UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

Modal expressions in Xhosa, part I : Necessity

Crane, Thera Marie; Savi, Stefan; Slater, Onelisa; Bernander, Rasmus

2024-12
Musée royal de l'Afrique Centrale
http://hdl.handle.net/10138/589561
Crane, T M, Savi, S, Slater, O & Bernander, R 2024, 'Modal expressions in Xhosa, part I : Necessity', Africana Linguistica, vol. 30, pp. 39–70. https://doi.org/10.2143/AL.30.0.3293835

Downloaded from Helda, University of Helsinki institutional repository. https://helda.helsinki.fi
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.

Modal expressions in Xhosa, part I: Necessity

Thera Marie Crane, Stefan Savić, Onelisa Slater & Rasmus Bernander

This study reports on an investigation of modal expressions of necessity in Xhosa, a Bantu language of South Africa. Combining analysis of occurrences in a small corpus and the native-speaker intuitions of one of the paper's authors, we describe patterns and restrictions in the distribution of two markers of necessity, auxiliary-like **fanele** and **funeka**; we also give some notes on the necessity marker **mele**. The most significant differences in the meanings and uses of these markers are in the participant-internal and epistemic domains, and the lexical meanings (still active in Xhosa, in addition to the markers' modal functions) are evident in their usage distributions.

Keywords: corpus linguistics, modality, Nguni Bantu, necessity

Introduction

Modal expressions of necessity and possibility, long understudied in Bantu languages, have received more attention in recent years. Recent studies of Bantu modality are based on corpus work (e.g. Kawalya *et al.* 2021), targeted elicitation (e.g. Gluckman & Bowler 2020), or surveys of published material (e.g. Bernander *et al.* 2022). Devos (2008), a pioneering work in the study of modal expressions in a Bantu language, employs a corpus that is partially made up of elicited sentences. In this study, we make use of all these resources – descriptive and documentary literature, a small, curated corpus, and native-speaker intuitions – to draw the first detailed picture of key expressions of necessity and possibility in Xhosa, and to suggest avenues of further targeted research. While the corpus sheds light on the overall distribution patterns of each modal expression's uses, the native-speaker intuitions of the third author allow us to explore the limits of each expression's functional distribution. We can thereby determine with some confidence whether certain uses do not appear in our corpus because they are infelicitous, or simply because they are rare.

This is the first article of a two-part study. In this article, we give background to our study and discuss our findings for necessity expressions. In the second part (*Africana Linguistica* 31, forthc.), we describe Xhosa possibility expressions in detail.

Xhosa (Guthrie [1970] classification code S41; endonym **isiXhosa**) is a Bantu language spoken as an official language in South Africa, with home language speakers numbering at least eight million. Xhosa belongs to the Nguni branch of Southern-Bantu (Hammarström *et al.* 2024). Other official Nguni languages spoken in South Africa include Zulu (S42), "Southern" Ndebele (S407), and Swazi (S43). Nguni languages are part of the Southern Bantu group, which comprises mostly "Guthrie Zone S" languages spoken in South Africa, Lesotho, Eswatini, Botswana, and Mozambique. Southern Bantu belongs to the Eastern Bantu genealogical subgroup (Grollemund *et al.* 2015). While scattered descriptions exist of modal expressions in Zone S languages, modality has until recently rarely been studied as a unified phenomenon in these languages or in Eastern Bantu as a whole. Crane *et al.* (2024b) discuss Xhosa possibility markers as part of a comparative study of Nguni and Tsonga, but to our knowledge, no published literature has dealt comprehensively with modal expressions in Xhosa.

The necessity markers occurring most frequently in our corpus are **fanele** (1) and **funeka** (2).[1] Another marker, **mele** (3), occurred only a few times in the corpus, so we do not give a full analysis here, but we do make some notes about its form and function.

1. Throughout the article, we gloss all possibility markers as POSSIB and all necessity markers as NEC so as not to pre-impose an analysis on their functions. Corpus examples are attributed to the corpora from which they were originally drawn (see Section 2.2); glossing, analysis, and translations are our own. Other examples are marked with their source or as (our own) constructed examples.

(1) **U-<u>fanele</u> uku-khathazek-a**.
SP_{2SG}-NEC INF-become_worried-FV
'You should be worried.' (Snyman *et al.* 2012)

(2) **lo mntu ku-<u>funeka</u> enz-e ingxelo** …
DEM_1.PROX 1.person SP_{17}-NEC SP_1.do-SBJV 9.report
'this person must make a statement …' (Eiselen & Puttkammer 2014)

(3) **Ku-<u>mele</u> u-hamb-e ngoku**.
SP_{17}-NEC SP_{2SG}-go-SBJV now
'You should go now.' (constructed)

We use the modal typology laid out in van der Auwera & Plungian (1998), supplemented with additional (sub)categories, to investigate the distribution of modal meanings (or 'flavours'; see Section 1.1 for descriptions) with the modal expressions in the corpus. We compare this distribution with the third author's intuitions to round out the picture of the expressions' functional scopes. Because most of the modal expressions we explore are transparently derived or at least reasonably reconstructable, we also look at lexical 'persistence' in the terms of Hopper (1991), that is, whether the lexical and constructional sources of modal expressions have an influence on their modal functions.

The remainder of this article is organised as follows: In the following section, we define the concept of modality used in this article and describe the modal subcategories we used in coding our corpus data (Section 1.1); discuss patterns of grammaticalisation, syntactic and semantic change, and corresponding 'subjectification' within the modal domain, both from a typological perspective and as they relate to our study (Section 1.2); and survey previous research on modality in Southern and Eastern Bantu (Section 1.3). Section 2 describes our methodology and its benefits and limitations. Section 3 details our findings about the necessity markers **fanele** (Section 3.2), **funeka** (Section 3.3), and – briefly – **mele** (Section 3.4). The article concludes with a summary of our key findings and recommendations for further research.

1. Background

1.1. Modality: definition and taxonomy

We define modality simply as the expression of necessity and possibility, following typological studies such as van der Auwera & Plungian (1998). Modal expressions mark the "factual status" of the proposition as "undetermined" (Narrog 2012: 6). For example, the English statement **we might come over later** expresses modal possibility, and **you must be exhausted** expresses modal necessity. In neither statement is the proposition asserted; the propositions are respectively described as possibly true, or likely or presumed to be true, rather than asserted as true.

For our current purposes, our definition excludes the category of evidentiality, which we understand as marking the source of evidence for a proposition. Evidentiality

is even less studied in Bantu than is modality, although some research suggests that it may be a more significant category than typically assumed (Givón 1982; Botne 1995, 1997; Roth 2018; Gibson & van der Wal 2023; Crane *et al.* 2024a; cf. Botne 2020 for further discussion). However, at present we do not know of any dedicated evidential markers in Xhosa, nor have we uncovered any obviously evidential uses of the modal markers we investigated in our corpus study.

We took as a starting point the typologically oriented modal taxonomy put forth in van der Auwera & Plungian (1998). Van der Auwera & Plungian's taxonomy has frequently been adopted – sometimes in modified form – in studies of Bantu modality (e.g. Nurse & Devos 2019; Bernander *et al.* 2022).[2] Van der Auwera & Plungian's modal maps, drawing on data from Bybee *et al.* (1994), aim to show not only semantic connections between what Kratzer (1981, 1991) terms modal 'forces' (such as possibility and necessity) and modal 'flavours' (such as non-epistemic vs. epistemic), but also pathways of diachronic development, from 'premodal' lexical sources to 'postmodal', non-lexical syntactic or semantic functions (see Section 2.2).

Van der Auwera & Plungian distinguish between possibility and necessity, also mentioning various strengths of necessity, and between four primary semantic flavours: 'participant-internal' (describing participant abilities or internal needs); 'participant-external' (describing situational possibilities or necessities, independent of participant characteristics); 'deontic' (a subtype of participant-external modality describing permission or requirements based on rules, expectations, or social norms); and 'epistemic' (describing possibilities or necessities based on knowledge about the situation). In addition to the macro 'force' distinction between necessity and possibility, an additional 'strength' distinction can be made between 'strong' and 'weak' necessity (and perhaps also possibility) (Kratzer 1991); in English, 'must' expresses strong necessity while 'should' marks weak necessity. In Xhosa, the necessity modals do not appear to distinguish strength, although some may imply a stronger necessity than others due to their lexical sources.

Although we took van der Auwera & Plungian's (1998) taxonomy as our starting point, we aimed to keep open minds in our categorisations, without forcing particular uses of modal expressions into categories into which they do not naturally fit. Indeed, it is precisely these challenging examples that can enlighten us as to which modal categories are significant in Xhosa, and how a Xhosa-specific model of modality might need to be developed, without losing cross-linguistic comparability. The categories in our modal taxonomy are described below, along with examples and the questions that we used to diagnose each flavour in our corpus analysis.

'Participant-internal (PI) modality' describes the abilities (PI possibility) or "internal need[s]" (PI necessity) of a discourse participant (van der Auwera & Plungian 1998: 80). As noted by Narrog (2012: 10), participant-internal modality is easier to conceptualise for possibility than for necessity, since it is hard to distinguish unambiguously internal needs beyond the need to consume and excrete food. Other

2. The extensive questionnaire laid out in Vander Klok (2022) is also frequently utilized. Vander Klok makes these distinctions, and also includes bouletic (possibility or necessity based on personal desire) as a sub-flavour of participant-internal modality. Earlier versions of the questionnaire (such as Vander Klok 2014) did not distinguish participant-internal and participant-external 'circumstantial' flavours, following Kratzer (1981, 1991).

needs associated with discourse participants tend to relate additionally to goals (like the need to drink coffee in order to accomplish the goal of focused work), desires, or expectations.

We used the diagnosing question in (4) to code participant-internal necessity. An illustrative (constructed) example is given in (5).

