BRILL

LANGUAGE DYNAMICS AND CHANGE 14 (2024) 1–73

brill.com/ldc

Sound change versus lexical change for subgrouping

*Word-final lenition of Proto-Bantu *ŋg in West-Coastal Bantu*

Sara Pacchiarotti | ORCID: 0000-0003-1360-5060
Department of Languages and Cultures, UGent Centre for Bantu Studies (BantUGent), Gent, Belgium
Sara.Pacchiarotti@UGent.be

Guy Kouarata | ORCID: 0000-0002-4549-2278
Department of Languages and Cultures, UGent Centre for Bantu Studies (BantUGent), Gent, Belgium
Guy.Kouarata@UGent.be

Koen Bostoen | ORCID: 0000-0003-2284-6165
Department of Languages and Cultures, UGent Centre for Bantu Studies (BantUGent), Gent, Belgium
Koen.Bostoen@UGent.be

Received 4 July 2023 | Accepted 9 May 2024 |
Published online 25 July 2024

Abstract

This article focuses on languages of the Kwilu-Ngounie subbranch within a branch of the Bantu language family known as West-Coastal Bantu. Within Kwilu-Ngounie, B70 and B80 languages emerge as paraphyletic in the most comprehensive lexicon-based phylogeny of the branch. We assess whether the impossibility to group them into lexicon-based monophyletic subgroups can be bypassed by using the phonological innovation of word-final loss of Proto-Bantu *ŋg as diagnostic of a new subgroup. It is hard to tell whether this new subgroup is a clade by descent or instead a taxon resulting from a contact-induced innovation affecting related varieties. The unconditioned reflexes of *ŋg across varieties signal that both language-internal lexical diffusion and contact-induced crosslinguistic spread of phonological innovation thwart the Neogrammarian axiom of flawlessly regular sound change. Beyond its rele-

Published with license by Koninklijke Brill BV | DOI:10.1163/22105832-bja10033
| ISSN: 2210-5824 (print) 2210-5832 (online)

vance for low-level Bantu subgrouping, this article contributes to the methodological issue of conflicting lexical and diachronic phonological evidence for internal classification.

Keywords

West-Coastal Bantu – Kwilu-Ngounie – lexicon-based phylogenetics – sound change – lexical diffusion – dialectal diffusion – language contact

1 Introduction

The so-called B70 Teke group in Guthrie's referential Bantu classification (Guthrie, 1971; Maho, 2009; Hammarström, 2019) is a cross-border cluster of languages straddling the Democratic Republic of Congo (DRC, Congo-Kinshasa), the Republic of the Congo (Congo-Brazzaville), and the Gabonese Republic (Gabon). Teke refers to the eponymous kingdom to which the speakers of these languages trace their origins. This polity is also known as Tio or Tyo (Vansina, 1973). As a matter of fact, "Teke" is an exonym of Kikongo origin (Matakumba Kanika, 1980: 1; Pacchiarotti et al., 2019: 169). With this term outsiders used to designate all people inhabiting the uplands to the north of Malebo Pool on both banks of the Congo River, even as far north as the mouth of the Nkeni River (Vansina, 1966: 102).

"Teke" is derived from the root *tég 'sell,' attested in several Bantu languages of Guthrie's Zones B, C, and H (Bastin et al., 2002). It underlines that Teke people were historically most famous for the commerce they controlled in the vicinity of the Malebo Pool as intermediaries between the merchant fleets coming from the Congo rainforest in the north and the trade caravans heading toward the Atlantic coast (Vansina, 1973: xv). In Kikongo, one of the main languages in that caravan trade, the root *tek* 'sell' gave rise to the glossonym *kiteke* and the ethnonym *muteke/bateke* (Lema, 1978: 25). Rather than using this term, Teke speakers refer to themselves with the root *tyo* or one of its many variants in the different Teke varieties, such as *tio*, *teo*, *tege*, *tsege*, *teye*, *tɛɛ*, or *tye* (Jacquot, 1965: 340; Vansina, 1966: 102; Boone, 1973: 295).

Genealogically speaking, languages of Guthrie's B70 Teke group belong to the branch of the Bantu family known as West-Coastal Bantu (WCB; Vansina, 1995; Bostoen et al., 2015; de Schryver et al., 2015; Pacchiarotti et al., 2019; Koile et al., 2022) or West-Western Bantu (Grollemund et al., 2015). In the most comprehensive lexicon-based WCB phylogeny to date (Pacchiarotti et al.,

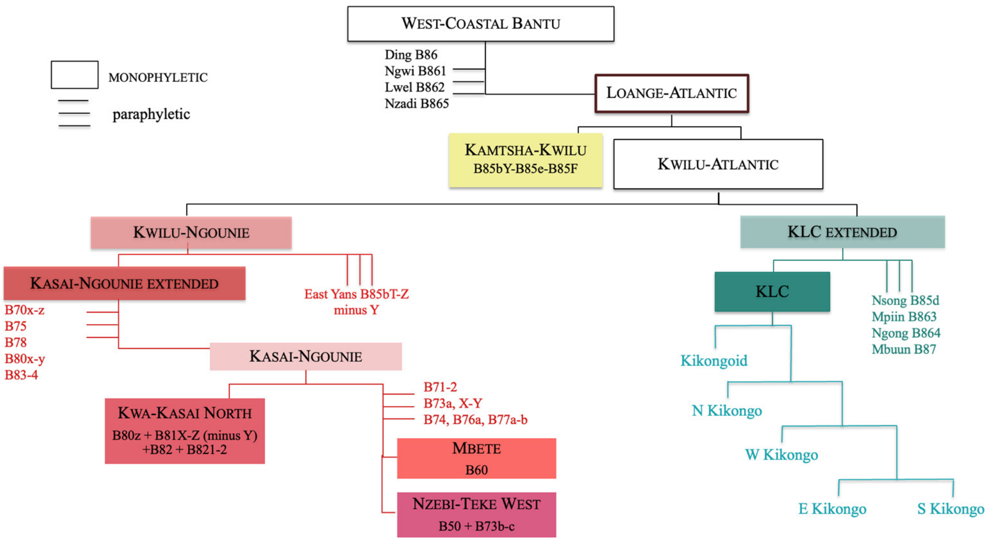
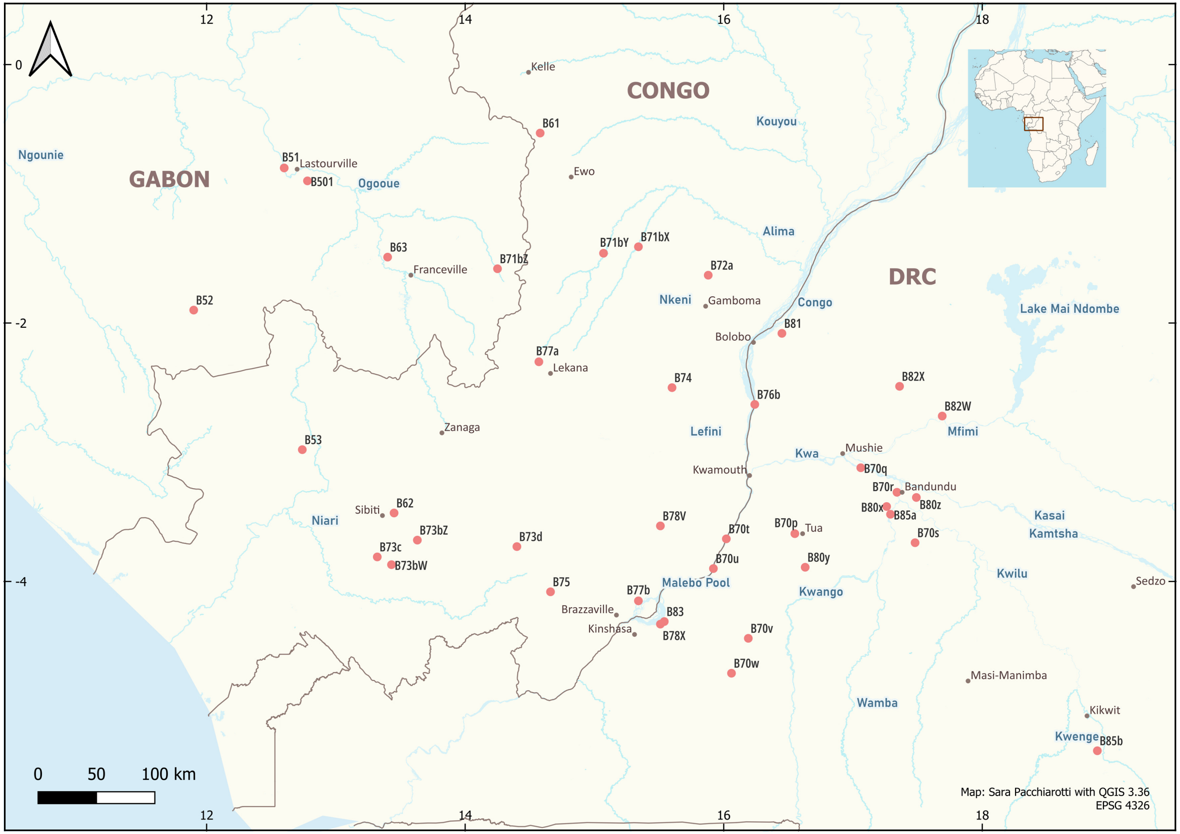


FIGURE 1 Internal lexicon-based phylogenetic classification of West-Coastal Bantu
DE SCHRYVER ET AL., 2015; BOSTOEN AND DE SCHRYVER, 2018A, 2018B; PACCHIAROTTI ET AL., 2019

2019), all B70 Teke languages included are subclassified in the so-called Kwilu-Ngounie clade (see Fig. 1), which comprises languages spoken between the Kwilu River in the DRC and the Ngounie River in Congo and Gabon (see Map 1).

Even if Teke people claim to speak a single language that does not require the intervention of an interpreter regardless of the variety being spoken (Kristensen et al., 1984: 1), Fig. 1 clearly shows that Guthrie's B70 Teke languages do not constitute a discrete subclade within Kwilu-Ngounie. Although they are unmistakably closely related, they are not more closely related to each other than they are to languages belonging to Guthrie's groups B50, B60, and (part of) B80. Quite the opposite, in fact: the closest relatives of Laali B73b and Yaa B73c are the B50 Nzebi languages, with which they form the Nzebi-Teke West subclade, and indeed Laali and Yaa are the only members of Guthrie's B70 Teke group which belong to a well-defined monophyletic unit within Kwilu-Ngounie.

While Kasai-Ngounie incorporates three monophyletic units, namely Kwa-Kasai North, Mbete, and Nzebi-Teke West, North Teke B71, Ngungwel B72a, Tsaayi B73a, West Teke B73, Eboo B74, Mosieno B76a, Kukuya B77a, and Fumu B77b all form a paraphyletic grade, that is, a cluster of individual languages whose most recent common ancestor is that of Kasai-Ngounie itself. All languages between Kasai-Ngounie and the ancestral level directly above, that is,



MAP 1 Geographic distribution of the Kwilu-Ngounie languages considered in this study

Kasai-Ngounie Extended, also form such a paraphyletic grade, incorporating several languages of Guthrie's B70 Teke group, namely Bibaana B70x, South Teke B70y, Bwala B70z, Tio Bali B75, Wuumu B78, Boma Nkuu B80x, South Boma [Nkuu] B80y, Mfinu B83, and Mpuono B84.

The fact that Guthrie's B70 Teke referential group and surrounding languages end up in several distinct paraphyletic grades instead of well-defined monophyletic groups is possibly due to intensive mutual contact between closely related varieties (Bollaert et al., 2021: 3). If the genealogical position of several B70 and B80 languages within the Kwilu-Ngounie branch of WCB cannot be resolved through basic vocabulary, the question is whether other types of data are better suited to shedding light on this issue.

To this end, we present in this article an in-depth quantitative and qualitative analysis of the development, conditioning, and chronology of one specific sound shift: the word-final loss of Proto-Bantu *ŋg in 39 Kwilu-Ngounie varieties. This sound change is unique to a subset of Kwilu-Ngounie languages which includes several B70 and B80 varieties that are paraphyletic in the phylogeny of Pacchiarotti et al. (2019). Methodologically, we address the problem of (partially) conflicting types of evidence, namely lexical versus phonological, for the internal classification of low-level Bantu subgroups.

Accordingly, this article is organized as follows. In Section 2, we present the data and methodology used for this comparative study. In Section 3, we show that within WCB the sound shift *ŋg > ŋ > ∅ in C2 position—that is, the position of the second stem consonant—is only attested in languages belonging to the Kwilu-Ngounie branch. In Section 4 and subsections therein, we offer a detailed account of the evolution of C2 *ŋg across 39 Kwilu-Ngounie varieties and assess the relevance of this sound shift for the internal classification of this subbranch, and especially for the genealogical position of varieties belonging to Guthrie's referential B70 and B80 groups. In Section 5, we first assess the extent to which Kwilu-Ngounie lexicon-based subgroups in the phylogeny of Pacchiarotti et al. (2019) match the subgroups based on the phonological innovation C2 *ŋg > ŋ > ∅ (Section 5.1). In an attempt to account for mismatches between diachronic phonology and lexicon-based phylogeny, we attribute in Section 5.2 greater diagnostic power to historical sound shifts for internal classification and assess how this impacts the phylogenetic subgroups. In Section 5.3, we give the multiple shared lexical innovations underlying the phylogeny of Pacchiarotti et al. (2019) greater weight and assess how the distribution of the diachronic sound change C2 *ŋg > ŋ > ∅ can be accounted for without overruling the lexicon-based subgroups. Conclusions are in Section 6.

2 Data and methodology

Although this study intends to better understand the genealogical position of the paraphyletic languages within Kwilu-Ngounie (Pacchiarotti et al., 2019), and especially those of Guthrie's B70 Teke group, it also includes, for comparative purposes, languages from Kwilu-Ngounie's monophyletic Kwa-Kasai North, Mbete, and Nzebi-Teke West subgroups, all part of the Kasai-Ngounie subbranch (see Fig. 1). Due to lack of data, not all potentially relevant WCB varieties included in the phylogeny of Pacchiarotti et al. (2019) are included in this study. Map 1 displays the geographic distribution of the 39 Kwilu-Ngounie language varieties on which our historical-comparative study focuses. Appendix 1 lists them with additional information on the country and the reference location (geocoordinates column) where they are spoken, as well as the sources from which data were obtained.

It is worth mentioning that many of the B70 varieties included here were previously neither documented nor inventoried in existing referential classifications of Bantu languages such as Guthrie (1971) and Maho (2009); see Kouarata et al. (2023). Mostly based on a perusal of their lexicon, we assume them to belong to the Kwilu-Ngounie branch, but we do not know where exactly. For doculects of languages or dialectal variants not inventoried in the referential lists of the Bantu languages (Guthrie, 1971; Maho, 2009; Hammarström, 2019), we use provisional codes which were already proposed in Pacchiarotti et al. (2019) or abide by the following principles established in that study:

- A code consisting of a decimal number where the second digit is zero followed by a lowercase letter (starting from the end of the alphabet)—for example, Tiimi (Bokala) B70q or Kikimi (Nganambo) B70r—refers to a variety inventoried in neither Guthrie (1971) nor Maho (2009) which we tentatively place in one of Guthrie's referential groups (in this case B70); the lowercase "q" and "r" in B70q and B70r indicate that we consider Tiimi spoken in Bokala and Kikimi spoken in Nganambo to be two distinct languages, rather than regiolects of a single language.
- A code consisting of a decimal number where the second digit is not zero followed by an uppercase letter (starting from the end of alphabet)—for example, North Boma (Bopaka) B82X or North Boma (Inongo) B82W—is used to distinguish varieties inventoried in Guthrie (1971) or Maho (2009) which we consider to be regiolects/dialects of the same language.[1]

---

[1] These two conventions can also be combined. It is possible to combine just these two, as in the cases of Boma Yumu (Pentane/Mondai) B80zX and Boma Yumu (Saio) B80zY, which

As can be seen in Appendix 1, for most varieties we used firsthand fieldwork data collected by the second author in 2021 and 2022 (see Kouarata et al., 2023) in the framework of the ERC-funded BantuFirst project led by the last author (see https://www.bantufirst.ugent.be/). When the data originate instead in existing documentation, we relied on specialized sources for individual languages. New data collected on varieties which were already documented, such as Laali B73b (Bissila, 1991) and North Boma B82 (Stappers, 1986), allowed us to test whether the distribution of reflexes emerging from existing sources holds true for other idiolects and examine factors behind possible differences. We also included several regiolects of more widely spoken WCB varieties, such as Yans (see Appendix 1), whose speakers number around 100,000 according to *Ethnologue* (Eberhard et al., 2023). Compared to varieties such as Laali B73b and Tiene B81, which respectively have around 2,000 and 24,500 speakers according to *Ethnologue*, Yans is certainly a bigger language in terms of number of speakers. However, based on fieldwork experience, the literature available to us (Swartenbroeckx, 1948; Rottland, 1977; Mayanga, 1985; Nguma, 1986; Impubi Mukwa, 1987), and previous research in the area (Pacchiarotti et al., 2019; Pacchiarotti and Bostoen, 2021b), we know that more widely spoken languages in the region, such as Yans B85 and Ding B86, can vary dramatically depending on (a) their geographic location and (b) the different populations among whom their speakers live interspersed.

