BRILL

LANGUAGE DYNAMICS AND CHANGE 13 (2023) 74–131

brill.com/ldc

Divergence and contact in Southern Bantu language and population history

A new phylogeny in cross-disciplinary perspective

Hilde Gunnink
Leiden University Centre for Linguistics, Leiden University, Leiden, The Netherlands; Department of Languages and Cultures, UGent Centre for Bantu Studies (BantUGent), Ghent, Belgium
h.gunnink@hum.leidenuniv.nl

Natalia Chousou-Polydouri
Department of Comparative Language Science and Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Zürich-Oerlikon, Switzerland
natalia.chousou-polydouri@uzh.ch

Koen Bostoen
Department of Languages and Cultures, UGent Centre for Bantu Studies (BantUGent), Ghent, Belgium
koen.bostoen@ugent.be

Abstract

In this paper we present a new, lexicon-based phylogeny of 34 Southern Bantu languages, and combine it with previous insights from linguistics, archaeology, and genetics to study the history of Southern Bantu languages and their speakers. Our phylogeny shows all Southern Bantu languages to derive from a single, direct ancestor, which contrasts with archaeological evidence indicating separate migrations of Bantu speakers into southern Africa. This suggests that the Bantu languages spoken by the first migrants became extinct, and the ancestor of present-day Southern Bantu languages only emerged in southern Africa during the second millennium CE. We also map the distribution of previously established or suspected Khoisan-derived linguistic features in Southern Bantu languages onto this phylogeny. Evidence for intensive contact with speakers of Khoisan languages also comes from population genetics, which shows that

Published with license by Koninklijke Brill NV | DOI:10.1163/22105832-bja10022
 | ISSN: 2210-5824 (print) 2210-5832 (online)

Khoisan linguistic influence is mainly seen in languages spoken by populations displaying a higher degree of genetic admixture.

Keywords

Bantu – southern Africa – linguistic phylogeny – historical linguistics – language contact – archaeology – population genetics

1 Introduction

Languages of the close-knit Bantu subfamily of the Niger-Congo phylum dominate the linguistic landscape of most of (sub)equatorial Africa. Their wide geographic distribution contrasts with their relatively recent origin in the Nigeria-Cameroon border area some 4,000 to 5,000 years ago. From there, Bantu speakers spread eastward and southward, ultimately reaching southern Africa some 2,000 years ago (Bostoen, 2018). Southern Bantu languages, which belong to the Eastern branch of Bantu, are commonly seen as the direct descendants of the earliest Bantu languages introduced in the southeastern part of southern Africa (Grollemund et al., 2015), a widely held belief which we challenge in this article. Southern Bantu languages have long been recognized to consist of six distinct language clusters: Shona, Venda, Sotho, Nguni, Tsonga, and Copi (Doke, 1954; Guthrie, 1948). However, existing classifications (e.g., Bastin et al., 1999; Currie et al., 2013; Grollemund et al., 2015) include a relatively small number of Southern Bantu languages. Uncertainty therefore persists on the internal classification of Southern Bantu, an inadequacy we wish to remedy in this article.

One of the factors that have contributed to diachronic change and diversification in Southern Bantu languages is contact with resident speakers of various Khoisan languages. Southern African Khoisan languages were traditionally seen as a single language family (Greenberg, 1963), but have been increasingly recognized as belonging to three different families: Tuu, Kx'a, and Khoe-Kwadi (Güldemann, 2014). The term "Khoisan" is maintained, however, to refer to the non-Bantu indigenous languages of southern Africa that share the typologically highly unusual feature of click phonemes (Witzlack-Makarevich and Nakagawa, 2019; Güldemann, 2014). Although recognized as the hallmark of Khoisan languages, clicks also occur in certain Southern Bantu languages. Their absence in Bantu languages outside southern Africa and their ubiquitous use in Khoisan languages make clicks a clear sign of (past) language contact between Khoisan and Bantu (Herbert, 1990; Pakendorf et al., 2017; Sands and

Güldemann, 2009; Vossen, 1997). Other, less typologically salient, phonological, morphological, and lexical developments in Southern Bantu languages are likely of Khoisan origin, but these are much harder to establish. As many Khoisan languages have become extinct with little to no documentation, possible contact-induced features are difficult to trace to an ultimate source. In addition, as the exact genealogical relationships of Southern Bantu languages are mostly unknown, inherited and contact-induced linguistic features are often difficult to distinguish. Even when it is possible to attribute a particular linguistic change to language contact with one or more Khoisan languages, a well-established Southern Bantu family tree that would allow us to link such contact-induced changes to either an individual language or an earlier common ancestor is missing, another deficiency we wish to address in this article. For instance, the attestation of clicks in both the entire Nguni cluster and one language of the Sotho cluster seems to suggest that these loan phonemes were adopted relatively early (Pakendorf et al., 2017). Establishing the internal relationships of the Nguni languages and understanding the historical relationships between the Nguni and Sotho clusters will facilitate the historical reconstruction of click acquisition as well as other possible contact-induced developments in the history of the Southern Bantu languages.

In this article, we provide a new, lexicon-based linguistic phylogeny of 34 Southern Bantu languages. Firstly, we show the genealogical unity of the Southern Bantu clade within Eastern Bantu. All Southern Bantu languages directly descend from a common ancestor that is unique to them. We also provide support for the widely accepted division of Southern Bantu into six subgroups, and add new insights into the relationships within and between subgroups. We discuss a number of previously described Southern Bantu sound changes in light of this lexicon-based phylogeny, and their significance for Southern Bantu subclassification. Secondly, we investigate the extent of influence from Khoisan languages on Southern Bantu languages, showing that some cases of Khoisan influence can be projected back to earlier nodes in the family tree, but others rather appear to be of more recent origin. Finally, we also compare our linguistic phylogeny with earlier findings from archaeology and genetics. The apparent mismatch between the archaeological and linguistic data provides evidence for a situation of multiple migration events into southern Africa involving language shift and language death. Furthermore, comparison with genetic data shows that Khoisan-derived linguistic changes are strongest in languages spoken by populations that also display a high degree of Khoisan genetic admixture.

This paper is structured as follows. In Section 2, we introduce the Southern Bantu languages and discuss the selection of languages included in this study

and the sources of our data. In Section 3, we introduce a new linguistic phylogeny of 34 southern African Bantu languages and language varieties, based on 100 basic lexical concepts. In Section 3.1, we provide an overview of earlier classifications which included Southern Bantu languages. The methodology is discussed in Section 3.2, and Section 3.3 describes the main results. In Section 4 we focus on suspected or proven contact-induced changes in Southern Bantu languages, and how these are distributed across the phylogenetic tree. In Section 5 we discuss the implications of our findings for the history of the Southern Bantu languages, also drawing on earlier archaeological and genetic research. Conclusions are presented in Section 6.

2 Southern Bantu languages

Bantu languages spoken in South Africa and their direct relatives spoken in the surrounding countries Lesotho, Eswatini, Zimbabwe, Mozambique and Botswana are usually referred to as "Southern Bantu languages" (Doke, 1954). Most of these languages are grouped into zone S according to the well-known referential classification of the Bantu languages originally devised by Malcolm Guthrie (Guthrie, 1971; Hammarström, 2019; Maho, 2009). This classification divides Bantu-speaking Africa into geographic zones. Each zone is indicated with a letter A-S, and groups of languages within each zone with multiples of 10. Individual languages are indicated with a two-digit number, or three-digit number in case of languages added in later updates of the classification. Varieties (or "dialects") of the same language can be indicated with letters following the number. To illustrate, the Nguni cluster of zone S is indicated as S40, with Zulu receiving the code S42, and the KwaZulu-Natal variety of Zulu the code S42A.[1] Guthrie subdivided Bantu languages into geographic zones and groups without the implication that languages of each zone and group also share a closer genealogical relationship with each other than with languages classified in other zones and groups. For Southern Bantu, however, Guthrie's referential classification, which is largely indebted to the historical-comparative research of the early twentieth-century linguist Clement Doke (as summarized in Doke, 1954), does reflect language genealogy quite well.

Southern Bantu languages are traditionally divided into six groups or language clusters: Shona (S10), Venda (S20), Sotho (S30), Nguni (S40), Tsonga

1 Varieties having a code ending in a capital letter were added to the referential classification of the Bantu languages by Maho (2003, 2009) in contrast to those ending in a small letter which were already in Guthrie's original inventory (Guthrie, 1948, 1971).

(S50), and Copi (S60). The languages within each of the individual clusters are very closely related or even mutually intelligible, and the reasons for considering certain varieties as separate languages may be sociopolitical rather than linguistic, especially for the Sotho and Nguni clusters (Alexander, 1998; Chebanne, 2003).

The Shona group (S10) includes a large number of languages and varieties that were harmonized into a single language, Shona, due to the efforts of Doke (cf. Doke and Chimhundu, 2005). Although this standardization has taken hold, especially in education and official policy, the different Shona languages are still well differentiated on the ground. They include Korekore (S11), Zezuru (S12), Manyika (S13), Karanga (S14), Ndau (S15), and Kalanga (S16),[2] each with numerous varieties.

S20 contains a single language, Venda (S21). Venda shares some linguistic characteristics with both Sotho (S30) and Shona (S10), which also fits its geographic position (Jones-Phillipson, 1972; Mulaudzi, 2004; Wentzel, 1981).

S30 is known as the Sotho or Sotho-Tswana group, and includes the larger languages Tswana (S31), Northern Sotho (S32), and Southern Sotho (S33), as well as a number of smaller languages and subvarieties (see Table 1). Lozi (K21) is spoken in Zambia and Namibia, and is the result of a migration of an originally Southern Sotho-speaking group into western Zambia, where they underwent extensive contact with a local Luyi-speaking population. Although this resulted in extensive restructuring of the language (Gowlett, 1989), its Sotho affiliation is still clear, and the language differs markedly from surrounding western Zambian languages (de Luna, 2010; Seidel, 2005). Our current classification also confirms the Sotho membership of Lozi (see Fig. 1).

S40 is known as the Nguni group and includes a large number of languages and varieties spoken in South Africa, Eswatini, and Zimbabwe. The larger Nguni languages are Xhosa (S41), Zulu (S42), Swati (S43), and Southern Ndebele (S407), which are official languages in South Africa, and Zimbabwean Ndebele (S44), which is the result of a nineteenth-century northward migration starting in South Africa. Many other, smaller Nguni languages have been listed in the literature (Doke, 1954; Hammarström, 2019; Maho, 2009), but only those on which enough data are available are included in this study, as listed in Table 1.

2 Kalanga, though it clearly has a close affiliation to the other Shona languages, was never integrated into the harmonization process, because it was partially spoken in Botswana (then the Bechuanaland Protectorate) rather than Zimbabwe (then Southern Rhodesia), and because the language policy at the time envisioned the use of Ndebele rather than Shona for the area where Kalanga was dominant (Doke and Chimhundu, 2005).

TABLE 1 Southern Bantu languages included in this study

Language	Guthrie code	Glottocode	ISO code	Source
Zezuru	S12	zezu1238		Borland (1970)
Manyika	S13	many1258	mxc	Borland (1970)
Karanga	S14	kara1480		Borland (1970)
Ndau	S15	ndau1241	ndc	Borland (1970)
Kalanga	S16	kala1384	kck	Mathangwane (1999)
Nambya	S16B	namb1291	nmq	Anderson et al. (2014)
Venda	S21	vend1245	ven	Wentzel and Muloiwa (1982)
Kutswe	S302	khut1241		Ziervogel (1954)
Pai	S303	paii1250		Taljaard (1997); Ziervogel (1954)
Pulana	S304	pula1264		Ziervogel (1954)
Lozi	K21	lozi1239	loz	Burger (1960)
Tswana	S31	tswa1253	tsn	Creissels (n.d.)
Tawana	S31c	tawa1269		Gunnink field notes (2019)
Kgalagadi	S311	kgal1244	xkv	Lukusa and Monaka (2008)
Tjhauba	S311	tjha1234		Gunnink field notes (2019)
Northern Sotho	S32	pedi1238	nso	Kriel (1976)
Lobedu	S32b	lobe1242		Kotzé (1995; 2001)
Southern Sotho	S33	sout2807	sot	Mabille and Dieterlen (1961)
Bhaca	S402	bhac1238		Hallowes (1942); Jordan (1953); Msimang (1989)
Phuthi	S404	phut1246		Donnelly (2007)
Nhlangwini	S405	nhla1239		Msimang (1989); P.J.N. Zungu (1989)
North Lala	S406	lala1263		E.M. Zungu (1999)
South Lala	S406	lala1263		E.M. Zungu (1999)
Southern Ndebele	S407	sout3270	nbl	Iziko Lesihlathululi-Mezwi Sesindebele (2006)
Northern Ndebele	S408	sout2808		Msimang (1989); Skhosana (2009); Ziervogel (1959)
Xhosa	S41	xhos1239	xho	Fischer et al. (1985)
Zulu	S42	zulu1248	zul	Doke et al. (1990)
Swati	S43	swat1243	ssw	Rycroft (1981)
Zimbabwean Ndebele	S44	nort2795	nde	Pelling (1966)
Tswa	S51	tswa1255	tsc	Bastin et al. (1999)
Tsonga	S53	tson1249	tso	Chatelain (1978)
Ronga	S54	rong1268	rng	Quintão (1951)
Copi	S61	chop1243	cce	dos Santos (1941); Junod (1930)
(Mozambican) Tonga	S62	gito1238	toh	Lanham (1955)

The Tsonga group (S50) includes Tswa (S51), Tsonga (S53, also known as Shangaan), and Ronga (S54).

Finally, the Copi group, S60, includes only Copi (S61) and Tonga (S62); the latter should be confused neither with Tsonga (S53), which is occasionally referred to with the exonym Thonga, nor with the Tonga spoken in Zambia (M64), nor that in Malawi (N15), both of which are distinct languages.

Table 1 presents all the Southern Bantu languages included in this linguistic phylogeny and the sources from which the data are taken. Guthrie codes and the names of subgroups are adopted from Hammarström (2019), and glottocodes are taken from Hammarström et al. (2020). Where language names are not attested in Hammarström (2019), the language name listed by the source (in the case of published data) or given by the speakers (in the case of field data) is used. In the case of identical language names for distinct languages, a geographic label was added in order to disambiguate.

Some of the data used in this study were collected through fieldwork in Botswana in 2019 by the first author, but most data are drawn from published sources. As such, the choice of languages to be included is restricted by the availability of data. Smaller languages, and especially languages that are considered "non-standard" varieties of larger, sociopolitically dominant languages, are often not well documented or not documented at all.

3 The phylogenetic classification of Southern Bantu

In this section, we present our phylogenetic classification of southern African Bantu languages. Section 3.1 presents earlier classifications that included Southern Bantu languages, showing that the sample of Southern Bantu languages included is often too small or unbalanced to provide a good insight into the internal classification of Southern Bantu. Section 3.2 discusses the methodology and the way the data were collected and coded. Section 3.3 presents and discusses the resultant linguistic phylogeny, and compares the results with earlier proposals for Southern Bantu (sub)classifications in conjunction with other types of historical linguistic evidence, such as common sound shifts and morphological innovations.

3.1 *Earlier classifications of Southern Bantu*

In this section, we review earlier classifications which focus on Southern Bantu languages or include Southern Bantu languages in a wider classification. For an overview and evaluation of Bantu-wide classifications, see also Philippson and Grollemund (2019).

Early classifications are concerned with the unity, or lack thereof, of the Southern Bantu languages. Ehret's (1972) lexicostatistic study of cognacy between the Shona, Venda, Sotho, Nguni, Tsonga, and Copi clusters prompts him to exclude Shona from Southern Bantu, despite a relatively high degree of cognacy with Venda, which is attributed to contact rather than inheritance. On the other hand, the lexicostatistic study by Borland (1986), which is based on cognacy in word lists of various sizes, argues for the inclusion of Shona within Southern Bantu, albeit as its most divergent branch.

More recent Bantu-wide classifications include various samples of Southern Bantu languages. Bastin et al. (1999) provide a lexicostatistic classification making use of cognacy for 92 basic semantic concepts in 542 word lists representing some 450 different Bantu languages. Another Bantu-wide classification based on the same data set, but making use of Bayesian phylogenetic methods, is Currie et al. (2013). In both classifications, the 23 Southern Bantu languages included group together as a single genetic cluster to the exclusion of all other Bantu languages. The traditionally recognized Shona, Venda, Sotho, Nguni, and Tsonga subgroups also emerge in both classifications. However, in Currie et al. (2013), Konde, which was originally thought to belong to the Tsonga cluster, is linked to the Nguni group, although with a low (0.72) support value. Unfortunately, no languages from the Copi group are included in this data set, and the relatively large Sotho and Nguni clusters are only represented by four and five languages respectively. Other classifications based on the data set of Bastin et al. (1999) are based on a smaller sample of languages, including even fewer Southern Bantu languages. Nonetheless, they yield very similar results. Holden (2002) applies maximum parsimony to obtain a phylogeny of 73 Bantu languages, including 12 Southern Bantu languages; Holden and Gray (2006) apply Bayesian methods to a slightly larger sample of 93 Bantu languages, including the same 12 Southern Bantu languages; and Rexová et al. (2006), which combine lexical data with phonological and morphosyntactic features, provide a phylogenetic analysis of 87 languages, including six Southern Bantu languages, using both a maximum parsimony and a Bayesian method. In all these classifications, Southern Bantu languages emerge as a single, monophyletic group to the exclusion of all other Bantu languages.

Grollemund et al. (2015) furnish the most comprehensive, up-to-date phylogenetic classification, including 409 Bantu languages, of which 11 belong to Southern Bantu.[3] Making use of a tailored list of 100 semantic concepts, trans-

3 At the time this article was sent off for production, the new phylogeographic analysis of the Bantu Expansion by Koile et al. (2022), published online August 1, 2022, was not available yet. It is in line with Grollemund et al. (2015) as far as Southern Bantu is concerned.

lation equivalents for each concept are collected for each language, and cognacy is established on the basis of the comparative method where possible, or impressionistically elsewhere. In this classification, the 11 Southern Bantu languages included also form a single monophyletic group; despite the small data sample, the Shona, Venda, Sotho, Nguni, and Tsonga clusters clearly emerge distinctly (no languages of the Copi group were included), with the Shona languages as a sister branch to the remaining Southern Bantu languages, and the Tsonga and Nguni clusters forming a subgroup.

Most recently, Semo et al. (2020) provide a phylogenetic classification of 21 Bantu languages spoken in and around Mozambique, including eight Southern Bantu languages, which are grouped together to the exclusion of the other Bantu languages included in the analysis. This linguistic phylogeny also supports Shona as the first branch within Southern Bantu, and provides support for the relationship between the languages of the Tsonga and Copi clusters.