(4) *Does it reflect the immediate subject's physical / mental / inherent needs?*

(5) **Ku-funeka** **ndi-y-e** **e-bathroom** **ngoku!**
SP$_{17}$-NEC SP$_{1SG}$-go-SBJV LOC-9.bathroom now
'I need to go to the bathroom now!' (constructed)

'Participant-external (PE) modality' relates to situational factors that make a predicate situation possible or necessary for the discourse participant(s) (van der Auwera & Plungian 1998: 80). It is sometimes referred to as 'participant-imposed modality'. Note that when we refer to participant-external modality in this article, we refer specifically to non-deontic participant-external modal functions, whereas van der Auwera & Plungian (1998) consider deontic modality a subtype of the more general participant-external category.

Non-deontic participant-external modality can further be divided into (at least) 'teleological', or goal-based, necessity and possibility; and 'situational', or more purely 'circumstantial', necessity and possibility. Van der Auwera & Plungian (1998) do not directly describe these subcategories, but they are noted in, for example, Narrog (2012). This distinction is not always straightforward to make in practice, and we therefore do not include it in our corpus statistics, although we refer to it in discussions of particular examples.

Example (6) gives our diagnosing question for teleological necessity, with a constructed example given in (7).

(6) *Does it reflect a necessary means to fulfil participant goals?*

(7) **Uku-y-a** **e-sitishi-ni** **ku-funeka** **u-khwel-e**
INF-go-FV LOC-7.station-LOC SP$_{17}$-NEC SP$_{2SG}$-ride-SBJV
ibhasi **66**.
9.bus 66
'To get to the station, you have to take bus 66.' [because it is the only possible way to get there] (constructed)

Our diagnosing question for circumstantial/situational participant-external necessity is given in (8), with a constructed illustrative example in (9). Although circumstantial examples may suggest goals, the goals are not stated outright as they are in (7), nor is the necessity in (9) overtly described as a mean to a goal; circumstantial modal expressions may also refer to goals themselves.

(8) *Does it reflect what is necessary given external situational constraints?*

(9) **Ku-funeka** **u-m-thwal-e**, **ngoba** **a-ka-kwazi**
SP$_{17}$-NEC SP$_{2SG}$-OP$_1$-carry-SBJV because NEG-SP$_1$-POSSIB
uku-hamb-a.
INF-walk-FV
'You have to carry him, because he cannot walk.' (constructed)

'Deontic modality' refers to possibilities or necessities governed by rules or norms (van der Auwera & Plungian 1998: 81). We include under deontic everything from official laws and enforced rules, to societal expectations, to personal judgments of what is right and proper made by the speaker of an utterance.

Our diagnosing question for deontic necessity is given in (10), with a constructed illustrative example in (11).

(10) *Does it reflect the speaker's or societal [etc.] rules or expectations?*

(11) **Ku-funeka** **u-nxib-e** **isigcina–ntloko**.
SP$_{17}$-NEC SP$_{2SG}$-wear-SBV 7.helmet
'You have to wear a helmet.' [it is the law] (constructed)

'Epistemic modality' involves the speaker's assessment of whether the proposition is possibly or probably/necessarily true based on knowledge and beliefs (van der Auwera & Plungian 1998: 81).

We diagnosed epistemic necessity with the question in (12). A constructed illustrative example is given in (13).

(12) *Does it reflect the speaker's [or matrix clause subject's] knowledge, predictions, beliefs about what must be or is probably true?*

(13) **Ku-fanele** **a-b-e-se-sikolweni**
SP$_{17}$-NEC SP$_1$-be-SBJV-LOC.COP-7.school.LOC
'[Sipho] must be at school.' [based on what we know] (constructed)

Table 1 summarises the taxonomy of modal flavours used in this study. We include both strong and weak necessity conditions in the necessity row of the table, as we have not found Xhosa necessity modals to clearly distinguish between them.

Participant-internal (PI)	**Participant-external (PE)**		**Deontic**	**Epistemic**
	Teleological	**Situational**		
Internal need, drive, preference	Exclusive / preferred means to goal	Necessity / preference given situational factors	Requirement / expectation / obligation	Proposition judged as necessarily / probably holding

Table 1. Modal taxonomy used in this study

We also occasionally mention 'existential' uses of necessity expressions, indicating that something must be / is 'always' true of an entity or class. We did not find evidence of existential flavour as a significant subcategory of necessity meanings in Xhosa. However, we judged that a few of the corpus examples have, as one of several senses, a plausible existential reading.

Importantly, as noted by Narrog (2005: 684, citing Coates 1983), a modal expression can lie at the border of several categories, and an utterance with a modal expression can simultaneously invoke several modal flavours. Thus, when dealing with real-world data, an unambiguous classification of modal uses into discrete categories is not merely impossible but also undesirable.

Even when allowing for the coding of multiple modal senses, classifying modal expressions in natural language is challenging, and we had significant inter-coder variation on our first rounds of coding the corpus examples.[3] This challenge also brings advantages: cross-coder inconsistencies helped draw our attention to the contexts in which meaning ambiguities tended to arise, which in turn led us to seek other relevant factors in the choice of modal expressions, to expand our taxonomy to include existential modality, and to dig deeper in to each modal expression's unique semantics.

1.2. Modal expressions, grammaticalisation, subjectification, and semantic change

Van der Auwera & Plungian (1998), based largely on typological data from Bybee *et al.* (1994), discuss a typical pathway of semantic change for modal expressions, shown in a simplified form in (14). Basically, participant-internal meanings expand or shift towards participant-external meanings (and to the participant-external subcategory deontic) and further to epistemic meaning. Depending on the lexical source or source construction, expressions can enter the modal pathway at different points, for example, moving straight from lexical to deontic or epistemic meaning. Van der Auwera & Plungian also discuss more complex connections between possibility and necessity domains, and the model has been refined in later work and with data from broader language samples (e.g. in van der Auwera *et al.* 2009).

(14) **[participant-internal] > [participant-external (> deontic)] > [epistemic]**

Van der Auwera & Plungian also describe 'postmodal' developments of modal markers, that is, modal markers that have expanded their meanings beyond the modal domain and gained extended non-modal functions (which are still conceptually

3. In an experiment, Rubinstein *et al.* (2013) found that expert annotators could distinguish between 'priority' (deontic, teleological, bouletic) and 'non-priority' (epistemic, ability, circumstantial) modal flavours with high inter-annotator agreement, but that inter-annotator reliability was low at finer-grained levels. They liken the key priority/non-priority distinction to the 'root' (non-epistemic) vs. 'epistemic' divide assumed in much generative semantic literature on modality. Despite the low degree of inter-annotator agreement, Rubinstein *et al.* suggest that the finer-grained judgments are also useful, because "[t]hey allow the researcher to distinguish between more and less ambiguous exemplars of each modality type and to investigate the grammatical and contextual properties of examples that are judged as more ambiguous" (2013: 8).

close to modality). Such functions include evidentiality and future tense marking, subordinated conditional, concessive and complementation marking, as well as non-declarative, illocutionary marking in the form of optative ('wishing mood', related to speaker desires), and imperative and interrogative marking. In our study of Xhosa, we noticed the expanded use of modals in conditional and concessive constructions as well as in illocutionary expressions, functioning as optatives and desideratives (the latter expressing participant rather than speaker wishes or desires) and as 'softeners' or face-saving hedging markers.

An important semantic driving force in the development of polysemous patterns within the modal domain, where one modal expression expands to express multiple flavours, is the process of subjectification. Following Traugott (1989, 2006, 2010), subjectification may be conceptualised as an extension in semantic scope from more event-oriented, objective descriptions of the qualities of the subject or the circumstantial constraints of a proposition to expressing the speaker's subjective view on a situation. This means expressing the speaker's own viewpoint – attitude and beliefs – towards what is conveyed. Subjective or subjectified meanings typically manifest themselves in the flavors of deontic and epistemic modality, as they are more 'attitudinal', expressing whether the state of affairs "exists or not, or is morally acceptable or not, or is 'agreeable' or not" (Nuyts 2015: 109). At the same time, subjectification is a gradient concept, cutting right through the subdomain of deontic modality, where expressions can be, but are not necessarily, subjective (van der Auwera & Plungian 1998: 83). This depends on whether they are 'performative' or 'descriptive' (Nuyts 2006: 15), that is, whether they have the speaker as the source of the force, explicitly conveying his/her subjective perspective, or whether they merely convey objective reports on the norms and attitudes of someone else.

It is often posited that further grammaticalisation of a form – that is, the shift along the spectrum from fully lexical to fully grammatical meaning, often with corresponding phonological reduction – correlates with increased subjectification (Narrog & Heine 2011), and we have largely found this to be the case with Xhosa modals, where more grammatical elements also occur more frequently with more subjective meanings. We do not attempt to pinpoint or 'prove' the exact extent of grammaticalisation of the modal markers,[4] but we do show differences between more auxiliary-like and more inflectional markers, as well as factors that point to the grammatical(ised) status of particular modal expressions. We also show that when lexical and constructional sources of modal expressions are transparent – and, in some cases, still actively used in the language – these meanings appear to have influenced the expressions' development paths and functional ranges, in line with Hopper's (1991) notion of 'persistence'.

On the topic of auxiliaries, it should be noted that we depart from the traditionally more narrow definition within South Africa Bantu research of what constitutes an auxiliary verb (where they are assigned certain functional and formal features to distinguish them from so-called 'deficient verbs'; see Anderson 2006: 11 and further references therein). Instead, we make use of a broad definition of auxiliaries and

4. Indeed, the findings in Bernander *et al.* (2022) suggest that modal verbs in Bantu generally do not exhibit much evidence of having undergone (formal) grammaticalisation.

auxiliary constructions – inspired by the work of Anderson (2006, 2011), and Heine (1993) – as constituting any instance where two (or more) verbs occur together and where one of them (the auxiliary verb) contributes some functional quality to the other one (the lexical verb). With regard to the necessity verbs, we arguably even transcend the liberal definition of Anderson (2011), as they typically occur as matrix verbs with the main verb in an embedded clause, thus not conforming to his 'mono-clausal structure' requirement for an auxiliary verb construction (Anderson 2011: 11). According to this requirement, an auxiliary verb and a lexical verb form a single verb phrase, a form–function combination in which the auxiliary merely functions as a (dependent) functional operator, while it is the lexical verb that acts as the head. We therefore refer to the modal forms **fanele**, **funeka**, and **mele** as 'auxiliary-like' to indicate their structural differences from affixes (some of which expess modal possibility), without taking any particular theoretical stance on their syntactic status.