In order to carry out the comparative study presented in this article, we compiled numerous cognate sets of relevant vocabulary in the selected set of 39 Kwilu-Ngounie varieties. To identify protoforms which contain *ŋg in C2 and are reconstructable to some node of Proto-WCB, we relied on a constantly updated synchronic lexical comparative database originating in the successive ERC-funded KongoKing (2012–2016) and BantuFirst (2018–2023) projects granted to the last author within the UGent Centre for Bantu Studies (BantUGent). This database is linked to a diachronic database known as Bantu Lexical Reconstructions 2/3 (BLR 2/3; Coupez et al., 1998; Bastin et al., 2002), containing nearly 10,000 protoforms with variable depth in the Bantu family. In Appendix 3, we present 48 widespread Bantu roots reconstructed with *ŋg in C2 position (Bastin et al., 2002) along with their attested reflexes in the Kwilu-Ngounie language varieties at the center of this comparative study.[2] Through-

---

we consider regiolects of the not yet inventoried Boma Yumu B80z; or use the conventions together with Guthrie (1971) or Maho (2009) codes, as with Laali (Mayeye) B73bZ vs. Laali (Kendi) B73bW, regiolects/dialects of the already inventoried Laali B73b.

[2] For reasons of space, cognates in WCB languages outside of Kwilu-Ngounie are not included in Appendix 3.

out this article, all reconstructed roots obtained from BLR2/3 are given together with the unique index number with which they are identified in that database. Whenever a reconstruction with a BLR index does not have tone marking, it is because none is provided in the BLR database. Reconstructed roots not accompanied by an index number are new, tentative reconstructions which we propose based on cognate sets attested across WCB languages and Central-Western Bantu (CWB) languages.

Because this study looks at the reflexes of C2 *ŋg in Kwilu-Ngounie languages, it is important to note that all relevant sources on languages which have both [ŋg] and [ŋ] as reflexes of C2 *ŋg clearly mark the difference between [ŋg] and [ŋ] orthographically by using ⟨ng⟩ vs. ⟨ŋ⟩ (Bissila, 1991; Ndouli, 2001). For Nduumo B63, Biton (1969: 556) uses ⟨ng⟩ for [ŋg] and ⟨ṅ⟩ for [ŋ]. For Tiene B81, Ellington (1977) uses ⟨n̵⟩ for [ŋ]. In these cases, we uniformized these different graphemes to ⟨ŋ⟩.

As for tone marking, we uniformized all data from secondhand sources to the conventions we used for firsthand fieldwork data: low tone [à], high tone [á], falling tone [â], and rising tone [ǎ]. For Yaa B73c, Mouandza (2001) marks an extra high tone as [a̋]. The following secondhand sources do not mark tone for the corresponding languages: Biton (1969) for Nduumo B63; Raharimanantsoa and Ntsiba Ngolo (2015) and Ntsiba Ngolo and Raharimanantsoa (2021) for Tyee B73d; Swartenbroeckx (1948) for West Yans (Mukonkie) B85a; and Nguma (1986) for East Yans (Niadi) B85b.

3 The word-final reflexes of Proto-Bantu *ŋg in WCB

The sound shift on which this article focuses is the complete loss of Proto-Bantu *ŋg at the end of noun and verb stems, also known as the C2 (consonant 2) position in a stem with the shape C1V1(N)C2V2. This is illustrated in (1) with original fieldwork data from Ŋiŋi B76b[3] and Kukuya B77a. Throughout this article, Bantu lexical reconstructions found in Bastin et al. (2002) are presented to

---

[3] In the phylogenetic classification of Pacchiarotti et al. (2019), the only B76 variety included was Mosieno B76a. As we argue in Kouarata et al. (2023: 15), Ŋiŋi probably corresponds to what is known as *ŋge-ŋge* in Vansina (1966: 131) or *ŋg'ee* in Maho (2009: 23), who attributes the code B76b to it. Since Ŋiŋi B76b was not included in the 2019 phylogeny, we do not know whether it would cluster with Mosieno B76a. Similarly, since our only data on Mosieno B76a is a 92-word list from Bastin et al. (1999), we do not have enough evidence to claim that Mosieno consistently displays the sound shift C2 *ŋg > ŋ > ∅ illustrated in (1). Nevertheless, the few reflexes of reconstructions with C2 *ŋg in the Mosieno list consistently show C2 *ŋg > ∅.

the left, their reflexes to the right.[4] The meaning of reflexes is specified only when it deviates from that of the reconstruction.

(1)	*bóŋgó ‘knee’	(BLR 275)	>	B76b *búɔ́*, B77a *búɔ̀*
	*cáŋgʊ́ ‘millet’	(BLR 486)	>	B76b *sìá*, B77a *lí-sáá* ‘maize’
	*dòŋg-à ‘advise, teach’	(BLR 1127)	>	B76b *ò-lùɛ̀*, B77a *kí-lúɔ̀*
	*dòŋgò ‘line, row’	(BLR 1133)	>	B76b *mò-lùɔ̀*, B77a *ú-lúɔ́*
	*dʊ́ŋgʊ́ ‘pepper’	(BLR 1223)	>	B76b *lè-lúú*, B77a *à-ndzú*
	*gàŋgà ‘medicine man’	(BLR 1332)	>	B76b *ŋ-gàà*, B77a *ŋ-gàà*
	*káŋgà ‘guinea fowl’	(BLR 1720)	>	B76b *ŋ-káá*, B77a *ŋ-káà*
	*tʊ́ŋg-à ‘build’	(BLR 3081)	>	B76b *ò-tùà*, B77a *kì-tsúà*

As shown in (2) with data from the same language varieties, this loss does not take place at the beginning of words, that is, in C1 position. Proto-Bantu *ŋg is preserved there.[5]

(2)	*N-gèmbʊ́á ‘bat’	(BLR 1357)	>	B76b *ŋ-gyèèmú*, B77a *ŋ-gèèmè*
	*N-gòmbè ‘cattle’	(BLR 1434)	>	B76b *ŋ-gɔ́mɛ̀*, B77a *ŋ-gɔ́mɛ̀*
	*N-gʊ̀dʊ́ ‘pig’	(BLR 1493)	>	B76b *ŋ-gùlù*
	*N-gʊ̀dʊ̀bè ‘pig’	(BLR 1494)	>	B77a *ŋ-gùlùpì*
	*N-gòì ‘leopard’	(BLR 7154)	>	B76b *ŋ-gɔ̀*, B77a *ŋ-gɔ̀*
	*N-gʊ́kʊ̀ ‘mother’[6]		>	B76b *ŋ-gú*, B77a *ŋ-gúɣù*

---

[4] Unlike in Bastin et al. (2002), reconstructed verb stems are presented with their default final vowel -à preceded by a hyphen; see also Appendix 3. To mark the historical morpheme boundary between a noun prefix and the stem, we separate the two with a hyphen in the reflexes, even if this may sometimes be at odds with present-day morphology due to the occasional integration of noun prefixes into the noun stem. We also systematically replace the notation of *ng in Bastin et al. (2002) by one that is closer to phonetic reality, i.e., *ŋg. Nasals in Bantu NC clusters usually assimilate to the following consonant’s place of articulation (Hyman, 2019: 136).

[5] To the reconstructed noun stems in (2) we add the nasal prefix of Proto-Bantu noun classes 9 and 10 (Bostoen, 2019: 313). It is written in small caps (i.e., as N-) because it is a homorganic nasal. It adapts to the place of articulation of the stem-initial consonant and is thus realized here as a velar nasal. Note that certain nouns manifest the common Bantu diachronic sound change known as Meinhof’s Rule (Meeussen, 1962; Dammann, 1972), whereby a NC cluster (i.e., a sequence of a nasal and a stop) in C1 position is reduced to a simple nasal (N) when C2 also contains a NC cluster and/or when it contains a simple nasal, e.g., *gòmà ‘drum’ (BLR 1429) > *ŋɔ̀mɔ̀* (B76b); *gòndé ‘crocodile’ (BLR 486) > *ŋwɔ̀ɔ̀ní* (B76b). In these two nouns, the word-initial consonant is the simple velar nasal /ŋ/ and not the prenasalized consonant cluster /ŋg/ that is shown in the examples in (2). Nonetheless, Meinhof’s Rule is systematic in neither Ŋiŋi nor Kukuya.

[6] This lexical reconstruction does not feature in Bastin et al. (2002); see Section 2. It is a ten-

Elsewhere in WCB outside of the Kwilu-Ngounie branch, the complete loss of *ŋg is unattested in C2 position, as Map 2 shows. For ease of exposition, we omit from Map 2 varieties which show multiple unconditioned reflexes (/ŋg/, /ŋ/, and/or /∅/) in almost equal proportions in C2 position; see Map 3 and the discussion in Sections 4.3 and 4.4 for details.

Zero reflexes of *ŋg in C2 position (indicated by the symbol ◦ in Map 2) occur neither in the two other main WCB branches, that is, Kamtsha-Kwilu and Kikongo Language Cluster (KLC) Extended, nor in the paraphyletic varieties of the homeland area that branch off first. Languages within these branches have a segmental reflex of *ŋg, that is, either the prenasalized voiced velar /ŋg/ (indicated by the symbol □ in Map 2) or the simple velar nasal /ŋ/ (indicated by the symbol ◇ in Map 2).

Within the KLC Extended branch, the KLC is conservative when it comes to *ŋg in C2 position. Its full retention as /ŋg/ is attested without exception across the KLC, as illustrated in (3) with data from Punu B43 (West Kongo; Nsuka-Nkutsi, 1980) and Bembe H11 (North Kongo; Kouarata, 2016).

(3) *dòŋg-à 'advise, teach' (BLR 1127) > B43 *ù-lɔ̌ŋgə̀*, H11 *kù-lɔ́ɔ́ŋgɔ̀*
*dʊ́ŋgʊ́ 'pepper' (BLR 1223) > B43 *núŋgù*, H11 *lù-ɲúúŋgù*
*gàŋgà 'medicine man' (BLR 1332) > B43 *ŋ-gǎŋgə̀*, H11 *ŋ-gááŋgà*
*kíŋgó 'neck' (BLR 1845) > B43 *kíŋgu*, H11 *ŋ-kíiŋgú*
*táŋg-à 'read, count' (BLR 2786) > B43 *ù-ráŋgə̀*, H11 *kù-tááŋgà*
*tʊ́ŋg-à 'build' (BLR 3081) > B43 *ù-rúŋgə̀*, H11 *kù-túúŋgá*

The retention of *ŋg is not restricted to the KLC. As we discuss in Section 4, the Nzebi-Teke West subgroup, which comprises the most western Kwilu-Ngounie languages, also preserved *ŋg in C2 position. In all B50 varieties and Yaa B73c, *ŋg in C2 is systematically preserved. In Laali B73b, the conservative /ŋg/ reflex shows evidence for lenition to /ŋ/ (see discussion in Section 4.1). This regular shared retention is illustrated in (4) with data from Duma B51 and Nzebi B52 (Mouélé, 1997), as well as Yaa B73c (Mouandza, 2001).

(4) *bìng-à 'chase' (BLR 213) > B51 *m-bìŋgù*, B73c *ú-bíŋgì*
*bóŋgó 'knee' (BLR 275) > B51 *∅-bɔ́ɔ́ŋgɔ́*, B73c *∅-bɔ̋ɔ̋ŋgɔ́*
*cìngà 'string' (BLR 622) > B51 *∅-sììŋgà*, B73c *mú-sííŋgì*
*dòŋg-à 'advise, teach' (BLR 1127) > B51 *lɔ̀ɔ̀ŋgɔ̀*, B73c *n-dóòŋgí*

---

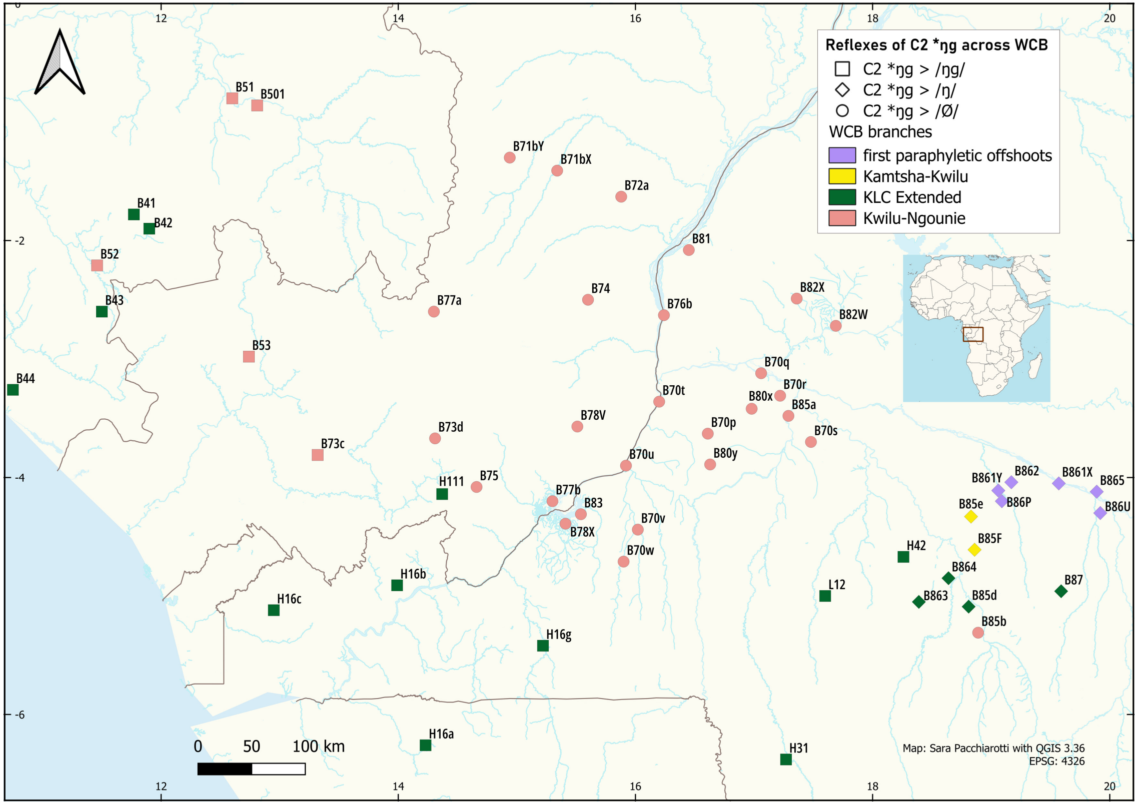
tative reconstruction proposed on the basis of comparative evidence from several Kwilu-Ngounie languages and possibly goes back to their most recent common ancestor.

*dʊ́ŋgʊ́ ‘pepper’	(BLR 1223)	>	B51 *n-dúúŋgú*, B73c *n-dű̃ű̃ŋgú*
*gàŋgà ‘medicine man’	(BLR 1332)	>	B51 *ŋ-gààŋgà*, B73c *ŋ-gáángà*
*kíŋgó ‘neck’	(BLR 1845)	>	B51 *∅-kííŋgú*, B73c *∅-kĩ̋ĩ̋ŋgí*
*táŋg-à ‘read, count’	(BLR 2786)	>	B52 *ù-ráŋgə̀*, B73c *ú-tã̋ã̋ngà*
*tʊ́ŋg-à ‘build’	(BLR 3081)	>	B51 *∅-tóóŋgà*, B73c *ú-tű̃ű̃ŋgù*

In the wider homeland area in the east, that is, roughly between the Kwilu and Kasai Rivers in the DRC, this archaism does not occur. Across the diverse subgroups represented there, the velar nasal, which is the outcome of the *ŋg > ŋ cluster reduction, is the prevalent reflex. It is attested in the small Kamtsha-Kwilu branch, the paraphyletic varieties branching off first within the KLC Extended branch, and the paraphyletic grade residing immediately under the ancestral WCB node and consisting of Ding B86, Ngwi B861, Lwel B862, and Nzadi B865. In (5), the systematic sound shift *ŋg > ŋ in C2 is illustrated with data from Nsambaan B85F (Kamtsha-Kwilu; Koni Muluwa and Bostoen, 2015), Nsong B85d (KLC Extended; Koni Muluwa and Bostoen, 2015, 2019), and Ngwi B861 (WCB; Pacchiarotti and Bostoen, 2021a, 2021b, 2022).

(5)	*bóŋgó ‘knee’	(BLR 275)	>	B85F *é-bɔ̀ŋ*, B85d *ɛ́-bɔ̀ŋ*, B861 *ì-bwɔ́ŋ*
	*cáŋgʊ́ ‘small seeds’	(BLR 487)	>	B85F *lá-sáŋ*, B85d *ɛ̀-sáŋ* ‘rattle’
	*dòŋgà ‘river’	(BLR 1128)	>	B85F *n-dɔ̀ŋ*, B861 *n-dwɔ̌ŋ*
	*dʊ́ŋgʊ́ ‘pepper’	(BLR 1223)	>	B85F *n-dúŋ*, B861 *è-lúŋ*
	*gàŋgà ‘medicine man’	(BLR 1332)	>	B85F *ŋ-gáŋ*, B85d *ŋ-gáŋ*, B861 *ò-ŋ-gɛ́àŋ*
	*gʊ̀ŋgà ‘bell’	(BLR 1514)	>	B85F *ŋ-gùŋ*, B85d *ŋ-gùŋ*
	*tʊ́ŋg-à ‘build’	(BLR 3081)	>	B85F *kà-túŋ*, B85d *kó-tùŋ*, B861 *tûŋ*
	*cʊ̀ŋgʊ̀ ‘sugarcane’	(BLR 5111)	>	B85F *má-súŋ*, B85d *mó-súŋ*, B861 *ò-ʃûŋ*
	*cèŋgè ‘main village’	(BLR 7720)	>	B85F *má-sɛ́ŋ*, B85d *mɔ̀-sɛ́ŋ*

Interestingly, East Yans varieties, which are an integral part of Kwilu-Ngounie in the lexicon-based phylogeny of Pacchiarotti et al. (2019), are major exceptions to this areal pattern. Although being surrounded by varieties having /ŋ/ as reflex of C2 *ŋg, East Yans has zero as the most common reflex, as shown in (6) with data from the Niadi B85b variety (Nguma, 1986). West Yans B85a, spoken considerably more to the northwest, displays zero as a reflex of C2 *ŋg in all cases; see Section 4.2.