While the genetic unity of Southern Bantu emerges from different studies, despite differences in sample size, data, or method, earlier studies of (Southern) Bantu classification are limited in the number of Southern Bantu languages that are included. As such, several uncertainties remain on the internal subclassification of Southern Bantu, such as the genealogical relationship of the six clusters with respect to each other, or the internal organization of the relatively large Sotho and Nguni clusters. Moreover, a larger, more representative sample of languages can considerably alter the typology of linguistic trees. For instance, Pacchiarotti et al. (2019) show that the internal classification of West-Coastal Bantu languages is substantially altered by the inclusion of a larger number of languages belonging to the B80 cluster, which turns out to host the branch's highest internal diversity. A phylogenetic classification based on the largest possible sample of languages, as we propose for Southern Bantu in this study, is therefore essential for a thorough understanding of the historical relationships of these languages, and certainly so if one wishes to juxtapose this with other types of evidence, from both inside and outside linguistics.

3.2 *Methodology*

The new phylogenetic classification presented here includes 79 Bantu languages. Of them, 34 are Southern Bantu—that is, Guthrie's zone S plus Lozi (K21)—from all six traditionally recognized Southern Bantu clusters. This is the largest sample of Southern Bantu languages used in a classification to date, and includes all Southern Bantu languages on which reliable data are available. The remaining 45 languages are from different Bantu subgroups, according to the Bantu-wide phylogeny of Grollemund et al. (2015): 18 belong to Eastern Bantu (of which Southern Bantu is also part), eight to South-Western Bantu, three

to West-Western Bantu, three to Central-Western[4] and one to North-Western. The remaining 12 languages (Fwe, Subiya, Totela, Mbukushu, Ila, Koti, Cuwabo, Humbe, Kwandu, Kuvale, Mbandja, Mbalanhu) were not included in the classification of Grollemund et al. (2015), but, considering their geographic location and the classification of their sister languages, are likely to be part of the South-Western and Eastern Bantu groups as well. Since we were interested in both the validity (monophyly) and placement of Southern Bantu as a genealogical subgroup, as well as in its internal classification, we sampled almost exhaustively our in-group (Southern Bantu) and extensively various potential out-groups, as well as representatives from all major Bantu branches according to Grollemund et al. (2015). In order to root the tree, we used a North-Western Bantu language, Duala (A24).

For all the languages in our sample we collected lexemes for 100 concepts. This list is partially based on the reduced Swadesh-100 list of 92 basic vocabulary items that was initially used for the lexicostatistic Bantu classification of Bastin et al. (1999), and in certain subsequent lexicon-based phylogenetic classifications of specific Bantu branches and subgroups (de Schryver et al., 2015; Pacchiarotti et al., 2019). Eleven concepts ('to fly,' 'good,' 'ground,' 'lie down,' 'red,' 'round,' 'say,' 'sleep,' 'star,' 'to swim,' 'warm') were removed from this list of 92 concepts, either because they were not attested in many less well-documented Southern Bantu languages, or because the concept was often not covered by a single lexeme, but rather semantically subdivided into different concepts, each with different, unrelated lexemes. For instance, the concept 'good' was found in many languages to be divided into 'correct,' 'morally right,' 'beautiful,' 'tasty,' and even others. Since many data sources do not provide detailed semantic information, there is a risk of comparing lexemes that have the same basic English translation yet express semantically distinct concepts in the languages in question. In order to avoid this possible methodological pitfall, concepts for which this semantic "vagueness" was frequently observed were excluded.

Another 19 concepts ('to dig,' 'elephant,' 'finger,' 'to fall,' 'fly,' 'to go to,' 'house,' 'to laugh,' 'milk,' 'to rain,' 'river,' 'short,' 'to be sick,' 'to sleep,' 'snake,' 'to speak,' 'to steal,' 'thirst,' 'wind') were added. These were impressionistically found to be well attested, even in less well-documented languages. They include concepts such as 'to dig,' 'to fall,' and 'finger' that are known to be relatively resistant to borrowing as attested by a low borrowability score (Haspelmath and Tadmor, 2009). Other concepts that were added are known to be relatively stable

4 Ntomba was not included in Grollemund et al.'s (2015) phylogenetic classification, but earlier lexicon-based classifications of Bantu have placed it in the Central-Western subgroup (Vansina, 1995; Bastin et al., 1999).

in the context of southern African Bantu languages, such as 'elephant,' 'snake,' 'milk,' and 'house.' The full 100-concept list for all languages, as well as the cognate judgments, can be found in the supplementary materials. The creation and curation of the CLDF data set, as well as the exportation of coded matrices for phylogenetic analyses, were done with Lexedata version 1.0.0-beta3 (Kaiping et al., 2021).

Although the list of concepts was tailored to focus on words most likely to be attested in Southern Bantu, the uneven state of documentation unfortunately led to concepts being unattested in certain languages. Out of 7,900 potential lexemes (100 concepts for 79 language varieties) in total, 606 (about 7.7 percent) were unattested in the available sources and were coded as missing data for the phylogenetic analysis.

Similarly, many sources provided more than one lexeme for a particular concept. Where possible, the most semantically basic lexeme was selected. For instance, the concept 'ashes' in Lozi is attested as *mulola* 'ashes' and *lupa* 'hot ashes' (Burger, 1960); since the meaning of *mulola* 'ashes' is the most basic, only this lexeme was included. Where no judgment could be made on which lexeme was the most basic, all were included.[5]

Finally, as our data collection focused on a targeted list of semantic concepts, cognate forms that have undergone semantic shifts are excluded from our data set or more rarely are included under a different concept. While it is possible to include as many of such forms as possible and correspondingly make the characters used in the analyses more independent (Michael et al., 2015; Birchall et al., 2016; Chousou-Polydouri et al., 2016), it is a huge undertaking for so many languages, as it involves extensive collection of near-synonymous or related concepts, searching for expected forms based on sound correspondences, and is to a large degree dependent on the quality and availability of documentation. Dictionaries are only available for 13 of the 34 Southern Bantu languages included in this study. For the remaining 21 languages, no full-sized dictionaries are available, and data are rather sourced from grammars, word lists, and MA or PhD theses. For these languages, available data is therefore limited. Despite this shortcoming, we consider our approach a valuable step towards a character-based lexical phylogeny of Southern Bantu languages, as our data set is the most complete to date in terms of languages included in the analysis. Cognacy judgments were performed based on resemblance in form within each concept.

5 This results in multiple cognate sets being associated with a single concept in a single language. A language that has two synonyms for concept X would be coded as "present" for the character X1 (associating one root with this concept) and also "present" for the character X2 (associating the other root with the same concept).

Where possible, forms were checked against known sound shifts in order to prove or disprove cognacy, but as the diachronic phonology of most languages involved in the study is not (thoroughly) known, many cognacy judgments had to be done impressionistically. More careful cognacy judgments, based on regular sound changes that are yet to be established, would possibly improve the analysis, but this is left for future research.

Each association between cognate class and concept was then coded as a binary character, resulting in a total of 1,431 characters. The evolution of the characters was modeled with a simple binary (restriction site) model (with "noabsencesites" ascertainment correction, since all absent characters are not observable), which allows asymmetry in the rates of gain and loss (Felsenstein, 1992). We also evaluated using Bayes factors, as approximated via stepping-stone sampling, the addition of across-character rate heterogeneity (modeled as a gamma distribution as approximated by four rate categories).

We analyzed the character matrix using Bayesian inference, as implemented in MrBayes version 3.2 (Ronquist et al., 2012; Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). We used default priors for all parameters: Dirichlet (1, 1) for the state frequencies, exponential ($\lambda = 1$) for the shape (alpha) parameter of the gamma distribution, uniform for the topology, and compound Dirichlet[6] for the branch lengths. We ran four independent runs of 10 million generations each, sampling every 1,000 generations for each analysis. In all cases, the average standard deviation of split frequencies fell below 0.01, ensuring that the sampling chains had converged. We also visually assessed convergence and mixing using Tracer version 1.7.1 (Rambaut et al., 2018) and summarized the posterior sample using majority-rule consensus trees with 25 percent of the sample discarded as burn-in. We visualized the summary trees using FigTree version 1.4.4 (Rambaut, 2018).[7] Finally, we performed and visualized parsimony ancestral state reconstructions on the majority-rule consensus tree for selected morphological and phonological features using Mesquite version 3.70 (Maddison and Maddison, 2021).[8] All features were modeled as unordered characters, except for the presence and productivity of the diminutive suffix, which was modeled as an ordered character (its states being absent, unproductive, productive).

6 More precisely, a uniform Dirichlet prior on branch length proportions, with the tree length drawn from a gamma distribution with mean 10. The corresponding setting is unconstrained:gammadir(1, 0.1, 1, 1).

7 Since we have not estimated divergence times, the branch lengths of our consensus trees represent expected amount of change (the product of rate and time).

8 For all visualizations of feature distribution and reconstructions we are using cladograms, so the branch lengths are uninformative.

3.3 *Results*

In this section, we present the results of our phylogenetic classification of Southern Bantu, and compare these to the findings of certain earlier classificatory works. The Bayes factor comparison showed decisive support for the inclusion of across-character rate heterogeneity: –14692.40 marginal likelihood in natural log units for the analysis including rate heterogeneity, vs. –14959.91 for the analysis without rate heterogeneity, resulting in a logarithmic Bayes factor of 2ln(BFhet/non-het) = 267.51 (Kass and Raftery, 1995). The majority-rule consensus tree based on the analysis which included gamma-distributed rate heterogeneity is shown in Fig. 1. In Section 3.3.1, we present the external classification of Southern Bantu, and the internal classification of its six traditionally recognized clusters. We subsequently focus on the internal subgrouping of the large and diversified Nguni (3.3.2) and Sotho (3.3.3) clusters.

3.3.1 The classification of Southern Bantu languages

Our new phylogeny situates Southern Bantu within the larger Eastern Bantu clade, which is in line with earlier lexicon-based classifications (e.g., Bastin et al., 1999; Grollemund et al., 2015, to cite only the most comprehensive ones). Furthermore, it clearly shows that all Southern Bantu languages share a direct and unique common ancestor, Proto-Southern Bantu. Earlier classifications also provided evidence for the coherence of Southern Bantu (see Section 3.1), and our phylogeny, on the basis of a larger sample of languages, strengthens this result.

Our linguistic phylogeny also confirms the subdivision of Southern Bantu into the six traditionally recognized subgroups Shona (S10), Venda (S20), Sotho (S30), Nguni (S40), Tsonga (S50), and Copi (S60). It further supports a close relationship between the Tsonga (S50) and Copi (S60) groups, which are spoken in contiguous areas in southern Mozambique (see Map 1). A link between the Nguni (S40) and Tsonga/Copi (S50–S60) clades is also suggested, though with a somewhat lower support value (0.91 posterior probability). Similarities between Nguni (S40) and Tsonga (S50) have been pointed out before, for instance in their phonological development (Baumbach, 1987: 2; Janson, 1991–1992). The Nguni and the Tsonga/Copi clades subsequently form a clade with Sotho (S30), which is sister to Venda (S20). A closer genealogical relationship between Venda (S20) and Sotho (S30), as suggested on the basis of certain shared phonological features (Janson, 1991–1992) or lexicon (Jones-Phillipson, 1972), is not in line with our classification. The observed similarities between Venda (S20) and Sotho (S30) may be the result of language contact rather than inheritance (Wentzel, 1981). Finally, Shona (S10) is sister to all other Southern Bantu languages, which supports earlier observations that Shona displays sig-

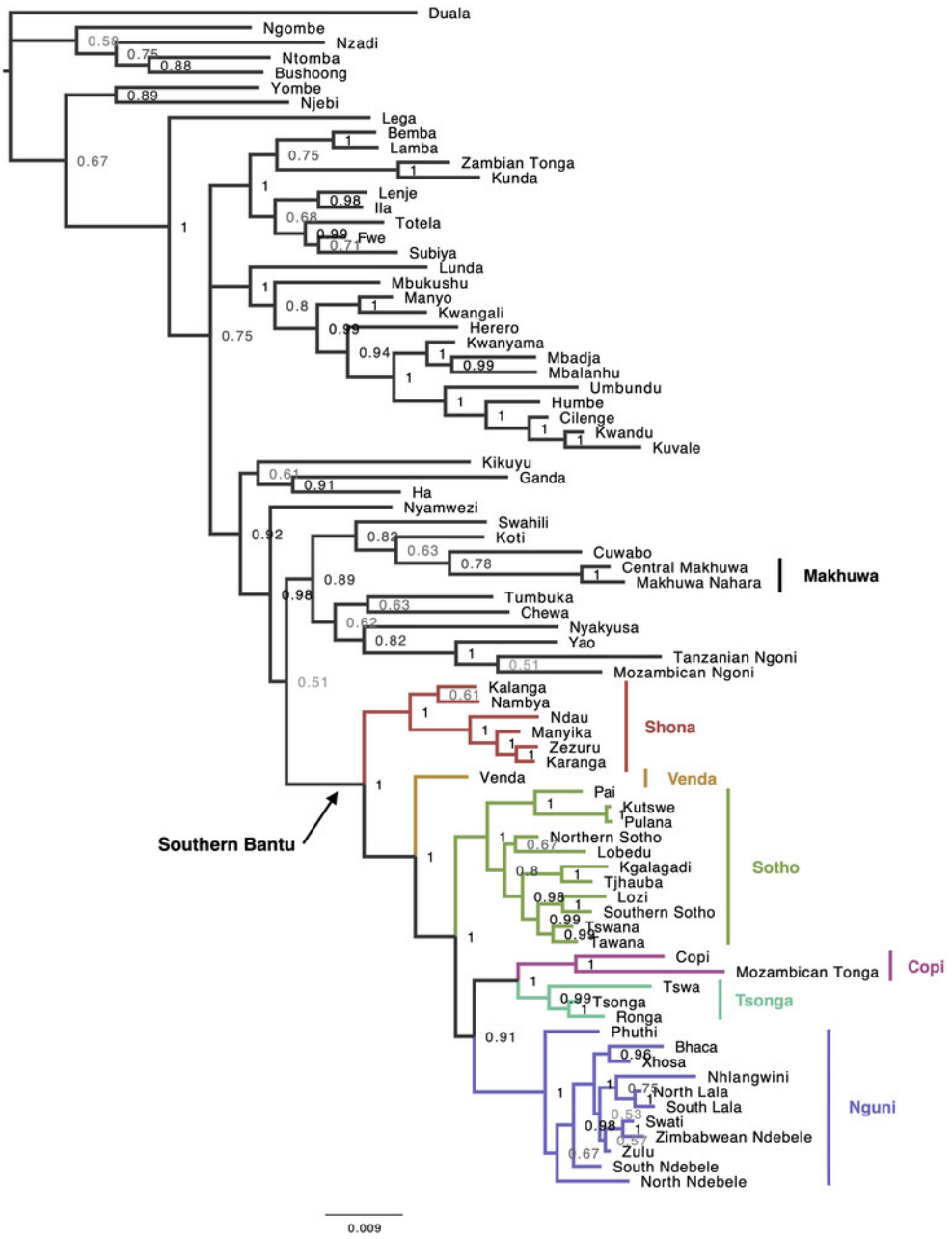
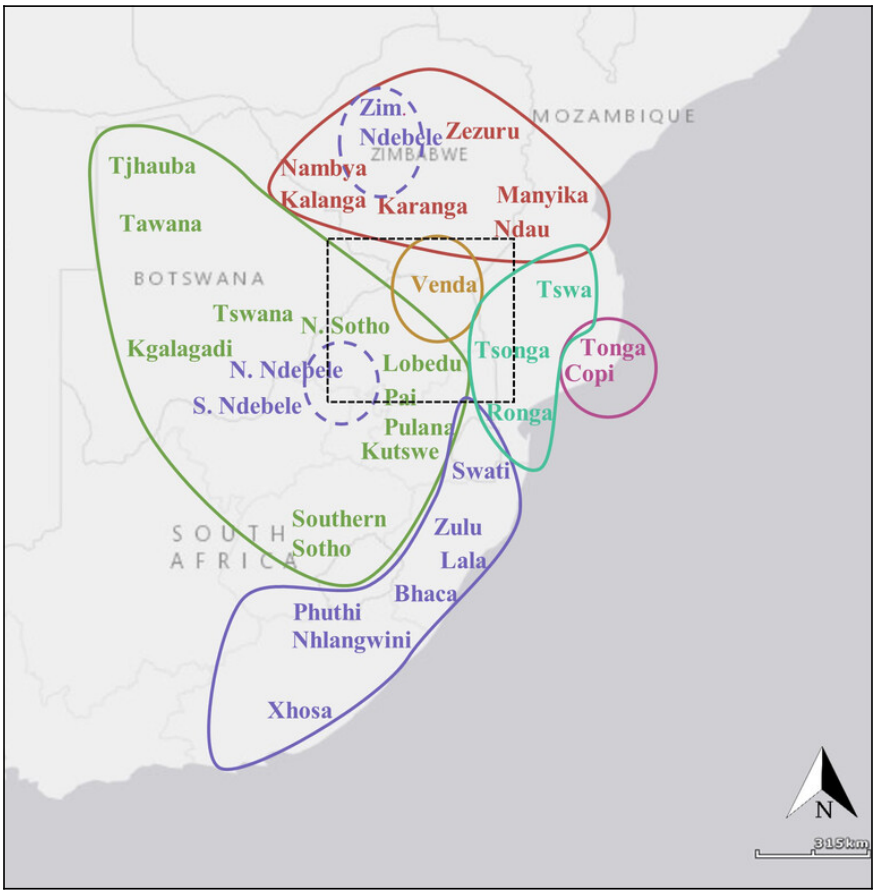


FIGURE 1 Majority-rule consensus tree based on the posterior sample of the phylogenetic analysis. Branch values are posterior probabilities. The main subgroups of Southern Bantu are highlighted in color.



MAP 1 The location of the six subgroups of Southern Bantu: Shona (S10; in red), Venda (S20; in yellow), Sotho (S30; in green), Nguni (S40; in purple), Tsonga (S50; in blue), and Copi (S60; in pink). The black square indicates the location where Proto-Southern-Bantu was most likely spoken.

nificant linguistic differences with respect to the remainder of Southern Bantu, which we will call "Nuclear Southern Bantu," that is, Southern Bantu minus Shona (S10). Being mainly spoken in Zimbabwe, Shona is at the northern border of the Southern Bantu-speaking region.

Our new lexicon-based phylogeny can be used as a baseline to compare with existing or future research on regular sound changes that define particular Southern Bantu subgroups. While an exhaustive comparison is beyond the scope of our research, a number of striking differences can be observed. Earlier studies in diachronic phonology did not succeed in identifying shared innovations that corroborate Southern Bantu as a discrete subclade within the

Bantu family (Janson, 1991–1992; van der Spuy, 1990). Although Southern Bantu languages share many phonological innovations suggesting common inheritance, their distribution across subgroups and their interaction with other sound changes do not support their reconstruction to Proto-Southern Bantu. This is mainly due to the distribution of Bantu Spirantization within Southern Bantu. Bantu Spirantization is a common change in Bantu languages where plosives changed into affricates or fricatives when directly followed by the Proto-Bantu high vowels *i and *u (Bostoen, 2008; Janson, 2007; Schadeberg, 1994–1995). This change was usually combined with a subsequent change in the vowel inventory, specifically the merger of the high vowels *i and *u with their near-high counterparts *ɪ and *ʊ, resulting in the reduction of the inherited system of seven vowel phonemes to only five. Spirantization and the subsequent seven-to-five-vowel merger affected all Southern Bantu languages except the Sotho group (S30; Janson, 1991–1992; van der Spuy, 1990). In the Southern Bantu languages that were affected by Bantu Spirantization, this change predated another series of changes which are specific to voiceless stops. This voiceless stop shift is also shared with the Sotho languages, which are not affected by Bantu Spirantization (Janson, 1991–1992: 82). The way Bantu Spirantization interacts with other sound changes therefore leads to a reconstruction of Proto-Southern Bantu segmental phonology which is virtually identical to that of Proto-Bantu (van der Spuy, 1990).