1.3. Previous studies of modality in Southern Bantu

Modality is understudied in Bantu languages, and Southern Bantu is one of few groups – if not the only group – within the Bantu family represented by early studies dedicated to modality, consisting of what appears to have been a small cluster of articles published in the late 1980s and early 1990s. Particularly Fourie's (1989) work on Zulu and Louwrens (1990) on Northern Sotho (S32) are worth mentioning, as both are studies on languages related to Xhosa: Zulu is a Nguni language, as mentioned above, and Northern Sotho belongs to the closely related Sotho-Tswana branch. Slightly further afield, Fourie (1991) deals with modality in the South-Western Bantu language Oshindonga (R22). Early works tend to situate modality within the general category of mood, and to focus largely on the latter (subjunctives, hortatives, imperatives, etc.); see Nurse & Devos (2019: 219). More recently, Creissels (2017) described the necessity auxiliary **tshwanela** and its grammaticalisation in the Sotho-Tswana language Setswana with an additional note on the possibility prefix **nga-** in Zulu and other Bantu languages. Treatments of modal expressions have also found their way into grammatical descriptions of Southern Bantu languages, and from there into larger comparative work. Often, they are described in relation to the elaborate and complex auxiliary inventories characteristic of Southern Bantu (cf. Gibson & Riedel 2021). Bernander *et al.*'s (2023) recent study of negative verbs in Bantu, for example, shows that Southern Bantu languages and other languages spoken in the South African region, such as Swazi and 'Northern Transvaal' Ndebele (S408)[5] use auxiliary verb constructions to express flavours of, in this case, specialised negative modality (e.g. a negative possibility '(be) unable'-function).

5. Bernander *et al.* (2023) refer to this language as S44 (Zimbabwean Ndebele), but the data are from Ziervogel's (1959) description of so-called 'Northern Transvaal' Ndebele (S408), a minority language spoken in South Africa (see Aunio & Fleisch 2019). Confusions between the three 'Ndebele' languages spoken in this region are regretfully common.

Despite these studies, there has been little to no targeted research on modal semantics as a cohesive category in Xhosa and its Nguni relations until recently. Crane *et al.* (2024b) offer a comparative look at possibility expressions across several Nguni languages and Tsonga, showing cross-varietal variation in how these expressions can be used. Bernander *et al.'s* (2022) typological study of modal auxiliary constructions in Eastern Bantu dealt with a large chunk of Eastern Bantu but omitted the subgroup of Southern Bantu. The present paper therefore helps fill this research gap with a detailed overview of modal expressions in Xhosa's verbal paradigm.

2. Methodology

2.1. Introduction and motivation

In line with the rising popularity of modality as a topic of investigation, corpus studies of modal expressions in Bantu languages have gained momentum in recent years, with works like Bostoen *et al.* (2012) and Mberamihigo (2014) on Rundi (JD62, Burundi), and Kawalya *et al.* (2014) on Ganda (JE15, Uganda), among others. Corpus studies are, generally speaking, less conceptually constrained than purely elicitation-based studies, since they show actual uses without prior consideration of categories, and they are inherently quantitative. Unfortunately, for Bantu languages, extensive and reliable corpora are somewhat rare. While numerous corpora of Xhosa and its close relatives exist, not all are publicly available. Those that are available can be of uneven quality and reliability, and most of them lack annotation or translation.[6] Web searches can be similarly problematic, despite the fortunate circumstance that Xhosa is used widely online. For one thing, we do not know of a straightforward way of excluding machine translations from our results. In addition, Xhosa's agglutinative nature makes it difficult to target single morphemes in a web search. For these and other reasons, it is not clear to us that meaningful quantitative statistics can be drawn from web searches for Xhosa expressions. As we had access to a ready-made, reliable Xhosa corpus sample (see Section 2.2), compiled by the second author for his reseach on tense and aspect (Savić 2020), we were eager to repurpose it for the study of modal expressions.

In addition to our corpus searches, we relied on the linguistic intuitions of the third author, a linguist who is also a native speaker of Xhosa. See van der Auwera & Diewald (2012) for further discussion of multifaceted approaches to studying modality. The third author translated and judged a set of constructed examples that aimed at uncovering the felicity of each modal expression across a range of modal flavours. These judgments prove an important addition to the corpus study, because a few functions not represented in the corpus data turn out to be felicitous (e.g. the use of **funeka** for participant-internal necessity). Corpus studies, especially with a relatively small corpus such as ours, do not provide negative data.

6. A happy exception is the newly released *Corpus of Spoken isiXhosa*, accessible at https://spraakbanken.gu.se/korp/?mode=xhosa#?lang=eng&cqp=%5B%5D&corpus=xhosa, which currently has more than 7000 tokens and includes both translations into English and detailed morpheme-by-morpheme annotations.

In the terms of Kawalya *et al.* (2014), who follow Tognini-Bonelli (2001), we offer a 'corpus-based' rather than a 'corpus-driven' study. In the latter, as Kawalya *et al.* (2014:62) put it, "the corpus itself is considered to be the sole source of hypotheses about language". Because of the size of our corpus, we find the corpus-based approach, incorporating speaker intuitions and tests with minimally ambiguous contexts, to be more enlightening for the present study, although future corpus-driven studies undertaken with larger corpora are also desirable.

2.2. Description of corpus

The material used for this study originates from three Xhosa corpora: the Genre Classification Corpus, compiled by Snyman *et al.* (2012); the AST Text Corpus, compiled by Roux *et al.* (2001), and the NCHLT Text Corpus, compiled by Eiselen & Puttkammer (2014); see Savić (2020: 26-29) for details. These corpora include diverse genres and are made up of both original Xhosa literature and translations into Xhosa. The data we examined consisted of excerpts from these three corpora, originally selected for a study by the second author (Savić 2020).[7] Savić selected the excerpts for the occurrence of a target verb surrounded by extensive context and used these excerpts to investigate tense and aspect contrasts. For the original study, each excerpt was verified for quality and discussed in detail with Xhosa-speaking consultants. It is on this excerpted, verified corpus that we base our current investigation.

The extracted corpus consists of approximately 60 000 words, including some repetitions and occasional extraneous characters such as page numbers. Around one-third of the words occur only once in the corpus, a high number attributable to the agglutinative nature of Xhosa.[8] Although this curated corpus gives us a much smaller sample than the original corpora (which comprise almost five million words combined), it has the advantage of verified quality, whereas the original corpora also include low-quality translations and some poorly digitalised texts (Savić 2020: 27). We consider the selection of texts in our corpus to be essentially random, first since the texts were extracted for Savić's (2020) study based on the occurrence of common verbs that are unrelated to modal expressions, and second, because they include extensive textual material surrounding these verbs.

---

7. Approximately 31% of the excerpts in this curated corpus are drawn from literature, consisting of published novels originally written in Xhosa. These tend to contain the 'deep' language described below. The remaining texts are all translations and can be divided into texts written in a 'formal' register, such as official guidelines and forms, making up about 7% of the excerpts, and 'neutral' register texts, including magazine articles and letters to magazines, speeches and announcements, and cartoons, making up the remaining 62% of the excerpts. Modal expressions were spread approximately evenly across these genres (approximately 25% came from literature, 14% from formal texts, and 61% from neutral texts), with slight overrepresentation of the 'formal' register, which is not necessarily unexpected, given that guidelines are likely to contain instructions about what is possible and necessary based on regulations.

8. Some of the single-token types are likely also due to typographical errors or misrecognition of characters in the scanning process; however, we believe that this effect is somewhat minor. For example, when we spotted and corrected a few typos, a search found that the corrected words still did not recur in the text.

It should be mentioned that our corpus contains data from written Xhosa and skews towards 'standard' Xhosa, represented by the written forms of the Tshiwo dialects Gcaleka and Rharhabe (see e.g. Nyamende 1994; Bloom Ström 2018 for details). Other, 'non-standard' dialects are mostly excluded from the original corpora from which ours is drawn, as are most informal spoken registers. The corpus further includes some literary sources that are written in so-called 'deep' Xhosa[9] and contain expressions that speakers from urban environments sometimes find difficult to decipher.

2.3. Study inclusion criteria

We view the umbrella category of 'modal expressions' or 'modal markers' as containing all lexical and grammatical means of expressing possibility and necessity. In this study, however, we restrict ourselves to verbal expressions of modality: verbal auxiliary-like markers (see Section 1.2) and inflectional markers (possibility expressions in Part II, forthc.). These markers were identified using the three-volume *Greater Dictionary of (isi)Xhosa* (Pahl *et al.* 1989; Mini *et al.* 2003; Tshabe & Shoba 2006) and confirmed as common through corpus searches.

2.4. Coding and analysis

Our extraction and coding methodology followed the following basic steps:

(i) Extraction and translation of examples; checking unclear contexts and translations.

We searched our corpus and extracted all examples of each expression. We included all morphological variants we were aware of. We placed the extracted examples in a shared online spreadsheet (one sheet per modal expression), including its source location and at least a sentence or two of context, along with translations. We occasionally corrected spelling and punctuation when we could do so confidently.

(ii) Initial coding

We divided the task of the initial coding amongst us by marker and coded each example for its possible modal flavours (see Section 2.1), referring to their wider contexts in cases of uncertainty. We aimed to include all possible interpretations, not restricting ourselves to a single flavour for each example. We excluded examples that we judged to be primarily lexical (e.g. the possibility verb **kwazi** in its literal

9. The concept of 'deep' language varieties is common across the southern, central, and eastern African environments we are familiar with from our research. It refers roughly to (imagined?) varieties, spoken in rural areas, that are perceived as older and as linguistically and culturally purer than more urban varieties. Deep language is often associated with traditional oral literature. Among the Xhosa speakers we have worked with, the concept of deep language is often invoked in cases where (especially younger, urban) speakers cannot understand a portion of a text, especially one that involves metaphorical language. Many Xhosa literary works employ 'deep' language. See Thompson (2013: 54-55) for discussion of ideologies surrounding 'deep' Shona.

meaning of 'know (something, someone)', without any implication of ability; or **fanele** meaning 'suit'). We also noted any postmodal uses and coded for polarity and the subject's animacy, and for subclasses of participant-internal (learnt vs. inherent; physical vs. mental), participant external (goal-oriented vs. situational), and deontic (based on source of authority) modality. Some of these additional features proved relevant for individual modal markers, but none appear to be of general significance for the modal system as a whole.