MAP 2 Overview of reflexes of Proto-Bantu C2 *ŋg across West-Coastal Bantu languages (except Mbete B60, Latege B71bZ, and Boma Yumu B80z)

(6) East Yans (Niadi) B85b

*càŋgà 'island'	(BLR 475)	>	*e-saa*
*càŋgò 'news'	(BLR 479)	>	*mu-saa*
*gàŋgà 'medicine man'	(BLR 1332)	>	*ŋ-gaa*
*gòŋgà 'bell'	(BLR 1514)	>	*ŋ-guu*
*jòŋgó 'cooking pot'	(BLR 1632)	>	*n-zuu*
*tóŋgá 'basket'	(BLR 3080)	>	*mu-lwɔ*
*tóŋg-à 'build'	(BLR 3081)	>	*o-twɔ*
*còŋgò 'sugarcane'	(BLR 5111)	>	*mu-swɔ*
*cèŋgè 'main village'	(BLR 7720)	>	*mu-syɛ*

In Map 2, this East Yans variety is not only a single pink circle in a sea of multicolored diamonds, but it is also geographically removed from all the other pink circles. This specific East Yans variety is fully representative of other East Yans varieties such as Nkara B85bT as documented by Koni Muluwa and Bostoen (2015) in that others also consistently lose C2 *ŋg. East Yans varieties are amongst the few WCB languages where geography and phylogeny do not align. Their closest relatives, that is, other Kwilu-Ngounie languages, are not their closest geographic neighbors. Nonetheless, they do share the highly distinctive *ŋg > ∅ sound shift in C2 position with their geographically remote but genealogically closest relatives. This is in contrast with their geographically closer but genealogically more distant relatives which all manifest *ŋg > ŋ.

Before moving to Section 4, it is worth noting that the consonant cluster reduction observed in *ŋg > ŋ is not unique for the voiced velar NC cluster. With the exception of the KLC and Nzebi-Teke West subgroups—see (3) and (4)—both situated in the extreme west of the WCB distribution area, all WCB languages simplified voiced bilabial (*mb) and alveolar (*nd) nasal clusters in favor of the nasal; that is, *mb > m and *nd > n.[7] Unlike the Bantu dissimilatory sound change known as Meinhof's Rule, whereby a NC cluster is reduced to N in C1 position if the word contains another NC cluster in C2 position (e.g., *N-dòŋgó 'pot' > *nòŋgó*), the NC reduction observed in WCB languages does not appear to have any conditioning and targets exclusively NC in C2 position, as shown in with data from Kwilu-Ngounie languages included in this study in (7) and (8).

---

[7] As can be seen in (7) and (8), in some languages /m/ and /n/ historically deriving from *mb and *nd are also disappearing, see e.g., B72a *òbũ̂*, B74 *bíɔ̃̀* in (7a); B72a *nzìá̃(n)/ànzìá̃(n)*, B74 *ndzaɔ̃́* in (7b); B70q *màlìĩ̀*, B72a *àlũ̏*, B74 *aliɔ̃* in (7c); B72 *ŋàá/àŋàá* in (8b); B70r *mùkàá̃/mìkàá̃, òŋkà̃á̃(n)/ìŋkà̃á̃(n)* in (8d). On this phenomenon, see Paulian (1994). It seems likely that the loss of *ŋg which characterizes these varieties pushed other reduced nasal clusters to disappear, leaving nasalized vowels as a trace.

(7) *mb > m
   a. *bímb 'swell' (BLR 240) > B70q *ò-bìmì*, B70t *ó-bímà*, B70u *ò-byúúmù*, B70v *ó-bímì*, B72a *ò-bɩɩ̂̀*, B73d *ɔ́-bíímè*, B74 *bɩ́ɔ̃̀*, B76b *ò-bìmà*, B80y *ì-m-bvíímɛ̀*, B83 *ó-bímì*
   b. *jàmbé 'God' (BLR 3196) > B70q *n-zyɛ̀ɛ̀mɛ́/bà-n-zyɛ̀ɛ̀mɛ́*, B70s *n-zyém/bá-n-zyém*, B70v *n-zyààmí/bá-n-zyààmí*, B72a *n-zɩ̀ã́(n)/à-n-zɩ̀ã́(n)*, B73d *n-zààmá*, B74 *n-dzaɔ̃́*, B76b *n-zyààmí|bà-n-zyààmí*, B80z *n-zɩ̀ɛ̀m*, B82W *n-zà:mí*, B83 *n-dzyààmì|bá-n-dzyààmì*, B85b *n-zyam*
   c. *dìmbò 'birdlime' (BLR 985) > B70q *mà-lɩ̀ɩ̃̀*, B70r *má-lém*, B70s *mà-líím*, B70u *n-dímà| má-límà*, B70v *∅-líímì/má-líímì*, B70w *∅-líímɛ̀*, B72a *à-lɩɩ̀̀*, B74 *à-lɩ̀ɔ̃̀*, B76b *mà-lìmà*, B80z *mà-liɛ́m*, B85a *mu-liim*

(8) *nd > n
   a. *ténd 'cut' (BLR 2844) > B70t *ò-tyɛ́ɛ́nɛ̀*, B70v *ɔ́-tyɛ́ɛ́nɛ̀*, B73d *∅-tyɛ́nɛ̀*, B74 *∅-tínà*, B77a *∅-tyɛ́nɛ̀*, B83Z *∅-tyénè*, B85b *o-ten*
   b. *gàndʊ́ 'crocodile' (BLR 1326) > B63 *ŋ-gánì*, B70s *ŋ-gáán/bá-ŋ-gáán*, B70u *ŋ-gàànà/bá-ŋ-gàànà*, B72 *∅-ŋàá/à-ŋàá*, B73d *n-gàànà*, B74 *∅-ŋààni*, B77b *n-gànù*, B80x *∅-ŋáán/bà-ŋáán*, B80z *n-gàn*, B83 *n-gàànà/bá-n-gààna*, B85b *ŋ-gaan*
   c. *gòndé 'crocodile' (BLR 1446) > B76b *∅-ŋwòòní/bà-ŋwòòní*, B81 *∅-ŋɔ̀ɔ̀nɔ́*, B82 *n-gɔ̀ːɳɛ́*
   d. *kàndá 'letter' (BLR 1706) > B70q *mò-kàànáá/mì-kàànáá*, B70r *mù-kàã́/mì-kàã́*, B70w *mú-kàànà/mí-kàànà*, B72 *-ò-ŋ-kã̀ã́(n)/ì-ŋ-kã̀ã́(n)*, B74 *ù-ŋ-káánáá*, B76b *mò-kàànáá/mè-kàànáá*, B80x *mò-kã̀ã́*, B81 *mò-kàànáá*

The fact that the low-level KLC and Nzebi-Teke West subgroups, both part of distinct major WCB branches (KLC Extended and Kwilu-Ngounie, respectively), escaped this widespread voiced NC cluster reduction suggests that this innovation only started after an initial phase of divergence within WCB. This pulse of fragmentation involved the expansion of both KLC Extended and Kwilu-Ngounie branches toward the Atlantic coast, the first in the south, the second further north (see also the discussion in Section 6).

4 Word-final loss of Proto-Bantu *ŋg within Kwilu-Ngounie

As shown in Section 3, within WCB, *ŋg > ŋ > ∅ in C2 position only occurs in languages belonging to Kwilu-Ngounie. Nonetheless, even within that branch, this highly distinctive sound shift is not omnipresent. Although all circles in Map 2 are pink (i.e., belonging to Kwilu-Ngounie), not all pink symbols are circles. There are also pink squares (*ŋg > ŋg). Hence, if Kwilu-Ngounie is a valid genealogical unit within WCB, the sound change *ŋg > ŋ > ∅ certainly did not happen at the level of the branch's most recent common ancestor, but at a later stage. To better delineate that period in terms of relative chronology, we undertook a detailed comparative study of *ŋg loss in C2 position within Kwilu-Ngounie. Since our dataset is unbalanced due to the fact that not all varieties have a comparable number of reflexes (some have up to 84 while others only 23), we organize our discussion in the following subsections around the varieties on which we could gather the most data.

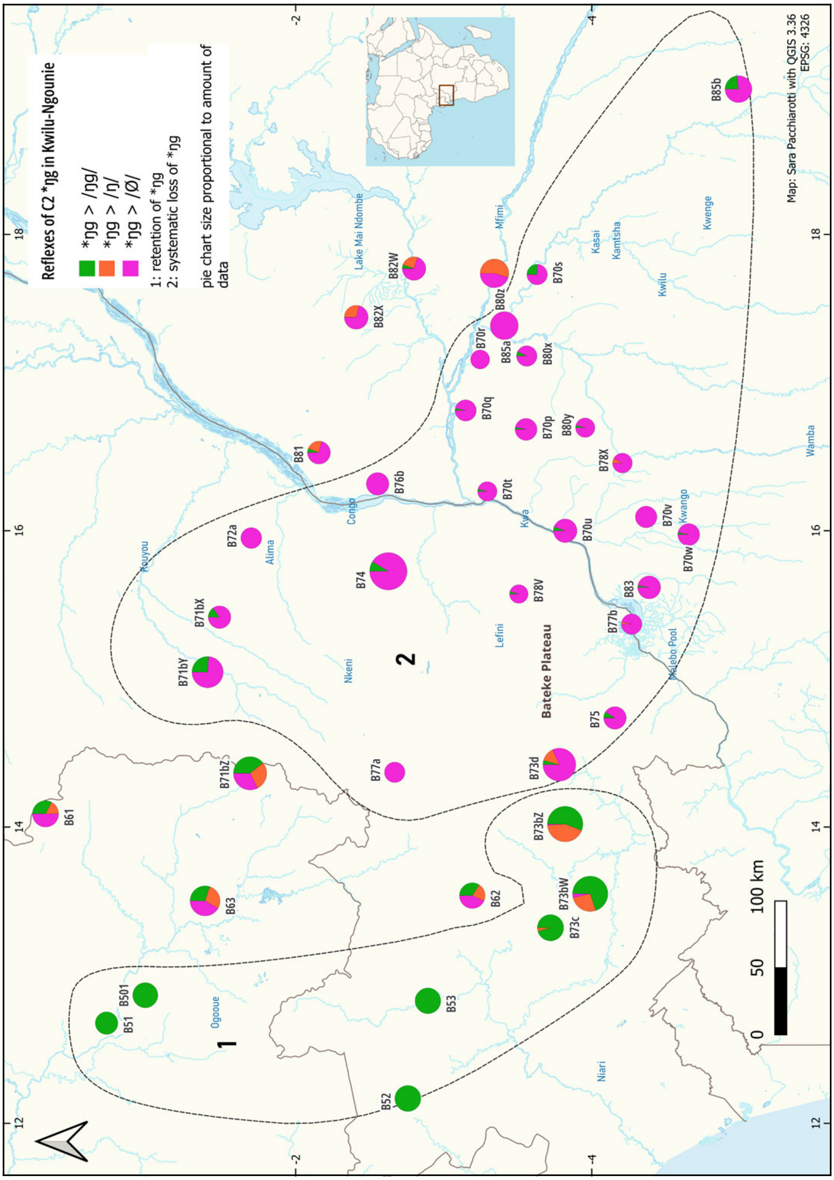
In Appendix 2, we show the reflexes of Proto-Bantu *ŋg in C2 position across 39 Kwilu-Ngounie varieties. We identified three reflexes for C2 *ŋg, namely /ŋg/, /ŋ/, and /∅/ (zero). The third column in Appendix 2 shows the total number of reflexes identified in each variety.[8] Varieties with more than one reflex in Appendix 2 display multiple unconditioned reflexes (see Pacchiarotti and Bostoen, 2022), that is, two or three reflexes of one and the same protosound (namely, *ŋg in C2 position) in the absence of any conditioning environment; see discussion in Section 4.4. The columns after the total number of reflexes are organized according to the number of lexical items with a given reflex and its corresponding percentage. To give an example, in Laali Mayeye B73bZ, we identified 84 relevant lexical items (including reflexes of protoforms containing *ŋg and synchronic forms containing /ŋg/ and /ŋ/ not linkable to any reconstruction). Of these, 47 had /ŋg/ in C2 (47/84 = 56 %), 36 had /ŋ/ in C2 (36/84 = 43 %), and only one among those linkable to a protoform showed /∅/ as a reflex (1/84 = 1 %).

Map 3 presents the data in Appendix 2 visually. The percentages of different *ŋg reflexes, that is, /ŋg/, /ŋ/, and /∅/, for each variety are depicted by means of pie charts. The size of each pie chart is proportional to the amount of data we were able to collect for each variety.

In the following subsections, we show that, on a continuum of variation across Kwilu-Ngounie, four distinct groups emerge based on the evolution of C2 *ŋg. The two extremities on that continuum are marked as distinct zones in Map 3. Zone 1 is a small group of conservative language varieties which have fully retained Proto-Bantu *ŋg in C2 (Section 4.1). This group overlaps entirely with monophyletic Nzebi-Teke West in the lexicon-based phylogeny of Pacchiarotti et al. (2019; see our Fig. 1). Zone 2 is a vast group of innovative language

---

[8] For languages having /ŋg/ or /ŋ/ in C2, we included in the count words displaying these sounds in that position even when we could not link them to existing reconstructions in the BLR2/3 database.



MAP 3 Reflexes of C2 *ŋg across Kwilu-Ngounie

varieties which have systematically lost C2 *ŋg (Section 4.2). Zone 2 comprises all varieties classified as paraphyletic grades within Kasai-Ngounie Extended based on lexical evidence (see Fig. 1).

Apart from these centers of retention (Zone 1) and innovation (Zone 2), enclosed by the dotted lines in Map 3, there are two groups of language varieties which are neither fully conservative nor fully innovative in terms of total *ŋg loss in C2 but seem to have reduced *ŋg in different ways.

A first intermediate group on the continuum of variation, mostly situated to the (north)east of the hub of innovation (Zone 2), consists of language varieties where the total loss of *ŋg in C2 is pervasive but not fully systematic (Section 4.3). This peripheral group either shared an ancestor with Zone 2 languages (Section 4.2) or partook in the lenition chain *ŋg > ŋ > ∅ due to contact with Zone 2 languages. Since this peripheral group contains only languages belonging to the monophyletic low-level Kwa-Kasai North group (see Fig. 1), we refer to it as Kwa-Kasai North.

A second intermediate group on the continuum of variation consists of languages varieties having all three reflexes of C2 *ŋg in very similar proportions (Section 4.4). We refer to this area, situated to the west of the center of innovation (Zone 2), as the "buffer zone," since we believe that the multiple unconditioned reflexes of C2 *ŋg here are mainly the outcome of lexical diffusion (Wang, 1969; Labov, 1981). The buffer zone includes the lexicon-based monophyletic low-level group Mbete (see Fig. 1) and one Teke doculect known as Latege B71bZ spoken in Gabon. In the following subsections, we discuss each zone in turn.

4.1 *Retention of C2 *ŋg*

As pointed out in Section 3, the Nzebi-Teke West clade within Kwilu-Ngounie is the most conservative in that it escaped the loss of *ŋg in C2 entirely Table 1. On Map 3, these are varieties with a pie chart that is (almost) completely green. As shown in Table 1, zero reflexes of *ŋg in C2 are nearly unattested in Nzebi-Teke West. Only Laali B73b is exceptionally innovative in showing the ongoing lenition process ŋg > ŋ (see further discussion below).

The monophyletic Nzebi-Teke West group in Fig. 1 has systematically retained /ŋg/ in C2. Nevertheless, in Yaa B73c, we have found two instances where Mouandza (2001) notes /ŋ/ instead of /ŋg/: *mbő̋ŋíŋì* 'elbow' and *kwắắŋí* 'cassava.' These lexical items are possibly the result of contact-induced influence or borrowings from Laali B73b, as discussed below. As the preservation of *ŋg in C2 position is a shared retention and not a shared innovation, it cannot corroborate the genealogical unity of Nzebi-Teke West in itself. Nonetheless, it is still genealogically relevant. First, the absence of *ŋg > ∅ indicates that this

TABLE 1 Kwilu-Ngounie varieties with retention of C2 *ŋg

Variety	**No. of reflexes**	**Reflex**					
		ŋg		**ŋ**		**∅**	
		No.	**%**	**No.**	**%**	**No.**	**%**
B501	42	42	100%	0	–	0	–
B51	34	34	100%	0	–	0	–
B52	45	45	100%	0	–	0	–
B53	44	44	100%	0	–	0	–
B73bZ	84	47	56%	36	43%	1	1%
B73bW	83	58	71%	22	27%	3	2%
B73c	46	44	94%	2	6%	0	–

distinctive sound shift only started within Kwilu-Ngounie after the Nzebi-Teke West subgroup branched off. Second, the nearly full retention of *ŋg in Yaa B73c and the B50 varieties indicates that Nzebi-Teke West also did not partake in a *ŋg > ŋ shift that very likely preceded total loss, that is, *ŋg > ŋ > ∅.

As can be observed in Table 1, unlike all other Nzebi-Teke West varieties, Laali Mayeye B73bZ and Laali Kendi B73W are the only ones currently in the process of undergoing the cluster simplification *ŋg > ŋ. The data in (9) show that it is impossible to find a (supra)segmental conditioning environment for this change.