Although Bantu Spirantization cannot be reconstructed as a shared innovation dating back to the subgroup's most recent common ancestor, the way it evolved as a parallel phonological shift is indicative of subgrouping within Southern Bantu. As has been shown elsewhere in Bantu (Labroussi, 2000; Bostoen, 2009; Bostoen and Goes, 2019), sound changes accompanying Bantu Spirantization, such as the (partial) merger of places of articulation and spirant devoicing, are often more diagnostic for the internal classification of lower-level language groups than the consonant mutation in itself. Table 2 summarizes the outcomes of Bantu Spirantization in Southern Bantu following Janson (2007). Sotho (S30) is not included as Janson (2007) has shown it to be absent in this subclade. Blanks represent gaps in the evidence due to a lack of sufficient data.

First of all, the relatively high frequency of affricates in Table 2 seems to suggest that Bantu Spirantization is indeed a relatively recent development within Southern Bantu, as affricates are often seen as one of the first stages in the development of Bantu Spirantization (Mpiranya, 1997; Hyman, 2003; Janson, 2007). Second, the way in which Bantu Spirantization has developed in Shona (S10) and Venda (S20) is clearly distinct from how it evolved in Nguni (S40), Tsonga (S50), and Copi (S60). This corroborates the hypothesis that Bantu Spi-

TABLE 2 Bantu Spirantization in Southern Bantu apart from Sotho, following Janson (2007)

Proto-Bantu	*pu	*tu	*ku	*bu	*du	*gu	*pi	*ti	*ki	*bi	*di	*gi
Korekore (S11)	fu	fu	pfu	bvu	bvu	bvu	si	tsi	si	zi	dzi	zi
Zezuru (S12)	fu	fu	pfu	bvu	bvu	bvu	si	tsi	si	zi	dzi	zi
Manyika (S13)	fu	fu	pfu	bvu	bvu	bvu	si	tsi	si	zi	dzi	zi
Karanga (S14)	fu	fu	pfu	bvu	bvu	bvu	si	tsi	si	zi	dzi	zi
Ndau (S15)	fu	fu	pfu	bvu	bvu	bvu	si	tsi	si	zi	dzi	zi
Kalanga (S16)	fu	fu	pfu	bvu	bvu	bvu	si	tsi	si	zi	dzi	zi
Venda (S21)	fu	fu	fu	vu	bvu	vu	s̬wi	tshi	t̬si	z̬wi	d̬zi	
Xhosa (S41)	fu	fu	fu	vu	vu	vu	fi	si	si	vi	zi	zi
Zulu (S42)	fu	fu	fu	vu	vu	vu	fi	si	si	vi	zi	zi
Swati (S43)	fu	fu	fu	vu	vu	vu	fi	si	si			
Zimbabwean Ndebele (S44)	fu	fu	fu	vu	vu	vu	fi	si	si	vi	zi	zi
Tswa (S51)	fu	fu	fu	vu	dzu	vu	fi	si	si	vi	ti	
Tsonga (S53)	fu	fu	fu	vu	dzu	vu	fi	si	si	vi	ti	
Ronga (S54)	fu	fu	fu	vu	dzu	vu	fi	si	si	vi	ti	
Copi (S61)	fu	fu	fu	bvu		bvu		si	si	si	ti	ti
Tonga (S62)	fu	fu	fu	bvu		bvu		si	si	si	ti	ti

rantization is a parallel innovation that was initiated at least twice in Southern Bantu. Third, regarding our phylogeny, the distribution of Bantu Spirantization provides evidence that within the large subclade uniting Sotho (S30), Nguni (S40), Tsonga (S50), and Copi (S60), the last three subgroups are indeed more closely related to each other than to Sotho. Not only do they share Bantu Spirantization, while Sotho lacks it, but they also share a number of further Bantu Spirantization-related changes which are clearly distinct from Shona (S10) and Venda (S20). It could be posited that the partial merger of places of articulation in front of both *u and *i, as summarized in Table 3, goes back to their most recent common ancestor.

As for Shona (S10) and Venda (S20), their specific Bantu Spirantization outcomes clearly set them apart as distinct clades, even if their respective systems could be considered to be derived from a shared ancestral system that is best conserved in Shona (S10). Positing a Proto-Shona-Venda would be at odds with

TABLE 3 Bantu Spirantization in the ancestor of Nguni, Tsonga, and Copi

Proto-Bantu	*pu	*tu	*ku	*bu	*du	*gu	*pi	*ti	*ki	*bi	*di	*gi
Proto-Nguni-Tsonga-Copi	°fu	°fu	°fu	°bvu	°dzu	°bvu	°fi	°si	°si	°vi	°zi	°zi

our phylogeny in which Shona (S10) and Venda (S20) share no other ancestor than Proto-Southern Bantu, for which Bantu Spirantization cannot be reconstructed at all. It rather suggests that Bantu Spirantization was initiated three times, at the level of Proto-Shona, Proto-Venda, and Proto-Nguni-Tsonga-Copi. Considered in this way, the phonetic details of Bantu Spirantization do not contradict the genealogical unity of Southern Bantu languages. As Bantu Spirantization is known to have been initiated convergently across the Bantu area, often with parallel outcomes, it is plausible that it also emerged independently thrice within Southern Bantu.

3.3.2 The subclassification of the Nguni cluster

The Nguni languages (S40) have always been assumed to be closely related (see, e.g., Ownby, 1981, for an overview of early classifications of Nguni languages). This is confirmed in the current phylogenetic classification: all languages traditionally subsumed under Nguni cluster into a single highly supported clade. This includes a number of Nguni languages that are known to be strongly influenced by neighboring non-Nguni Bantu languages. Phuthi (S404) has undergone strong influence from Southern Sotho (S33; Donnelly, 1999; Mzamane, 1949), and Southern (S407) and Northern Ndebele (S408) have also been influenced by languages of the Sotho cluster (Grünthal et al., 2019; Wilkes, 1997). However, in spite of the effects of language contact, Phuthi, Northern Ndebele, and Southern Ndebele clearly retain enough of their Nguni basic vocabulary to be classified as Nguni.

Whereas the unity of the Nguni languages as a single cluster has long been recognized, the internal classification of Nguni languages is much more uncertain. Early classifications (Bryant, 1905; Doke, 1954) make use of phonological criteria to divide the Nguni languages into two groups, Zunda and Tekela, based on the observation that /z/ in Zunda regularly corresponds to /t/ in Tekela, as illustrated by the cognates in Table 4.

Another phonological correspondence that distinguishes Zunda languages from some of the Tekela languages is the one between alveolar stops and affricates (see Table 5). Unlike the /t/-/z/ correspondence, this sound correspondence is not homogeneously spread across all Tekela languages. Swati,

TABLE 4 The correspondence between /t/ and /z/ in Nguni languages

Zunda	**'parent'**	**'tooth'**	**'call'**
Zulu	*um-zali*	*i-zinyo*	*biza*
Xhosa	*um-zali*	*i-zinyo*	*biza*
Zimbabwean Ndebele	*um-zali*	*i-zinyo*	*biza*
Southern Ndebele	*um-zali*	*i-zinyo*	*biza*
Tekela	**'parent'**	**'tooth'**	**'call'**
Swati	*um-tali*	*li-tinyo*	*bita*
Phuthi	*mu-tadi*	*li-tinyo*	*bita*
Bhaca	*um-tali*	*ili-tinyo*	*bita*
Lala	*u-tali*	*li-tinyo*	*bita*
Northern Ndebele	*mu-tali*	*li-tinyo*	*bita*
Nhlangwini	*um-tali*	*i-tinyo*	*bita*

TABLE 5 The correspondence between alveolar stops and affricates in Nguni languages

Zunda	**'stick'**	**'three'**	**'long'**	**'matter'**	**'knee'**
Zulu	*ulu-t^{h}i*	*t^{h}at^{h}u*	*de*	*in-daba*	*i-dolo*
Tekela	**'stick'**	**'three'**	**long'**	**'matter'**	**'knee'**
Swati	*lu-ts^{h}i*	*ts^{h}atfu*	*dze*	*in-dzaba*	*li-dvolo*
Phuthi	*lu-ts^{h}i*	*ts^{h}atfu*	*dze*	*i-dzaba*	*li-dvolo*
Bhaca	*ulu-ts^{h}i*	*ts^{h}atfu*	*dze*	*in-dzaba*	*i-dvolo*
Lala	*lu-ts^{h}i*	*ts^{h}ats^{h}u*	*dze*	*in-dzaba*	*li-dzolo*
Northern Ndebele	*lu-t^{h}i*	*t^{h}at^{h}u*	*nde*	*in-daba*	*li-dolo*
Nhlangwini	*ulu-t^{h}i*	*t^{h}at^{h}u*	*de*	*in-daba*	*i-dolo*

Phuthi, and Bhaca use two allophones: a labiodentalized, heterorganic affricate /tf, dv/ before back vowels and an alveolar, homorganic affricate /ts^h, dz/ elsewhere. However, Lala exclusively uses the homorganic alveolar affricates /ts^h, dz/.[9] Northern Ndebele and Nhlangwini do not use affricates at all, but follow the Zunda pattern of using the alveolar plosives /t^h, d/. Zunda languages, on the other hand, invariably use plosives, so Table 5 only lists Zulu as a representative of the Zunda group.

It has also been suggested that the use of a single phonemic click type characterizes the Tekela languages, where Zunda languages distinguish multiple click types (Louw, 2013: 437; Louw and Ngidi, 1962: 46). While it is true that the Zunda languages distinguish three contrastive click types /ǀ, ǃ, ǁ/ (although this contrast is now being lost in Southern Ndebele; Schulz et al., 2019), the Tekela languages Phuthi, Bhaca, Lala, and Nhlangwini distinguish the same three contrastive click types (see among others Msimang, 1989). Only Swati uses a single click type /ǀ/, whereas Northern Ndebele does not make use of click phonemes at all (Ziervogel, 1959).

In sum, the Zunda-Tekela divide is not so well supported by diachronic phonological evidence as once was believed. The only phonological property that really distinguishes Tekela as a whole from Zunda is the regular correspondence of /t/ to /z/. Moreover, even this shared innovation is not unique to Tekela, as it also occurs in the Tsonga-Copi clade (Baumbach, 1987: 2; Msimang, 1989: 75). Hence, the opposition between Tekela languages and Zunda languages can be considered irrelevant for genealogical classification, which is confirmed by our phylogeny. Figure 2 shows that the Tekela languages do not share a common ancestor to the exclusion of Zunda languages. In fact, many of the closest and best supported clades unite Zunda and Tekela languages, for instance Bhaca (Tekela) and Xhosa (Zunda), and Swati (Tekela) and Zimbabwean Ndebele (Zunda). The phonological properties that characterize the Zunda-Tekela distinction are therefore more likely to have spread areally, which also explains their occurrence in the non-Nguni languages of the Tsonga and Copi groups, which border on Tekela Nguni languages such as Swati.

More recent classifications of Nguni languages have made use of lexicostatistics (Ownby, 1985; Jimenez, 2017). Similar to our linguistic phylogeny, these confirm the early divergence of Phuthi and Northern Ndebele, and in the case of Ownby (1985), Southern Ndebele. Jimenez (2017) divides Nguni

9 Wilkes (1981) describes affricates only for North Lala, and describes South Lala as using plosives. Zungu (1999), however, describes both North and South Lala as having affricates.

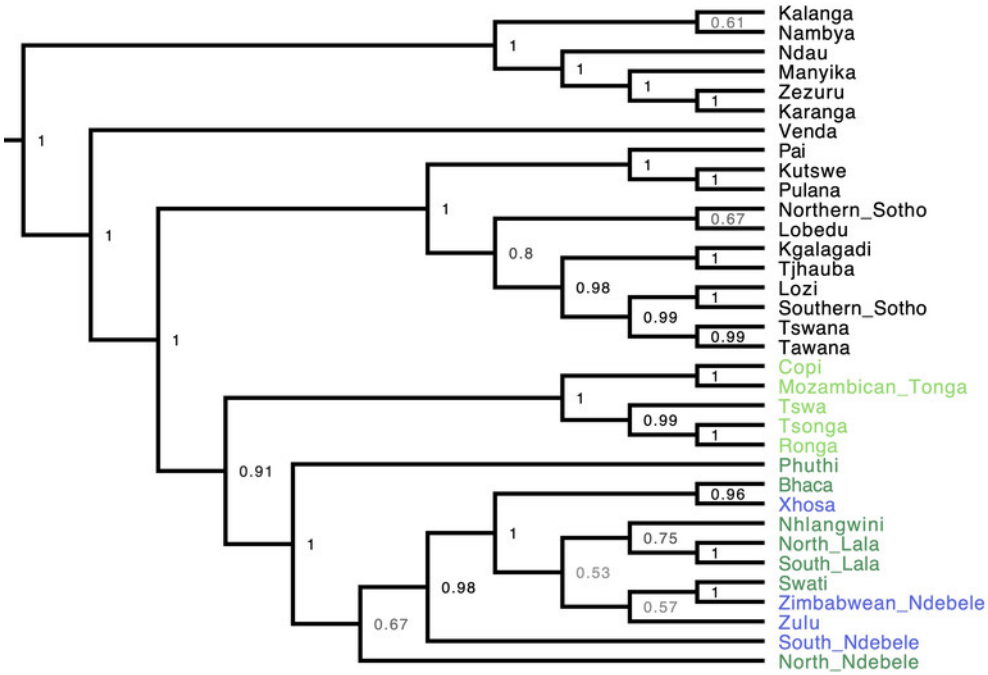


FIGURE 2 Distribution of Zunda (blue) and Tekela (dark green) languages in the Nguni subgroup. Languages in the Tsonga-Copi subgroup that also pattern with the Tekela languages are highlighted in light green.

into Northern Nguni (including Swati, Zimbabwean Ndebele, Southern Ndebele, and Hlubi) and Southern Nguni (including Bhaca, Mpondomise, Thembu, Xhosa, and Zulu). In our new phylogeny, a clade such as Southern Nguni is not ruled out, but the position of Southern Ndebele provides clear support against a Northern Nguni clade. Furthermore, the possibly early branching off of Lala, as suggested by Ownby (1985), is not supported by our linguistic phylogeny, which includes Lala in a clade with Bhaca, Xhosa, Nhlangwini, Swati, Zimbabwean Ndebele, and Zulu. Our phylogeny further weakly (0.75 posterior probability) supports Lala as a sister clade to Nhlangwini.

These discrepancies between earlier lexicostatistic classifications and our current linguistic phylogeny may be due to differences in both the data and the methodology. The method of lexicostatistics groups languages according to overall similarity, rather than focusing on shared innovations to establish genetic relationships. Furthermore, Donnelly (2007: 41), in his extensive work on Phuthi phonology, notes a number of errors in the Phuthi data used in Ownby (1985), which could affect her cognacy judgments. Discrepancies can also be observed between the data in Ownby (1985) and other more recent

sources. The Southern Ndebele dictionary (Iziko Lesihlathululi-Mezwi Sesindebele, 2006), for example, reveals different lexemes for eight concepts.

An explanation for the difficulty in classifying Lala may lie in the dubious status of the language. Aside from being a label for an ethnic group speaking a distinct language, Lala has also been used to refer to people living in a certain location, and to a particular social class (Wright, 2012). Even if a distinct speech variety called Lala does exist, there is some regional variation (Wilkes, 1981; E.M. Zungu, 1999), and there is ongoing influence from surrounding, more dominant languages, especially Xhosa and Zulu (E.M. Zungu, 1999). Lala also shows similarities to Tsonga (S50), for instance in its use of simple nasals instead of prenasalized voiceless stops (Msimang, 1989: 149–151). Our phylogeny, however, shows Lala to be affiliated with the other Nguni languages with very high posterior probability.

Another lexicostatistic study by Miti (1996) specifically focused on identifying the relationship between Nguni and the so-called Ngoni languages spoken in northern Mozambique, Malawi, and southern Tanzania. Ancestors of present-day Ngoni speakers migrated north due to a prolonged period of political disruption and warfare in South Africa in the 1820s and 1830s known in Zulu as *mfecane*. These Ngoni migrants maintained their Nguni language but did not impose it on the people they subjugated, who maintained their own East African Bantu languages. When the Ngoni lost their political power, they shifted to the language of their former serfs, but maintained the label Ngoni (Ngonyani, 2001). This Ngoni language is most closely related to East African rather than southern African Bantu languages, which is also shown in Miti's (1996) study: only Malawian Ngoni shows significant similarities with the South African Nguni languages Zulu and Swati, whereas Zambian and Tanzanian Ngoni do not. Our phylogeny, which only includes Tanzanian Ngoni (N12) and Mozambican Ngoni (N10), situates these two varieties outside the Nguni cluster. Both cluster with geographically adjacent Yao (P21). This confirms that modern-day Ngoni languages do not descend from Southern Bantu.

3.3.3 The subclassification of the Sotho cluster

Like Nguni, the Sotho or Sotho-Tswana languages (S30) have long been recognized as forming a single, closely related cluster (van Warmelo, 1935; van Wyk, 1969). This is confirmed by our phylogeny: all Sotho languages share a common ancestor that is unique to them. The Sotho languages Kutswe (S302), Pai (S303), and Pulana (S304) are known to have been influenced by the Nguni language Swati (S43; Taljaard, 1997; Ziervogel, 1954), but our phylogeny still recognizes them as Sotho, although they are sister to the remainder of the subclade.

Kgalagadi (S311) was originally seen as a variety of Tswana (S31; Doke, 1954), but later as a separate Sotho language (Auer, 1977; van Wyk, 1969: 176), which is confirmed by our classification. The Tjhauba variety of Kgalagadi (S311), spoken in northwestern Botswana, differs significantly from other varieties of Kgalagadi in lexicon and phonology (Lukusa and Monaka, 2008: 8; Gunnink, 2022b), but in our classification Tjhauba and Kgalagadi are still closely related sister languages.