(iii) Cross-checking and synthesis of coding

The authors blindly re-coded the (respective) other markers that they had not coded in the first round, for the purpose of cross-checking. In cases of inter-coder divergence, we re-examined the uses and their contexts and came to a consensus based on extended context from the corpus and further discussion. We incorporated all judgments that seemed plausible within the context into the final coding.

(iv) Statistical and qualitative analysis

We assigned one numerical 'point' for each modal flavour associated with an occurrence of a modal marker. We also distinguished between what we judged to be the 'primary' sense or senses, which seemed to us to be the main semantic function(s) of the marker in that context, and 'secondary' senses, which seemed to us to be plausible additional interpretations, but not the main intended meaning. Obviously, personal judgment is involved in this coding procedure, and readers may well disagree with some of our judgments.

For example, (15) shows what we judged to be a straightforwardly participant-external context, without a participant-internal sense.

(15) **U-<u>nako</u>** **uku-theng-a**
SP$_{2SG}$-POSSIB INF-buy-FV
'You can buy
[**into yokujika ibala lokutya nakuyiphi na ivenkile okanye isuphamakethi**.]
[food colourant in any shop or supermarket.]' (Roux *et al.* 2004)

In contrast, (16) shows an ambiguous participant-internal–participant-external possibility reading, in which the children's inherent abilities intersect with their circumstances (having eaten breakfast) to produce the conditions for learning. We considered both participant-internal and participant-external possibility to be primary senses in this example.[10]

(16) [**Abantwana abasityayo isidlo sakusasa baqhuba kakuhle esikolweni kuba**]
'[Children who eat breakfast excel in school because]
ba-b-a **<u>nako</u>** **uku-yi-gqal-a** **into**
SP$_2$-be-FV POSSIB INF-OP$_9$-attend_to-FV 9.thing

10. Here and in the remainder of this article, we only provide full interlinear glosses for the relevant modal expression and its immediate context in each example.

aba-yi-fund-is-w-a-yo.
REL.SP$_2$-OP$_9$-learn-CAUS-PASS-FV-DJ.REL
'they can attend to what they are being taught.' (Roux *et al.* 2004)

In (17), we considered participant-external (largely teleological) possibility to be the primary sense, with a secondary participant-internal sense, since the action described has something to do with internal ability, although not in the most prototypical meaning of knowing how to / being able to do something.

(17) [**Kube ngenxa yokuqonda ukuba ukukhathalela umzimba nempilo yam**]
'[It's because I realised that taking care of my body and my health,]
lolona **thando**
COP$_{11}$.PRON$_{11}$.SUPERL 1.love
ndi-noku-zi-nik-a **lona**.
SP$_{1SG}$-POSSIB-REFL-give-FV.SUBORD PRON$_{11}$
it's the best kind of love I can give to myself.' (Roux *et al.* 2004)

Because of the inherent subjectivity of the coding process, we do not aim to give a precise statistical analysis. Therefore, despite our attempts at remaining consistent and our cross-checking, the numbers we give should be understood as indicative rather than the final quantitative word. Even this imprecise statistical analysis is enlightening, as it allows a visual overview of each modal expression's usage tendencies and uncovers interesting patterns.

3. Xhosa expressions of necessity

3.1. Introduction and overview

Our corpus searches uncovered only two common expressions of necessity: **fanele** (Section 3.2) and **funeka** (Section 3.3). A third expression, **mele**, was found only rarely in the corpus, and we did not include it in our statistical analysis, although we discuss it briefly in Section 3.4. The intuitions of the third author regarding the possible modal functions associated with each marker are shown in Table 2. **Funeka** can be used with participant-internal, participant-external, and deontic meaning, although we did not find any participant-internal uses in the corpus. **Fanele** can be used with participant-external, deontic, and epistemic meaning, and in our judgment always contains at least some sense of deonticity in addition to any primary meanings.[11] **Mele** has a similar distribution to **fanele**, but possibly without the consistent deontic overtones. None of the markers can be used in the straightforward existential contexts we attempted to construct (e.g. 'it must be / is always cold'), but all can be used in an example where existential necessity is one possible sense out of several senses (example 18; see Section 3.3 for discussion). This distribution suggests that existential necessity may not be a significant category in Xhosa's modal system, at least not with the markers we investigated.

11. **Fanele** cannot be used in the most straightforward participant-internal necessity context, needing to go to the bathroom. Participant physical needs, however, may be part of the situational necessity expressed by **fanele**, as in (19).

	PI	PE	Deontic	Epistemic
funeka	✓	✓	✓	
fanele		✓	✓	✓
mele		✓	✓	✓

Table 2. Overall functional distribution of flavours in positive-polarity contexts, by necessity marker, in Xhosa, based on the third author's intuitions

The counts of corpus occurrences we found of these two markers are shown in Figure 1.

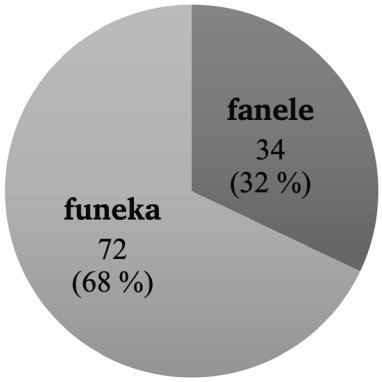


Figure 1. Number of occurrences of each necessity marker in corpus (total = 106)

Most interpretations associated with necessity in the corpus related to deontic and participant-external modality, as shown in Figure 2, which shows the modal senses (both primary and secondary; see Section 2.4) associated with either **funeka** or **fanele** in the corpus, with each sense (i.e., possible form–meaning pairing in a context) being assigned one point. As we discuss in Sections 3.2 and 3.3, there is significant interpretive overlap between deontic and participant-external senses. Note that because occurrences of a modal expression can be associated with more than one flavour, the number of senses is greater than the number of occurrences.

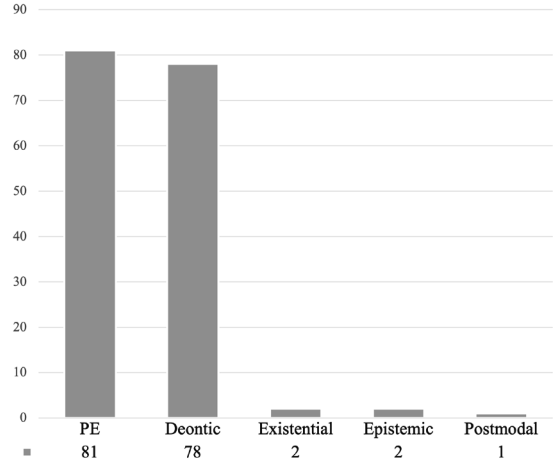


Figure 2. Overall distribution of necessity flavours, both primary and secondary senses

Figure 3 shows the same modal senses as they were distributed across the markers. **Fanele** has undergone significantly more semantic expansion than **funeka** (and is also syntactically more integrated, in that it can take agreeing subject marking in addition to 'expletive' marking with class 17 prefix **ku**-), but both markers centre around participant-external and deontic readings, with only minor differences in the distribution of possible interpretations.

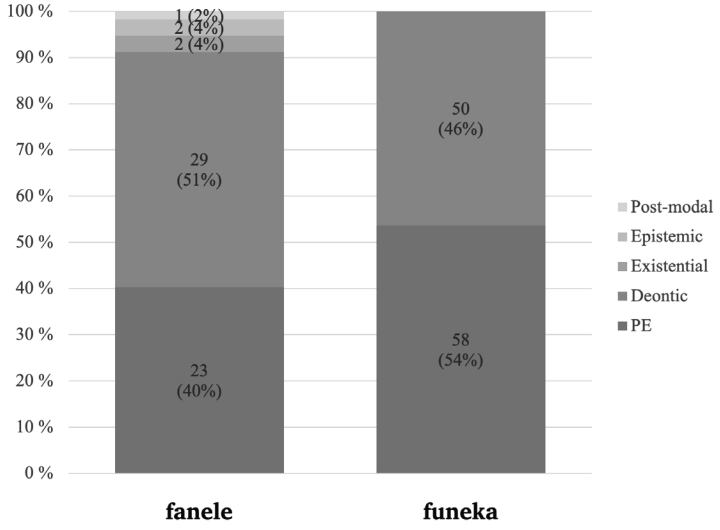


Figure 3. Total modal senses (both primary and secondary) associated with each necessity marker, as a proportion of total senses. The numbers in the graph show the number of occurrences of each sense with the respective marker, along with the percentage (rounded) of total senses this represents.

3.2. **Fanele**

3.2.1. Forms and source

Fanele is the perfective form of the lexical verb **fánèlà** ‘suit, be suitable or fit for; be proper, right, becoming, worthy, deserving’ (Tshabe & Shoba 2006: 492). It still frequently occurs with lexical meaning, especially when used with the neuter suffix (as in e.g. **kufanelekile** ‘it is appropriate’). The use of verbs originally denoting ‘be fitting’, ‘be suitable’ and similar meanings as markers of necessity is widespread in the rest of the Eastern Bantu group (Bernander *et al.* 2022), with many languages even making use of the cognate etymon, ultimately linked to the Proto-Bantu stem ***pú-an** ‘be fitting, resemble each other’ (Bastin *et al.* 2002). Necessity verbs with similar etymologies have also been attested cross-linguistically (Bybee *et al.* 1994: 182-183; van der Auwera & Plungian 1998; Kuteva *et al.* 2019: 414-415). As mentioned in Section 1.3, Creissels (2017) offers a detailed discussion, with ample reference to Southern Bantu languages, of the polysemous pattern of similative verbs developing the meaning ‘suitable’, and from there, modal meanings.

