gaan*. This includes words expressing emotions or emotional states, such as *skrik* 'to have a fright'. Verbs indicative of a transfer, such as *kry* 'to get', *vat* 'to take', *haal* 'to fetch' and *dra* 'to carry', tend to prefer collocations with *sal* rather than with *gaan*. Likewise, mental or cognitive verbs such as *dink* 'to think', *agterkom* 'to notice', and *vergeet* 'to forget' prefer to collocate with *sal*. Finally, sensory verbs, like *sien* 'to see' and *hoor* 'to hear', and abstract verbs, which do not reflect the action that accompanies it in the meaning of the word, such as *saamstaan* 'to stand together/join forces', *help* 'to help' and *herstel* 'to repair', collocate with *sal* rather than with *gaan*.

Although these verbs prefer to collocate with *sal*, the difference in the collocational strength of the main verbs that collocate with *sal* and *gaan* respectively is noteworthy. Only four of the words that prefer to collocate with *sal* have a collocation strength higher than 150. The remaining 26 verbs all have a collocation strength lower than 60. In contrast, 11 main verbs that

prefer to collocate with *gaan* have a collocational strength higher than 150 and only 10 verbs have a collocation strength lower than 50. These results indicate that these verbs have much stronger relational forces of attraction with *gaan* than with *sal*, but that they more frequently occur within user-based language with *sal*.

Sal is thus more commonly used in Afrikaans to express modal meanings, but it is not as strongly connected to the main verb of the sentence as *gaan* is. This supports what is predicted by grammaticalisation theory. *Sal* is thus more grammaticalised than *gaan* because it does not have lexical meaning and can, therefore, be used in a wider variety of contexts than *gaan,* while the retention of the lexical meaning of *gaan* ensures that main verbs have stronger collocational strengths with *gaan* in general.

5.   Conclusion

This article explored the interaction and competition between *sal* and *gaan* for modal functions within the scope of grammaticalisation theory. The results from the corpus study indicate that the modal uses of *sal* and *gaan* are greatly influenced by register. Although *gaan* occurred less frequently that *sal* as a marker of modality, it is the preferred modal auxiliary in the informal register included in the Commentary Corpus. It is therefore remarkable that *gaan* is very rarely used to denote deontic modality in either corpus, especially considering that deontic modality is often used in discussions of rules or laws – themes that are expected to occur often in newspaper articles. It seems that *sal* is more readily used in these contexts than *gaan*.

In contrast, *sal* occurs much more often in both corpora than *gaan*. This tendency is attributed to the fact that *gaan* still retains some of its lexical and inchoative uses, while *sal* exclusively expresses grammatical meaning. Although *sal* is generally more commonly used in modal functions than *gaan,* there seems to be a noticeable preference for *gaan* in the informal register of the Commentary corpus. This finding corresponds with those of Kirsten (2018) and Erasmus (2019).

The collexeme and distinctive collexeme analyses prove that *gaan* has stronger verb collocations overall, indicating that *sal* has fewer usage restrictions and is, therefore, more grammaticalised than *gaan*. The collexeme analyses also indicated that *sal* often occurs with prepositional verbs, because *sal* requires the prepositional verb to occur as a single word unit. In contrast *gaan* does not have the same syntactic requirement.

It was further found that both *sal* and *gaan* generally collocate with verbs denoting negative actions or experiences, and moral disapproval. Finally, *gaan* also collocates strongly with verbs that contain an inherent meaning denoting movement, because these verbs are suggestive of the lexical meaning of *gaan,* namely 'to go'.

References

Breed, A. 2012. *Die grammatikalisering van aspek in Afrikaans: 'n Semantiese studie van perifrastiese progressiewe konstruksies.* Potchefstroom: Noordwes-Universiteit.

Bybee, J.L. & W. Pagliuca. 1985. Cross-linguistic comparison and the development of grammatical meaning. In J. Fisiak (ed.), *Historical semantics, historical word formation*. The Hague: Mouton de Gruyter. pp. 59-83. https://doi.org/10.1515/9783110850178.59

Bybee, J.L., R.D. Perkins & W. Pagliuca. 1994. *The evolution of grammar: tense, aspect and modality in the languages of the world.* Chicago: University of Chicago.