One of the few earlier systematic attempts at subclassifying Sotho is the lexicostatistic study by Auer (1977), in which Kgalagadi is classified as sister to the other Sotho languages. The remaining languages fall into two clusters, one including, among others, Tswana and Southern Sotho, and the other including, among others, Northern Sotho and Lobedu. Although the set of Sotho languages included by Auer (1977) differs considerably from the set of Sotho languages included in our study, there are some similarities in the classification, for instance the relatively close relationship between Southern Sotho and Tswana. A significant difference is the position of Kgalagadi, which our classification shows to form a sister clade to Tswana and Southern Sotho, rather than an early branch. This may be due to a different methodology, as explained in the previous subsection. There is also a difference in data: about one-fourth of the lexical data used by Auer (1977) does not match the data used in this study. Kgalagadi is spoken over a vast area and displays extensive regiolectal differences, which might explain the discrepancies between the data sets. The Kgalagadi data used in this study, from Lukusa and Monaka (2008), represent the Ngologa variety spoken in the north of Botswana's Kgalagadi district, where the language is still vibrant (Monaka, 2013: 43). The data in Auer (1977) were collected in the South African village Bray, on the southern border of Botswana and relatively far from the core Kgalagadi-speaking area. A better regional coverage of Kgalagadi would clarify if certain varieties have been influenced by Tswana more than others.

A genealogical relationship has been proposed between the Sotho languages and the Makhuwa language cluster (P30), spoken in Southern Tanzania and Northern Mozambique (Janson, 1991–1992; Batibo et al., 1997). The evidence for this is mainly phonological, as Makhuwa and Sotho share a number of sound changes that are relatively rare in Bantu languages: the shift of reconstructed prenasalized voiced stops *mb, *nd, and *ng to simple voiceless stops /p, t, k/, and the dental reflexes of *c and *j. Our linguistic phylogeny does not support a close relationship between the Makhuwa and Sotho clusters: Koti (P311), Cuwabo (P34), and the two Makhuwa varieties included in this study (P31A and P31E) form a cluster that is a sister clade to Swahili (see Fig. 1). A possible explanation for these shared sound changes would be a shared, non-Bantu

substrate (as also suggested by Bailey, 1995: 47). Khoisan languages typically do not have prenasalized consonants, nor do they allow nasal-consonant clusters, so Khoisan influence could account for the loss of prenasalization. However, most Khoisan languages do contrast voiced and voiceless plosives, which makes the devoicing of voiced prenasalized stops in both Sotho and Makhuwa difficult to account for. Hence, if the loss of prenasalized stops in both Sotho and Makhuwa can be attributed to a shared non-Bantu substrate, the phonology of this unknown language would differ from those of attested southern African Khoisan languages.

4 Language contact in the history of the Southern Bantu languages

Having presented the genealogical relationships between Southern Bantu languages, we now turn to a number of phonological and morphological characteristics that are typical of Southern Bantu,[10] and that are known or suspected to have developed under influence from one or more Khoisan languages. We show that these features are widespread among different branches of Southern Bantu, but that it is often not possible to link their development to a single node in the phylogenetic tree; their distribution rather suggests horizontal transmission. The implications of these contact-induced changes for the history of the Southern Bantu languages are discussed in Section 5.

4.1 *Phonological features of Southern Bantu languages*

In this section, we discuss four typologically unusual phonological features in Southern Bantu, summarized in Table 6: lateral obstruents and dental stops as reflexes of the reconstructed Proto-Bantu phonemes *c and *j, labial velarization or palatalization, tonal depressors, and click phonemes. We discuss how these phonological features are distributed across Southern Bantu and show that this distribution mostly does not map directly onto our phylogenetic tree. Possible or established contact with Khoisan languages appears to have played a role in the spread of these phonological characteristics across languages of separate subgroups of Southern Bantu.

10 Lozi (K21) has been excluded from this section. While our lexicon-based phylogeny confirms the Sotho affiliation of this language, its phonology and morphology have been heavily restructured due to contact with neighboring western Zambian languages, particularly Luyana, resulting in the loss of almost all characteristically Southern Bantu features. As this is the result of a recent (nineteenth-century) and well-documented contact situation (cf. Gowlett, 1989), we do not consider Lozi in this section.

TABLE 6 Widespread phonological developments in Southern Bantu

	Reflexes of *c, *j	Palatalization or velarization of labials	Tonal depression	Clicks
Shona (S10)	s, ʃ, z, ʒ	velarization	yes	no
Venda (S20)	t̪, t̪ʰ, d̪	palatalization (with maintenance of labial)	unknown	no
Sotho (S30)	t̪, t̪ʰ, d̪ (Kgalagadi, Tjhauba, Lobedu, Tawana) kʟ, tɬʰ, ɬ (Pai, Pulana, Kutswe) ɬ, tɬ, tɬʰ (Southern Sotho, Northern Sotho) tɬ, tɬʰ (Tswana)	palatalization (with loss of labial)	no	no (except Southern Sotho, Tjhauba)
Nguni (S40)	ɬ, tɬ, ɮ	palatalization (with loss of labial) (except Lala?)	yes	yes (except Northern Ndebele)
Tsonga (S50)	ɬ, tɬ, ɮ	palatalization (with maintenance of labial)	yes	no
Copi (S60)	?	velarization	yes	no

4.1.1 Reflexes of *c and *j

The reconstructed Proto-Bantu phonemes *c and *j, as well as their prenasalized counterparts *nc and *nj, are usually represented as palatal plosives, although this was not necessarily their phonetic realization (Meeussen, 1967: 83). Common reflexes in Bantu languages are /s/ for *c and /z/, /j/, or /dʒ/ for *j (Hyman, 2019: 128). In Southern Bantu, however, these common Bantu reflexes are only seen in the Shona (S10) group, for instance in Kalanga (S16; Mathangwane, 1999: 213). In Nguni (S40), Tsonga (S50), and the Sotho (S30) languages Tswana (S31), Northern Sotho (S32), and Southern Sotho (S33), the reflexes of *c and *j are lateral fricatives /ɬ, ɮ/ or lateral affricates /tɬ, tɬʰ, dɮ/. The Sotho languages Pai (S303), Pulana (S304), and Kutswe (S302) use a velar lateral affricate /kʟ/ in addition to an alveolar lateral fricative or affricate (Taljaard, 1997; Ziervogel, 1954). Three other Sotho languages, Kgalagadi (S311), Lobedu (S32b), and Tawana (S31c), do not make use of lateral obstruents at all, but rather have dental stops as reflexes of *c and *j. The same is true for Venda (S21). The situation in Copi (S60) is unclear, but no dental stops are attested, and lateral obstruents are explicitly analyzed as loan phonemes (Gowlett, 2003: 615; Lanham, 1955: 30). An overview of the (known) reflexes of *c and *j in Southern Bantu is given in Table 6,[11] along with that of other common phonological evolutions.

11 This represents a slightly oversimplified picture, mainly that of the reflexes of *c and *j

The fairly homogenous development of *c and *j into lateral obstruents in Nguni (S40) and Tsonga (S50) languages suggests that this may be a shared innovation dating back to their most recent shared ancestor. Our linguistic phylogeny supports classifying Tsonga and Nguni into a single clade, but this clade is shared with Copi (S60), which does not have lateral obstruent reflexes of *c and *j. The development of lateral obstruents can also not be projected back to the clade unifying Sotho with Nguni/Tsonga/Copi, because of the heterogeneous nature of the reflexes of *c and *j in the Sotho group. While the innovation of a velar lateral affricate /kʟ/ seems to define the Pai/Pulana/Kutswe node, the dental reflexes of *c and *j, found in Kgalagadi (S311) and Lobedu (S32b), clearly do not follow the internal subbranching of the Sotho subgroup. Furthermore, these dental reflexes are also found outside the Sotho group, in Venda (S21).

Inheritance can therefore not explain the distribution of the reflexes of *c and *j across Southern Bantu, especially not in the Sotho cluster. On the other hand, the relatively unique status of the Southern Bantu reflexes of *c and *j, especially the lateral obstruent reflexes, makes it unlikely that their proliferation in genealogically related and geographically contiguous languages is due to chance. Language contact with a non-Bantu language has also been proposed (Güldemann, 2019: 292; Janson, 1991–1992: 86), for instance with Southern Cushitic languages spoken in Tanzania, which make use of lateral obstruents (Louw and Finlayson, 1990: 406). Although the ancestors of Southern Bantu speakers likely migrated via areas in East Africa where Southern Cushitic languages are spoken, so far no (other) Cushitic influence has ever been identified in Southern Bantu languages. In the absence of evidence for language contact between the ancestors of Southern Bantu speakers and speakers of Cushitic languages, it is unlikely that Southern Bantu lateral obstruents result from Cushitic influence.

Contact with Khoisan languages is more likely to have played a role, as lateral obstruents are attested in various Khoisan languages: the now-extinct Tuu language ǁXegwi, once spoken in northern South Africa and Eswatini (Lanham and Hallowes, 1956b); the Kx'a language ǂHoan (a variety of ǂ'Amkoe), spoken in Botswana;[12] in some varieties of the Khoe-Kwadi languages Glui and Gǁana

before non-front, non-high vowels. Before high vowels, Bantu Spirantization has led to different reflexes in most languages, and before front vowels, some languages have developed sibilants (for a more complete overview, see Janson, 1991–1992: 85).

12 Sands (2007), on the basis of her own fieldwork, notes velar lateral affricates in ǂHoan, but Gerlach (2016: 104–105), who collected data on the N!aqriaxe variety, found the lateral affricate to be very rare.

spoken in the central Kalahari (Sands, 2007; Nakagawa, 1996: 112); in the Kx'a language !Xung spoken in Angola (Traill and Vossen, 1997); and in Kwadi, a virtually extinct Khoe-Kwadi language once spoken in Angola (Güldemann, 2013). Lateral obstruents in Khoisan languages often result from click loss, for instance the loss of a lateral click in Kwadi (Fehn, 2020a) or in !Xung varieties spoken in Angola (Fehn, 2020b, Traill and Vossen, 1997), or the loss of a lingual palatal click in ǁXegwi (Sands, 2007). This raises the possibility that some Southern Bantu lateral obstruents also derive from earlier clicks.

Contact with different Khoisan languages in which lateral obstruents are attested may have facilitated the parallel development of lateral obstruents in different Southern Bantu subgroups. However, in the case of ǁXegwi, some lateral obstruents may actually be the result of Bantu influence, due to intensive contact with Swati, Zulu, and Sotho (Lanham and Hallowes, 1956a). More research on the prevalence and origin of different types of lateral obstruents in Khoisan languages is needed, as well as a better understanding of the phonetic motivation behind the development of palatal plosives into lateral obstruents in Southern Bantu languages.[13]

4.1.2 Palatalization of labials

Another peculiar sound change that affected many Southern Bantu languages concerns velarization or palatalization in clusters of a bilabial consonant and the glide /w/, which in turn is the reflex of a reconstructed back vowel when followed by a non-back vowel. Velarization occurs in Shona (S10) and Copi (S60): after a bilabial consonant, /w/ has become a velar fricative or plosive. Palatalization of bilabial consonants before /w/ occurs in Venda (S20), Sotho (S30), Nguni (S40), and Tsonga (S50). These developments are illustrated in (1) for the different reflexes of Proto-Bantu *bóà 'dog' (Bastin et al., 2002).

13 Lateral obstruents in Southern Bantu languages occur as reflexes of Proto-Bantu phonemes, but also in lexemes that do not have a Bantu origin. The Nguni languages Zulu (S42), Zimbabwean Ndebele (S44), and Swati (S43) make use of the velar lateral affricate /kʟ/, which has been suggested to be of Khoisan origin (Lanham and Hallowes, 1956b: 103), and does not occur in words with a Bantu etymology. Certain Tswana words with /tɬ/ are cognate with Xhosa words with clicks, for instance Xhosa *ǁoǁa* 'chat' and Tswana *tɬotɬa* 'chat' (Louw and Finlayson, 1990: 406), although /tɬ/ in Tswana also occurs in words of Bantu origin.

(1)	Proto-Bantu	*N-bóà	'dog'
	Kalanga (Shona, S16)	*m-bga*	'dog'
	Venda (Venda, S21)	*mm-bya ~ mm-bɣwa*	'dog'
	Tswana (Sotho, S31)	*n-tʃa*	'dog'
	Zulu (Nguni, S42)	*i-n-dʒa*	'dog'
	Tsonga (Tsonga, S53)	*yi-m-bya*	'dog'
	Tonga (Copi, S62)	*yi-m-bɣa*	'dog'

The data in (1) show that palatalization is most extreme in Sotho and Nguni, where the bilabial place of articulation is lost completely, whereas in Tsonga and Venda, the bilabial is maintained and a palatal coarticulation is added. Palatalization is also still active as a morphophonological alternation observed with passive, diminutive, and locative suffixes (Louw, 1975; Ohala, 1978). As a morphophonological process, palatalization also has the most far-reaching consequences in the Nguni and Sotho languages, occurring in more morphological and phonological contexts than in the other groups (Louw, 1975).

The only language for which palatalization or velarization of bilabial consonants followed by glides is unclear is the Nguni language Lala (S406). As a morphophonological process, neither palatalization nor velarization occur (Van Dyk, 1960: 16; E.M. Zungu, 1999: 69–70). Diachronically, some lexemes do show the effect of palatalization, for example *dì-bòè 'stone' > Lala *li-tʃ'e* 'stone,' whereas others do not, for example **m-bóà* 'dog' > Lala *i-m-bwa* 'dog.' More data would be needed to understand if palatalization has occurred in Lala and if so, how the exceptions can be explained.

The phonetic and phonological motivation behind the creation of palatals out of a sequence of two labial phonemes has incited much research (Bennett and Braver, 2015; Kotzé and Zerbian, 2008; Naidoo, 2002; Ohala, 1978, among many others), and is typologically highly unusual (Bateman, 2011). Interestingly, many Khoisan languages have also undergone processes of "unnatural" palatalization, that is palatalization in the absence of a conditioning palatal consonant or high front vowel (Nakagawa, 1998). Unconditioned changes from an alveolar into a palatal consonant are seen in the Khoe languages Glui and Gǁana (Nakagawa, 1998), in western varieties of the Kx'a language ǂ'Amkoe, and in the Bantu language Kgalagadi (S311, Sotho; Gerlach, 2018). These languages are all in contact with each other, so their shared process of palatalization has been analyzed as an areal feature (Gerlach, 2018; Traill, 1980).

This process of palatalization is quite different from labial palatalization in Southern Bantu, both in its conditioning (unconditioned instead of conditioned by a following /w/) and its target (alveolars instead of bilabials). Nonetheless, both processes share the unusual characteristic of creating palatal

consonants outside the context of a conditioning palatal or front vowel. This might therefore indicate that the two processes are related, and that there may once have been a wider and more varied tendency towards palatalization that could account for the palatalization in Southern Bantu as well as that in certain Khoisan languages.

4.1.3 Depressor consonants

Many Southern Bantu languages have "depressor consonants," a set of consonants that have a lowering effect on the tone of the following vowel. Crosslinguistically, voiced obstruents are common as depressor consonants, but in Southern Bantu the set of depressor consonants appears more varied, and many depressor consonants do not exhibit any recognizable voicing (Downing, 2009; Downing and Gick, 2005; Schachter, 1976; Traill, 1990; Traill et al., 1987). The precise extent of depressor consonants in Southern Bantu is difficult to establish as many languages lack detailed phonetic and tonal descriptions, but it is clear that they occur in languages of the Shona (S10), Nguni (S40), and Tsonga (S50) groups, while they are absent in Sotho (S30). Tonal depression is also noted for Copi (S61; Gowlett, 2003: 618), but the situation for Tonga (S62) is unclear. Figure 3 represents an overview of the occurrence of tonal depression in Southern Bantu languages.

The wide distribution of tonal depression across Southern Bantu languages suggests that this feature may already have been present in Proto-Southern Bantu. However, this would entail that the Sotho languages have lost this feature. Loss of contrast between depressor and non-depressor consonants is often phonologized through tonogenesis, the development of an additional (low) tone level (Michaud and Sands, 2020), but this is not attested for Sotho languages, which maintain the two contrastive tone levels that are also used in other Southern Bantu languages. Furthermore, a comparison of the depressor consonant inventories of different Southern Bantu languages (Kula, 2018) shows that they vary from language to language, so even if the phenomenon of tonal depression could be reconstructed to Proto-Southern Bantu, it is unclear what the reconstructed inventory of depressor consonants would be like.

Instead of having developed at the Proto-Southern Bantu level, tonal depression may also have a more recent contact-induced origin. Many depressor consonants are not regular reflexes of Proto-Bantu consonants, as for instance in Zulu (Schadeberg, 2009). Tonal depression has been reported for various Khoe languages. It can be a phonologically active process, such as described in most detail for the Khoe language Tsua, spoken in eastern Botswana (Mathes, 2015). Interestingly, the composition of the set of Tsua depressor consonants is typologically unusual, as it combines both voiced and aspirated obstruents, as well

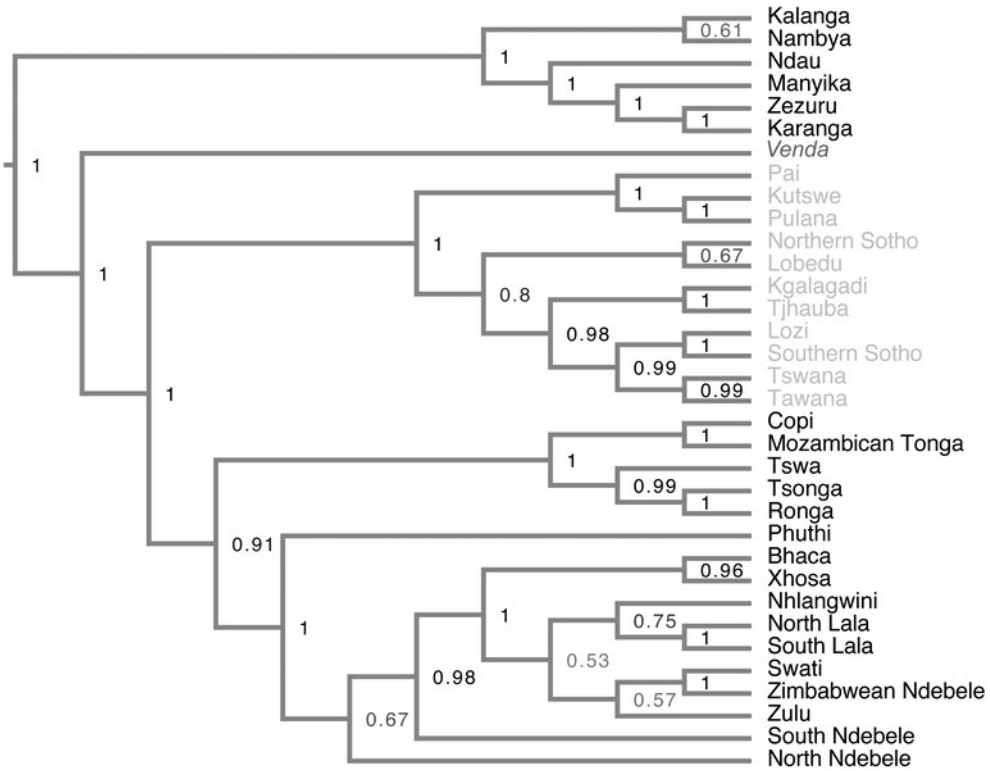


FIGURE 3 Distribution of presence (black) or absence (light gray) of tonal depression in Southern Bantu languages. The presence of tonal depression in Venda is not known.

as /h/ (Mathes, 2015: 94). In other Khoe languages, historical tonal depression can be reconstructed on the basis of newly developed tone levels (Elderkin, 2004, 2008). The Kx'a language Jul'hoan displays evidence of phonologization of tonal depression, and synchronically, pitch lowering of epiglottalized vowels is still active (Miller, 2013). In the Tuu language !Xóõ (also known as Taa), non-modal phonation, which includes glottalization, pharyngealization, and breathy voice, has a strong lowering effect on pitch (Traill, 1985: 38–42). Breathy voiced consonants also act as depressors in certain Southern Bantu languages (Liu and Kula, 2018). It is possible, as has been previously suggested (Downing, 2009: 183, 2018: 117), that the occurrence of tonal depression in Southern Bantu languages is the result of Khoisan influence. In order to test this hypothesis, a systematic comparison between the depressor consonant inventories of Khoisan and Southern Bantu languages is required.