(9) Laali (Mayeye) B73bZ
a. /ŋg/ *bìŋgá 'green pigeon' (BLR 216) > *m-byɛ́ɛ̀ŋgɛ̀*
*jʊ̀ŋgʊ́ 'pot' (BLR1632) > *n-zúúŋgɔ̀*
*káŋgà 'guinea fowl' (BLR 1729) > *∅-káàŋgá* 'black-crowned crane'[9]

---

[9] The original French translation for *káàŋgá* in Bissila (1991: 41) is *grue couronnée*, known in English as 'black-crowned crane.' While no scientific name is given in the original source, this common name usually refers to *Balearica pavonina*. However, as pointed out by an anonymous reviewer, *Balearica pavonina* is unlikely to be the right species denoted by the term *káàŋgá* in Laali B73b, because cranes are arid savanna birds, and Laali is not spoken in that environment. Possibly, *grue couronnée* in Bissila (1991) refers to the helmeted guinea fowl (*Numida meleagris*) or to the plumed guinea fowl (*Guttera plumifera*). Both species are found

	*jóŋgò ‘bile’	(BLR 3573)	>	*ɲ-ɔ́ɔ́ŋgɔ̀*
	*boŋgo ‘money’		>	*m-bɔ̀ɔ̀ŋgɔ́*
	*dʊŋgʊ ‘canoe’		>	*∅-lúúŋgɔ̀*
	*daŋga ‘taro’		>	*∅-lááŋgà*
b. /ŋ/	*bìŋg-à ‘hunt’	(BLR 213)	>	*ɔ́-byɛ̀ŋɛ̀*
	*bóŋgó ‘knee’	(BLR 275)	>	*bɔ́ɔ̀ŋɔ̀*
	*gòŋgò ‘back’	(BLR 1450)	>	*mɔ́-ŋ-gɔ́ɔ́ŋɔ̀*
	*káŋg-à ‘fry’	(BLR 1719)	>	*ɔ́-kááŋa*
	*tʊ̀ŋgí ‘corner’	(BLR 5091)	>	*ɛ́-tsúúŋɔ̀*
	*dʊ̀ŋgá ‘ring’		>	*∅-lúúŋɔ̀*

Considering the percentages of words with C2 /ŋg/ (56 %) vs. /ŋ/ (43 %) in the most innovative variety of Laali, that is, Laali Mayeye B73bZ, the simplification of C2 *ŋg is either fully ongoing or has stopped before reaching completion. Note that out of 84 words from the variety of Laali documented in Bissila (1991), that is, Laali Mayeye B73bZ, only one had zero as a reflex of C2 *ŋg, namely *ɔ́-dzí* ‘to roll up’ (< *díng ‘wrap up,’ BLR 1062). This is despite the geographical proximity to B70 varieties where C2 *ŋg disappeared entirely, see for example, Tyee B73d in Map 3. Given that (a) nowhere within Kwilu-Ngounie does *ŋg > ŋ occur as frequently as in Laali Mayeye B73bZ, and (b) no neighboring variety manifests the same shift, conceivably only the first wave of the change *ŋg > ŋ reached the Mayeye variety, while ŋ > ∅ either did not reach it or was stopped. The data from Laali Kendi B73bW, suggests that different varieties are at different stages within the lenition chain: Laali Kendi appears to be slightly more conservative than Laali Mayeye. The very low percentage of zero reflexes in these Laali varieties suggests that zero might be the next natural development in this chain. Complete loss might be influenced by the presence of surrounding B70 languages which regularly underwent the change *ŋg > (ŋ) > ∅ (see Section 4.2).

4.2 *Systematic loss of C2 *ŋg*

Varieties which systematically lost *ŋg in C2 are geographically separated from the conservative Nzebi-Teke West subgroup (see Section 4.1) by a buffer zone where C2 *ŋg shows highly irregular reflexes (see Map 3 and the discussion in Section 4.4). Languages with systematic loss of C2 *ŋg cover a vast area, starting in the center of the Bateke Plateau in the Republic of the Congo and extending

---

in the Republic of the Congo where Laali is spoken and the latter has a particularly conspicuous plume of black feathers on its head which could perhaps explain the use of *couronnée* ‘crowned’ in the French translation.

TABLE 2 Kwilu-Ngounie varieties with systematic loss of C2 *ŋg

Variety	No. of reflexes	Reflex					
		ŋg		ŋ		∅	
		No.	%	No.	%	No.	%
B70p	32	1	1%	0	–	31	99%
B70q	30	1	4%	0	–	29	96%
B70r	24	0	–	0	–	24	100%
B70s	28	7	25%	0	–	21	75%
B70t	25	1	4%	0	–	24	96%
B70u	37	2	5%	0	–	35	95%
B70v	31	0	–	0	–	31	100%
B70w	32	1	4%	0	–	31	96%
B71bX	34	5	15%	0		29	85%
B71bY	65	17	26%	0	–	48	74%
B72a	29	0	–	0	–	29	100%
B73d	72	3	4%	10	14%	59	82%
B74	93	8	9%	0	–	85	91%
B75	34	3	9%	0	–	31	91%
B76b	34	0	–	0	–	34	100%
B77a	28	0	–	0	–	28	100%
B77b	28	0	–	1	4%	27	96%
B78V	23	1	4%	0	–	22	96%
B78X	26	0	–	2	8%	24	92%
B80x	28	2	5%	0	–	26	95%
B80y	25	1	4%	0	–	24	96%
B83	35	1	4%	0	–	34	96%
B85a	52	0		0		52	100%
B85b	48	11	23%	0	–	37	77%

all the way up to the Kwilu River in southwestern DRC. They are listed in Table 2. On Map 3, these are the varieties whose pie chart is almost entirely pink, in Zone 2. Although we do not address this issue here, only in a handful of the varieties listed in Table 2 has the loss of C2 *ŋg resulted in the development of phonological nasal vowels (see Hombert, 1986, for a detailed discussion). As can be seen in Appendix 3, in most languages the loss resulted in a CVV structure, occasionally shortened to CV.

Somewhat arbitrarily, we have placed a first tentative cutoff point at 90%; that is, varieties with /∅/ as a reflex in 90% or more of all identifiable reflexes of reconstructions with *ŋg in C2 position are considered as varieties which have systematically (i.e., regularly) innovated the loss of the velar nasal cluster in this phonotactic position. We also include in Table 2 a handful of cases where the percentage of *ŋg > ∅ is slightly lower (88–74%). While we discuss each individual case in this section, we do not believe these data justify lowering the cutoff point below 90%.

In all languages with 90% or more of zero reflexes, the exceptions to the *ŋg > ∅ innovation—most commonly /ŋg/, with /ŋ/ present only in B77b and B78X—are lexemes which are often shared across varieties in Table 2 and the wider region, specifically the area around Lake Mai Ndombe, see Map 3 and B80z, B81, and B82X reflexes in (10).

(10) a. *tʊ̀ŋgá 'basket' (BLR 3082) > B70w *í-tùŋgà/bí-tùŋgà*, B74 *ì-túnga*, B75 *í-tùngà/bí-tùngà*, B78V *íntúngà/bíntúngà*
b. *dòŋgà 'plate' (BLR 1131) > B70p *lóngá/mà-lóngá*, B70t *lùŋgá/mà-lùŋgá*, B70u *lóóŋgà/mà-lóóŋgà*, B80y *lùŋgá/mà-lùŋgá*, B80z *í-lóng*, B85b *lɔnga*
c. *páŋgò 'cave' (BLR 2404) > B70s *lé-pááŋg/má-pááŋg* 'chief's enclosure,' B85b *le-pango* 'enclosure,' B74 *lì-pángù* 'cave, enclosure,' B71bY *lè-pángì* 'cave, enclosure'
d. *gʊ̀ŋgà 'bell' (BLR 1514) > B70s *ŋ-gùŋg/bà-ŋgùŋg*, B74 *n-gùngà*, B75 *n-gúúngà*, B80z *n-gùngà*, B82X *ŋúŋà*, B83Z *ŋ-gɔ́ɔ́ŋ*, B85b *ŋ-gunga*
e. *doŋgodoŋgo 'okra' (*Abelmoschus esculentus*) > B77b *dódóŋó*, B74 *dɔ̀ngɔ́-dɔ̀ngɔ́*, B73d *dɔ̀ngɔ́dɔ̀ngɔ́*, B71bY *dɔ̀ngɔ́dɔ̀ngɔ́*, B85b *dongo-dongo*
f. *daŋgɪ 'bottle' > B71bY *ò-làngì*, B74 *ù-làngì*, B80z *mù-làng*, B81 *mò-làŋè*, B82X *mù-láŋì/mì-láŋì*, B85b *mu-langi*

While these lexemes are widely attested elsewhere in Kwilu-Ngounie languages, they are formally irregular (in that they have C2 /ŋg/ instead of zero) only in the varieties in Table 2 and some of those in Section 4.3. All the words in (10) are found with identical meanings in Lingala and/or Kongo Ya Leta, the two lingua francas of the region. Hence, they are in all likelihood borrowings.

In some varieties, borrowings containing C2 /ŋg/ in the vehicular languages Lingala, Kongo Ya Leta, or French may undergo nativization by deleting the velar nasal plus consonant; see, for example, the varieties in Table 2 where the reflex of a given reconstruction does not feature /ŋg/ but zero. In addition, in Tyee B73d the borrowed word for 'mango' has been nativized to *maã* (Ntsiba Ngolo and Raharimanantsoa, 2021: 33); compare with *mangulu* in Eboo B74, where no nativization has taken place (Raharimanantsoa, 2021). The same process is observed in Eboo B74 (Raharimanantsoa, 2021), where the word *nzùŋgù* 'pot' borrowed from Lingala according to Raharimanantsoa (2021: 198) can be alternatively realized as *nzuu*.[10] This might suggest that the innovative loss of C2 *ŋg has become a distinctive linguistic feature of at least some B70 languages, one which is implemented to nativize borrowings.

Strikingly, in so far as these varieties regularly attesting total *ŋg loss were included in the phylogenetic study of Pacchiarotti et al. (2019)—most are not—they belong to a paraphyletic grade at some level of internal Kwilu-Ngounie classification (see Fig. 1). Ngungwel B72a, Kukuya B77a, and Fumu B77b, all of which manifest C2 *ŋg loss in 100 % of reflexes, belong to the paraphyletic grade within Kasai-Ngounie. Wuumu B78, South Boma B80y, Mfinu B83, Eboo B74, and Boma Nkuu B80x, where C2 *ŋg loss ranges between 96 % and 92 %, are part of the paraphyly within Kasai-Ngounie Extended, which is parallel to all Kasai-Ngounie. In West Yans B85a, not included in Pacchiarotti et al. (2019), word-final *ŋg loss is also complete.[11]

Hence, the innovation *ŋg > ∅ appears to be a good candidate to group the paraphyletic varieties in the lexicon-based phylogeny of Pacchiarotti et al. (2019). Instead of being paraphyletic grades within Kwilu-Ngounie, the languages manifesting *ŋg > ∅ in C2 position could descend from a most recent common ancestor from which they inherited this diagnostic sound shift. This would mean that they constitute a discrete subgroup within Kwilu-Ngounie based on a uniquely shared phonological innovation. If word-final *ŋg loss is indeed indicative of a new genealogical subgroup resolving all paraphyly within Kwilu-Ngounie as defined by Pacchiarotti et al. (2019), the question is how this newly identified subgroup based on a shared phonological innovation relates to the subgroups emerging from the lexicon-based phylogeny in Fig. 1.

We now turn to the discussion of the varieties in Table 2 which display a percentage of zero reflexes for C2 *ŋg lower than 90 %, but still in the range of 88–74 %. We start with the two easternmost varieties Kaan B70s and Yans B85b. Both of these show roughly 75 % zero reflexes and 25 % /ŋg/ reflexes. Kaan B70s is a so-called Teke variety not included in referential classifications

---

[10] The very same process happens with the same word in Latege B71bZ; see Linton (2013a).

[11] All East Yans varieties in Pacchiarotti et al. (2019) end up in a paraphyletic grade sister to Kasai-Ngounie Extended within Kwilu-Ngounie. While it is possible that West Yans would occupy the same position, this is not to be taken for granted because in the phylogeny of Pacchiarotti et al. (2019) many doculects that are commonly presumed to be varieties of the same language (i.e., labeled with the same glossonym) end up in different places within the tree.

of Bantu languages (Guthrie, 1971; Maho, 2009). As far as we can tell, it was documented for the first time by the second author during a fieldtrip to the DRC in 2021 (Kouarata et al., 2023). Like several other Teke languages spoken in the vast plateau northeast of Kinshasa (Pacchiarotti and Bostoen, 2021b), Kaan undergoes non-systematic final vowel loss (see Appendix 3 for examples). The items which have [ŋg] instead of zero in our limited dataset are in (11).

(11) Kaan B70s
n-tsààŋg/bà-n-tsààŋg 'news'
ì-lààŋg/bì-lààŋg 'field'
ì-lɔ́ŋgɔ́k 'to learn'
lɛ́-kɔ̀ŋgá/má-kɔ̀ŋgá 'spear'
ŋ-gùŋg/bà-ŋ-gùŋg 'bell'
lé-pááŋg/má-pááŋg 'chief's enclosure'

These words look like recent borrowings, possibly from Kongo Ya Leta or Lingala, for at least two reasons: (a) those which have lost their final vowels still preserve [ŋg], that is, there has not been a lenition such as ŋg > ŋ found everywhere in languages which have lost their final vowels systematically (those found around the WCB homeland area; Pacchiarotti and Bostoen, 2021b); and (b) lexical items for concepts such as 'bell' in (11) are also borrowings in East Yans varieties, or in those having 90% or more of zero reflexes of *ŋg—see (10c) and (10d). Unlike all other Kwilu-Ngounie varieties discussed in this paper, Yans varieties have systematically undergone the diachronic sound change of final vowel loss (Pacchiarotti and Bostoen, 2021b). As shown in (12), in the Niadi variety of East Yans B85b, all words with C2 [ŋg] are easily recognizable borrowings, probably from vehicular Kongo Ya Leta or Lingala because: (a) they preserve their final vowel, and (b) some of them are also borrowings in other Kwilu-Ngounie varieties or Kaan B70s, such as 'bell,' 'plate,' 'bottle,' 'enclosure,' and 'to teach'; see (10), but also (14)–(16) and (18). Note that in East Yans B85b some borrowed words are undergoing nativization by losing their final vowel; see the alternation between *∅-lɔŋga ~ le-lɔŋg* 'plate,' *o-lɔŋg* 'to teach' (but *mu-lɔŋgi* 'teacher'), *o-yuŋgul* 'to sieve,' and *∅-fuŋgul* 'key, padlock' (also a borrowing in other varieties; see, e.g., Tyee B73d *∅-fùŋgúrò*, Mbaama B62 *∅-fúngúlá/á-fúngúlá*, Boma Yumu B80z *∅-fùngúlà*).

(12) East Yans B85b
∅-lɔŋga ~ le-lɔŋg 'plate'
ŋ-guŋga 'bell'
le-paŋgo 'enclosure'

le-vuŋga 'loincloth'
mu-laŋgi 'bottle'
ke-yoŋgo 'beggar'
mu-m-boŋgo 'commerce' (cf. Lingala/Kongo Ya Leta *boŋgo* 'money')
∅-fuŋgul 'key'
o-lɔŋg 'to teach'
o-yuŋgul 'to sieve'
kɛ-laŋg 'field'

In sum, both Kaan B70s and East Yans B85b are geographically removed from the center of innovation of Kwilu-Ngounie (see Map 3) but still preserve a considerable majority of zero reflexes despite borrowings. Perhaps especially for Kaan, the limited amount of data available to us has caused the number of borrowings to have a greater impact on the total number of reflexes of C2 *ŋg.

We now turn to Tyee B73d, a variety where *ŋg > ∅ (82 %), *ŋg > ŋ (14 %), and *ŋg > ŋg (4 %). First, we note that Tyee B73d and geographically close Tyɔɔ B74c, spoken in Kingoué (latitude – 3.77, longitude 14.17) but not included in this study due to scantiness of data, are the only two varieties we are aware of where /ɣ/ is being inserted to break up the sequence of two vowels created by the loss of *ŋg in C2; for example, *búng-à 'mix' (BLR 385) > B73d *ó-bvúɣù ~ ó-bvúútò*, B74c *ɔ́-bvúɣù*; *kíngó 'neck' (BLR 1805) > B73d *ŋ-kíí ~ ŋ-kíɣí*, B74c *ŋ-kíɣí*; *kíngá 'mountain' (BLR 5706) > B74c *ŋ-kíɣì*; *dʊ́ngʊ́ 'pepper' (BLR 1223) > B73d *n-dwú ~ n-dúɣú*, B74c *n-dúɣú*; *cʊ̀ngʊ̀ 'sugarcane' (BLR 5111) > B73d *mú-sṹũ̀/mí-sṹũ̀,* B74c *mú-súɣù.*

The Tyee lexical items featuring /ŋg/ and /ŋ/ instead of the expected /∅/ reflex are listed in (13). Throughout this section, the absence of an etymon for a reflex containing /ŋ/ or /ŋg/ means that we do not have enough comparative evidence in our database (see Section 2) to set up a reconstruction for that reflex.