The syntax of the modal verb **fanele** is complex and has been explored at length for Nguni languages (e.g. Zeller 2006; Halpert 2012, 2016). **Fanele** occurs both with expletive subject marking and in raised form, with agreeing subject marking (Zeller 2006). Agreeing subject marking was a little more than twice as common as expletive-marked **fanele** in our sample, with 24 agreeing subject markers (including one that agreed with a nominal infinitive) and 10 expletive-marked occurrences.

Expletive-marked **fanele** is usually followed by an inflected verb in the subjunctive form (18),[12] optionally with a complementiser in between (19).

(18) [**Amaphulo oluntu – ingabi yinto yomntu omnye.**]
‘[Community campaigns – don’t be a one-person thing.]
Ku-fanele **i-b-e** **liphulo** **e-li-vuselel-a**
SP$_{17}$-NEC SP$_{9}$-be-SBJV COP$_{5}$.campaign REL-SP$_{5}$-rouse-FV
It should be a campaign that stimulates the spirit
umoya **wo-luntu** **ekuhlaleni**.
3.spirit PP$_{3}$.CONN-11.human LOC.community.LOC
of the people in the community.’ (Roux *et al.* 2004)

12. Other tense-aspect-negation marking can be used to communicate different temporal and polarity configurations; for example, a present-tense imperfective form could indicate ‘should be doing X’.

(19) [**UNtokozo Mbambo ubebanjwe ngumkhuhlane ngoko**]
'[Ntokozo Mbambo had a fever so]
beku-fanele **ukuba** **u-phuml-e**
SP_1.PST.IPFV-NEC COMP SP_1-rest-SBJV
she should have rested
[**kodwa uzile naye ukuzokuthetha neBONA.**]
[but she came to talk to BONA.]'[13] (Snyman *et al.* 2012)

When **fanele** is inflected with an agreeing subject, it can be followed with an optional complementiser and a subjunctive verb, like the expletive form. It can also be followed by an infinitive verb (20) or by a non-subjunctive verb, if other TAM marking is involved (21).

(20) **U-fanele** **uku-khathazek-a.**
SP_{2SG}-NEC INF-become_worried-FV
'You should be worried.' (Snyman *et al.* 2012)

(21) [**U-ISAAC MOKOENA oneminyaka engama-38 nekuyiminyaka yobudala**]
[ISAAC MOKOENA who is 38 years of age]
ebe-fanele **ukuba** **u-sa-phangel-a**
SP_1.PST.IPFV-NEC COMP SP_1-PERS-work-FV
should still have been working
[**kodwa ingxaki yesifuba yamenza wangabinako ukuqhubeka nokuphangela**].
[but a chest problem prevented him from continuing to work.]'
(Snyman *et al.* 2012)

3.2.2. Corpus statistics and analysis

Corpus statistics point to the primacy of deontic meaning for **fanele**. We found 35 occurrences of **fanele** in our corpus. The primary interpretations (in our judgment) lean heavily towards the deontic, but participant-external functions are also common (Figure 4).

13. A reviewer suggests a primarily participant-internal reading of this example. Indeed, under some conceptualisations of participant-internal necessity, such as that in Gluckman & Bowler (2020: 214), in which world circumstances affect bodily needs, this is a plausible reading. However, we judge this example to describe mainly what is needed given the circumstances, including the teleological necessity of resting to get better, and the speaker's (subjective, deontic) assessment of what is the best thing to do in such circumstances. **Fanele** is infelicitous when describing bodily needs such as urination that have minimal teleological or deontic associations.

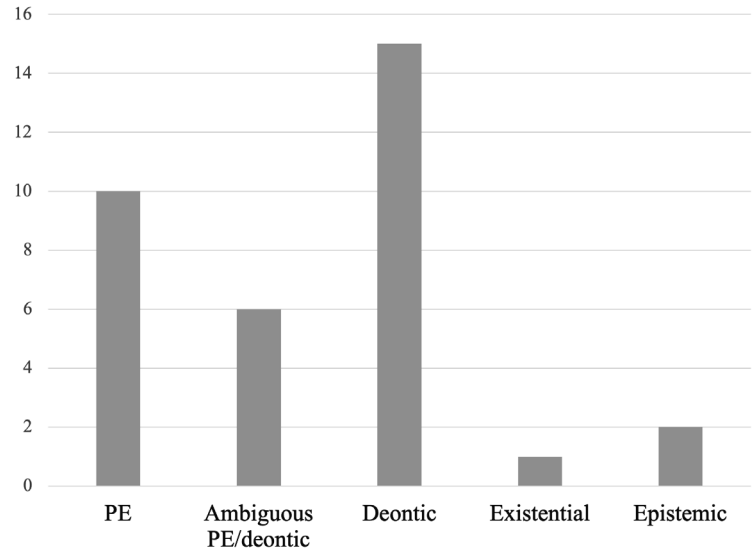


Figure 4. Distribution of primary senses of **fanele** (y-axis indicates number of occurrences as a primary meaning)

However, most of the uses of **fanele** had two or more plausible readings (Figure 5), and the only examples we judged to express a single modal flavour were deontic. If we classify the speaker's subjective view of what is right and proper as deontic, all examples had a plausible deontic interpretation. For example, the speaker's judgment of what should happen or be done can be seen in all examples in Section 3.2.1 (18-21), in addition to their other senses, although none of these refers to an externally imposed rule. Corpus functions thus suggest an important deontic component to **fanele**, not unexpected given its source in 'suit, be proper [etc.]'.

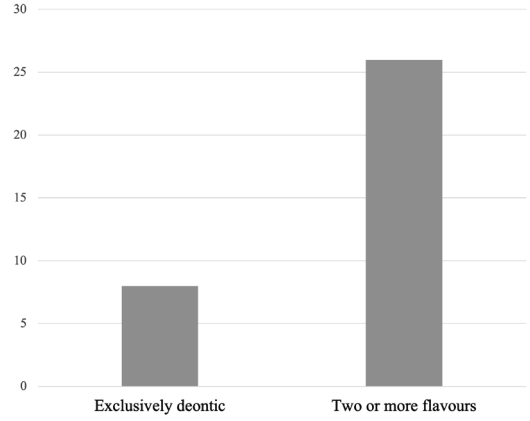


Figure 5. Distribution of senses of **fanele**, when both primary and secondary interpretations are considered (y-axis indicates number of occurrences as a primary or secondary meaning)

Many of the occurrences are counterfactual, describing something that should have happened or that the speaker or participant wishes had happened (22). Of the two modal markers, only **fanele** was used with counterfactual meaning in the corpus, although such uses are also felicitous with **funeka**.

(22) [**Waphaphama xa ephulukana neemoto zakhe ezimbini iBMW 325 CineGolf 5 kwaza kwanyanzeleka ukuba athengise nomzi wakhe. Uthe**,]
'[He woke up when he lost his two cars BMW 325Ci and Golf 5 and he had to sell his house. He said,]
"E-ndaw-eni **yo-ku-theng'** **iimoto** **ezi-mbini**
LOC-9.place-LOC PP_9.CONN-INF-buy[FV] 10.car NP_{10}-two
umzi **kwaku-fanele**[14] **ndi-hlawul-ele.**
3.home SP_{17}.REMPST.IPFV-NEC SP_1-pay-APPL.PFV.DJ
Kwaku-fanele **nd-ong-e** **imali**
SP_{17}.REMPST.IPFV-SP_{17}-NEC SP_{1SG}-save-SBJV 9.money
kodwa **azange** **ndi-y-enz-e** **loo** **nto!"**
but PST.NEG(never) SP_{1SG}-OP_9-DO-SBJV DEM_9.MED 9.thing
"Instead of buying two cars I should have paid for the house. I should have saved money, but I never did that!"' (Snyman *et al.* 2012)

Some corpus examples might also hint at epistemic functions for **fanele**. Example (23), which is also counterfactual,[15] depicts the speaker's view of what would be proper and desirable (deontic), but it might also be construed as suggesting a prediction about what would be the case were the reputation of the area better. The third author also finds **fanele** to be appropriate in straightforwardly epistemic contexts (e.g. **Kufanele abesesikolweni** '[Sipho] must be at school [based on what I know]'), but judges that even these have deontic overtones based on the speaker's personal judgment of what is right or on the participant's obligations. These examples and judgments suggest that **fanele** can be used epistemically, but that deontic judgment is part of its base semantics.

14. **Kwaku-** is an aspectually marked class 17 remote past prefix, a contracted form of **kwaye kufanele**, which consists of the auxiliary **kwa-ye** (class 17 remote past subject prefix **kwa-** and auxiliary perfective **ye** 'go') and **kufanele,** with the class 17 situative subject marker **ku-**. See Savić (2020:164-179) for discussion. In this article, we gloss such grammaticalised aspectually marked prefixes (also in the near past paradigm) with the functional labels that we judge to most nearly approximate their tempo-aspectual roles.

15. A reviewer judges (23) as primarily circumstantial rather than epistemic. The analysis of counterfactual modals is highly debated in linguistics and philosophy and is beyond the scope of this article; see e.g. Condoravdi (2002) for discussion of metaphysical vs. epistemic modal bases. Our tentative judgment that the meaning is at least somewhat epistemic rests on the fact that it refers to what would be the speaker's expectations given circumstances in the relevant imagined world (where the beautiful area does not suffer from a bad reputation), although the proposition is not epistemically necessary or likely (or, indeed, possible) if judged from the real-world point of evaluation. Further study of the semantics of counterfactual necessity in Xhosa is called for.