4.1.4 Clicks

The occurrence of phonemic clicks in certain Southern Bantu languages is well known to have resulted from Khoisan influence (Herbert, 1990; Louw, 2013; Pakendorf et al., 2017; Sands and Gunnink, 2019). Although the ultimate Khoisan origin of clicks is undisputed, the pattern of distribution of clicks across the subbranches of Southern Bantu is informative of how and when clicks were copied. Clicks occur in all Nguni languages except Northern Ndebele (S408), which lost clicks relatively recently, most likely due to contact with clickless Bantu languages (Schulz et al., 2019). Clicks also occur in the Sotho language Southern Sotho (S33), and the Tjhauba variety of Kgalagadi (S311). Clicks in Nguni languages may have been incorporated at the Proto-Nguni level, because all Nguni languages have clicks, or had these until recently. Furthermore, clicks occur in cognate vocabulary exhibiting regular sound changes which allow their reconstruction to Proto-Nguni (Gunnink, 2022a). The occurrence of clicks in Southern Sotho has been attributed to horizontal transmission from Nguni (Pakendorf et al., 2017; Sands and Gunnink, 2019), as many click words are shared between Nguni and Southern Sotho (Bourquin, 1951; Doke and Mofokeng, 1957: 23). In Tjhauba, clicks were incorporated through contact with neighboring languages with clicks, most notably the Bantu language Yeyi and the Khoe language Khwe, while no click words are attested that appear to be cognate with Southern Sotho click words (Gunnink, 2021). The possibility that clicks were already acquired at the node connecting the Sotho and Nguni language clusters is unlikely, as this is also the ancestor of the clickless Tsonga (S50) and Copi (S60) languages. Furthermore, Sotho languages other than Southern Sotho and Tjhauba do not make use of clicks, so if clicks were present in Proto-Sotho, click loss would have occurred in all other Sotho languages. So, whereas the presence of clicks in Proto-Nguni is likely, clicks are unlikely to have occurred in Proto-Sotho or further back in Proto-Sotho-Nguni-Tsonga-Copi. A parsimony reconstruction of click phonemes across Southern Bantu languages also shows three independent acquisition events (see Fig. 4).

In conclusion, while the four phonological developments surveyed here are relatively widespread across multiple Southern Bantu subgroups, none can be reliably reconstructed to Proto-Southern Bantu, but rather seem to have been innovated more than once in separate subgroups or languages. Contact with Khoisan languages can be hypothesized or, in the case of clicks, even proven, accounting for the non-genealogical distribution of these features. In Section 5 we discuss the implications of these findings in more detail.

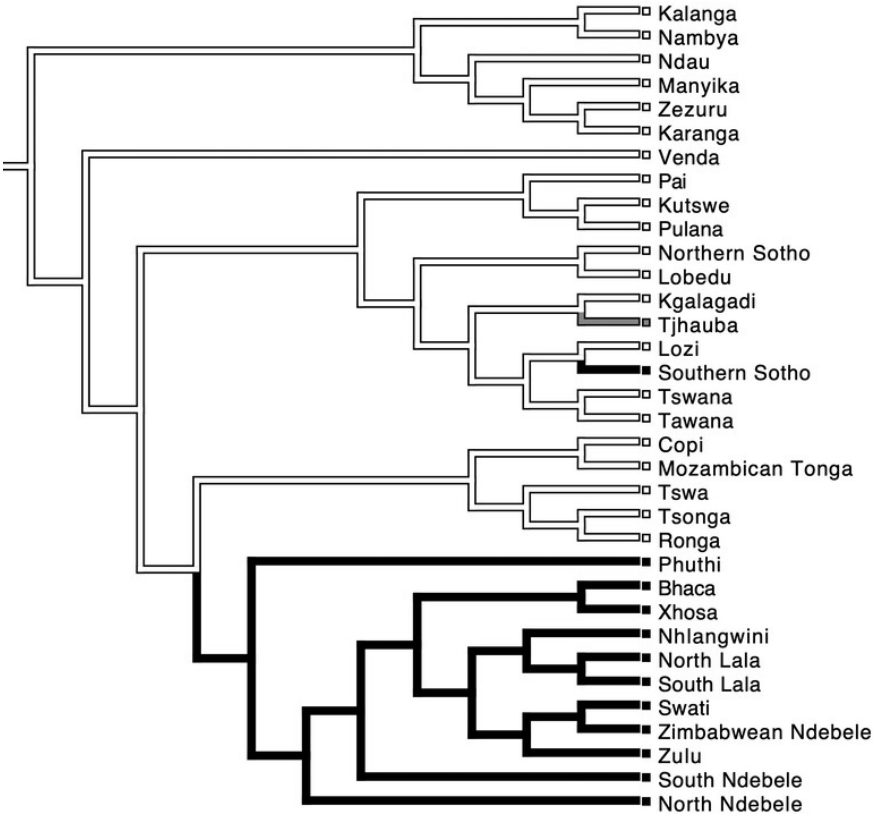


FIGURE 4 Parsimony reconstruction of the presence of click phonemes across Southern Bantu languages showing three independent acquisitions. Southern Sotho has many shared click words with Nguni, so it has been coded with the same state (black), while the click words of Tjhauba are entirely different (gray).

4.2 *Morphological features of Southern Bantu languages*

Several morphological features occur in Southern Bantu that set them apart from Bantu languages spoken elsewhere, and which have been previously shown to result from Khoisan contact (Engelbrecht, 1925; Güldemann, 1999). In this section we review how some of them map onto the genealogical substructure of Southern Bantu: the loss of prefixal diminutive and locative marking, and the innovation of suffixal diminutive, locative, and feminine marking. Their distribution across Southern Bantu is summarized in Table 7.

A pervasive feature of Bantu morphology is its use of genders, or noun classes, which are marked by nominal prefixes on nouns and agreement prefixes on modifiers and verbs. Bantu languages distinguish up to 23 classes, and some of these are used with a particular semantic designation. This is the case

TABLE 7 Morphological features of Southern Bantu languages

	Diminutive class 12/13 prefix	Diminutive suffix	Locative class 16/17/18 prefixes	Locative suffix	Feminine suffix
Shona (S10)	retained	unproductive	productive	no	no
Venda (S20)	lost	absent	unproductive	yes	no
Sotho (S30)	lost (reintroduced in Lozi)	productive (lost in Lozi)	unproductive (reintroduced in Lozi)	yes (lost in Lozi)	only Southern Sotho
Nguni (S40)	lost	productive	unproductive	yes	only Xhosa, Zulu
Tsonga (S50)	lost	productive	unproductive	yes	no
Copi (S60)	lost	productive	unproductive	yes	no

for classes 12 (singular) and 13 (plural), both of which express diminutive meanings, and for classes 16, 17, and 18, which express locative meanings. Prefixes of these classes can be added to a noun to derive a diminutive or locative.

Many Southern Bantu languages have lost part of this inherited system. Reflexes of the class 12/13 diminutive prefixes *ka and *tu are only seen in Shona languages (e.g., Zezuru [S12] *mu-nhu* 'person' > *ka-mu-nhu* 'small person'; Fortune, 1955: 94). No reflexes of the class 12 or 13 prefixes are seen in the Nuclear Southern Bantu languages: Fig. 5 presents the reconstruction of class 12/13 prefixes across Southern Bantu languages. Instead of class 12 or 13, Venda (S21) uses a prefix of class 7 *tʃʰi-* (e.g., *tʃʰi-tukana* 'small and short boy') or class 20 *ku-* (e.g., *ku-ɖi* 'small village'; Ziervogel et al., 1972: 3–4). Ronga (S54) and Tsonga (S53) combine a prefix of class 7 *ʃi-* with a diminutive suffix *-ana* or *-ɲana* (e.g., Ronga *ᵐbuti* 'goat' > *ʃi-ᵐbutw-ana* 'small goat'; Baumbach, 1970: 16). In Sotho (S30), Nguni (S40), and Copi (S60) a suffix is the only possible diminutive marking, and no prefixes are used (e.g., Zulu [S42] *in-ɮu* 'hut' > *in-ɮw-ana* 'small hut'; Poulos and Msimang, 1998: 102).

Similarly, the prefixal locative marking has also been lost to various degrees in Southern Bantu languages (also shown in Fig. 5). Outside southern Africa, Bantu languages typically use reflexes of the prefixes of classes 16 *pa-, 17 *ku-, and 18 *mu- to derive a locative. The productive use of these prefixes in Southern Bantu languages is again only seen in Shona (S10), for example Zezuru (S12) *mhiri* 'other side' > *ku-mhiri* 'to the other side,' *mauro* 'evening' > *mu-mauro* 'in the evening' (Fortune, 1955: 101–102). In Nuclear Southern Bantu languages, the former prefixes of class 16, 17, and 18 are only retained in lexicalized forms, for example Zulu (S42) *pʰaⁿsi* 'on the ground,' *kude* 'far,' *muva* 'behind' (Poulos and Msimang, 1998: 76).

The loss of prefixal locative and diminutive marking in Nuclear Southern Bantu was combined with the development of new, suffixal locative and

FIGURE 5 Parsimony reconstruction of the presence of class 12/13 diminutive prefixes and locative prefixes showing loss at Proto-Nuclear Southern Bantu and a subsequent regain in Lozi via borrowing. The inverse pattern is reconstructed for the locative suffix, showing its acquisition at the Proto-Nuclear Southern Bantu node and its subsequent loss from Lozi (replaced by the re-borrowed locative prefix).

diminutive marking. To express a locative, Venda (S20), Sotho (S30), Nguni (S40), Tsonga (S50), and Copi (S60) use a suffix with the shape *-ini*, *-ni*, *-ŋ*, or similar, for example Venda (S21) *maṱo* 'eyes' > *maṱo-ni* 'in the eyes' (Ziervogel et al., 1972: 40). To express a diminutive, Sotho (S30), Nguni (S40), Tsonga (S50), and Copi (S60) use a suffix with the shape *-ana*, *-ɲana*, or similar, for instance Tswana (S31) *tʰipa* 'knife' > *tʰip-ana* 'small knife' (Cole, 1955: 106). Figure 6 shows which Southern Bantu languages make use of a diminutive suffix. The productivity of the diminutive suffix is highest in Sotho (S30) and Nguni (S40; Güldemann, 1999), which also make use of another nominal suffix, expressing feminine gender, for example Zulu (S42) *inkosi* 'king' > *inkosi-kazi* 'queen' (Poulos and Msimang, 1998: 113). Such a feminine suffix of similar phonological

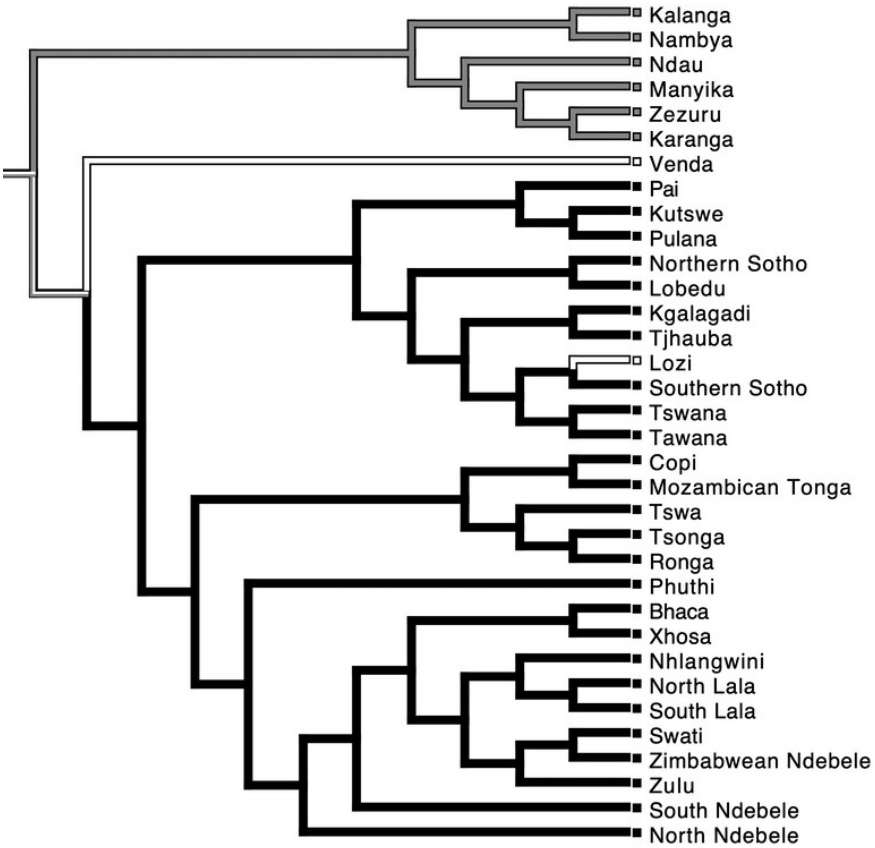


FIGURE 6 Parsimony reconstruction of the presence of a productive (black) or unproductive (gray) diminutive suffix across Southern Bantu languages. Double shading indicates uncertainty in the reconstruction.

shape is productive only in the Sotho language Southern Sotho (S33) and the Nguni languages Xhosa (S41) and Zulu (S42; Güldemann, 1999: 58), pointing to three independent borrowing events, as shown in Figure 7.

As shown above, Bantu languages mainly use nominal prefixes, and nominal suffixes are fairly uncommon. Aside from Southern Bantu languages, diminutive suffixes are only attested in some languages of the Kikongo language cluster, part of the West-Western Bantu subgroup (Goes and Bostoen, 2021), which most likely represents a separate development. Locative suffixes are also attested in certain Eastern Bantu languages (Güldemann, 1999; Schadeberg, 2003), belonging to the same subgroup as Southern Bantu. In spite of these exceptions, nominal suffixes in Bantu are clearly rare and typologically unexpected, given the overall head-initial structure of the language family, and as

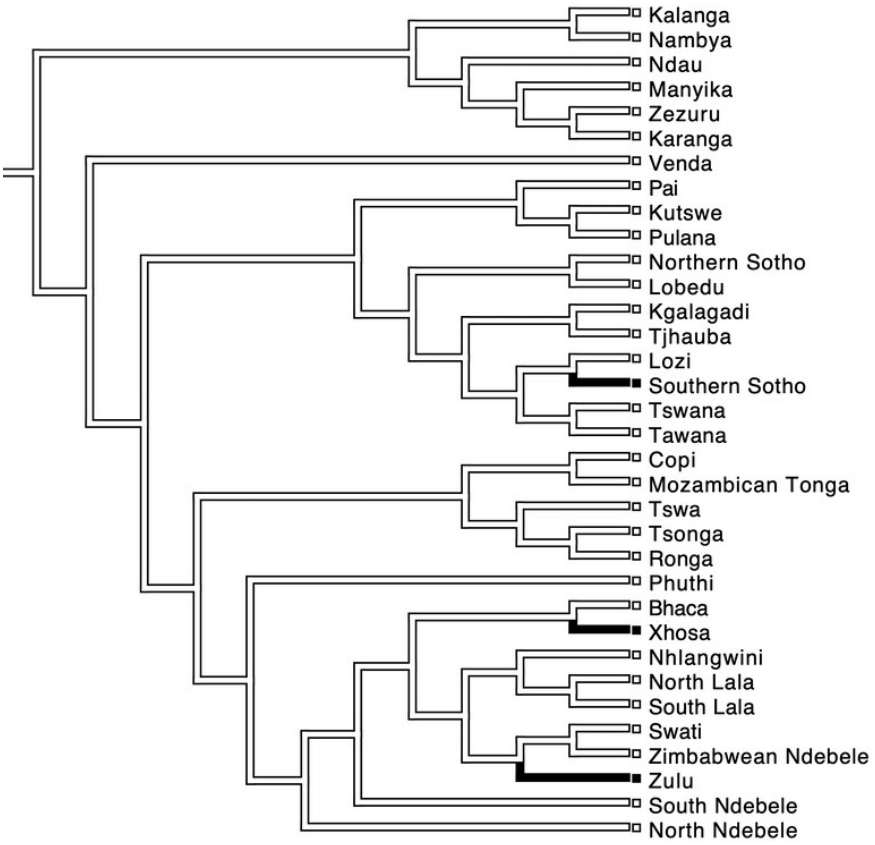


FIGURE 7 Parsimony reconstruction of the presence of the feminine suffix across Southern Bantu, showing three independent borrowing events

such have been attributed to contact with Khoisan languages, where nominal suffixes are common (Engelbrecht, 1925; Güldemann, 1999). The current linguistic phylogeny makes it possible to link the loss of the class 12/13 diminutive prefixes and lexicalization of the class 16/17/18 locative prefixes to the most recent common ancestor of Venda, Sotho, Nguni, Tsonga, and Copi, that is to Proto-Nuclear Southern Bantu. The different degrees of productivity of nominal suffixes, from fairly low in Venda (S20) to highest in Nguni (S40) and Sotho (S30), shows that, once innovated, nominal suffixes continued to grammaticalize. The development of the feminine suffix is attested only in the Sotho language Southern Sotho (S33) and the Nguni languages Xhosa (S41) and Zulu (S42). Although a more thorough description of Nguni languages may uncover the use of a feminine suffix in more languages, the current distribution of this suffix in three geographically contiguous Southern Bantu languages belonging

to distinct clades suggests that it is the result of areal spread and points to a much shallower time depth than the diminutive and locative suffixes.

5 Discussion

In this section, we discuss new insights resulting from our phylogeny into the migration of Southern Bantu speakers to and throughout southern Africa (Section 5.1), and their interactions with each other and, especially, with speakers of various Khoisan languages (Section 5.2). We also combine our linguistic findings with previous research in the fields of archaeology and population genetics.