(13) Tyee B73d

a.	/ŋ/	*táng-à 'read, count'	(BLR 2786)	>	*ó-táŋà ~ ó-táã̀*
		*nyʊ̀ŋg-à 'move (intr.)'	(BLR 4446)	>	*ó-nyíŋì ~ o-nyiĩ*
		*taŋg-à 'flow, drip'	(BLR 8732)	>	*ó-táŋà*
		*déŋgam-à 'float'	(BLR 7664)	>	*ó-lɛ́ŋɛ́nɛ̀*
		*bàŋgan-à 'quarrel'	(BLR 9679)	>	*ó-báŋánà*
		*cambʊgʊ 'shoulder'		>	*e-saŋama*
		*kʊáŋgà 'fermented manioc'		>	*∅-kwáŋà*
					mú-tɔ́ŋɔ̀ 'caterpillar, worm'

					í-kyàŋánà ‘heat’
					ó-sɛ́ŋɛ́nɛ̀ ‘to shine’
					e-ŋ-gaŋana ‘saw’
b.	/ŋg/	*gìŋgì ‘fly’	(BLR 1406)	>	*ŋ-gingi*
		*góŋgòdó ‘centipede’	(BLR 1453)	>	*ŋ-gɔ̀ngɔ́nɔ̀*
		*doŋgodoŋgo ‘okra’		>	*∅-dɔ̀ngɔ́dɔ̀ngɔ́*

The words with C2 /ŋg/ in (13b) look like borrowings: the word for ‘okra’ is a widespread borrowing in other Kwilu-Ngounie languages, see (10e), while ‘fly’ and ‘centipede’ also irregularly have /ŋg/ in C2 in the following varieties: Latege B71bY *ŋ-gɔ́ŋgɔ̀, o-ŋ-giŋgi/a-ŋ-giŋgi,* Eboo B74 *∅-giŋgi,* Mbete B61Z *∅-ŋíŋgì/ à-ŋíŋgì.*

As can be seen with the first two entries in (13a), disyllabic roots such as *táŋà* and *nyíŋì* can also be pronounced with a long nasalized vowel (where in the orthography only the last vowel is marked as nasal), that is, [tã́ã̀] and [nyĩ́ĩ̀] respectively.[12] Raharimanantsoa and Ntsiba Ngolo (2015: 9) note that all disyllabic roots containing /ŋ/ in C2 can be alternatively realized with a long nasalized vowel instead of /ŋ/, but that this alternative pronunciation is never an option for trisyllabic roots such as *lɛ́ŋɛ́nɛ̀* and *báŋánà*. Although evidence is scanty, it seems that some /ŋ/ in trisyllabic roots can be the result of nasal harmony, vowel harmony, and metathesis, for example, *cambʊgʊ > *sambuŋu* > *sambaŋa* > *samaŋa* > *saŋama*. The fact that words such as *ndwu* ‘pepper’ (82 % of the reflexes in our dataset) are no longer attested with either an alternative pronunciation such as *ŋduŋu* or nasalized vowels suggests that these underwent the change *ŋg > ŋ > ∅ before those which are still found with /ŋ/ in C2 and/or nasalized vowels as an alternative pronunciation. This also shows that nasalization (and subsequent loss) might be an intermediate step in the chain *ŋg > ŋ > ∅. Interestingly, it is only words with *ŋg > ∅ that /ɣ/ is starting to be inserted to break a word-final long vowel.

The last two varieties to be discussed in this section are Latege (Okoyo) B71bY and Nzinii B71bX spoken in the Republic of the Congo.[13] As happens with previously discussed varieties such as Kaan B70s and East Yans B85b, several

---

[12] Interestingly, this alternative pronunciation is also available for words with a C2 /ŋ/ which does not derive historically from C2 *ŋg but from nasal harmony, e.g., *Ǹ-jʊkɪ ‘bee’ (BLR 1622) > *nyuŋu* > *nyũũ*; *mòk ‘chat (v.)’ (BLR 2205) > *emoŋo* ‘chat (n.)’ > *emoõ*.

[13] Note that these two varieties have the same code but different names. In the updated referential classification of Maho (2009), B71b is variously called Kateghe, Njining’i, or Nzikini. To these names, Linton (2013b) adds Latege. At present, we do not know whether B71bY, B71bX, and B71bZ in our dataset are all dialectal varieties of the same language.

words with the unexpected /ŋg/ reflex in both varieties are likely borrowings. We list all words with a C2 /ŋg/ in both varieties in (14) and (15).

(14) Nzinii B71bX
è-sàŋgà 'island'
ŋ-gúŋgwɔ̀ɔ̀ 'millipede'
ŋ-gòŋgà 'bell'
ò-n-tòŋgá/è-n-tòŋgá 'needle'
ò-m-bàŋgá/è-m-bàŋgá 'testicle'

(15) Latege (Okoyo) B71bY
kè-sàŋgà 'island'
lè-sàŋgà 'dance, especially of joy'
∅-sáŋgí 'maize, millet'
lè-sùŋgà 'to help'
ŋ-gìŋgì 'fly'
ŋ-gɔ́ŋgɔ̀ 'millipede'
lè-kàŋgà 'to attach, take'
n-tòŋgà 'needle'
ŋ-kóŋgá 'fish sp.'
mw-ɛ́ŋgɛ̀ 'fish sp.'
lè-ŋ-gèŋgè 'to shine'
ò-làŋgì 'bottle'
∅-dɔ̀ŋgɔ́dɔ̀ŋgɔ́ 'okra'
á-táŋgà 'mourning'
n-zùŋgù 'pot'
lè-páŋgì 'enclosure'
∅-yɛ̀ŋgɛ̀sɛ̀/ɛ̀-yɛ̀ŋgɛ̀sɛ̀ 'sieve'

All words having /ŋg/ instead of zero in B71bX, listed in (14), have been shown to be borrowings in many other varieties discussed in this section, with the exception of *ò-m-bàŋgá/è-m-bàŋgá* 'testicle.' Note that the word for 'island' has an irregular reflex in B71bX and B71bY. In the case of B71bY, among the words listed in (15), those which do not appear to be borrowings shared with other varieties are: *mw-ɛ́ŋgɛ̀* 'fish sp.,' *ŋ-kóŋgá* 'fish sp.,' *∅-sáŋgí* 'maize, millet,' *lè-sùŋgà* 'to help,' *lè-kàŋgà* 'to attach, take,' *∅-yɛ̀ŋgɛ̀sɛ̀/ɛ̀-yɛ̀ŋgɛ̀sɛ̀* 'sieve,' *lè-ŋ-gèŋgè* 'to shine,' and *á-táŋgà* 'mourning.' It is worth noting that *mw-ɛ́ŋgɛ̀* 'fish sp.' (*Hepsetus odoe* or African pike characin) is also *myɛ̀ŋgɛ̀/à-myɛ̀ŋgɛ̀* (/ŋg/ instead of /∅/) in B74. It could well be that some of these words also feature /ŋg/ in other Kwilu-Ngounie varieties discussed in this section, but that our uneven data do not show this.

TABLE 3 Kwilu-Ngounie varieties with pervasive loss of C2 *ŋg, plus Boma Yumu B80z

Variety	**No. of reflexes**	**Reflex**					
		ŋg		**ŋ**		**∅**	
		No.	**%**	**No.**	**%**	**No.**	**%**
B80z	55	0	–	30	55%	25	45%
B81	35	2	6%	8	22%	25	71%
B82X	38	0	–	11	29%	27	71%
B82W	37	2	–	9	26%	26	74%

4.3 *Pervasive loss of C2 *ŋg*

In this section, we discuss varieties found (north)east of the vast area characterized by the systematic loss of C2 *ŋg discussed in Section 4.2. In these varieties, the loss of *ŋg is pervasive, in the range of 71–74%, as shown in Table 3. Unlike the varieties discussed in Section 4.2, they preserve 20–30% of original *ŋg in C2 as either [ŋg] or [ŋ]. For ease of exposition, we discuss the peculiar case of the Kwa-Kasai North language variety Boma Yumu B80z in this section, along with the other two Kwa-Kasai North varieties Tiene B81 and North Boma B82, even though Boma Yumu B80z does not display pervasive loss of C2 *ŋg.[14]

We first address Tiene B81 and the two North Boma B82 varieties in Table 3. These belong to the monophyletic Kwa-Kasai North subgroup (see Fig. 1) which additionally includes Mpe B821, Nunu B822 (both excluded from this study due to insufficient data), and Boma Yumu B80z (see discussion below). With the exception of Boma Yumu B80z, the Kwa-Kasai North languages are separated from the cluster of languages with systematic C2 *ŋg loss not only by the Kwa and Kasai Rivers in the south but also by the Congo River in the west (see Map 3). Nevertheless, in the Mai Ndombe province they are in close contact with Ŋiŋi B76b, a variety that displays the systematic loss of *ŋg in C2 position (see Table 2 and Map 3).

---

[14] Although not included in the present study, data collected within our research group on Sakata C34 varieties (Maselli et al., 2023) spoken in the same area within the Mai Ndombe province where North Boma B82 is also spoken suggests that some of these also have pervasively lost *ŋg. While not included in Pacchiarotti et al. (2019), Sakata C34 is part of WCB in the Bantu-wide phylogenies of Grollemund et al. (2015) and Koile et al. (2022). Lexically, the Sakata group is very closely related to Kwa-Kasai North.

For North Boma B82, we have used firsthand fieldwork data collected in different years by different team members working with different speakers and we have additionally relied on the grammar sketch of Stappers (1986). It is striking that the two North Boma varieties in Table 3 show nearly identical percentages of lexical items where *ŋg is lost or simplified to [ŋ]. In both B82 varieties, [ŋ] is present in words which are unmistakably borrowings. This is illustrated in (16) with the variety of North Boma spoken in Bopaka. Some of these were probably borrowed at an ancestral stage within Kwa-Kasai North as they have [ŋ] instead of zero as a reflex of *ŋg in Tiene B81 as well (see, e.g., *mò-láŋè* 'bottle') or are also borrowings in varieties of the wider region (see [10d] and [10f] in Section 4.2).

(16) North Boma (Bopaka) B82X
∅-fɛ̀àŋà 'money' (from French *franc*)
∅-ŋúŋà/∅-ŋúŋà 'bell'
mù-láŋì/mì-láŋì 'bottle'

In this respect, Ellington (1977: 24) notes that in Tiene B81 the velar nasal [ŋ] is rare and occurs in words which are possibly borrowings from Lingala or Bobangi such as *mò-bàŋà* 'head rest,' *kè-sàŋà* 'island,' *kè-yèŋà* 'Sunday,' and *ò-táŋà* 'to read.' In the Tiene variety documented by Motingea Mangulu (2004), we found additionally *∅-tìːŋà* 'pull,' *dzíŋà* 'take with trap,' and two words with /ŋg/ which are also likely borrowings, perhaps of more recent introduction as [ŋg] has not been simplified to [ŋ], namely *mù-n-tsíŋgí* 'belt, waistband' and *mù-n-dìŋgì* 'tree sp.'

While some words with C2 [ŋ] are synchronically recognizable borrowings in the Kwa-Kasai North languages, others are either older borrowings or words which escaped the pervasive loss of C2 *ŋg; see (17). Only a few lexemes in (17) can be linked to an existing reconstruction.

(17) North Boma (Bopaka) B82X

*káŋgà 'guinea fowl'	(BLR 1720)	>	*è-káŋà/ŋ-káŋà*
*póŋgò 'fat'	(BLR 6806)	>	*m-pɔ́ŋì/m-pɔ́ŋì* 'marrow'
*bòŋgó 'brain'	(BLR 274)	>	*bɔ̀ŋɔ̀/bɔ̀ŋɔ̀* 'skull, brain'
			mù-ŋìŋà/mì-ŋìŋà 'lightning'
			bò-nːáŋà 'beautiful'
			mù-bàŋí/mì-bàŋí 'wild cat'
			è-sàŋí 'be equal'
			n-sàŋú/n-sàŋú 'basket to prepare fermented manioc'

The words in (17) include animals but also specialized cultural vocabulary such as *nsàŋú*, a basket especially designed to prepare fermented manioc. One thing all words in (17) have in common is that, unlike all other words with a historical nasal cluster in C2 position, they do not show automatic vowel lengthening. This phenomenon is extremely common in Bantu when a vowel precedes a NC cluster (Hyman, 2019). In North Boma B82, this is observable in all words whose word-final simple nasal is a reflex of either *mb or *nd, such as *n-zàːmí* 'God' < *jàmbé (BLR 3196), *ŋ-gɔ̀ːɳɛ́* 'crocodile' < *gòndé (BLR 1446).[15] The lack of vowel lengthening in (17) is thus evidence that these forms are irregular.

As discussed at the beginning of this section, Boma Yumu B80z does not display pervasive loss of C2 *ŋg. Instead, it is the only Kwa-Kasai North variety to show two reflexes of C2 *ŋg in an almost 1:1 ratio, namely [ŋ] and /∅/.[16] Geographically, it is removed from Tiene B81 and North Boma B82 (see Map 3). As was the case for Laali Mayeye B73bZ in Section 4.2, there seems to be no conditioning environment that could tease apart the two reflexes of C2 *ŋg observed in Boma Yumu B80z; see (18).

(18) Boma Yumu B80z

a.	/ŋ/	*báŋgá 'stone'	(BLR 8633)	>	*ì-báŋá* *ò-báŋ* 'to stop'
		*dòŋg-à 'teach'	(BLR 1127)	>	*ó-lɔ́ŋ*
		*baŋgʊ 'twin'		>	*bàŋɔ̀* 'second-born twin'
		*káŋg-à 'fry'	(BLR 1719)	>	*ò-káŋ*
		*kàŋg-à 'tie up, seize'	(BLR 1715)	>	*ò-kàŋ* 'to close'
		*dòŋgà 'plate'	(BLR 1131)	>	*í-lɔ́ŋ*
		*gàŋgà 'medicine man'	(BLR 1332)	>	*mù-ŋáŋà*
		*bʊ́ŋgò 'shore'	(BLR 341)	>	*lè-bóŋ* 'harbor'

---

[15] In the case of *ŋg, it is hard to determine where lengthening comes from. A form such as *báŋgá 'jaw' (BLR 108) has *ɛ̀-báː* 'chin' as a reflex. There are at least two possible pathways to account for this outcome. The first is that the erstwhile automatically lengthened V1 in front of *ŋg is shortened again once the conditioning factor, i.e., the following *ŋg, is lost, to avoid an illicit sequence of a long and a short vowel, i.e., *báŋgá > báːŋá > báː$_{\mathrm{V1}}$á$_{\mathrm{V2}}$ > bá$_{\mathrm{V1}}$á$_{\mathrm{V2}}$. The second is that it is V2 that is deleted after the loss of *ŋg to avoid the same illicit sequence, i.e., *báŋgá > báːŋá > báː$_{\mathrm{V1}}$á$_{\mathrm{V2}}$ > báː$_{\mathrm{V1}}$.

[16] For Boma Yumu B80z, Hochegger (1972: 199) states that ⟨ng⟩ corresponds to [ŋ]. We do not know whether this is also true in those words which preserve a final vowel in (18). It is possible that in these words, ⟨ŋg⟩ corresponds to [ŋg], and that the lenition chain ŋg > ŋ > ∅ starts only when words lose their final vowel. Unfortunately, we do not have any firsthand data to assess this hypothesis.

	*dɪŋg-à ‘do, become’	(BLR 5664)	>	*ò-zíŋ-à* ‘to live’
	*dɪŋg-à ‘do, become’	(BLR 5664)	>	*lè-zíŋ* ‘life’
b. /∅/	*báŋgá ‘jaw’	(BLR 108)	>	*m-bà* ‘lower jaw (animal)’
	*dòŋg-à ‘teach’	(BLR 1127)	>	*í-lwɔ̀* ‘advice’
	*dòŋgò ‘lineage’	(BLR 1135)	>	*ké-lwɔ̀*
	*káŋgà ‘guinea fowl’	(BLR 1720)	>	*n-káː*
	*bʊ̀ŋgò ‘calabash’	(BLR 4460)	>	*m-bwɔ̀*
	*bʊ̀ŋgʊ́ ‘hyena’	(BLR 4462)	>	*kè-mbwɔ̀*
	*díŋg-à ‘search for’	(BLR 997)	>	*ò-lì-à*
	*dìŋgà ‘smoke’	(BLR 1071)	>	*mú-dʒìà*

As can be seen in (18), Boma Yumu is the only Kwa-Kasai North variety to undergo final vowel loss. Unlike in languages spoken in the WCB homeland area, this change did not happen systematically in Boma Yumu, but is rather the result of intense contact with languages where it is fully regular, such as West Yans B85a (Pacchiarotti and Bostoen, 2021b). If one takes the WCB phylogeny of Pacchiarotti et al. (2019) and the resulting Kwa-Kasai North affiliation of Boma Yumu B80z as a point of departure, then one needs to explain why the other two Kwa-Kasai North varieties considered in this study, North Boma B82 and Tiene B81, have roughly 75 % of zero reflexes of C2 *ŋg and 25 % of /ŋ/ reflexes of C2 *ŋg, while in Boma Yumu B80z there is almost an equal ratio of zero reflexes and /ŋ/ reflexes for C2 *ŋg.

To conclude this section, we note that morphophonological evidence suggests that in North Boma B82 the sound shift *ŋg > ∅ must be older than the loss of other C2 consonants such as *d. In verb roots with a historical *d in C2, such as *læ̀ː* ‘to cry’ (< *dìd-à, BLR 959), the historical *d re-emerges in the remote past tense form *lìlí*. By contrast, in verb roots with a historical *ŋg in C2, such as *kàː* ‘to tie up’ (< *kàng-à, BLR 1715), the historical *ŋg does not surface in the remote past tense form *kàí*. We take this as evidence that *ŋg was lost earlier than *d in North Boma. In turn, as we show in Pacchiarotti and Bostoen (2021b: 449–451), loss of C2 consonants such as *d is older than diachronic sound changes such as phonologically unconditioned final vowel loss in WCB varieties spoken around the homeland area.