(23) [**Indawo yaseMolweni intle kangangokuba**]
'[The area of Molweni is so beautiful that]
ibi-fanele **uku-tsal-a** **umdla** **ku-bakhenkethi**
SP_9.PST.IPFV-NEC INF-pull-FV 3.interest 17-2.traveler
it should attract tourists
[**ize ingenisele imali abahlali bakhona – kodwa akunakwenzeka oko xa inegama elibi lokuba ngummandla kafayayo**]
[and generate income for its residents – but that's impossible when it has the bad reputation of being a dying area.]' (Roux *et al.* 2004)

In addition to deontic (the speaker's personal judgment) and teleological senses, (24) also suggests a possible existential interpretation ('change **always** comes from inside [the] person [themselves, if it comes at all]'). However, the third author deems **fanele** as infelicitous in the more straightforward existential necessity contexts we tested (e.g. 'it has to be / is **always** cold there in the winter').[16]

(24) [**Xa ucing' ukuba indoda ungayitshintsha udlala ngexesha lakho kuba iza kukuvisa obukhulu ubuhlungu**.]
'[If you think you can change a man you are wasting your time because he will only cause you great pain.]
Utshintsho **yi-nto** **e-ku-fanele** **i-suk-e** **ku-ye**
1A.change COP_9-9.thing REL-SP_{17}-NEC SP_9-come_from-SBJV 17-$PRON_1$
umntu.
1.person
Change is something that has to come from a person.' (Snyman *et al.* 2012)

The only modal flavour in which **fanele** neither occurred in the corpus, nor was judged acceptable by the third author, is participant-internal necessity. For participant-internal necessity, **funeka** (Section 3.3) is required. **Fanele**, then, as would be expected based on its lexical source, entered the domain of modal necessity via deontic modality, and has expanded to include epistemic and more general participant-external uses, although both flavours still seem – at least based on our investigations – to have deontic overtones when expressed with **fanele**.

3.3. **Funeka**

3.3.1. Overview

The modal **funeka** is the neuter (**-ek**) marked form of the lexical verb **fûnà**, which Tshabe & Shoba (2006: 533) define as 'seek, search for look for; desire, want, require, need, be inclined, have an inclination, tend' (among other meanings).[17] All corpus examples of **funeka** are marked with expletive **ku-** in the subject-marker

16. The third author judges that (24) would also be acceptable with **funeka** (Section 3.3), which is infelicitous with epistemic uses. **Funeka**, like **fanele**, cannot be used in existential contexts such as our test sentence.

17. Because **funa** is inherently non-factive (cf. Givón 2001: 304), there are essentially no semantic differences between its lexical and modal uses, and we have therefore not included it in our corpus study.

position and followed by the main verb in subjunctive form. The lexical subject, when it appears, usually comes before **kufuneka**, as in (25), but it can also appear after **kufuneka** and before the main verb, as in (26). (See Zeller 2006: §5 for syntactic analysis of this type of "long left dislocation of the embedded subject").

(25) **lo** **mntu** **ku-funeka** **enz-e** **ingxelo** …
DEM$_1$.PROX 1.person SP$_{17}$-NEC SP$_1$.do-SBJV 9.report
'this person must make a statement …' (Eiselen & Puttkammer 2014)

(26) **Ku-funeka** **isidlo** **sa-kho** **si-b-e** **yi-80grams**
SP$_{17}$-NEC 7.meal PP$_7$.CONN-POSS$_{2SG}$ SP$_7$-be-SBJV COP$_9$-9.80$_g$
'Your meal should be 80 grams…' (Snyman *et al.* 2012)

The use of **funeka** as a necessity marker instantiates a common pattern witnessed throughout the Eastern Bantu group where an erstwhile desiderative verb, comprising the type of lexical meanings provided above, gets derived with an agent-demoting extension – either with the passive or, as in the case with Xhosa, the anti-causative and de-transitivising neuter suffix (cf. Schadeberg & Bostoen 2019). As argued in Bernander *et al.* (2022), this type of morpho-syntactic modification can be taken to reflect the semantic specifics of the derived verb, where **funeka** principally refers to externally imposed forces or necessities affecting the first participant of the proposition.

As common as it is throughout Eastern Bantu, this strategy, as far as we know, is not discussed in the more general typological literature (although parallels can be drawn to, e.g. English 'be supposed to', which is also an agent-demoted verb employed for participant-external necessity).

3.3.2. Corpus statistics

Funeka is used in contexts of non-epistemic necessity. In our corpus searches, we found 72 occurrences of **funeka**, all with either participant-external (mostly teleological) or deontic interpretations, and with most examples plausibly having both interpretations. Ten of the occurrences had inanimate subjects, and only one had negative polarity.

Figure 6 shows our coding of what we interpret as the strongest interpretation of each occurrence; these are more or less evenly split between participant-external and deontic interpretations.

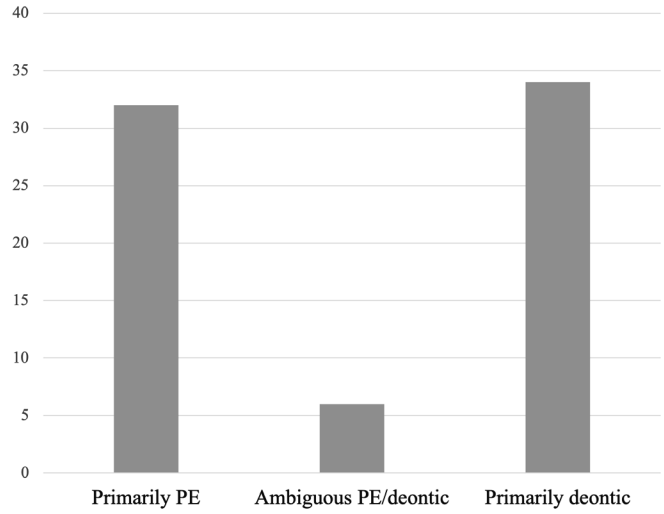


Figure 6. Distribution of primary senses of **funeka** (y-axis indicates number of occurrences as a primary meaning)

Illustrating the modal ambiguity of most examples, Figure 7 shows all interpretations we assigned as being plausibly expressed (whether as a primary or a secondary sense) in occurrence of **funeka**. We judged 22 of the examples to be exclusively participant-external, and 14 to be exclusively deontic, with most corpus examples – 36 – being plausibly associated with both deontic and participant-external (situational, teleological) interpretations, even if we judged one of the readings to be stronger.

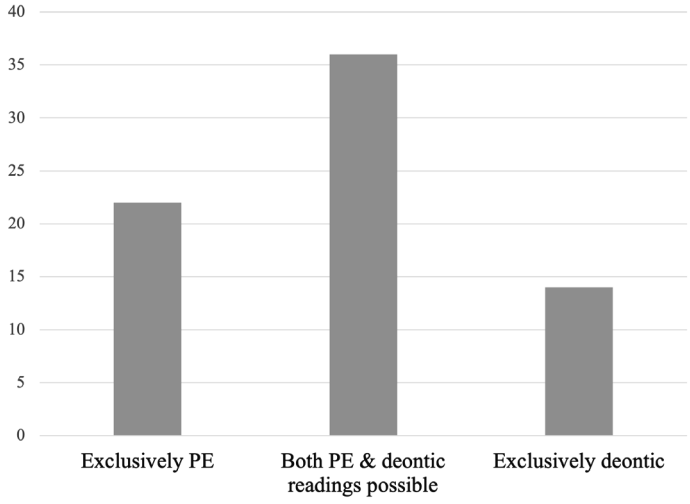


Figure 7. Distribution of senses of **funeka**, when both primary and secondary interpretations are considered (y-axis indicates number of occurrences as a primary or secondary meaning)

Example (27) is associated with both deontic and participant-external necessity. The law is explicitly invoked, making the example clearly deontic, but a goal – voting – is also overtly stated. Other examples we judged as having both readings involved topics like advice on how to achieve a goal or a better life situation, with an additional sense of subjective moral judgment on the part of the speaker or writer.

(27) **Umthetho** **u-thi** **ku-funeka** **u-b-e** **ne-minyaka**
3.law SP_3-say SP_{17}-NEC SP_{2SG}-be-SBJV COM-4.year
e-li-18 **ukuze** **u-b-e** **noku-vot-a.**
REL_4-COP_5-5.eighteen so_that SP_{2SG}-be-SBJV POSSIB-vote-FV
'The law says that you should be 18 in order to vote.' (Roux *et al.* 2004)

We did not find any examples that are straightforwardly participant-internal, nor did we find epistemic uses. The third author's intuitions rule out epistemic interpretations, but judge **funeka** as the modal marker of choice for participant-internal contexts, as in (28). It is not surprising that such contexts did not arise in our corpus data, since they are few in general. As noted above, we believe that **funeka** follows the Eastern Bantu pattern of marking externally imposed necessity, which seems to pose a functional challenge to its use in participant-internal necessity contexts. However, the contrast between **funa** and neuter-marked **funeka** in this context suggests a possible solution: **ndifuna ukuya ebathroom ngoku** means 'I want to go to the bathroom now', while **kufuneka** in (28) indicates a need, perhaps 'imposed' on the speaker by their own body. The development of participant-internal readings for **kufuneka** deserves further study.

(28) **Ku-funeka** **ndi-y-e** **e-bathroom** **ngoku!**
SP_{17}-NEC SP_{1SG}-go-SBJV LOC-9.bathroom now
'I need to go to the bathroom now!' (translated example)

As with **fanele**, we were unable to construct any purely existential necessity contexts that were felicitous with **funeka**. Example (24), with its existential component, is also felicitous with **funeka**. Together, these facts suggest that existential necessity does not play a major role in Xhosa's modal necessity system, and only emerges in conjunction with other modal meanings. We therefore analyse **funeka** as expressing participant-internal and participant-external necessity, including deontic necessity.

3.4. **Mel(w)e**

Because it occurred rarely in our corpus, this section gives an abbreviated description of the modal verb **mel(w)e**; additional research is needed.