Celle, A. 2004. The French future tense and English *will* as markers of epistemic modality. *Languages in Contrast,* 5(2): 181-218. https://doi.org/10.1075/lic.5.2.02cel

Conradie, C.J. 2016. Willens en wetens: Perspektiewe op die Afrikaanse werkwoord *wil. Tydskrif vir Geesteswetenskappe,* 56(1): 7-24. https://doi.org/10.17159/2224-7912/2016/v56n1a2

Conradie, C.J. 2018. Modale kettings in Afrikaans. *Tydskrif vir Geesteswetenskappe,* 58(2): 258-275. https://doi.org/10.17159/2224-7912/2018/v58n2a4

Conradie, C.J. 2020. *Modality*. https://taalportaal.org/taalportaal/topic/pid/topic-20170328102748593. [Accessed 11 November 2020]

De Villiers, M. 1971. *Die grammatika van tyd en modaliteit*. Kaapstad: Balkema.

Dik, S.C. 1987. Copula auxiliarization: how and why. In M. Harris et al. (eds.), *Historical development of auxiliaries*. De Gruyter Mouton. pp. 53-84.

Erasmus, C.J. 2019. *Op soek na 'n omvattende tipologiese beskrywing van modaliteit: die Afrikaanse modale sisteem*. Potchefstroom: North-West University.

Fleischman, S. 1982. *The future in thought and language. Diachronic evidence from Romance*. Cambridge Studies in Linguistics London.

Fischer, O. 2007 *Morphosyntactic change. Functional and formal Perspectives*: Oxford: Oxford University Press.

Frawley, W. 1992. *Linguistic Semantics*. New York: Routledge, Taylor & Francis.

Gries, S.T. & A. Stefanowitsch. 2020. *Collostructional analysis*. http://www.stgries.info/teaching/groningen/index.html. [Accessed 11 November 2020]

Gries, S.T. & A. Stefanowitsch. 2004a. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics,* 9(1): 97-129. https://doi.org/10.1075/ijcl.9.1.06gri

Gries, S.T. & A. Stefanowitsch. 2004b. Covarying collexemes in the Into-causative. In Achard, M. & S. Kemmer (eds.), *Language, Culture, and Mind*. Stanford: CSLI Publications. pp. 225-236.

Heine, B. 1993. Auxiliaries: Cognitive forces and grammaticalization. New York: Oxford University Press.

Hopper, P.J. & E.C. Traugott. 2003. *Grammaticalisation*. Cambridge: Cambridge University Press.

Huddleston, R.D. & G.K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316423530

Kiefer, F. 1997. Modality and pragmatics. *Folia Linguistica* XXXI/3-4: 241-253.

Kirsten, J. 2018. "Dit gaan nog belangrik word": Veranderinge in grammatikale toekomsverwysing in Afrikaans. *Tydskrif vir Geesteswetenskappe,* 58(2): 277-292. https://doi.org/10.17159/2224-7912/2018/v58n2a5

Lehman, C. 2015. *Thoughts on grammaticalisation*. Berlin: Language Science Press.

Nuyts, J. 2016. Surveying modality and mood: An introduction. In J. Nuyts & J. Van der Auwera (eds.), *The Oxford handbook of modality and mood*. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199591435.013.37

Palmer, F.R. 1977. Modals and actuality. *Journal of Linguistics,* 13: 1-23. https://doi.org/10.1017/S002222670000517X

Palmer, F.R. 2001. *Mood and modality*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139167178

Ponelis, F.A. 1979. *Afrikaanse sintaksis*. Kaapstad: J.L. van Schaik.

Salkie, R. 2010. *Will*: tense or modal or both? *English Language and Linguistics,* 14(2): 187-215. https://doi.org/10.1017/S1360674310000055

Select Statistic Services. 2019. *Population Proportion – Sample Size*. https://select-statistics.co.uk/calculators/sample-size-calculator-population-proportion/. [Accessed 30 August 2019]

Siewierska, A. 1991. *Functional grammar*. London: Routledge.

Traugott, E.C. 2011. Modality from a historical perspective. *Language and Linguistics Compass,* 5(6): 381-396. https://doi.org/10.1111/j.1749-818X.2011.00280.x

Traugott, E.C. & R.B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486500

Van der Auwera, J. & V.A. Plungian. 1998. *Linguistic Typology*. Berlin: Walter de Gruyter.

VivA. 2020. *Korpusportaal*. http://korpus.viva-afrikaans.org/whitelabversoek/explore/corpus. [Accessed 11 November 2020]

Wärnsby, A. 2006. *(De)coding modality. The case of must, may, måste and kan.* Lund: Lund University.