4.4 *The buffer zone*

What we define as the buffer zone is a transitional zone of varieties found geographically between those where C2 *ŋg is preserved (see Section 4.1) and those where C2 *ŋg was systematically lost (see Section 4.2). Unlike those with pervasive *ŋg loss (see Section 4.3), varieties in the buffer zone are characterized by

TABLE 4 Buffer zone varieties

Variety	**No. of reflexes**	**Reflex**					
		ŋg		**ŋ**		**∅**	
		No.	**%**	**No.**	**%**	**No.**	**%**
B61	46	15	33%	8	17%	23	50%
B62	44	15	34%	10	23%	19	43%
B63	60	18	30%	17	28%	25	42%
B71bZ	74	29	38%	21	29%	24	33%

the presence of three reflexes of C2 *ŋg—/ŋg/, /ŋ/, and /∅/—without any identifiable conditioning environment to tease apart their distribution. In general, all varieties discussed in this section have numerous lexical items attesting one of the two stages in the lenition chain *ŋg > ŋ > ∅. All provide strong evidence for the existence of multiple unconditioned reflexes of C2 *ŋg, "a situation where one and the same protosound appears to have two or more synchronic reflexes in a given language without any phonological conditioning environment to tease them apart" (Pacchiarotti and Bostoen, 2022: 384). Although this is not the main focus of this article, as we argue in Pacchiarotti and Bostoen (2022: 409–411; see also Section 5.3 below), one of the possible explanations for the existence of multiple unconditioned reflexes in the buffer zone (and elsewhere in WCB and Bantu more generally) is lexical diffusion (Wang, 1969; Labov, 1981).

As shown in Table 4, in languages with three reflexes, these are in a 4:3:3 ratio (B71bZ), 3:3:4 ratio (B63), 3:2:4 ratio (B62), and 3:2:5 ratio (B61). In this section, we organize the discussion around varieties for which we have most data, namely Nduumo B63 and Latege (Léconi) B71bZ.

We consider Nduumo B63 to be representative of the situation in other languages within the Mbete B60 subgroup included in this study (see Fig. 1), namely Mbete B61 and Mbaama B62. Mbete languages are found to the northwest of the main cluster of languages with systematic loss of C2 *ŋg, predominantly in Gabon, although the B62 variety included here is spoken in the Republic of the Congo where it is known as Embaama.[17] In the Mbete sub-

---

[17] The Mbaama B62 variety spoken in Gabon and documented in Okoudowa (2016) seems to abide by the pattern observed in other B60 languages where three reflexes of C2 *ŋg are attested. However, the lexical data in that source are insufficient to make any further claims.

group, 40–50 % of lexical items have zero as the reflex of C2 *ŋg. In the variety of Nduumo B63 discussed here, there are 42 % zero reflexes, 30 % with /ŋg/, and the remaining 28 % show /ŋ/. As observed earlier with the two reflexes for Laali Mayeye B73bZ in (9) and for Boma Yumu B80z in (18), there does not seem to be a conditioning environment to tease apart these three reflexes; see (19).

(19) Nduumo B63

a. /∅/	*bìŋg-a ‘hunt’	(BLR 213)	>	*m-bi*
	*gàŋgà ‘medicine man’	(BLR 1332)	>	*ŋ-gaa*
	*káŋgà ‘guinea fowl’	(BLR 1720)	>	*ŋ-ka*
	*bóŋgó ‘knee’	(BLR 275)	>	*a-bwo*
	*gòŋgò ‘back’	(BLR 1450)	>	*oŋ-gwo*
	*jòŋgá ‘spear’	(BLR 3567)	>	*ywo*
	*cʊ̀ŋgʊ̀ ‘sugarcane’	(BLR 5111)	>	*o-tʃu*
b. /ŋ/	*gàŋgá ‘root’	(BLR 1335)	>	*o-kaŋa*
	*bèŋg-à ‘be red’	(BLR 151)	>	*a-byeŋe*
	*bèŋge ‘splendor’	(BLR 7649)	>	*li-beŋi*
	*tʊ́ŋgʊ̀dʊ́ ‘vegetable’	(BLR 5097)	>	*e-tuŋu*
	*góŋgòdó ‘millipede’	(BLR 1453)	>	*ŋ-gonoŋo*
c. /ŋg/	*bìŋg-i ‘hunter’	(BLR 5660)	>	*o-biŋgi*
	*bìŋgá ‘pigeon’	(BLR 216)	>	*m-biŋga*
	*káŋg-à ‘fry’	(BLR 1719)	>	*gi-kaŋga*
	*gʊ̀ŋgà ‘bell’	(BLR 1514)	>	*o-kuŋgu*
	*jóŋgò ‘bile’	(BLR 3573)	>	*n-djuŋgu*
	*gʊ̀ŋgʊ́ ‘caterpillar’	(BLR 5062)	>	*o-kuŋgu*

Strikingly, this pattern of multiple unconditioned reflexes of C2 *ŋg is present even in the Congolese variety of Mbaama B62 spoken hundreds of kilometers away from other Mbete varieties in northeastern Gabon. This might suggest that zero reflexes of C2 *ŋg are reconstructable at an ancestral stage within Proto-Mbete. Despite the patchiness of our comparative data for B60, Table 5 shows that this hypothesis is confirmed by several lexical items.

All the words in Table 5 point to the complete loss of C2 *ŋg at Proto-Mbete stage. Table 6 shows that at that ancestral stage, some lexical items escaped this innovation and preserved *ŋg, while others started to undergo the lenition chain *ŋg > ŋ > ∅ but stopped at ŋ.

So-called Teke B70 varieties spoken close to Mbete languages, such as Latege (Léconi) B71bZ (see Map 3), show a distribution of reflexes almost identical to that of Nduumo B63. Clear borrowings in the variety of Latege spoken in Léconi

TABLE 5 Lexical items where C2 *ŋg > ∅ at Proto-Mbete stage. A question mark indicates a lack of data; "–" that the language uses a different root for this concept

Lexical item	**B61**	**B62**	**B63**
*báŋgá 'jaw' (BLR 108)	–	*báá/á-báá*	*gi-ba*
*bóŋgó 'knee' (BLR 275)	*vú-wjóː*	*vúɔ́/á-vúɔ́*	*a-bwo*
*dʊ́ŋgʊ́ 'pepper' (BLR 1223)	*n-dwó/a-n-dwó*	*n-dúó/à-n-dúó*	*n-du*
*gàŋgà 'medicine man' (BLR 1332)	*ŋ-gà:*	*n-gáà/á-n-gáà*	*n-gaa*
*gòŋgò 'back' (BLR 1450)	*ŋ-gùwò*	*ó-n-gwɔ́ɔ̀/é-n-gwɔ́ɔ̀*	*on-gwo*
*káŋgà 'guinea fowl' (BLR 1720)	*ŋ-ká*	?	*n-ka*
*kíŋgó 'neck' (BLR 1845)	*ŋ-kíì*	*n-kíí/á-n-kíí*	*li-tsii*
*tʊ́ŋg-à 'build' (BLR 3081)	*twá*	*ké-túá*	*ho-twa*
*jòŋgá 'spear' (BLR 3567)	*yúwɔ́*	*yúɔ́/á-yúɔ́*	*ywo*
*cʊ̀ŋgʊ̀ 'sugarcane' (BLR 5111)	*ʃùù*	*ó-sùù/é-sùù*	*o-tʃu/e-tʃu*

TABLE 6 Lexical items where C2 *ŋg > ŋg/ŋ at Proto-Mbete stage. A question mark indicates a lack of data

Change	**Lexical item**	**B61**	**B62**	**B63**
*ŋg > ŋg	*káŋg-à 'fry, roast' (BLR 1719)	*káŋglà*	*ké-kááŋgà*	*gi-kaŋga*
	*táŋg-à 'read, count' (BLR 2786)	*táŋgà*	*ké-tááŋgà*	*o-taŋgala*
	*páŋg-à 'act, make' (BLR 2397)	*wáŋgà*	?	*gi-paŋga*
	*gʊ̀ŋgʊ́ 'caterpillar' (BLR 5062)	*kóŋgó*	?	*o-kuŋgu*
	*gʊ̀ŋgʊ́ 'lion'	*ŋ-gú:ŋgù*	*n-gòòŋgó*	*gi-ŋ-guŋgu*
*ŋg > ŋ	*gàŋgá 'root' (BLR 1335)	*ò-kàŋà/è-kàŋà*	?	*o-kaŋa*
	*tʊ́ŋgʊ̀dʊ́ 'vegetable' (BLR 5097)	*tú:ŋù*	?	*e-tuŋu*
	*taŋg-à 'flow, drip' (BLR 8732)	*táŋíní* 'drop'	*táŋí* 'drop'	*taŋi* 'drop'
	*jáŋg-à 'smoke (meat)'	*yáŋà*	?	*yaŋa*

have [ŋg] in C2; for example, *òlàŋgì/èlàŋgì* 'bottle,' *kàtúŋgà/ètúŋgà* 'basket to go to the market,' and *fùŋgúlà* 'key'; see also Section 4.2. However, many lexical items which are less obvious borrowings also preserve [ŋg], such as *bɛ̀ŋgɛ́* 'palm oil,' *okùŋgá/ekùŋgá* 'manioc stick (prepared when manioc is not yet cooked),' *okíŋgí/ekíŋgí* 'wood for pounding manioc,' *wùŋgù/àwùŋgù* 'hunting net for wild boars,' *múŋgù/amúŋgù* 'canoe,' and *ŋgùŋgù* 'roof.' Similarly, numerous lexical items in Linton's (2013a) dictionary for which we could find no corresponding reconstruction in BLR2/3 have [ŋ] in C2; for example, *òmjáŋá/èmjáŋá* 'little intestine,' *kàpà:ŋí/èpà:ŋí* 'ripe mushroom,' *bòŋ'/àbòŋí* 'ringworm,' *jíŋà* 'to set a trap,' and *àmbwɔ̀ŋɔ̀* 'offering to the ancestors in order to obtain their favor (e.g., while hunting).'

Some illustrative Latege B71bZ data to be compared to the (Proto-)Mbete data in Tables 5 and 6 are given in (20) and (21), respectively.

(20) Latege (Léconi) B71bZ

*báŋgá 'jaw'	(BLR 108)	>	*bá/à-bá* 'cheek'
*bóŋgó 'knee'	(BLR 275)	>	*∅-kúnàbwóŋó/à-kúnàbwóŋó*
*dóŋgó 'pepper'	(BLR 1223)	>	*n-dú/à-n-dú*
*gàŋgà 'medicine man'	(BLR 1332)	>	*ŋ-gà/à-ŋ-gà*
*gòŋgò 'back'	(BLR 1450)	>	*e-ŋ-gwò*
*káŋgà 'guinea fowl'	(BLR 1720)	>	*ŋ-ká/à-ŋ-ká*
*kíŋgó 'neck'	(BLR 1845)	>	*ŋ-kí/à-ŋ-kí*
*tóŋg-à 'build'	(BLR 3081)	>	*tsúà*
*jòŋgá 'spear'	(BLR 3567)	>	*yùó*
*còŋgò 'sugarcane'	(BLR 5111)	>	*ò-fù/è-fù*

As can be seen by comparing the data in Table 5 with that in (20), all reconstructions targeted by the shift *ŋg > ∅ in Proto-Mbete also have a zero in Latege B71bZ. On the other hand, comparing the data in Table 6 with the overlapping roots in (21) shows that only the reflexes of *tóŋgòdó 'vegetable' (BLR 5097) have [ŋ] as a reflex in both Proto-Mbete and Latege B71bZ.

(21) Latege (Léconi) B71bZ

*káŋg-à 'fry, roast'	(BLR 1719)	>	*káà*
*káŋg-à 'shut, close'	(BLR 2075)	>	*kàŋgà*
*gòŋgó 'lion'		>	*kà-ŋ-gùŋù/è-ŋ-gùŋù*
*tóŋgòdó 'vegetable'	(BLR 5097)	>	*ò-n-túŋú/è-n-túŋú*
*taŋg-à 'flow, drip'	(BLR 8732)	>	*tàí/à-tàí*

Positing an ancestral scenario of prolonged and intense contact likely involving intermarriage between Mbete speakers and speakers of Teke varieties at the center of the innovation, yielding the massive adoption in Mbete of vocabulary where *ŋg > ∅, is hardly believable because all words in Tables 5 and 6 qualify as core vocabulary.

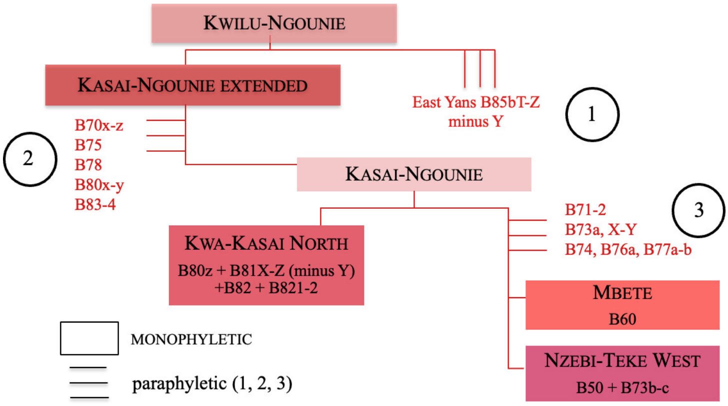


FIGURE 2 Kwilu-Ngounie phylogenetic subgroups based on lexical innovations

PACCHIAROTTI ET AL., 2019

5 Discussion

In Section 4, we showed the outcomes and distribution of the lenition chain C2 *ŋg > ŋ > ∅ and how its completion varies across the Kwilu-Ngounie branch of WCB. In this section, we first show (Section 5.1) how this shared diachronic sound shift is at odds with Kwilu-Ngounie subgroupings established in the lexicon-based phylogeny of Pacchiarotti et al. (2019). We then try to account for these mismatches with two alternative historical accounts, that is, a first account where diachronic phonology is given more diagnostic genealogical power than basic vocabulary-driven phylogenetics (Section 5.2), and then a second in which the phylogeny is given genealogical prominence and phonological innovation is primarily interpreted as a contact-induced sound change (Section 5.3).

5.1 *Kwilu-Ngounie subgrouping: Lexicon-based phylogeny versus diachronic phonology*

For ease of argumentation, Fig. 2 zooms in on the relevant phylogenetic structures within Kwilu-Ngounie as presented in Fig. 1; see also Pacchiarotti et al. (2019: 184, 186).

In the WCB phylogeny of Pacchiarotti et al. (2019: Fig. 1), Kwilu-Ngounie has a posterior probability of 0.9. Within this branch, as can be seen in Fig. 2 above, several East Yans doculects (paraphyletic grade 1; posterior probability rate 0.9)[18] are parallel to the Kasai-Ngounie Extended clade, in which para-

---

[18] For the sake of the argument, we assume that West Yans B85a (included in this study but

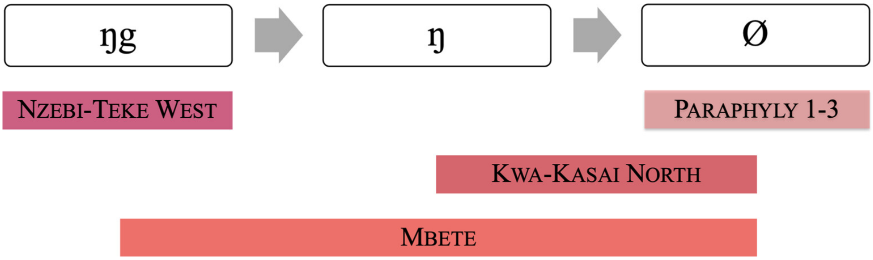


FIGURE 3 Distribution of outcomes of the lenition chain *ŋg > ŋ > ∅ across Kwilu-Ngounie

phyletic grade 2 (posterior probability rates of 0.54 and 0.51; including Boma Nkuu B80x, South Boma B80y, Mfinu B83, Mpuono B84, Bibaana B70x, South Teke B70y, Bwala B70z, Tio Bali B75, Wuumu B78) is parallel to Kasai-Ngounie. The latter subclade also contains paraphyly (paraphyletic grade 3 with posterior probability rate 0.81; including North Teke B71, Ngungwel B72, Tsaayi B73a, West Teke B73X and B73Y, Eboo B74, Kukuya B77a, and Fumu B77b), along with three monophyletic groups, namely Kwa-Kasai North, Mbete, and Nzebi-Teke West. Note that, based on lexical innovations, Kwa-Kasai North is the first to split off within Kasai-Ngounie, while Mbete and Nzebi-Teke West do so much later. This suggests that the latter two subgroups are more closely related to each other than they are to Kwa-Kasai North within Kasai-Ngounie.

Let us now consider to what extent this lexicon-based Kwilu-Ngounie subgrouping matches with possible subgroups signaled by the diachronic sound shifts *ŋg > ŋ > ∅. We remind the reader that, as stated in Section 2, the language sample in Pacchiarotti et al. (2019) and the one used in this study do not overlap entirely (see also Appendix 1), which makes comparability imperfect. Nonetheless, even under these suboptimal conditions, it is straightforward that diachronic phonology is strongly at odds with lexicon-driven phylogeny when it comes to how the outcomes of the lenition chain *ŋg > ŋ > ∅ are distributed across present-day Kwilu-Ngounie languages, as schematically represented in Fig. 3.