5.1 *Migration into southern Africa*

Our linguistic phylogeny confirms that Southern Bantu languages are part of the larger Eastern Bantu branch of Bantu. Its closest relatives are Bantu languages spoken to the northeast in Mozambique, Malawi, and Tanzania. As the spread of Bantu languages across Africa is known to be the result of human migration (Li et al., 2014; Pakendorf et al., 2011), this suggests that the ancestors of Southern Bantu speakers migrated into southern Africa from a northeasterly direction, which is in line with earlier studies of the Bantu expansion (Phillipson, 2005; Sinclair, 1991). Our linguistic phylogeny further indicates that all present-day South African Bantu languages, including Shona languages (S10), descend from a single ancestral language, to the exclusion of all other Bantu languages. This language was most likely once spoken in the borderland between Botswana, Zimbabwe, Mozambique, and South Africa (see Map 1). This is the zone with the highest linguistic diversity in the Southern distribution area, in that languages belonging to the different subgroups meet there.

This conclusion is rather surprising against the background of archaeological evidence which suggests that several successive waves of immigrating agriculturalist communities reached southern Africa between the beginning of the first millennium CE and the first half of the second millennium CE (Mitchell and Whitelaw, 2005; Huffman, 2007; Mitchell, 2013; Schoeman, 2013). Of particular relevance to the historical interpretation of our phylogeny in terms of population dynamics is the "complete disjunction" (Mitchell and Whitelaw, 2005: 226), especially in terms of ceramic production, which the southern African archaeological record manifests between what Mitchell (2013) and Schoeman (2013) call Early Farming Communities (first millennium CE) and Late Farming Communities (second millennium CE; cf. Seidensticker et al., 2021, for a

similar disconnection between Early and Late Iron Age in the Congo rainforest). The archaeological signature of the spread of the earliest farmers through eastern Africa during the first millennium CE is the so-called Chifumbaze Complex (Phillipson, 1977), mainly referring to a shared way of pot-making having its origins in the Urewe ceramics of Great Lakes Africa (Mitchell, 2013: 658). It comprises three traditions: Kalundu, Nkope, and Kwale (Huffman, 2007). These would reflect three distinct waves of migration of Eastern Bantu speakers (Huffman and Herbert, 1994–1995), all having reached different parts of southern Africa in the course of the first millennium CE (Mitchell, 2013: 221–222). Although archaeologists may diverge on exactly which ceramic groups belong to which tradition, they seem to agree on one point that is crucial with regard to our linguistic phylogeny: while the Nkope and Kwale traditions became extinct, Kalundu persists in the pottery made by present-day Shona speakers. The latter is traceable back to Happy Rest, the first Kalundu attestation in southern Africa, according to Huffman (2007: 335) and Mitchell (2013: 658). In contrast, the ceramics made by present-day speakers of Nuclear Southern Bantu languages, such as those of the Sotho, Nguni, and Tsonga clusters, cannot be traced back so far in time. Their ancestry does not extend further back than the second millennium CE (Huffman, 2007; Schoeman, 2013).

These discrepancies in the material cultural heritage of Shona speakers as opposed to Nuclear Southern Bantu speakers led archaeologists to conclude that not only their pottery but also their languages have distinct origins (Huffman, 2004, 2007; Parsons, 2008; Mitchell, 2013; Schoeman, 2013). Following this idea, Shona would be a continuation of the Eastern Bantu languages introduced by Kalundu-pottery-producing Early Farming Communities in the first millennium CE.[14] Unjustifiably conflating archaeology and linguistics, Huffman (2007: 335) concludes from the continuity in the archaeological record

14 As pointed out in Bostoen (2018), the hypothesis of Huffman (2007: 335) that "the makers of all Kalundu tradition spoke early forms of Eastern Bantu" conflicts with his assumption that the tradition's putative origins are situated in Benfica, south of Luanda, in Angola (Huffman, 2007: 349) or possibly even further north in Gabon (Huffman, 2007: 359), because no Eastern Bantu languages are spoken there today and, as far as is known, never were. The Bantu languages spoken there belong to either West-Western (also known as West-Coastal) or South-Western Bantu (Grollemund et al., 2015; Pacchiarotti et al., 2019). An anonymous reviewer pointed out that Thomas Huffman does currently discern Western Bantu influence in the Early Iron Age of southern Africa. According to a recent theory of his (Huffman, 2021), Bambata pottery would have its origins in the current-day Democratic Republic of Congo and would have been produced by Western Bantu speakers. Whatever the validity of this new hypothesis, it has little bearing on our present study as Bambata pottery has never been linked with any of the Southern Bantu speech communities discussed here.

that "Shona is the only known surviving language to have evolved directly out of the Early Iron Age in southern Africa." Nuclear Southern Bantu languages, on the other hand, would also have been introduced by Eastern Bantu-speaking farmers, but not earlier than the second millennium CE, with the possible exception of Venda, which is commonly seen as an interaction between Shona and Sotho (see, e.g., Schoeman, 2013: 935). What is more, it is commonly assumed that the ancestral languages of the two main Southern Bantu subgroups, Shona and Nuclear Southern Bantu, were introduced through separate East African immigrant communities, because the ceramics of their current-day speakers cannot be derived from one ancestral tradition.

Our phylogeny confirms that indeed all Southern Bantu languages are part of Eastern Bantu, but it also clusters them as a discrete branch within Eastern Bantu. In other words, all descend from a unique common ancestor and none of them has closer relatives within Eastern Bantu than the other Southern Bantu languages—and this includes Shona and Venda. Our lexicon-based phylogenetic classification is at odds with a scenario in which Southern Bantu languages would have been introduced in southern Africa in different millennia and from different homeland regions. In order to satisfy such a scenario, one would have to assume that Proto-Nuclear Southern Bantu remained in the homeland for a long time after the split with Proto-Shona and subsequently moved and diversified in southern Africa, without leaving any other trace (i.e., other related languages in the homeland region or elsewhere). Additionally, this would still not explain why the pottery traditions of Shona-speaking groups are not related to those of groups speaking Nuclear Southern Bantu languages. It would also not account for the complete disconnection between the material cultures of Early and Late Farming Communities in southern Africa more generally. We therefore consider such a scenario implausible.

Our linguistic findings instead support a scenario of divergence subsequent to arrival in southern Africa. In our view, the apparent mismatch between archaeology and historical linguistics points towards a "spread-over-spread scenario of Bantu expansion," as also recently argued by Seidensticker et al. (2021) for the Congo rainforest. Given the manifest rupture between Early and Late Farming Communities in most of the South African archaeological record, we postulate that the Southern Bantu branch of Eastern Bantu only emerged and started to spread from its place of origin in the borderland between Botswana, Zimbabwe, Mozambique, and South Africa in the second millennium CE. By that time, earlier Eastern Bantu languages introduced by Early Farming Communities in the first millennium had already disappeared together with the material culture of their speakers, or they started to vanish together with the ceramics of their speakers because the latter shifted to both the immaterial and

material heritage of the newcomers. Given the sharp divide between the pottery traditions of the first and second millennium in most of southern Africa, without any coexistence or hybridization of the two, the first scenario is the most likely one, except for Shona. Present-day Shona speakers still make pottery whose origins are traceable to that of the Early Farming Communities, while their language descends from an Eastern Bantu language introduced by Late Farming Communities. If such is indeed true, it would mean that Eastern Bantu-speaking communities who lived in Zimbabwe since the first millennium shifted to the newly introduced Eastern Bantu language—an ancestral form of what we today call "Southern Bantu"—but without giving up their ancestral pottery production. It would be a case of language shift without a shift in material culture.

Genetic data also shed light on the history of Southern Bantu-speaking populations, as discussed in more detail in the next section. However, with the currently available data, it is not possible to distinguish the number of migration events that brought Bantu-speaking communities into southern Africa. Larger and more detailed data sets may answer this question in the future (Choudhury et al., 2021). Shona-speaking populations in particular are not well represented in the currently available genetic studies, not even in Semo et al. (2020), which does contain some Mozambican Shona samples. These therefore cannot be used to test the hypothesis that Shona-speaking populations entered southern Africa before other Southern Bantu-speaking groups.

5.2 *Contact with Khoisan-speaking populations*

After their arrival in southern Africa, contact with speakers of different Khoisan languages played an important role in the diversification of the Southern Bantu languages. Our new linguistic phylogeny allows us to link certain Khoisan-derived linguistic features to specific nodes in the linguistic phylogeny. The loss of inherited locative prefixes and the concomitant innovation of locative suffixes occurred in Proto-Nuclear Southern Bantu, the ancestor of all the Southern Bantu languages but the Shona group. The innovation of a diminutive suffix occurred slightly later, at the node that is ancestral to the Sotho, Nguni, Tsonga, and Copi groups. Click phonemes were adopted even later, in the ancestor of the Nguni languages. Other phonological features of possible Khoisan origin were innovated more than once in the history of the Southern Bantu languages: tonal depression likely characterized Proto-Shona, Proto-Venda, and Proto-Nguni-Tsonga-Copi, and palatalization could be attributed to Proto-Sotho, Proto-Nguni, and Proto-Tsonga. Khoisan influence in individual languages is likely to account for the use of clicks in Tjhauba, and the development of feminine suffixes in Zulu, Xhosa, and Southern Sotho. The use of

clicks in Southern Sotho is also a recent contact-induced change, but in this case not necessarily with Khoisan languages but rather with Bantu languages of the Nguni cluster.

The distribution of (possible) Khoisan-derived features across Southern Bantu languages sheds light on the interactions between Khoisan-speaking populations and early Bantu-speaking groups. The fact that no Khoisan influence seems to be reconstructible for Proto-Southern Bantu suggests that, at this early stage of Bantu settlement in southern Africa, interactions with Khoisan-speaking communities either did not exist or were not of the nature or intensity to result in lasting linguistic changes. Rather, most cases of Khoisan influence on Southern Bantu languages are linked to lower nodes in the tree, suggesting that Khoisan-Bantu interactions that led to lasting linguistic changes in Bantu languages mostly took place once Southern Bantu languages had already diversified into different clusters. Some contact continued, at least in certain languages, once clusters had diversified into different languages. This is in line with interactions between speakers of certain Bantu and certain Khoisan languages lasting up to historic times (e.g., Khoekhoe and Xhosa; Harinck, 1969), or even up to the present (e.g., Kgalagadi and various Khoisan languages of Botswana; Ikeya, 2000).

The distribution of (putative) Khoisan influence across the subbranches of Southern Bantu established in our linguistic phylogeny also reveals parallel development in multiple subbranches of a number of contact-induced features, namely tonal depression, palatalization, and lateral obstruents. This may be indicative of multiple contact situations involving the same or similar Khoisan donor languages. This suggests a wider distribution of the Khoisan donor language(s) or language families than is attested.

Genetic studies provide further valuable insights into the history of interactions between Southern Bantu-speaking groups and their Khoisan-speaking neighbors. Southern Bantu-speaking populations show varying degrees of Khoisan admixture (Choudhury et al., 2021), which are low in populations speaking languages that belong to the Tsonga, Copi, Venda, and Shona clusters, but much higher in populations speaking languages of the Sotho and Nguni clusters (see Table 8). This admixture took place within the last 1,300 years (Sengupta et al., 2021; Semo et al., 2020). There are indications that more northern populations, speaking languages of the Venda, Tsonga, and Copi groups, show earlier admixture than populations living further south, speaking languages of the Sotho and Nguni groups (Sengupta et al., 2021; Semo et al., 2020). However, these early admixture dates may also be due to a limitation in dating techniques, as these populations also show very low degrees of admixture (Choudhury et al., 2021). Furthermore, there is a clear sex bias in this admix-

TABLE 8 Khoisan genetic admixture in the speakers of Southern Bantu languages and Khoisan linguistic influence in the languages

	Khoisan admixture	Female Khoisan admixture	Male Khoisan admixture	Number of phonological features of (putative) Khoisan origin	Number of morphological features of (putative) Khoisan origin
Shona cluster					
Ndau	2%[a]			1/4	0/5
Kalanga		35%[b]	0%[b]	1/4	0/5
Manyika	1%[a]			1/4	0/5
Venda cluster					
Venda	6.45%[c]	21.9%[c]	3.4%[c]	2/4	3/5
Sotho cluster					
Northern Sotho	10.61%[c]	25.2%[c]	3.8%[c]	2/4	4/5
Southern Sotho	14.65%[c]; 24%[a]	30%[d]; 29.6%[c]	11.2%[d]; 6.7%[c]	3/4	5/5
Tswana	20.49%[c]	30%[b]; 38.3%[c]	0%[b]; 5%[c]	3/4	4/5
Kgalagadi		53%[b]	15%[b]	1/4	4/5
Nguni cluster					
Swazi	8.69%[c]	19.2%[c]	7%[c]	4/4	4/5
Xhosa	17.62%[c]	44%[d]; 28.2%[c]	1.8%[d]; 4.9%[c]	4/4	5/5
Zulu	13.58%[c]; 24%[a]	30%[d]; 27.8%[c]	1.9%[d]; 2.6%[c]	4/4	5/5
Southern Ndebele		33%[d]	3.8%[d]	4/4	4/5
Tsonga cluster					
Tsonga	1.56%[c], 4%[a]	11%[c]	6.5%[c]	3/4	4/5
Tswa	3%[a]			3/4	4/5
Ronga	5%[a]			3/4	4/5
Copi cluster					
Tonga	4%[a]			1/4	4/5
Copi	5%[a]			1/4	4/5

[a] Semo et al. (2020). [b] Bajić et al. (2018). [c] Sengupta et al. (2021). [d] Marks et al. (2015)

ture, with much higher degrees of Khoisan admixture in the female line than in the male line (Choudhury et al., 2021), although the extent of this bias varies between populations (Sengupta et al., 2021).

Earlier studies (Pakendorf et al., 2017; Barbieri et al., 2013) found a correlation between contact-induced linguistic change and genetic admixture, but were focused on the presence of click phonemes, an easily recognizable fea-

ture of Khoisan influence in Bantu languages. This revealed an interesting anomaly, whereby specifically those populations speaking languages of the Sotho cluster displayed very high degrees of genetic admixture, yet their languages make no or little use of phonemic clicks. In Table 8, we compare the degree of Khoisan genetic admixture in various Southern Bantu-speaking populations (including information on sex-biased admixture, where available), with the phonological and morphological changes of putative Khoisan origin that were discussed in Section 4. This confirms that Khoisan genetic and linguistic influence are somewhat correlated. The lower degree of Khoisan linguistic influence in the Shona, Venda, and Copi clusters correlates with lower genetic admixture. The high degrees of Khoisan genetic admixture in speakers of Sotho and Nguni languages is also reflected in strong linguistic influence; while click phonemes are mostly absent in Sotho languages, other phonological and morphological features of potential Khoisan origin are frequent. This suggests that the close contact between Sotho-speaking and Khoisan-speaking populations did result in linguistic changes, even though the acquisition of click phonemes was, for most Sotho languages, not part of this contact situation.

Extensive contact-induced morphological restructuring, however, also extends to languages spoken by populations who have undergone relatively little Khoisan admixture, particularly those of the Tsonga and Copi clusters. As some of these morphological innovations likely characterized earlier nodes in the linguistic phylogeny, their presence in lower-level clusters is not necessarily the result of individual contact situations, but rather of inheritance from an earlier ancestral language.

6 Conclusions

In this paper we have presented a new, lexicon-based phylogenetic classification of 34 Southern Bantu languages as part of wider Eastern Bantu. This phylogeny supports the long-recognized division of Southern Bantu into six discrete subgroups—Shona, Venda, Sotho, Nguni, Tsonga, and Copi—but also reveals significant higher and lower structure. Tsonga and Copi languages form a single clade that is sister to the Nguni languages, which in turn is sister to the Sotho languages. This Sotho-Nguni-Tsonga-Copi clade is sister to Venda, with which it forms Nuclear Southern Bantu, and the Shona subgroup is sister to all other Southern Bantu subgroups. Some of these insights tally with earlier studies, for instance on the relative outsider status of Shona and Venda, and the relatively close relationship between the Nguni and Tsonga clusters. In

other cases, our linguistic phylogeny contradicts earlier findings; for instance, our findings do not support a close relationship between Sotho and Venda, or between Venda and Shona.

Our study also provides strong evidence that Southern Bantu languages share a single direct common ancestor, Proto-Southern Bantu. Its point of origin is most likely to be what is now the borderland between South Africa, Botswana, Zimbabwe, and Mozambique, where the different Southern Bantu subgroups meet. This suggests that Southern Bantu languages only started to diversify after their arrival in southern Africa. This is at odds with insights from archaeology, which instead support multiple migrations of Bantu speakers into southern Africa. This apparent mismatch can be interpreted as a spread-overspread scenario, where modern-day Southern Bantu languages are the descendants of languages introduced in a later migration, and those spoken by earlier Bantuphone migrants have disappeared. This is particularly the case for Shona speakers: their material culture descends from that of the earliest southern African Bantu speakers, but their language has been introduced through a later migration, suggesting a shift to the language of the new immigrant group. According to our new hypothesis, the most recent common ancestor of present-day Southern Bantu would only have emerged in southern Africa at the beginning of the Late Iron Age, that is, not before the second millennium CE. There is no direct historical link between present-day Southern Bantu languages and the ancestral Eastern Bantu languages that were introduced in southern Africa at the beginning of the Early Iron Age in the early first millennium CE. That clade of Eastern Bantu languages probably became extinct before the start of the Late Iron Age, except perhaps for the ancestors of present-day Shona speakers, who possibly only completely shifted to ancestral Southern Bantu in the course of the second millennium CE.

Our linguistic phylogeny also provides insights into interactions between migrating Bantu speakers and resident speakers of various Khoisan languages. The distribution of known or suspected linguistic features of Khoisan origin across Southern Bantu languages suggests that these were acquired after Southern Bantu languages started to diversify into different subgroups: the oldest (putative) cases of Khoisan influence can be linked to Proto-Nuclear Southern Bantu at the earliest, but most are linked to even lower nodes, especially to the ancestors of the six individual subgroups. Furthermore, several Khoisan-derived features appear in multiple Southern Bantu subgroups or languages but cannot be attributed to a single node in their classification. This parallel development could be indicative of influence from typologically or genealogically similar Khoisan languages. A comparison between linguistic and genetic data shows that languages that have undergone extensive Khoisan linguis-

tic influence are spoken by populations that display evidence for extensive Khoisan genetic admixture as well.

These findings also give rise to questions for further research. First, while some putative contact-induced developments in Southern Bantu languages have already been argued to result from Khoisan influence, such as clicks and nominal suffixes, this is not yet equally clear for others, such as lateral obstruents, tonal depression, and palatalization. A more detailed analysis of how these phenomena arose in Southern Bantu is needed, especially with regard to the role of language contact, and which Khoisan languages, if any, functioned as the donor. Language contact between Bantu languages should also be considered, as proposed for instance for the acquisition of clicks in Southern Sotho, which are hypothesized to be an adoption from Nguni. Finally, even if the sharing of particular features between Bantu and Khoisan can be attributed to contact, the direction of borrowing needs to be considered; in certain cases, Bantu influence on Khoisan is a possibility, for instance for the occurrence of lateral obstruents in the Tuu language ǁXegwi. More detailed studies of contact-induced changes from Khoisan languages in Southern Bantu languages can further improve our understanding of how these changes contributed to the diversification of Southern Bantu languages, and how Southern Bantu and Khoisan speech communities interacted in the past.