Mélà is defined by Mini *et al.* (2003: 328) as 'be expected to, ought to, or should do or say something'. This verb may be related to an identical form meaning 'stand, represent (etc.)' (see Mini *et al.* 2003: 237), the applicative-marked form of **ma** 'stand (etc.)'. Regarding Zulu, Doke *et al.* (2014) list 'wait [on behalf of] (a person)' as one meaning of **mela**, and a semantic pathway to the modal meaning via passivisation ('be waited /expected to…') is at least imaginable. Note that a similar formal variation between an active and a non-active rendering with seemingly no

difference in function has been noted for other Bantu necessity verbs, as in the **teeka** vs. **teekwa** variation in Luganda (Kawalya *et al.* 2019), and the possibility of indexing the first participant as either the object or the subject on the necessity verb **bidi** in Swahili (Bernander *et al.* 2022).

Like **fanele**, **mel(w)e** is used with perfective aspect morphology.

Used modally, **mel(w)e** occurred only five times in our corpus, once as **kumele** and four times as passive-marked **kumelwe**. Four of the occurrences had speaker-oriented deontic (and possibly teleological) flavour, while one was epistemic. Epistemic uses are also frequent in an online corpus of translations,[18] and the third author finds epistemic modality to be an important function of **mel(w)e**, as well. More detailed research on **mel(w)e**, including its source, its typical usage patterns, and possible reasons for its rarity in our corpus, is needed.

Conclusion: Necessity verbs

In this first part of our two-part study on modal expressions in Xhosa, we have investigated the semantic range and corpus distribution of two auxiliary-like modal verbs in Xhosa, **fanele** and **funeka**. Both are used primarily for non-epistemic participant-external (situational, teleological, and deontic) modal functions, but they are not fully synonymous. **Fanele** can be used to express epistemic modality, but **funeka** cannot; conversely, only **funeka** can be used in the description of participant-internal (bodily) needs. **Fanele** is both more syntactically integrated than **funeka** (the former allowing both class 17 expletive marking and agreeing subject marking, while the latter always occurs with expletive subject marking) and more subjective in its semantics, being used for epistemic and postmodal meanings, unlike **funeka**. **Mel(w)e** has a similar distribution to **fanele**, but initial investigations suggest it may not be as strongly deontic.

The functions and distribution of Xhosa necessity verbs confirm the conceptual separation of participant-internal modality from participant-external 'circumstantial' modality. Even though participant-internal necessity is a somewhat restricted category, it is conceptually distinct in Xhosa, as shown by the third author's judgment that only **funeka** is felicitous in participant-internal necessity contexts. The participant-internal/participant-external distinction is not made in some theoretical literature on modality (e.g. Kratzer 1991), but Xhosa and other Bantu languages provide evidence for its importance and distinctiveness: see, for example, Gluckman & Bowler (2020) on Logoori (JE21, Kenya and Tanzania), Bernander *et al.* (2022) on other Eastern Bantu languages, and Crane *et al.* (2024b) on a selection of Southern Bantu languages, including Xhosa.

The persistence of lexical meaning (and the meaning of constructions such as comitative plus infinitive) is evident throughout our corpus results. **Fanele**, with its lexical source in 'be suitable (etc.)' seems in our investigations to have consistent (often subjective) deontic overtones, even when the primary contextual sense is not deontic. **Mel(w)e**, which may be related to '(be) wait(ed) for' (and by extension 'be expected (to)'), is frequently used with epistemic meaning in online translated

18. See https://app.glosbe.com/xh/en/kumele

corpora and in the third author's intuitive assessment. Neuter-marked **funeka**, from **funa** 'want/need' seems to refer to imposed needs (even when referring to bodily needs).

The semantics and functional limits of forms of **mel(w)e** should be explored in more depth. Study with a larger number of speakers is also needed to confirm our hypothesis that **fanele** usually or always has deontic overtones. An examination of neuter-marked **faneleke/fanelekile** (roughly, 'be appropriate') and its modal functions might also be fruitful. Most significantly, more extensive corpus studies are called for. Ideally, these would involve a larger dataset for more statistical precision, including greater attention to information structural issues and the inclusion of spoken data in addition to written sources.

Abbreviations and glosses

1A	noun class 1a
2SG	2nd person singular (etc.)
3	noun class 3 (etc.)
APPL	applicative
COM	comitative
COMP	complementiser
CONN	connective
COP	copula
DEM	demonstrative
DJ	disjoint
EPEN	epenthetic form
FV	final vowel
INF	infinitive
IPFV i	imperfective
LOC	locative
MED	medial (distance)
NEC	modal necessity marker
NEG	negative
OP_9	object prefix (class 9, etc.)
PE	participant-external
PERS	persistive
PFV	perfective
PI	participant-internal
PL	plural
POSS	possessive
POSSIB	modal possibility marker
PP	possessive prefix
PRON	pronoun
PST	(near) past
REFL	reflexive
REL	relative
REMPST	remote past

SBJV	subjunctive
SG	singular
SIT	situative
SP_1	subject prefix (class 1, etc.)
SUBORD	subordinate form
SUPERL	superlative
TEMP	temporal marker

References

Anderson, G. 2006. *Auxiliary verb constructions.* Oxford: Oxford University Press.

Anderson, G. 2011. 'Auxiliary verb constructions in the languages of Africa'. *Studies in African Linguistics* 40(1-2): 1-409. DOI: 10.32473/sal.v40i1.107282

Aunio, L. & Fleisch, A. 2019. 'Introduction: Venturing into Ndebele language research'. In L. Aunio & A. Fleisch (eds), *Linguistic diversity research among speakers of isiNdebele and Sindebele in South Africa*. (*Studia Orientalia* 120), Helsinki: Finnish Oriental Society, pp. 1-16.

Bastin, Y., Coupez, A., Mumba, É. & Schadeberg, T. (eds) 2002. *Bantu lexical reconstructions 3 / Reconstructions lexicales bantoues 3*. Tervuren: Royal Museum for Central Africa, online database. http://www.africamuseum.be/collections/browsecollections/humansciences/blr (accessed 7 August 2024).

Bernander, R., Devos, M. & Gibson, H. 2022. 'Modal auxiliary verb constructions in East African Bantu languages'. *Journal of African Languages and Literatures* 3: 22-85.

Bernander, R., Devos, M. & Gibson, H. 2023. 'Bantu negative verbs: a typological-comparative investigation of form, function and distribution'. *Linguistique et langues africaines* 9(1). DOI: 10.4000/lla.9818

Bloom Ström, E.-M. 2018. 'Linguistic and sociolinguistic aspects of variation in the Eastern Cape: Complexities of Xhosa language use'. *Studia Orientalia Electronica* 6:90-120. DOI: 10.23993/store.75138

Bostoen, K., Mberamihigo, F. & De Schryver, G.-M. 2012. 'Grammaticalization and subjectification in the semantic domain of possibility in Kirundi (Bantu, JD62)'. *Africana Linguistica* 18: 5-40.

Botne, R. 1995. 'The Pronominal Origin of an Evidential'. *Diachronica* 12: 201-221.

Botne, R. 1997. 'Evidentiality and Epistemic Modality in Lega'. *Studies in Language* 21: 509-532.

Botne, R. 2020. 'Evidentiality in African Languages'. In C. Lee & J. Park (eds), *Evidentials and Modals*, Leiden: Brill, pp. 460-501.

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

Coates, J. 1983. *The semantics of the modal auxiliaries*. London: Croom Helm.

Condoravdi, C. 2002. 'Temporal interpretations of modals. Modals for the present and for the past'. In D. Beaver, S. Kaufmann, B. Clark & L. Casillas (eds), *Stanford papers on semantic*, Stanford: CSLI Publications, pp. 59-88.

Crane, T.M., Gunnink, H., Kanijo, P. & Roth, T. 2024a. 'Aspect and evidentiality in four Bantu languages'. In A. De Wit, F. Brisard, C. Madden-Lombardi, M. Meeuwis & A. Patard (eds), *Beyond Aspectual Semantics: Explorations in the Pragmatic and Cognitive Realms of Aspect*. Oxford: Oxford University Press, pp. 177-206.

Crane, T.M., Lubambo, R., Mabena, M.P., Nkwinika, C., Sibisi, M. & Slater, O. 2024b. 'What can be said? Variation among expressions of modal possibility in a South African language cluster.' *Studies in Language*. DOI: 10.1075/sl.22074.cra

Creissels, D. 2017. 'Similarity, suitability, and non-epistemic modalities (volitionality, ability, and obligation)'. In Y. Treis & M. Vanhove (eds), *Similative and equative constructions: A cross-linguistic perspective*. John Benjamins Publishing Company, pp. 79-89.

Creissels, D. 2023. 'The circumstantial forms of Tswana verbs and the notion of participle'. *Studies in African Linguistics* 52 (Supplement 13): 327-338. https://journals.flvc.org/sal/article/view/133512 (accessed 7 August 2024).

Devos, M. 2008. 'The expression of modality in Shangaci'. *Africana Linguistica* 14: 3-37.

Eiselen, E.R. & Puttkammer, M.J. 2014. 'Developing text resources for ten South African languages'. *Proceedings of the 9th International Conference on Language Resources and Evaluation*, Reykjavik, Iceland: 3698-3703.

Fourie, D. J. 1989. 'Modality in isiZulu'. *Logos* 9: 45-56.

Fourie, D. J. 1991. 'On the modal system of Ndonga'. *South African Journal of African Languages* 11(3): 88-91.

Gibson, H. & Riedel, K. 2021. 'Auxiliary constructions in Southern Bantu languages'. *Southern African Linguistics and Applied Linguistics Society virtual conference*.

Gibson, H. & van der Wal, J. (eds). 2023. 'Mirativity and evidentiality in Bantu'. Special volume of *Studies in African Linguistics*.

Givón, T. 1982. 'Evidentiality and epistemic space'. *Studies in Language* 6(1): 23-49.

Givón, T. 2001. *Syntax: An Introduction I*. Amsterdam & Philadelphia: John Benjamins.

Gluckman, J. & Bowler, M. 2020. 'The expression of modality in Logoori'. *Journal of African Languages and Linguistics* 41(2). DOI: 10.1515/jall-2020-2010, pp. 195-238.

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C. & Pagel, M. 2015. 'Bantu expansion shows that habitat alters the route and pace of human dispersals'. *Proceedings of the National Academy of Sciences* 112(43): 13296-13301 and Supporting Information (SI).