As shown in Fig. 3, the only real match with lexicon-based Kwilu-Ngounie subgroups is the full conservation of *ŋg in C2 in Nzebi-Teke West, except for Laali B73b where *ŋg > ŋ is attested irregularly (Section 4.1). However, this is a shared retention and not a shared innovation, and therefore has little diagnos-

---

not in the 2019 phylogeny) belongs together with East Yans varieties in paraphyletic grade 1.

tic value in terms of subgrouping. The first step in the lenition chain, that is, *ŋg > ŋ (part of the broader change *NC > N; see Section 3) is not only shared with WCB languages outside of Kwilu-Ngounie but also with certain CWB languages of the Lower Kasai region such as Bushong C83 (Daeleman, 1977). The second and last stage of the lenition chain, that is, ŋ > ∅, is restricted to Kwilu-Ngounie, but its distribution conflicts with the internal genealogical structure of this WCB branch based on innovations in basic vocabulary. The full loss of *ŋg in C2 is totally absent from Nzebi-Teke West (Section 4.1), but it is attested to variable degrees everywhere else within Kwilu-Ngounie. It is systematic in all paraphyletic grades (Section 4.2), that is, 1–3 in Fig. 2, pervasive within most of Kwa-Kasai North (Section 4.3), and significantly present in Mbete languages where *ŋg in C2 has comparable numbers of /ŋg/, /ŋ/, and /∅/ reflexes (Section 4.4).

In the following subsections, we come up with two possible historical accounts for these mismatches between diachronic phonology and lexicon-based phylogeny: one in which sound change overrules lexical innovations for subgrouping (Section 5.2) and one in which lexical innovations are given more diagnostic weight than historical sound shifts for genealogy (Section 5.3).

5.2 *Scenario 1: Sound change overrules lexical change*

In this subsection, we consider how the internal classification of Kwilu-Ngounie would look like if we attribute higher diagnostic power for genealogy to the sound change C2 *ŋg > ŋ > ∅ rather than to the parallel lexical innovations in basic vocabulary on which the subgroups in the phylogeny of Pacchiarotti et al. (2019) are based. To this end, we attempt to use the last step in the lenition chain C2 *ŋg > ŋ > ∅ to form a new subgroup by descent (Babel et al., 2013), that is, by positing that the innovation ŋ > ∅ happened at an ancestral node that did not surface in the lexicon-based phylogeny.

Apart from the aprioristic rejection of quantitative lexicon-based classifications (Nurse and Philippson, 2003), there are several reasons one might find to claim that these approaches have weak predictive power for internal subgrouping in Bantu. A first common assumption with tree-like representations of Bantu languages is that present-day languages descend directly from the ancestral languages of the first Bantu-speaking settlers in a given region. However, multidisciplinary evidence (de Filippo et al., 2011; Ansari Pour et al., 2013; Seidensticker et al., 2021) suggests that present-day Bantu languages might be the complex product of multiple superimposed historical layers rather than having evolved from a single ancestral language (see also Möhlig, 1977, 1979, 1981). Similarly, ongoing modeling of Bantu language dispersal in inferential statistics (Yanovich et al., 2021) confirms that Bantu language history was very much non-tree-like (see also Schadeberg, 2003) and heavily characterized by lexical replacement and borrowing (see also Hinnebusch, 1999).

Second, different manipulations of the same lexical dataset and the same cognacy judgments through different phylogenetic approaches can lead to significantly different topologies. To give an example, the two most comprehensive Bantu-wide phylogenies to date, Grollemund et al. (2015) and Koile et al. (2022), are based on exactly the same datasets of basic vocabulary, the same cognacy judgments, and the same sample of Bantu languages, but differ in their specific phylogenetic methods. This produces several differences in topology, amongst others for cwb within the rainforest and wcb at its southern margins. In Grollemund et al. (2015), wcb branches off after cwb, while in Koile et al. (2022), wcb and cwb are parallel branches. Additionally, portions of what is cwb in Grollemund et al. (2015) cluster more closely with wcb in Koile et al. (2022).[19]

Third, even when the statistical method used is exactly the same, the way the language sample is assembled (which languages are included, which are not, and how many representatives of a group are present) can have a significant effect on the topology of the resulting phylogenetic tree (Pacchiarotti et al., 2019: 178). To give an example, in studies covering the entire Bantu domain, whether lexicostatistical (Bastin et al., 1999) or phylogenetic (Grollemund et al., 2015; Koile et al., 2022), South-Western Bantu (swb) languages are clearly separate from wcb. However, in phylogenetic studies focusing on a single Bantu branch such as wcb (de Schryver et al., 2015; Pacchiarotti et al., 2019), the few swb languages included cluster with the low-level wcb subgroup known as the klc (see Fig. 1). The node that unites the klc with the swb languages included in Pacchiarotti et al. (2019) has a posterior probability rate of 0.97. Nevertheless, this should not be taken as evidence for swb languages being part of wcb, but rather as an artifact of a narrow-scope and thus unbalanced language sample where other swb and Eastern Bantu languages with which the sampled swb languages could cluster are absent.

Finally, any lexicon-based phylogeny is bound to the quality of the dataset and the knowledge of historical sound change of those who performed the cognacy judgments at any given point in time. The Kwilu-Ngounie subgroups in Pacchiarotti et al. (2019) seen in Fig. 2 result from the heavy use of secondhand

---

[19] Moreover, linguistic phylogenies are sometimes enlarged using non-linguistic data. For instance, Koile et al. (2022) “augment” their phylogeny with geographic information from languages that are missing in their phylogeny. Obviously, the phylogenetic position of these “augmented” languages (i.e., the clade to which they allegedly belong) is not established as they claim, but assumed and probably sometimes incorrectly assumed.

data from Bastin et al. (1999) and cognacy judgments performed with a limited knowledge of sound change. Moreover, Pacchiarotti et al. (2019) used the "Tervuren 92" list (Bastin et al., 1999), a reduced 100-word Swadesh list. However, it is becoming increasingly clear within circles of Bantuists dealing with phylogenies of specific branches (see, e.g., Gunnink et al., 2023) that several concepts from reduced versions of the 100-word Swadesh list should be replaced by concepts with stronger diagnostic power for lexical innovation in specific Bantu areas.

Besides these shortcomings, language groups such as WCB where most languages underwent severe degrees of phonological erosion might posit additional difficulties while trying to perform the lexical cognacy judgments needed for computational cladistics such as phylogenies (on this problem, see Scarborough, 2019).

Considering the limitations of phylogenetic approaches relying on basic vocabulary, it is certainly worth considering whether diachronic phonology can shed new light on the internal classification of Kwilu-Ngounie. This is all the more so if one looks at the high degree of paraphyly that its topology manifests in Pacchiarotti et al. (2019), possibly due to prolific contact between closely related varieties (Bollaert et al., 2021: 3) and their non-tree-like evolution (Schadeberg, 2003; Yanovich et al., 2021).

As discussed in Section 5.1, the final step in the lenition chain *ŋg > ŋ > ∅, that is, the complete loss of *ŋg, turns out to be systematic in all languages belonging to paraphyletic grades 1–3 in Fig. 2 (see also Section 4.2 and Fig. 3). These languages not only share this distinctive phonological innovation but are also geographically clustered on the Bateke Plateau (see Map 3). If one considered only sound change, the most parsimonious explanation would be to assume that this shared innovation only occurred once, namely in the most recent common ancestor of the languages concerned. On these diachronic phonological grounds, the languages of the lexicon-based paraphyletic grades 1–3 in Fig. 2 would form a discrete subgroup, which we tentatively call "Plateau" in Fig. 4 because its languages are mainly spoken on the Bateke Plateau.

Within Kwilu-Ngounie, the initial step *ŋg > ŋ might also be indicative of subgrouping, since it is attested everywhere except in Nzebi-Teke West, as shown in Fig. 3. Hence, it might point to a closer genealogical affiliation between Mbete, Kwa-Kasai North, and Plateau than between those subgroups and Nzebi-Teke West. However, as Fig. 3 also shows, /ŋ/ is nowhere the fully regular reflex of *ŋg. In Kwa-Kasai North, it is less common than the zero reflex (Section 4.3), while in Mbete it co-occurs equally with the archaic /ŋg/ and the more innovative zero reflex (Section 4.4). The buffer zone (Section 4.4), with Mbete and a few Teke varieties consistently displaying three unconditioned

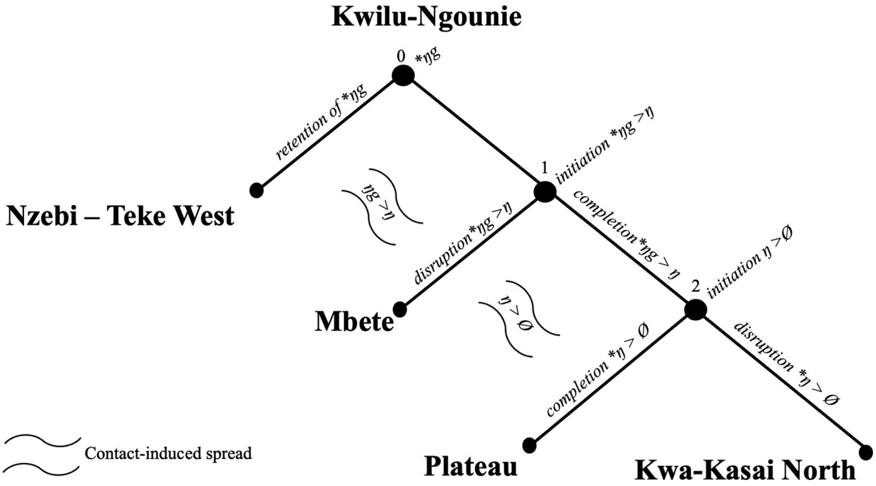


FIGURE 4 Genealogical interpretation of *ŋg > ŋ > ∅ outcomes across Kwilu-Ngounie

reflexes of *ŋg, represents an obvious problem for the Neogrammarian axiom of flawlessly regular sound change. Nonetheless, if one factors in the language-internal process of lexical diffusion at ancestral nodes and the subsequent contact-induced spread of sound change, as we do in Fig. 4, the distribution of the outcomes of *ŋg lenition across Kwilu-Ngounie (see Fig. 3) may have diagnostic power for subgrouping.

Figure 4 posits that the sound shifts *ŋg > ŋ and ŋ > ∅ may have been initiated at two successive ancestral nodes, that is, at nodes 1 and 2, respectively.[20] First, *ŋg > ŋ started in the most recent common ancestor of Mbete, Kwa-Kasai North, and Plateau (node 1). Thereafter, ŋ > ∅ began in the most recent common ancestor of Kwa-Kasai North and Plateau (node 2). In neither of these two ancestral stages were these phonological innovations fully regular. They were changes of the lexical diffusion type, that is, changes which are phonetically abrupt but lexically gradual (Labov, 1981, 1999). Several studies offer evidence

---

[20] The attentive reader might wonder why we decide not to posit that the lenition chain C2 *ŋg > ŋ > ∅ was just a single event (instead of being split up in two phases) that started at node 1 in Fig. 4. Developments of C2 *ŋg outside of the Kwilu-Ngounie branch offer evidence for an initial stage where C2 *ŋg > ŋ without ever being completely lost. The reason not to posit *ŋg > ŋ > ∅ at node 1 (which would include Mbete) is that in all languages where C2 ŋ > ∅ took place, there is no evidence for /ŋg/ reflexes, except when the word with /ŋg/ in C2 is a recognizable borrowing. However, words with C2 *ŋg in Proto-Mbete do not appear to be borrowings. This suggests that in Mbete the initial stage *ŋg > ŋ was never completed and it is thus unlikely that ŋ > ∅ was initiated because this stage usually takes place once all C2 *ŋg > ŋ.

for the fact that lenitions (such as C2 *ŋg > ŋ > ∅) are sensitive to lexical diffusion conditioned by frequency, with the most frequent words affected by the change first (Phillips, 1984, 2006, 2020; Bybee, 2002).

The partial retention of *ŋg in Mbete could be accounted for by the fact that this subgroup split off from node 1 before *ŋg > ŋ had targeted all eligible lexemes, after which the sound shift came to a standstill. The diffusion of *ŋg > ŋ only continued to its fullest extent between nodes 1 and 2. Subsequently, the further lenition to zero was initiated at node 2. Once again, though, the full lexical diffusion of ŋ > ∅ did not happen before Kwa-Kasai North broke away from node 2. The sound shift had only spread incompletely and was then disrupted in Kwa-Kasai North, which would explain the partial retention of the /ŋ/ reflex of C2 *ŋg in that subgroup.

Even though the diachronic-phonologically inspired genealogy in Fig. 4, with two successive ancestral phases of incomplete lexical diffusion, can explain the partial retention of /ŋg/ in Mbete and /ŋ/ in Kwa-Kasai North, it accounts for neither the presence of zero reflexes in Mbete nor for the attestation of /ŋ/ in Laali B73b, which is part of Nzebi-Teke West in terms of basic vocabulary. To uphold the validity of *ŋg > ŋ and ŋ > ∅ as indicators of, respectively, ancestral nodes 1 and 2 in Fig. 4, one needs to invoke language contact to explain the presence of /ŋ/ and /∅/ in languages descending from ancestral nodes which did not undergo *ŋg > ŋ (node 0) and ŋ > ∅ (node 1), respectively.

While lexical diffusion is a gradual process of phonological innovation that primarily happens within a given language, in this case at ancestral nodes 1 and 2, /∅/ in Mbete and /ŋ/ in Laali would instead be due to contact-induced horizontal transfer between languages. As discussed in Section 4.1, all present-day neighbors of Laali which do not retain /ŋg/ have the zero reflex but not /ŋ/. Consequently, either *ŋg > ŋ is a parallel language-internal innovation in Laali or the language contact triggering it must be old. The same holds for Mbete. Because words with zero reflexes of C2 *ŋg are reconstructable to Proto-Mbete, ŋ > ∅ must have started through contact in the most recent common ancestor of Mbete (and the Teke variety included in the buffer zone). The source of this external influence must have been the innovative languages which underwent the total loss of *ŋg rather than the present-day geographically contiguous but conservative Nzebi-Teke West languages.

Figure 4 is an attempt to resolve the lexicon-based paraphyly within Kwilu-Ngounie in Pacchiarotti et al. (2019) by positing the total loss of C2 *ŋg as a shared phonological innovation diagnostic of a discrete Plateau subgroup. This new subgroup would lump together paraphyletic grades 1–3 (see Fig. 2) and be most closely related to Kwa-Kasai North. This hypothesis, driven by diachronic phonology, brings about at least two analytical problems.

First, under the scenario hypothesized in Fig. 4, Nzebi-Teke West would need to split off first. However, how does one then explain the innovations in basic vocabulary that made Nzebi-Teke West emerge as the closest relative of Mbete and the paraphyletic grade 3 in the phylogeny of Pacchiarotti et al. (2019) seen in Fig. 2? Since phylogenetic methods create subgroups based on shared innovations, not retentions, one would need to argue that Nzebi-Teke West later innovated or borrowed vocabulary identical to that of Mbete and the paraphyletic grade 3. However, there is no evidence pointing to borrowing in the basic lexicon of Nzebi-Teke West. The idea that Nzebi-Teke West would have innovated exactly the same basic vocabulary as Mbete and the paraphyletic grade 3 would also not be a parsimonious solution. One could then argue alternatively that Nzebi-Teke West as a lexicon-based monophyletic subgroup is not accurate and should not be trusted. Nevertheless, the other two low-level monophyletic groups within Kasai-Ngounie in Pacchiarotti et al. (2019)—that is, Mbete and Kwa-Kasai North—are confirmed as distinct groups through their different developments in the phonological innovation C2 *ŋg > ŋ > ∅ (see Sections 4.3 and 4.4). Why would Nzebi-Teke West (also identified via phonology; see Section 4.1) be the only one so far off track?

Second, the broader lenition chain *NC > N, of which *ŋg > ŋ is part, is a sound change also attested outside of Kwilu-Ngounie. As discussed in Section 3, *NC > N happened in a geographically contiguous area and affected different WCB subgroups that had emerged after the initial divergence of WCB as well as some neighboring CWB languages. As we further argue in Section 6, this NC cluster reduction is best understood as a contact-induced areal innovation (Thomason and Kaufman, 1988; Thomason, 2001, 2008). If *ŋg > ŋ is part of a broader contact-induced areal change, can it have genealogical diagnosticity within Kwilu-Ngounie and point toward a closer relatedness between Mbete, Kwa-Kasai North, and Plateau (i.e., paraphyletic grades 1–3) as Fig. 4 suggests?

As an anonymous reviewer has pointed out, if the *NC > N change at WCB level is accounted for as a contact-induced areal innovation, which went further to total loss in the case of *ŋg and occasionally *mb and *nd in specific subbranches (see footnote 7 in Section 3), why should the next stage within the lenition chain (ŋ > ∅) in Kwilu-Ngounie not be accounted for in the same way? To account for ∅ reflexes in Mbete, we already posit in Fig. 4 the contact-induced spread of ŋ > ∅ at the Proto-Mbete stage. Instead of positing a disrupted lexical diffusion of ŋ > ∅ in the most recent common ancestor of Kwa-Kasai North to account for the non-systematicity of ŋ > ∅ (as we do in Fig. 4), one could also assume that it is due to contact with neighboring Plateau varieties where the loss of C2 *ŋg was systematic.