Supplementary materials

Supplementary materials are available at https://osf.io/7waus/. These include the lexical data set in CLDF format, as well as XLSX format, including the sources used for the 45 out-group (that is, not Southern Bantu) languages, the nexus files used for the phylogenetic analyses and the ancestral state reconstructions, as well as the full posterior sample of the best-fitting model.

Acknowledgments

We are grateful to Gereon Kaiping for helpful discussions on model selection, and to Dhriti Sengupta and Carina Schlebusch for help with interpreting the genetic research. We would also like to thank three anonymous reviewers, and the audiences of the Kalahari Basin Area Network and the 10th World Congress of African Linguistics for useful comments and questions on earlier versions of this research. Hilde Gunnink's research is funded by the Research Foundation Flanders (FWO) under grant number 12P8419N. Koen Bostoen's research

is funded by the Special Research Fund (BOF) of Ghent University and by his ERC Consolidator's Grant (nº 724275) for the BantuFirst project (https://www.bantufirst.ugent.be/).

References

Alexander, Neville. 1998. The political economy of the harmonisation of the Nguni and the Sotho languages. *Lexikos* 8: 269–275. https://doi.org/10.5788/8-1-957.

Anderson, Gregory D.S., Bengani Ncube, Tabona Nyoni, Pediatrice Sibanda, and K. David Harrison. 2014. Nambya talking dictionary. Version 1.0. Living Tongues Institute for Endangered Languages. http://www.talkingdictionary.org/nambya.

Auer, Richard Steven. 1977. *Proto-Sotho and the Sotho Group*. PhD dissertation, University of Cape Town.

Bailey, Richard. 1995. Sociolinguistic evidence of Nguni, Sotho, Tsonga and Venda origins. In Raj Mesthrie (ed.), *Language and Social History: Studies in South African Sociolinguistics*, 39–50. Cape Town: David Philip.

Bajić, Vladimir, Chiara Barbieri, Alexander Hübner, Tom Güldemann, Christfried Naumann, Linda Gerlach, Falko Berthold, Hirosi Nakagawa, Sununguko W. Mpoloka, Lutz Roewer, Josephine Purps, Mark Stoneking, and Brigitte Pakendorf. 2018. Genetic structure and sex-biased gene flow in the history of southern African populations. *American Journal of Physical Anthropology* 167: 656–671.

Barbieri, Chiara, Anne Butthof, Koen Bostoen, and Brigitte Pakendorf. 2013. Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia. *European Journal of Human Genetics* 21: 430–436. https://doi.org/10.1038/ejhg.2012.192.

Bastin, Yvonne, André Coupez, and Michael Mann. 1999. *Continuity and Divergence in the Bantu Languages: Perspectives from a Lexicostatistic Study*. Tervuren: Royal Museum for Central Africa.

Bastin, Yvonne, André Coupez, Evariste Mumba, and Thilo C. Schadeberg (eds.). 2002. Bantu lexical reconstructions 3. Tervuren: Royal Museum for Central Africa. https://www.africamuseum.be/en/research/discover/human_sciences/culture_society/blr (accessed 21 June 2021).

Bateman, Nicoleta. 2011. On the typology of palatalization. *Language and Linguistic Compass* 5: 588–602. http://doi.org/10.1111/j.1749-818X.2011.00294.x.

Batibo, H.M., J. Moilwa, and N. Mosaka. 1997. The historical implications of the linguistic relationship between Makua and Sotho languages. *PULA Journal of African Studies* 11: 23–29.

Baumbach, E.J.M. 1970. Introduction to the grammar of Xironga. *Limi* 9: 8–54.

Baumbach, E.J.M. 1987. *Analytical Tsonga grammar*. Pretoria: University of South Africa.

Bennett, William G. and Aaron Braver. 2015. The productivity of "unnatural" labial palatalization in Xhosa. *Nordlyd* 42: 33–44. http://doi.org/10.7557/12.3738.

Birchall, Joshua, Michael Dunn, and Simon J. Greenhill. 2016. A combined comparative and phylogenetic analysis of the Chapacuran language family. *International Journal of American Linguistics* 82: 255–284.

Borland, C.H. 1970. *Eastern Shona: A Comparative Dialect Study*. PhD thesis, University of Cape Town.

Borland, C.H. 1986. Internal relationships in Southern Bantu. *South African Journal of African Languages* 6: 139–141. https://doi.org/10.1080/02572117.1986.10586665.

Bostoen, Koen. 2008. Bantu spirantization: Morphologization, lexicalization and historical classification. *Diachronica* 25: 299–356. https://doi.org/10.1075/dia.25.3.02bos.

Bostoen, Koen. 2009. Shanjo and Fwe as part of Bantu Botatwe: A diachronic phonological approach. In Akinloye Ojo and Lioba Moshi (eds.), *Selected Proceedings of the 39th Annual Conference of African Linguistics*, 110–130. Somerville: Cascadilla Proceedings Project.

Bostoen, Koen. 2018. The Bantu expansion. *Oxford Research Encyclopedia of African History*. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780190277734.013.191.

Bostoen, Koen and Heidi Goes. 2019. Was Proto-Kikongo a 5 or 7 vowel language? Bantu Spirantization and vowel merger in the Kikongo language cluster. *Linguistique et langues africaines* 5: 25–68.

Bourquin, W. 1951. Click-words which Xhosa, Zulu and Sotho have in common. *African Studies* 10: 59–81. https://doi.org/10.1080/00020185108706837.

Bryant, Alfred T. 1905. *A Zulu-English Dictionary*. Pinetown, Natal: Marianhill Mission Press.

Burger, J.P. 1960. *An English-Lozi vocabulary*. Morija: Morija Printing Works.

Chatelain, C.W. 1978. *English-Tsonga, Tsonga-English pocket dictionary*. Braamfontein: Sasavona Books.

Chebanne, Andy. 2003. Reviving the Sotho-Tswana language: Some considerations for the union linguistic development of Sotho-Tswana variants. In Andy Chebanne, Mbulelo Jokweni, Makali Isabella Mokitimi, and Sihawukele Ngubane (eds.), *Unifying Southern African Languages: Harmonization and Standardization*, 75–85. Cape Town: CASAS.

Choudhury, Ananyo, Dhriti Sengupta, Michèle Ramsay, and Carina Schlebusch. 2021. Bantu-speaker migration and admixture in southern Africa. *Human Molecular Genetics* 30: 56–63.

Chousou-Polydouri, Natalia, Joshua Birchall, Sérgio Meira, Zachary O'Hagan, and Lev Michael. 2016. A test of coding procedures for lexical data with Tupi-Guarani and Chapacuran languages. In Christian Bentz, Gerhard Jäger, and Igor Yanovich (eds.),

Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics. University of Tübingen. https://publikationen.uni-tuebingen.de/xmlui/handle/10900/68558.

Cole, Desmond T. 1955. *An Introduction to Tswana grammar*. London: Longmans, Green and Co.

Creissels, Denis. n.d. Dictionnaire Tswana-Français. Unpublished manuscript.

Currie, Thomas E., Andrew Meade, Myrtille Guillon, and Ruth Mace. 2013. Cultural phylogeography of the Bantu languages of sub-Saharan Africa. *Proceedings of the Royal Society B* 280: 20130695. https://doi.org/10.1098/rspb.2013.0695.

de Luna, Kathryn. 2010. Classifying Botatwe: M60 languages and the settlement chronology of south central Africa. *Africana Linguistica* 16: 65–96. http://dx.doi.org/10.3406/aflin.2010.988.

de Schryver, Gilles-Maurice, Rebecca Grollemund, Simon Branford, and Koen Bostoen. 2015. Introducing a state-of-the-art phylogenetic classification of the Kikongo language cluster. *Africana Linguistica* 21: 1–76. http://dx.doi.org/10.2143/AL.21.0.3122579.

Doke, Clement M. 1954. *The Southern Bantu languages*. London: Oxford University Press for the International African Institute.

Doke, Clement M. and Herbert Chimhundu. 2005. *The Unification of the Shona Dialects: A Photographic Reprint with an Introduction by Herbert Chimhundu*. Oslo: ALLEX Project.

Doke, Clement M. and S. Machabe Mofokeng. 1957. *Textbook of Southern Sotho Grammar*. Cape Town: Longmans.

Doke, Clement M., D.M. Malcolm, J.M.A. Sikakana, and B.W. Vilakazi. 1990. *English-Zulu, Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.

Donnelly, Simon. 1999. Southern Tekela Nguni is alive: Reintroducing the Phuthi language. *International Journal of the Sociology of Language* 136: 97–120. http://doi.org/10.1515/ijsl.1999.136.97.

Donnelly, Simon. 2007. *Aspects of Tone and Voice in Phuthi*. PhD dissertation, University of Illinois.

Downing, Laura J. 2009. On pitch lowering not linked to voicing: Nguni and Shona group depressors. *Language Sciences* 31: 179–198. http://doi.org/10.1016/j.langsci.2008.12.015.

Downing, Laura J. 2018. Grounding Nguni depressor effects. In Wolfgang Kehrein, Björn Köhnlein, Paul Boersma, and Marc van Oostendorp (eds.), *Segmental Structure and Tone*, 108–146. Berlin: De Gruyter.

Downing, Laura J. and Bryan Gick. 2005. Voiceless tone depressors in Nambya and Botswana Kalang'a. *Berkeley Linguistic Society* 27: 65–80. https://doi.org/10.3765/bls.v27i1.1115.

Ehret, Christopher. 1972. Outlining southern African history: A re-evaluation A.D. 100–

1500. *Ufahamu: A Journal of African Studies* 3: 9–28. https://doi.org/10.5070/F7310163 84.

Elderkin, E.D. 2004. The starred tones of Central Khoisan. *Afrika und Übersee* 87: 3–77.

Elderkin, E.D. 2008. Proto-Khoe tones in Western Kalahari. In Sonja Ermisch (ed.), *Khoisan Languages and Linguistics: Proceedings of the 2nd International Symposium, January 8–12, 2006, Riezlern/Kleinwalsertal*, 87–136. Cologne: Köppe.

Engelbrecht, Jan. 1925. Suffixbildung in den südafrikanischen Bantu-sprachen. *Mitteilungen des Seminars für orientalische Sprachen* 28: 86–131.

Fehn, Anne-Maria. 2020a. Click loss in Khoe-Kwadi. In Bonny Sands (ed.), *Click Consonants*, 291–335. Leiden: Brill.

Fehn, Anne-Maria. 2020b. Click loss and replacement in Ju. In Bonny Sands (ed.), *Click Consonants*, 336–355. Leiden: Brill.

Felsenstein, Joseph. 1992. Phylogenies from restriction sites: A maximum-likelihood approach. *Evolution* 46: 159–173. https://doi.org/10.1111/j.1558-5646.1992.tb01991.x.

Fischer, A., Elisabeth Weiss, Ernest Mdala, and Sonwaba Tshabe. 1985. *Oxford English Xhosa Dictionary*. Cape Town: Oxford University Press Southern Africa.

Fortune, G. 1955. *Grammar of Shona*. London: Longmans, Green and Co.

Gerlach, Linda. 2016. *N!aqriaxe: The Phonology of an Endangered Language of Botswana*. Wiesbaden: Harrassowitz.

Gerlach, Linda. 2018. Palatalization in ǂ'Amkoe. *Africana Linguistica* 24: 31–43. http://doi.org/10.2143/AL.24.0.3285490.

Goes, Heidi and Koen Bostoen. 2021. Typology and evolution of diminutives in the Kikongo language cluster. *Africana Linguistica* 27: 59–102. https://doi.org/10.2143/AL .27.0.3290072.

Gowlett, D.F. 1989. The parentage and development of Lozi. *Journal of African Languages and Linguistics* 11: 127–149. https://doi.org/10.1515/jall.1989.11.2.127.

Gowlett, Derek. 2003. Zone S. In Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*, 609–638. London: Routledge.

Greenberg, Joseph H. 1963. *The Languages of Africa*. Den Haag: Mouton.

Grollemund, Rebecca, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti, and Mark Pagel. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences* 112: 13296–13301. https://doi.org/10.1073/pnas.1503793112.

Grünthal, Riho, Sami Honkasalo, and Markus Juutinen. 2019. Language sociological trends in South African Ndebele communities: A pilot survey. *Studia Orientalia* 120: 17–62.

Güldemann, Tom. 1999. Head-initial meets head-final: Nominal suffixes in Eastern and Southern Bantu from a historical perspective. *Studies in African Linguistics* 28: 49–91. http://dx.doi.org/10.32473/sal.v28i1.107378.

Güldemann, Tom. 2013. Kwadi phonology. In Rainer Vossen (ed.), *The Khoesan Languages*, 87–88. London: Routledge.

Güldemann, Tom. 2014. "Khoisan" linguistic classification today. In Tom Güldemann and Anne-Maria Fehn (eds.), *Beyond "Khoisan": Historical Relations in the Kalahari Basin*, 1–40. Amsterdam: John Benjamins.

Güldemann, Tom. 2019. The linguistics of Holocene high Africa. In Yonatan Sahle, Hugo Reyes-Centeno, and Christian Bentz (eds.), *Modern Human Origins and Dispersal*, 285–313. Tübingen: Kerns Verlag.

Gunnink, Hilde. 2021. The Tjhauba and their place in southern African prehistory. "Extracting the Past from the Present": International and Interdisciplinary Conference on African Precolonial History, online, March 1–5.

Gunnink, Hilde. 2022a. The early history of clicks in Nguni. *Diachronica* 39: 311–368. https://doi.org/10.1075/dia.19061.gun.

Gunnink, Hilde. 2022b. A grammar sketch of the Shetjhauba variety of Shekgalagadi. *Studies in African Linguistics* 51(1): 29–55. https://doi.org/10.32473/sal.v51i1.129091.

Guthrie, Malcolm. 1948. *The Classification of the Bantu Languages*. London: Oxford University Press for the International African Institute.

Guthrie, Malcolm. 1971. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg Press.

Hallowes, D.P. 1942. *A Grammar of Baca and its Relation to Swazi, Zulu and Xhosa*. MA thesis, University of the Witwatersrand.

Hammarström, Harald. 2019. An inventory of Bantu languages. In Mark Van de Velde, Koen Bostoen, Derek Nurse, and Gérard Philippson (eds.), *The Bantu Languages*, 2nd edn, 17–78. London: Routledge.

Hammarström, Harald, Robert Forkel, Martin Haspelmath, and Sebastian Bank. 2020. Glottolog 4.4. Leipzig: Max Planck Institute for the Science of Human History.

Harinck, Gerrit. 1969. Interaction between Xhosa and Khoi: Emphasis on the period 1620–1750. In Leonard M. Thompson (ed.), *African Societies in Southern Africa*, 145–170. London: Heinemann.

Haspelmath, Martin and Uri Tadmor (eds.). 2009. World loanword database. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wold.clld.org.

Herbert, Robert K. 1990. The sociohistory of clicks in Bantu. *Anthropological Linguistics* 32: 295–315.

Holden, Clare Janaki. 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: A maximum-parsimony analysis. *Proceedings of the Royal Society of London Series B* 269: 793–799. https://doi.org/10.1098/rspb.2002.1955.

Holden, Clare Janaki, and Russell D. Gray. 2006. Rapid radiation, borrowing and dialect continua in the Bantu languages. In Peter Forster and Colin Renfrew (eds.), *Phylogenetic Methods and the Prehistory of Languages*, 19–31. Cambridge: McDonald Institute for Archaeological Research.

Huelsenbeck, John P. and Fredrik Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755. https://doi.org/10.1093/bioinformatics/17.8.754.

Huffman, Thomas N. 2004. The archaeology of the Nguni past. *Southern African Humanities* 16: 79–111.

Huffman, Thomas N. 2007. *Handbook to the Iron Age: The Archaeology of Pre-colonial Farming Societies in Southern Africa*. Pietermaritzburg: University of KwaZulu-Natal Press.

Huffman, Thomas N. 2021. *Bambata* pottery and Western Bantu: Re-interpreting the Early Iron Age in southern Africa. *Southern African Humanities* 34: 1–17.

Huffman, Thomas N., and R.K. Herbert. 1994–1995. New perspectives on Eastern Bantu. *Azania* 29–30: 27–36.

Hyman, Larry M. 2003. Sound change, misanalysis and analogy in the Bantu causative. *Journal of African Languages and Linguistics* 24: 55–90. http://doi.org/10.1515/jall.2003.004.

Hyman, Larry M. 2019. Segmental phonology. In Mark Van de Velde, Koen Bostoen, Derek Nurse, and Gérard Philippson (eds.), *The Bantu Languages*, 2nd edn, 128–149. London: Routledge.

Ikeya, Kazunobu. 2000. The historical dynamics of the socioeconomic relationships between the nomadic San and the rural Kgalagadi. *Botswana Notes and Records* 31: 19–32.

Iziko Lesihlathululi-Mezwi Sesindebele. 2006. *isiNdebele-English/English-isiNdebele isihlathululi-mezwi/Bilingual and Explanatory Dictionary*. Cape Town: Phumelela.

Janson, Tore. 1991–1992. Southern Bantu and Makua. *Sprache und Geschichte in Afrika* 12–13: 63–106.

Janson, Tore. 2007. Bantu Spirantisation as an areal change. *Africana Linguistica* 13: 79–116. http://doi.org/10.3406/aflin.2007.972.

Jimenez, Raevin F. 2017. *Rites of Reproduction: Gender, Generation and Political Economic Transformation among Nguni-Speakers of Southern Africa, 8th–19th Century CE*. PhD dissertation, Northwestern University.

Jones-Phillipson, Rosalie. 1972. *Affinities between Venda and Other Southern Bantu Languages*. PhD thesis, University of London.

Jordan, A.C. 1953. *Some Features of the Phonetic and Grammatical Structure of Baca*. MA thesis, University of Cape Town.

Junod, Henri Philippe. 1930. *Éléments de grammaire tchopi* (*M'chopes*). Lisboa: Carmona.

Kaiping, Gereon A., Melvin S. Steiger, and Natalia Chousou-Polydouri. 2021. Lexedata: Lexical data editing tools. Version 1.0.0-beta3. Zenodo. https://doi.org/10.5281/zenodo.5541168.

Kass, Robert E. and Adrian E. Raftery. 1995. Bayes factors. *Journal of the American Statistical Association* 90: 773–795.

Koile, Ezequiel, Simon J. Greenhill, Damián E. Blasi, Remco Bouckaert, and Russell D. Gray. 2022. Phylogeographic analysis of the Bantu language expansion supports a rainforest route. *Proceedings of the National Academy of Sciences* 119(32): e2112853119. https://doi.org/10.1073/pnas.2112853119.

Kotzé, Albert E. 1995. *'n Fonologiese en morfologiese beskrywing van Lobedu*. PhD thesis, University of South Africa.

Kotzé, Albert E. 2001. Prominent features of Lobedu nouns and pronouns. *South African Journal of African Languages* 21(1): 89–107. https://doi.org/10.1080/02572117.2001.10587468.