Guthrie, M. 1970. *Comparative Bantu: An introduction to the comparative linguistics and prehistory of the Bantu language*s. *Vol. 4.* Farnborough: Gregg Press.

Halpert, C. 2012. 'Argument licensing and agreement in Zulu'. Ph.D. dissertation, MA: Massachussetts Institute of Technology, Cambridge. https://core.ac.uk/download/pdf/10129479.pdf (accessed 7 August 2024).

Halpert, C. 2016. 'Raising parameters'. In K. Kim, P. Umbal, T. Block, Q. Chan, T. Cheng, K. Finney, M. Katz, S. Nickel-Thompson & L. Shorten (eds), *Proceedings of the 33rd West Coast Conference on Formal Linguistics*, Somerville, MA: Cascadilla Proceedings Project, pp. 186-195. https://www.lingref.com/cpp/wccfl/33/paper3238.pdf (accessed 7 August 2024).

Hammarström, H., Forkel, R., Haspelmath, M. & Bank, S. 2024. 'Spoken L1 language: Xhosa'. *Glottolog 4.8*. https://glottolog.org/resource/languoid/id/xhos1239 (accessed 7 August 2024).

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

Hopper, P. 1991. 'On some principles of grammaticization'. In E. Closs Traugott & B. Heine (eds), *Approaches to grammaticalization, vol.* 1. Amsterdam: John Benjamins, pp. 17-35.

Kawalya, D., Bostoen, K. & De Schryver, G.-M. 2014. 'Diachronic semantics of the modal verb *-sóból-* in Luganda: A corpus-driven approach'. *International Journal of Corpus Linguistics* 19: 60-93.

Kawalya, D., Bostoen, K. & De Schryver, G.-M. 2021. 'A diachronic corpus-driven study of the expression of possibility in Luganda (Bantu, JE15)'. *International Journal of Corpus Linguistics* 26(3): 336-369.

Kawalya, D., De Schryver, G.-M. & Bostoen, K.. 2019. 'A corpus-driven study of the expression of necessity in Luganda (Bantu, JE15)'. *Southern African Linguistics and Applied Language Studies* 37(4): 361-38. DOI:10.2989/16073614.2019.1697624

Kratzer, A. 1981. 'The notional category of modality'. In H-J. Eikmeyer & H. Rieser (eds), *Words, worlds, and contexts: New approaches in word semantic*s. Berlin: De Gruyter, pp. 38-74.

Kratzer, A. 1991. 'Modality'. In A. von Stechow & D. Wunderlich (eds), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*. Berlin: De Gruyter, pp. 639-650.

Kuteva, T., Heine, B., Hong, B., Long, H., Narrog, H., & Rhee, S. 2019. *World lexicon of grammaticalization*. 2nd ed. Cambridge: Cambridge University Press.

Louwrens, L.J. 1990. 'Mood and modality in Northern Sotho'. *South African Journal of African Languages* 10(1): 10-17.

Mberamihigo, F. 2014. 'L'expression de la modalité en kirundi. Exploitation d'un corpus électronique'. PhD dissertation, Université libre de Bruxelles (ULB)/ Université de Gand (UGent), Brussels/Ghent.

Mini, B.M., Tshabe, S. L., Shoba, F.M. & van der Westhuizen, P.N. 2003. *The greater dictionary of isiXhosa. Vol 2: K to P*. Alice: University of Fort Hare.

Narrog, H. 2005. 'Modality, mood, and change of modal meanings: A new perspective'. *Cognitive Linguistics* 16(4): 677-731.

Narrog, H. 2012. *Modality, subjectivity, and semantic change: A cross-linguistic perspective.* Oxford: Oxford University Press.

Narrog, H. & Heine, B. 2011. 'Introduction'. In H. Narrog, & B. Heine (eds), *The Oxford handbook of grammaticalization*. Oxford: Oxford University Press, pp. 1-15.

Nurse, D. & Devos, M. 2019. 'Aspect, tense and mood'. In M. Van de Velde, K. Bostoen, D. Nurse & G. Philippson (eds), *The Bantu languages [2nd ed]*. London: Routledge, pp. 204-236.

Nuyts, J. 2006. 'Modality: Overview and linguistic issues'. In W. Frawley, E. Eschenroeder, S. Mills & Nguyen, T. (eds), *The expression of modality*, 1-26. Berlin: Mouton de Gruyter.

Nuyts, J. 2015. 'Subjectivity: Between discourse and conceptualization'. *Journal of Pragmatics* 86: 106-110.

Nyamende, A. 1994. 'Regional variation in Xhosa'. *Stellenbosch Papers in Linguistics Plus* 26: 202-216.

Pahl, H.W., A.M. Pienaar & T.A. Ndungane. 1989. *The greater dictionary of Xhosa*. Vol 3: Q to Z. Alice: University of Fort Hare.

Roth, T. 2018. 'Aspect in Ikoma and Ngoreme: a comparison and analysis of two Western Serengeti Bantu languages'. PhD dissertation, University of Helsinki, Helsinki.

Roux, J.C., Louw, P.H. & Botha, E.C. 2001. 'African Speech Technology telephone speech databases'. *Proceedings of the 12th Annual Symposium of the Pattern Recognition Association of South Africa* (PRASA), University of Pretoria, pp. 115-119.

Rubinstein, A., Harner, H., Krawczyk, Simonson, E.D., Katz, G. & Portner, P. 2013. 'Toward fine-grained annotation of modality in text'. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS)*, Workshop on annotation of modal meanings.

Savić, S. 2020. 'Tense and aspect in Xhosa'. PhD dissertation, Rhodes University, Makhanda.

Schadeberg, T.C. & Bostoen, K. 2019. 'Word formation'. In M. Van de Velde, K. Bostoen, D. Nurse & G. Philippson (eds), *The Bantu languages [2nd ed]*, New York: Routledge, pp. 172-203.

Snyman, D., van Huyssteen, G.B. & Daelemans, W. 2012. 'Cross-lingual genre classification for closely related languages'. *Proceedings of the 22nd Annual Symposium of the Pattern Recognition Association of South Africa*, 132-137.

Thompson, K.D. 2013. *Zimbabwe's cinematic arts: Language, power, identity*. Bloomington/Indianapolis: Indiana University Press.

Tognini-Bonelli, E. 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins.

Traugott, E. 2010. 'Revisiting subjectification and intersubjectification'. In K. Davidse, L. Vandelanotte & H. Cuyckens (eds), *Subjectification, intersubjectification and grammaticalization*. Berlin: De Gruyter Mouton, pp. 29-70.

Traugott, E. 2006. 'Historical aspects of modality'. In W. Frawley (ed.), *The expression of modality*. Berlin: Mouton de Gruyter, pp. 107-139.

Traugott, E.C. 1989. 'On the rise of epistemic meaning in English: an example of subjectification in semantic change'. *Language* 65: 31-55.

Tshabe, S.L. & Shoba, F.M. 2006. *The greater dictionary of isiXhosa*. Vol 1: A to J. Alice: University of Fort Hare.

van der Auwera, J. & Diewald, G. 2012. 'Methods for modalities'. In A. Ender, A. Leemann & B. Wälchli (eds) *Methods in contemporary linguistic*s, Berlin: de Gruyter Mouton, pp. 121-142.
van der Auwera, J., Kehayov, P. & Vittrant, A. 2009. 'Acquisitive modals'. In L. Hogeweg, H. de Hoop, & A. Malchukov (eds), C*ross-linguistic semantics of tense, aspect and modality*, Amsterdam: John Benjamins, pp. 271-302.
van der Auwera, J. & Plungian, V.A. 1998. 'On modality's semantic map'. *Linguistic Typology* 2: 79-124.
Vander Klok, J. 2014. Modal questionnaire for cross-linguistic use. https://www.eva.mpg.de/lingua/tools-at-lingboard/pdf/Modal_Questionnaire_CrossLing_JVK.pdf (accessed 7 August 2024).
Vander Klok, J. 2022. Revised modal questionnaire for cross-linguistic use. http://tulquest.huma-num.fr/en/node/70 (accessed 7 August 2024).
Zeller, J. 2006. 'Raising out of finite CP in Nguni: The case of *fanele*'. *Southern African Linguistics and Applied Language Studies* 24(3): 255-275.

Acknowledgements

We would like to thank the editors and staff of *Africana Linguistica* for their wonderful help throughout the publication process. We also extend heartfelt thanks to two anonymous reviewers for their detailed and thoughtful comments, which greatly improved this paper. Rasmus Bernander's contribution was supported by the Riksbankens Jubileumsfond (grant number P23-0101), here gratefully acknowledged. Thera Crane's work was supported by a grant from the Finnish Research Council.

Authors' addresses

Thera Marie Crane
Department of Languages
University of Helsinki, Finland
Department of African Languages
University of Pretoria
thera.crane@helsinki.fi

Stefan Savić
Department of German and Scandinavian Studies
University of Zurich, Switzerland
stefan.savic@ds.uzh.ch

Onelisa Slater
Department of Linguistics and Applied Language Studies
Rhodes University, South Africa
onelisaslater@outlook.com

Rasmus Bernander
Department of Languages and Literatures
University of Gothenburg, Sweden
rasmus.bernander@sprak.gu.se

Résumé

Cette étude est le fruit d'une recherche consacrée aux expressions modales de nécessité du xhosa, une langue bantu d'Afrique du Sud. Combinant une analyse d'occurrences réalisée dans un corpus limité et les intuitions de locuteur natif de l'un des auteurs de ce rapport, nous décrivons les modèles et les restrictions de la distribution de deux marqueurs de nécessité de type « auxiliaires » : **fanele** et **funeka**. Nous formulons également quelques notes sur le marqueur de nécessité **mele**. Les différences les plus significatives dans les significations et les utilisations de ces marqueurs se situent dans le domaine interne du participant et dans le domaine épistémique, et les significations lexicales (toujours actives en xhosa, en plus des fonctions modales des marqueurs) sont évidentes dans les distributions de leur utilisation.