“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

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

1This 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 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.

http://spilplus.journals.ac.za
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 am going home now.

(3) *Ek gaan vanaand huiswerk doen.*
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).
In example (4a,b), *sal* and *gaan* are both future tense markers, 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.

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).
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 > base modal > epistemic modal > subordinate/subjunctive**

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 proposition/utterance 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 (Nuys 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, 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. 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

<table>
<thead>
<tr>
<th></th>
<th>TK-newspaper sub-corpus</th>
<th>Commentary sub-corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus size (words)</td>
<td>16 537 677</td>
<td>14 748 603</td>
</tr>
<tr>
<td>Register</td>
<td>Formal</td>
<td>Informal</td>
</tr>
<tr>
<td>Text types</td>
<td>Newspaper articles</td>
<td>Commentary</td>
</tr>
<tr>
<td></td>
<td>Edited and censored</td>
<td>Unedited and uncensored</td>
</tr>
<tr>
<td>Corpus build</td>
<td>Stratified</td>
<td>Spontaneous</td>
</tr>
</tbody>
</table>

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

Wie gaan dit keer?
who go it stop
Who’s going to stop it?

Hoekom sal jy dan ‘n oornagtassie pak?
why shall you then an overnight bag pack
Why would you then pack an overnight bag?

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

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

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

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

<table>
<thead>
<tr>
<th>Modalities</th>
<th><em>Gaai</em> (363)</th>
<th><em>Sal</em> (649)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epistemic</td>
<td>305</td>
<td>540</td>
</tr>
<tr>
<td><em>Impossible</em></td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td><em>Possible</em></td>
<td>63</td>
<td>85</td>
</tr>
<tr>
<td><em>Hypothetical</em></td>
<td>46</td>
<td>91</td>
</tr>
<tr>
<td><em>Probable</em></td>
<td>24</td>
<td>43</td>
</tr>
<tr>
<td><em>Certain</em></td>
<td>142</td>
<td>290</td>
</tr>
<tr>
<td>Deontic</td>
<td>5</td>
<td>43</td>
</tr>
<tr>
<td><em>Obligatory</em></td>
<td>2</td>
<td>34</td>
</tr>
<tr>
<td><em>Recommended</em></td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td><em>Permissible</em></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td><em>Prohibited</em></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dynamic</td>
<td>53</td>
<td>66</td>
</tr>
<tr>
<td><em>Ability</em></td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td><em>Volition</em></td>
<td>48</td>
<td>55</td>
</tr>
</tbody>
</table>

The epistemic modal functions of *sal* (81.81%) are slightly more frequent than those of *gaan* (81.11%) in the TK-corpus. *Gaai* (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)
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

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

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

<table>
<thead>
<tr>
<th>Modalities</th>
<th>Gaan (out of 301)</th>
<th>Sal (out of 604)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epistemic:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impossible</td>
<td>47</td>
<td>52</td>
</tr>
<tr>
<td>Possible</td>
<td>72</td>
<td>65</td>
</tr>
<tr>
<td>Hypothetical</td>
<td>45</td>
<td>150</td>
</tr>
<tr>
<td>Probable</td>
<td>10</td>
<td>52</td>
</tr>
<tr>
<td>Certain</td>
<td>115</td>
<td>220</td>
</tr>
<tr>
<td>Deontic:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obligatory</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Recommended</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Permissible</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Prohibited</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Dynamic:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Volition</td>
<td>3</td>
<td>44</td>
</tr>
</tbody>
</table>

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 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 are going to publish it in our second edition of 2007.*  
(Dynamic: TK-korpus)

(32) *Ons sal onmiddellik daarmee begin.*

*We will begin immediately.*  
(Dynamic: Commentary 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 *uitloop* ‘to flow out’, *terugkyk* ‘to look back’ and *aflewer* ‘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 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.
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

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


http://spilplus.journals.ac.za