Kotzé, Albert E. and Sabine Zerbian. 2008. On the trigger of palatalization in the Sotho languages. *Journal of African Languages and Linguistics* 29: 1–28. https://doi.org/10.1515/JALL.2008.001.

Kriel, T.J. 1976. *The New English-Northern Sotho Dictionary*. King William's Town: Educum.

Kula, Nancy C. 2018. How does the Proto-Bantu reconstructed consonantal system help us explain zone S consonantal inventories (viewed through a comparative study of depressors)? International conference on Reconstructing Proto-Bantu Grammar, Ghent University, November 19–23.

Labroussi, Catherine. 2000. Spirant-devoicing and genetic classification in Eastern Bantu. In H. Ekkehard Wolff and Orin D. Gensler (eds.), *Proceedings of the 2nd World Congress of African Linguistics*, 323–358. Cologne: Rüdiger Köppe.

Lanham, L.W. 1955. *A Study of Gitonga of Inhambane*. Johannesburg: Witwatersrand University Press.

Lanham, L.W. and D.P. Hallowes. 1956a. Linguistic relationships and contacts expressed in the vocabulary of Eastern Bushman. *African Studies* 15: 45–48. https://doi.org/10.1080/00020185608706982.

Lanham, L.W. and D.P. Hallowes. 1956b. An outline of the structure of Eastern Bushman. *African Studies* 15: 97–118. https://doi.org/10.1080/00020185608706989.

Li, Sen, Carina Schlebusch, and Mattias Jakobsson. 2014. Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. *Proceedings of the Royal Society B* 281: 20141448. https://doi.org/10.1098/rspb.2014.1448.

Liu, Xiaoxi, and Nancy Kula. 2018. A comparative study of depression in Bantu, Khoisan and Chinese Wu—Laryngeal settings and feature specifications. *Stellenbosch Papers in Linguistics Plus* 54: 17–43.

Louw, J.A. 1975. Palatalization of bilabials in the passive, diminutive and locative in Xhosa and Tsonga. *Afrika und Übersee* 59: 241–278.

Louw, J.A. 2013. The impact of Khoesan on Southern Bantu. In Rainer Vossen (ed.), *The Khoesan Languages*, 435–444. London: Routledge.

Louw, J.A. and Rosalie Finlayson. 1990. Southern Bantu origins as represented by Xhosa

and Tswana. *South African Journal of African Languages* 10: 401–410. https://doi.org/10.1080/02572117.1990.10586873.

Louw, J.A. and J. Ngidi. 1962. On the segmental phonemes of Zulu. *Afrika und Übersee* 46: 43–93.

Lukusa, Stephen T.M. and Kemmonye C. Monaka. 2008. *Shekgalagari Grammar: A Descriptive Analysis of the Language and its Vocabulary*. Cape Town: CASAS.

Mabille, A. and H. Dieterlen 1961. *Southern Sotho-English Dictionary*. Basutoland: Morija Sesuto Book Depot.

Maddison, Wayne P. and David R. Maddison. 2021. Mesquite: A modular system for evolutionary analysis. Version 3.70. http://www.mesquiteproject.org.

Maho, Jouni F. 2003. A classification of the Bantu languages: An update of Guthrie's referential system. In Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*, 639–651. London: Routledge.

Maho, Jouni F. 2009. NUGL online: The online version of the new updated Guthrie list, a referential classification of the Bantu languages (4 juni 2009). https://brill.com/fileasset/downloads_products/35125_Bantu-New-updated-Guthrie-List.pdf (accessed 31 October 2019).

Marks, Sarah J., Francesco Montinaro, Hila Levy, Francesca Brisighelli, Gianmarco Ferri, Stefania Bertoncini, Chiara Batini, George B.J. Busby, Charles Arthur, Peter Mitchell, Brian A. Stewart, Ockie Oosthuizen, Erica Oosthuizen, Maria Eugenia D'Amato, Sean Davison, Vincenzo Pascali, and Cristian Capelli. 2015. Static and moving frontiers: The genetic landscape of southern African Bantu-speaking populations. *Molecular Biology and Evolution* 32(1): 29–43.

Mathangwane, Joyce T. 1999. *Ikalanga Phonetics and Phonology: A Synchronic and Diachronic Study*. Stanford: CSLI Publications.

Mathes, Timothy K. 2015. *Consonant-Tone Interaction in the Khoisan Language Tsua*. PhD dissertation, New York University.

Meeussen, A.E. 1967. Bantu grammatical reconstructions. *Africana Linguistica* 3: 80–121.

Michael, Lev David, Natalia Chousou-Polydouri, Keith Bartolomei, Erin Donnelly, Sérgio Meira, Vivian Wauters, and Zachary O'Hagan. 2015. A Bayesian phylogenetic classification of Tupí-Guaraní. *LIAMES: Línguas Indígenas Americanas* 15(2): 193–221.

Michaud, Alexis and Bonny Sands. 2020. Tonogenesis. In Mark Aronoff (ed.), *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.748.

Miller, Amanda. 2013. Northern Khoesan phonetics and phonology. In Rainer Vossen (ed.), *The Khoesan Languages*, 45–50. London: Routledge.

Mitchell, Peter. 2013. Early Farming Communities of southern and south-central Africa. In Paul J. Lane and Peter Mitchell (eds.), *The Oxford Handbook of African Archaeology*, 657–670. Oxford: Oxford University Press.

Mitchell, Peter and Gavin Whitelaw. 2005. The archaeology of southernmost Africa from c. 2000 BP to the early 1800s: A review of recent research. *Journal of African History* 46: 209–241.

Miti, L.M. 1996. Subgrouping Ngoni varieties within Nguni: A lexicostatistical approach. *South African Journal of African Languages* 16: 83–93. https://doi.org/10.1080/02572117.1996.10587123.

Monaka, K.C. 2013. A sociolinguistic study of Shekgalagari: Issues of survival in the shadow of Setswana. *Nawa: Journal of Language and Communication* 7: 42–53.

Mpiranya, Fidèle. 1997. Spirantisation et fusion vocalique en bantu: Essai d'interpretation fonctionelle. *Linguistique africaine* 18: 51–77.

Msimang, Christian Themba. 1989. *Some Phonological Aspects of the Tekela Nguni Dialects*. PhD dissertation, University of South Africa.

Mulaudzi, P.A. 2004. Tshimbedzi variety: A link between the Venda language and the Kalanga dialect cluster? *South African Journal of African Languages* 24: 132–139. http://doi.org/10.1080/02572117.2004.10587230.

Mzamane, Godfrey Isaac Malunga. 1949. *A Concise Treatise on Phuthi with Special Reference to its Relationship with Nguni and Sotho*. MA thesis, University of South Africa.

Naidoo, Shamila. 2002. The palatalisation process in isiZulu revisited. *South African Journal of African Languages* 22: 59–69. https://doi.org/10.1080/02572117.2002.10587498.

Nakagawa, Hiroshi. 1996. An outline of |Gui phonology. *African Study Monographs, Supplementary Issue* 22: 101–124.

Nakagawa, Hiroshi. 1998. Unnatural palatalization in |Gui and ||Gana? In Mathias Schladt (ed.), *Language, Identity, and Conceptualization among the Khoisan*, 245–262. Cologne: Rüdiger Köppe.

Ngonyani, Deo S. 2001. The evolution of Tanzanian Ngoni. *Sprache und Geschichte in Afrika* 16–17: 321–353.

Ohala, John J. 1978. Southern Bantu vs. the world: The case of palatalization of labials. *Berkeley Linguistic Society* 4: 370–386. http://doi.org/10.3765/bls.v4i0.2218.

Ownby, Carolan Postma. 1981. Early Nguni history: Linguistic suggestions. *South African Journal for African Languages, Supplement* 1: 60–81. https://doi.org/10.1080/02572117.1981.10586450.

Ownby, Carolan Postma. 1985. *Early Nguni History: The Linguistic Evidence and its Correlation with Archaeology and Oral Tradition*. PhD dissertation, University of California, Los Angeles.

Pacchiarotti, Sara, Natalia Chousou-Polydouri, and Koen Bostoen. 2019. Untangling the West-Coastal Bantu mess: Identification, geography and phylogeny of the Bantu B50–80 languages. *Africana Linguistica* 25: 155–229. http://doi.org/10.2143/AL.25.0.3287234.

Pakendorf, Brigitte, Koen Bostoen, and Cesare de Filippo. 2011. Molecular perspectives

on the Bantu expansion: A synthesis. *Language Dynamics and Change* 1: 50–88. https://doi.org/10.1163/221058211X570349.

Pakendorf, Brigitte, Hilde Gunnink, Bonny Sands, and Koen Bostoen. 2017. Prehistoric Bantu-Khoisan language contact: A cross-disciplinary approach. *Language Dynamics and Change* 7: 1–46. https://doi.org/10.1163/22105832-00701002.

Parsons, Neil. 2008. South Africa in Africa more than five hundred years ago: Some questions. In Natalie Swanepoel, Amanda Esterhuysen, and Philip Bonner (eds.), *Five Hundred Years Rediscovered: Southern African Precedents and Prospects*, 41–54. n.p.: Wits University Press.

Pelling, J.N. 1966. *A Practical Ndebele Dictionary*. Bulawayo: Daystart Publications.

Phillipson, D.W. 1977. *The Later Prehistory of Eastern and Southern Africa*. London: Heinemann.

Phillipson, D.W. 2005. *African Archaeology*. 3rd edn. Cambridge: Cambridge University Press.

Philippson, Gérard and Rebecca Grollemund. 2019. Classifying Bantu languages. In Mark Van de Velde, Koen Bostoen, Derek Nurse, and Gérard Philippson (eds.), *The Bantu Languages*, 335–354. London: Routledge.

Poulos, George and Christian T. Msimang. 1998. *A Linguistic Analysis of Zulu*. Pretoria: Via Afrika.

Quintão, José Luiz. 1951. *Dicionários xironga-português e português-xironga*. Lisboa: Agência Geral das Colónias.

Rambaut, Andrew. 2018. FigTree. Version 1.4.4. https://github.com/rambaut/figtree/releases/.

Rambaut, Andrew, Alexei J. Drummond, Dong Xie, Guy Baele, Marc A. Suchard. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67: 901–904. https://doi.org/10.1093/sysbio/syy032.

Rexová, Katerina, Yvonne Bastin, and Daniel Frynta. 2006. Cladistic analysis of Bantu languages: A new tree based on combined lexical and grammatical data. *Die Naturwissenschaften* 93: 189–194.

Ronquist, Fredrik and John P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574. https://doi.org/10.1093/bioinformatics/btg180.

Ronquist, Fredrik, Maxim Teslenko, Paul van der Mark, Daniel L. Ayres, Aaron Darling, Sebastian Höhna, Bret Larget, Liang Liu, Marc A. Suchard, and John P. Huelsenbeck. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542. https://doi.org/10.1093/sysbio/sys029.

Rycroft, D.K. 1981. *A Concise SiSwati Dictionary*. Pretoria: J.L. van Schaik.

Sands, Bonny. 2007. The contribution of language documentation to historical phonology. In Peter K. Austin, Oliver Bond, and David Nathan (eds.), *Proceedings of*

Conference on Language Documentation and Linguistic Theory, 209–219. London: SOAS.

Sands, Bonny and Tom Güldemann. 2009. What click languages can and can't tell us about language origins. In Rudolf P. Botha and Chris Knight (eds.), *The Cradle of Language*, 204–218. Oxford: Oxford University Press.

Sands, Bonny and Hilde Gunnink. 2019. Clicks on the fringes of the Kalahari Basin Area. In Emily Clem, Peter Jenks, and Hannah Sande (eds.), *Theory and Description in African Linguistics: Selected Papers from the 47th Annual Conference on African Linguistics*, 703–724. Berlin: Language Science Press.

dos Santos, Luiz Feliciano. 1941. *Gramática da língua chope*. Lourenço Marques: Imprensa Nacional de Moçambique.

Schachter, Paul. 1976. An unnatural class of consonants in Siswati. *Studies in African Linguistics, Supplement* 6: 211–220.

Schadeberg, Thilo C. 1994–1995. Spirantization and the 7-to-5 vowel merger in Bantu. In Marc Dominicy and Didier Demolin (eds.), *Sound Change*, 73–84. Amsterdam: John Benjamins.

Schadeberg, Thilo C. 2003. Derivation. In Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*, 71–89. London: Routledge.

Schadeberg, Thilo C. 2009. Appendix: On the origin of Zulu depressor consonants. *Language Sciences* 31: 193–198.

Schoeman, Alex. 2013. Southern African Late Farming Communities. In Peter Mitchell and Paul J. Lane (eds.), *Oxford Handbook of African Archaeology*, 928–941. Oxford: Oxford University Press.

Schulz, Stephan, Antti Olavi Laine, Lotta Aunio, and Nailya Philippova. 2019. Click variation and reacquisition in two South African Ndebele varieties. *Studia Orientalia* 120: 213–282.

Seidel, Frank. 2005. The Bantu languages of the Eastern Caprivi: A dialectometrical analysis and its historical and sociolinguistic implications. *South African Journal of African Languages* 4: 207–242. https://doi.org/10.1080/02572117.2005.10587261.

Seidensticker, Dirk, Wannes Hubau, Dirk Verschuren, Cesar Fortes-Lima, Pierre de Maret, Carina M. Schlebusch, and Koen Bostoen. 2021. Population collapse in Congo rainforest from 400 CE urges reassessment of the Bantu expansion. *Science Advances* 7: eabd8352. https://doi.org/10.1126/sciadv.abd8352.

Semo, Armando, Magdalena Gayà-Vidal, Cesar Fortes-Lima, Bérénice Alard, Sandra Oliveira, João Almeida, António Prista, Albertino Damasceno, Anne-Maria Fehn, Carina Schlebusch, and Jorge Rocha. 2020. Along the Indian Ocean coast: Genomic variation in Mozambique provides new insights into the Bantu expansion. *Molecular Biology and Evolution* 37: 406–416.

Sengupta, Dhriti, Ananyo Choudhury, Cesar Fortes-Lima, Shaun Aron, Gavin Whitelaw, Koen Bostoen, Hilde Gunnink, Natalia Chousou-Polydouri, Peter Delius, Stephen

Tollman, F. Xavier Gómez-Olivé, Shane Norris, Felistas Mashinya, Marianne Alberts, AWI-Gen Study, H3Africa Consortium, Scott Hazelhurst, Carina M. Schlebusch, and Michèle Ramsay. 2021. Genetic substructure and complex demographic history of South African Bantu speakers. *Nature Communications* 12: 2080. https://doi.org/10.1038/s41467-021-22207-y.

Sinclair, Paul J.J. 1991. Archaeology in eastern Africa: An overview of current chronological issues. *Journal of African History* 32: 179–219. https://doi.org/10.1017/S0021853700025706.

Skhosana, Philemon Buti. 2009. *The Linguistic Relationship between Southern and Northern Ndebele*. PhD thesis, University of Pretoria.

Taljaard, Petrus Cornelius. 1997. *The Pai Language of Eastern Mpumalanga and its Relation to Swati*. PhD dissertation, University of South Africa.

Traill, Anthony. 1980. Phonetic diversity in the Khoisan languages. In J.W. Snyman (ed.), *Bushman and Hottentot Linguistic Studies*, 167–189. Pretoria: University of South Africa.

Traill, Anthony. 1985. *Phonetic and Phonological Studies of !Xóõ Bushman*. Hamburg: Helmut Buske.

Traill, Anthony. 1990. Depression without depressors. *South African Journal of African Languages* 10: 166–172. http://dx.doi.org/10.1080/02572117.1990.10586849.

Traill, Anthony, J.S.M. Khumalo and P. Fridjhon. 1987. Depressing facts about Zulu. *African Studies* 46: 255–274. https://doi.org/10.1080/00020188708707678.

Traill, Anthony and Rainer Vossen. 1997. Sound change in the Khoisan languages: New data on click loss and click replacement. *Journal of African Languages and Linguistics* 18: 21–56. https://doi.org/10.1515/jall.1997.18.1.21.

van der Spuy, Andrew. 1990. Phonological relationships between the Southern Bantu languages. *African Studies* 49: 119–147. https://doi.org/10.1080/00020189008707720.

Van Dyk, Philippus Reynoldus. 1960. *'n Studie van Lala: Sy fonologie, morfologie en sintaksis*. PhD dissertation, University of Stellenbosch.

van Warmelo, N.J. 1935. *A Preliminary Survey of the Bantu Tribes of South Africa*. Pretoria: Government Printer.

van Wyk, E.B. 1969. Die indeling van die Sotho-taalgroep. *Ethnological and Linguistic Studies in Honour of N.J. van Warmelo*, Ethnological Publications 52, 169–179. Pretoria: Department of Bantu Administration and Development.

Vansina, Jan. 1995. New linguistic evidence and "the Bantu Expansion". *Journal of African History* 36(2): 173–195.

Vossen, Rainer. 1997. What clicks sounds got to do in Bantu. In Brigit Smieja and Meike Tasch (eds.), *Human Contact Through Language and Linguistics*, 353–366. Frankfurt am Main: Peter Lang.

Wentzel, P.J. 1981. *The Relationship between Venda and Western Shona*. PhD dissertation, University of South Africa.

Wentzel, P.J. and T.W. Muloiwa. 1982. *Improved Trilingual Dictionary: Venda Afrikaans English*. Pretoria: University of South Africa.

Wilkes, A. 1981. 'n Beknopte weergawe van enkele belangrike fonologiese en morfologiese verskille tussen Noord- en Suid-Lala. *Studies in Bantoetale* 8: 93–102.

Wilkes, A. 1997. Language contact and language change—The case of Southern Transvaal Ndebele. *South African Journal of African Languages* 17: 75–80. https://doi.org/10.1080/02572117.1997.10587164.

Witzlack-Makarevich, Alena and Hiroshi Nakagawa. 2019. Linguistic features and typologies in languages commonly referred to as "Khoisan." In H. Ekkehard Wolff (ed.), *The Cambridge Handbook of African Linguistics*, 382–416. Cambridge: Cambridge University Press.

Wright, John. 2012. A.T. Bryant and the "Lala." *Journal of Southern African Studies* 38: 355–368. https://doi.org/10.1080/03057070.2012.682838.

Ziervogel, Dirk. 1954. *The Eastern Sotho*. Pretoria: Van Schaik.

Ziervogel, Dirk. 1959. *A Grammar of Northern Transvaal Ndebele*. Pretoria: Van Schaik.

Ziervogel, Dirk, P.J. Wentzel, and T.N. Makuya. 1972. *A Handbook of the Venda Language*. Pretoria: University of South Africa.

Zungu, Elphas Mphunyuzwa. 1999. *A Comparative Phonological and Morphological Analysis of the North and South Lala Dialects of Tekela Nguni*. PhD dissertation, University of South Africa.

Zungu, Phyllis Jane Nonhlanhla. 1989. *Nhlangwini, a Tekela-Nguni Dialect and its Relationship to "Standard Zulu" and Other Nguni Dialects*. MA thesis, University of Natal.