It is hard to tell whether the sound shift ŋ > ∅ is a phonological innovation reconstructable to a specific ancestral node or instead an areal contact-induced change within the center of innovation. Assuming that ŋ > ∅ is only diagnostic for subgrouping in those languages where the sound shift is fully systematic (Section 4.2), that is, the Plateau subgroup in Fig. 4, is also problematic, because this new Plateau subgroup lumps together paraphyletic grades with disparate topologies and variable posterior probability rates in the lexicon-based phylogeny of Pacchiarotti et al. (2019), namely: (a) grade 1 with posterior probability rate 0.9; (b) grade 2 with low posterior probability rates between 0.54 and 0.51; and (c) grade 3 with posterior probability rate 0.81 (see Fig. 2).

In biology, where phylogenetic methods originate, trustworthy clades usually have a posterior probability rate of 0.9 or even 0.95 (see, e.g., Huelsenbeck and Rannala, 2004). In accordance with standard practice in the use of phylogenetics for linguistics where rates between 0.8 and 0.85 are usually considered trustworthy (Natalia Chousou-Polydouri, personal communication), Pacchiarotti et al. (2019) consider only nodes with posterior probability rates greater than 0.85 as valid for establishing monophyletic groups. Hence, paraphyletic grade 1 at the top of the Kwilu-Ngounie topology has high lexicon-based support (0.9) but consists exclusively of varieties of the same language, that is, East Yans. However, as pointed out in Section 5.1 (footnote 18), Yans Niadi B85b, the only East Yans variety in our dataset, displays pervasive and not systematic loss of C2 *ŋg, which is not strong evidence for inclusion in the Plateau subgroup based on *ŋg > ∅. West Yans B85a does attest /∅/ systematically, but does not feature in the phylogeny of Pacchiarotti et al. (2019). Conversely, paraphyletic grade 2 has very low lexicon-based support (around 0.5). Overruling it on the basis of diachronic phonological evidence in favor of a Plateau subgroup as in Fig. 4 is thus unproblematic.

Paraphyletic grade 3 within Kasai-Ngounie with a posterior probability of 0.81 is a borderline case in terms of lexicon-based support. This rate, lower than the 0.85 limit, could be an argument to exclude it from the new Plateau subgroup established on the basis of the phonological innovation C2 *ŋg > ŋ > ∅ (see Fig. 4). If one sets the limit for trustworthy rates below 0.85, then the rate of 0.81 can be used as an argument to include it in the new Plateau subgroup as paraphyletic grade 3 is, according to the lexicon-based phylogeny, more closely related to the monophyletic groups within Kasai-Ngounie in terms of basic vocabulary than to paraphyletic grade 2 (and 1).

5.3 *Scenario 2: Lexical change overrules sound change*

In this section, we argue that the internal structure of Kwilu-Ngounie in Fig. 2 emerging from the lexicon-based phylogeny of Pacchiarotti et al. (2019) is to be considered as the frame of reference for interpreting the different developments of the phonological innovation C2 *ŋg > ŋ > ∅ discussed in Section 4 without affecting its topology. To this end, we argue that the innovation C2 *ŋg > ŋ > ∅ did not originate in an ancestor language but rather diffused across the boundaries of different lexicon-based subgroups (Babel et al., 2013).

Despite the criticisms discussed in Section 5.2, most basic lexical concepts used in phylogenetic studies are quite universal and have a low borrowability score across the world's languages (Tadmor et al., 2010). Given that phylogenetic methods infer on the basis of cognacy judgments the most likely genealogical model to account for synchronic variation in thousands of arbitrary form-meaning associations (Dunn, 2015), they remain a robust method to signal deep divergence within a language family. Present-day diversity should of course be sufficiently represented to obtain the most likely family tree (Pacchiarotti et al., 2019). Additionally, one should always reckon that ancient diversity that went extinct cannot be factored in (Bostoen, 2018; Bostoen et al., forthcoming). Further, sound change might have less predictive power for internal classification than received wisdom has it, due to the fact that at least some sound changes "are usually so 'natural' that they can easily be repeated in different lines of descent" (Ringe et al., 2002: 66, cited in Babel et al., 2013).

Thus, in this section we consider an alternative scenario, where the innovation C2 *ŋg > ŋ > ∅ is not a shared phonological innovation reconstructable to a node which would lump together some paraphyletic grades within Kwilu-Ngounie with Kwa-Kasai North or all paraphyletic grades with Kwilu-Ngounie, as argued in Section 5.2. If one wants to preserve all lexicon-based internal structure within Kwilu-Ngounie (see Fig. 2), the only possibility is to posit that the total loss of C2 *ŋg after its lenition is, just like the broader innovation *NC > N in C2 position, a contact-induced areal sound change (Thomason, 2008), which is geographically bound and crosscuts different monophyletic subgroups and paraphyletic grades within Kwilu-Ngounie.

The center of innovation of this lenition chain is situated in Zone 2 (see Map 3) and comprises the varieties with systematic loss of C2 *ŋg in Section 4.2. Languages in Zone 1 (see Map 3), that is, the Nzebi-Teke West varieties in Section 4.1, were virtually unaffected by the areal lenition chain, although Laali B73b varieties clearly show evidence for either an ongoing or an interrupted lenition *ŋg > ŋ, perhaps due to their geographical proximity to varieties which are part of the center of innovation area. The absence of the innovation *C2 ŋg > ŋ > ∅ in most of Nzebi-Teke West might mean that this group was less in contact with the Teke varieties which propagated the change.

The buffer zone (see Section 4.4) is also geographically contiguous to the center of innovation (see Map 3), but in striking contrast to Nzebi-Teke West, was highly affected by the lenition chain, to the point that words with /∅/ and /ŋ/ as a reflex of C2 *ŋg can be reconstructed to Proto-Mbete. As we argued in Section 5.2, the fragmentation of C2 *ŋg reflexes in Proto-Mbete makes it a good case study for lexical diffusion (Wang, 1969). The first step in the lenition chain *ŋg > ŋ came to a halt; that is, it did not target all suitable lexical items (see Fig. 4). The second step in the lenition chain, namely ŋ > ∅, never reached all suitable targets either. It might have started at the Proto-Mbete stage as a contact-induced areal innovation ignited by the desire to imitate the speech habits of the Teke kingdom (represented in this study by the B70–80 varieties in Section 4.2), geographically contiguous to Mbete speakers (see Map 3) and considered prestigious in earlier times (Vansina, 1966). As discussed in Section 1, Teke speakers were highly mobile traders who traveled beyond their kingdom to sell raffia, art objects, and metalwork products, sometimes in exchange for slaves. Due to their trade-driven mobility, Teke people in all likelihood also contributed to the contact-induced propagation of the ŋ > ∅ innovation. Vansina (1973–1974: 336–337) conjectures that prestige no doubt played a major role in the diffusion of "some obvious manners of speech" of Teke people, probably because they were found to be more elegant, for example "a reduction of final nasal complexes and the loss of final vowel or consonant."[21]

The Kwa-Kasai North languages in Section 4.3 show pervasive loss of C2 *ŋg. While separated from the center of innovation by geographic barriers such as the Congo and Kwa Rivers, the area in which these languages are spoken today also hosts speakers of B76 Teke varieties where C2 *ŋg loss was systematic. Based on the morphophonological evidence we provide for North Boma B82 in Section 4.3, the change *ŋg > ŋ > ∅ is old, likely older than the loss of other C2 consonants. Unlike varieties of the buffer zone, Kwa-Kasai North displays two reflexes for *ŋg, namely a predominant zero reflex and to a much lesser extent /ŋ/. Words preserving C2 *ŋg are demonstrably borrowings in one or more Kwa-Kasai North languages.

---

[21] However, according to Lane (1989: 37), since colonial times the Mbete group of the Haute Ogouée region of Gabon has been socially and linguistically more prestigious than speakers of Teke B70 varieties in the same region and in the bordering Cuvette Ouest region of the Republic of the Congo. Lane (1989) states that many Teke speakers claim to be and/or speak Mbete in both Gabon and Congo because Mbete people are viewed historically as the politically and socially superior group. We do not know how old Proto-Mbete is, but it is in all likelihood older than first colonial contacts. If the sociolinguistic situation described by Lane (1989) were to be older, one might posit that the presence of zero reflexes in Proto-Mbete is attributable to Teke speakers who shifted to Mbete due to the higher social prestige of the latter group.

An alternative to positing that Kwa-Kasai North partook in the innovation *ŋg > ŋ > ∅ at an ancestral stage which also included varieties with systematic loss of C2 *ŋg (see Section 5.2) is to argue that zero reflexes of C2 *ŋg were acquired as a contact-induced areal innovation. Just like in the buffer zone, contact with the once prestigious Teke kingdom was the catalyzer for the introduction of zero reflexes at a time when the first stage of the lenition chain *ŋg > ŋ was already complete in Kwa-Kasai North. Lexical diffusion would again be responsible for the fact that not all phonotactically suited lexical items were affected by the second stage of the chain ŋ > ∅.

Under this second scenario, the 50:50 ratio in Boma Yumu B80z might be explained by positing that this group moved away from the area where North Boma B82 and Tiene B81 are currently spoken and resettled south of the Kwa River around Bandundu at a stage when C2 ŋ > ∅ had already started but was interrupted (and thus targeted less lexical items), due to the speakers resettling in an area where only the change C2 *ŋg > ŋ had taken place. According to historical sources (Tonnoir, 1970: 38), Boma Yumu speakers moved away from the area where North Boma B82 and Tiene B81 are still located today (see Map 3) and settled to the southwest, along the Kasai River on the other side of the mouth of the Kwango River (see Map 3) during the sixteenth or seventeenth century CE.

The question that then arises is: what different contact scenarios are responsible for the different outcomes of lexical diffusion in Kwa-Kasai North—where most languages have 75% of zero reflexes and 25% of /ŋ/ reflexes—and the buffer zone—where all languages have three reflexes, /ŋg/, /ŋ/, and /∅/ in almost equal proportions?

Whatever the answer to this question might be, C2 *ŋg > ŋ > ∅ fits well with the dialectal diffusion pattern of centrality vs. peripherality (Trudgill, 2011) or core vs. periphery (Winford, 2003), just like other WCB diachronic sound changes such as final vowel loss (Pacchiarotti and Bostoen, 2021b: 456–458). Varieties with systematic loss of C2 *ŋg are located in the center of the area affected by the contact-induced innovation (see Map 3). The conservative Nzebi-Teke West group constitutes the extreme western geographic periphery which was unaffected by the innovation. The buffer zone, in between the extreme western periphery and the center of innovation, and Kwa-Kasai North, representing the northeastern periphery and bordering languages further east which did not undergo the change, were affected to different degrees by the contact-induced spread of C2 *ŋg > ŋ > ∅. While the extreme western periphery was virtually impermeable to the change, the northeastern side of the periphery was not.

Finally, there is the question of how and why this lenition chain leading to the complete loss of C2 *ŋg emerged in the center of innovation. On the basis of lexicon, the varieties with systematic loss of C2 *ŋg (Section 4.2) end up in different paraphyletic grades within the phylogenetic tree of Pacchiarotti et al. (2019), although they could arguably be grouped together under a most recent common ancestor based on the innovation C2 *ŋg > ŋ > ∅ (see Fig. 4 in Section 5 and the discussion in Section 5.2).

Outside of WCB, the unconditioned cluster reduction NC > N in C2 position is widespread in Central-Western C60 languages spoken in the Congo rainforest as well as in A70 and A80 North Western Bantu languages spoken in the Republic of the Congo and Cameroon (Bostoen et al., 2023). The complete loss of C2 *ŋg (but not other NC) is attested only in some A80 languages (namely Shiwe A803, Njem A84, and Bomwali A87) to varying degrees (Guy Kouarata, fieldwork in 2023). Njem A84 also attests final vowel loss just like WCB languages from the homeland area. In Pacchiarotti and Bostoen (2021b), we attribute the emergence of final vowel loss to substrate interference. Given (a) the partial overlap of these features in A80 languages of Cameroon, where hunter-gatherer communities are still present, as well as in certain WCB languages of western DRC, where hunter-gatherer communities are no longer present; and (b) the absence of these features elsewhere in surrounding areas, total loss of C2 *ŋg could be a non-Bantu substrate feature as well.

6 Conclusions

Attempts to subgroup Bantu languages genealogically, whether on a comprehensive or more local level, have been dominated by quantitative approaches based on basic vocabulary (Philippson and Grollemund, 2019), first lexicostatistics (Coupez, 1956; Heine, 1973; Heine et al., 1977; Nurse and Philippson, 1980; Bastin et al., 1983, 1999; Batibo, 1998; Bastin and Piron, 1999) and more recently phylogenetics (Holden, 2002; Holden et al., 2005; Holden and Gray, 2006; Rexová et al., 2006; de Schryver et al., 2015; Grollemund et al., 2015; Pacchiarotti et al., 2019; Gunnink et al., 2023; Koile et al., 2022).

The tree-like structures representing diversification within the entire Bantu language family (approximately 500 members) have so far been based solely on shared lexical innovations. This is because building a family tree relying on shared phonological/morphological innovations is hard to achieve without first performing diachronic phonological reconstruction through a rigorous application of the comparative method on low-level subgroups. This task has been achieved for very few Bantu subgroups, WCB certainly being one of them (see, e.g., Daeleman, 1977; Rottland, 1977; Bostoen and Koni Muluwa, 2011, 2014; Koni Muluwa and Bostoen, 2012; Bostoen and Goes, 2019; Goes and Bostoen, 2019; Pacchiarotti and Bostoen, 2020, 2021b, 2022; Pacchiarotti et al., 2021).

Establishing a tree-like model based on shared phonological innovations can be further complicated by the lack of Neogrammarian regularity in sound change in (at least some areas of) the Bantu domain. This fact is concretely observable in the widespread phenomenon of multiple unconditioned reflexes for one single protosound not straightforwardly attributable to borrowing events, as seen in Section 4.4 (see Pacchiarotti and Bostoen, 2022, for a detailed account). Despite this pervasive irregularity in sound change, the comparative method can still be successfully applied to identify phonological innovations diagnostic of higher-level nodes (Pacchiarotti and Bostoen, 2022).

Hence, challenges and limitations of lexicon-based phylogenies when applied to the specific historical evolution of Bantu languages notwithstanding (see discussion in Section 5.2), we believe that tree models of Bantu language diversification, whether based on lexical or phonological innovations, are still an essential tool to understand processes of inheritance, shared innovation, and detectable borrowing (Geisler and List, 2013; Chacon and List, 2016; Jacques and List, 2019).

However, in telling the history of languages, tree-like structures have their limits (François, 2014; Kalyan and François, 2019; Kalyan et al., 2019), especially at shallow-time and restricted-space levels. The inclusion of too many closely related varieties which have been in intensive contact, as was the case with the so-called B70 Teke varieties in Pacchiarotti et al. (2019), may result in a lack of phylogenetic resolution. When lexical innovations shared between such close relatives crosscut each other rather than overlapping with each other, several paraphyletic grades may emerge in a lexicon-based phylogeny. This was the case for the topology of the Kwilu-Ngounie branch in the WCB phylogeny of Pacchiarotti et al. (2019).

In this article, we have assessed possible ways to deal with clashes between lexicon-based phylogenetic subgroups and subgroups emerging from a single phonological innovation identified through a careful application of the comparative method on a well-defined subset of Bantu languages (Section 5.1). Despite the clashes (Section 5.2), lexical and phonological evidence can be combined to shed new light on the internal structure and evolution of low-level Bantu subgroups. The picture that emerges is one where the phonological innovation (namely the stage ŋ > ∅ in the chain C2 *ŋg > ŋ > ∅) diffuses across different lexicon-based phylogenetic boundaries yielding a series of divergence and convergence pulses (Weinreich, 1953; Trudgill, 1986; Hinskens et al., 2009).

At the same time, there is evidence that this phonological innovation can be used to establish a subgroup which was previously invisible by looking at lexicon alone. Figure 5 summarizes the successive phases of divergence and

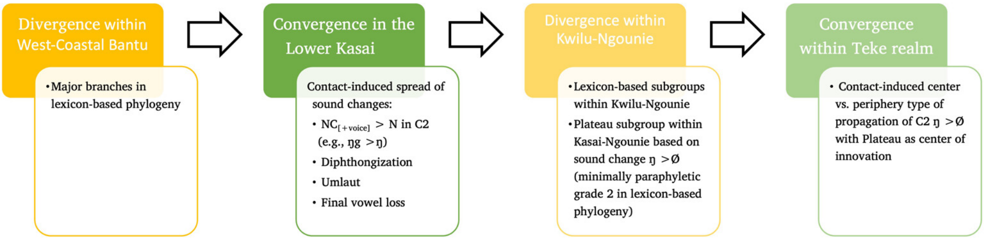


FIGURE 5 Consecutive phases of divergence and convergence in WCB and Kwilu-Ngounie

convergence within WCB in general and its Kwilu-Ngounie branch more specifically. These phases are distinguished by conjointly interpreting the lexicon-based phylogeny of Pacchiarotti et al. (2019) and the historical sound shifts which Proto-Bantu *ŋg underwent in C2 position along with other diachronic phonological evidence.

After a first stage of divergence within the WCB homeland (phase 1 in Fig. 5), which is for the time being exclusively signaled in the lexicon-based phylogeny, the same Lower Kasai region was subject to a period of convergence (phase 2 in Fig. 5), defined, among other phenomena, by the reduction of voiced NC clusters in word-final position. As shown in this study, the first step in the lenition chain C2 *ŋg > ŋ > ∅ is widespread within WCB and part of a broader phenomenon of word-final *NC$_{[+voice]}$ > N.

Except for the KLC and Nzebi-Teke West subgroups (and partially Mbete), both situated in the extreme west of the WCB distribution area and belonging to different major WCB branches (KLC Extended and Kwilu-Ngounie, respectively), all WCB languages simplified voiced NC clusters to N; that is, *mb > m, *nd > n, and *ŋg > ŋ. As the innovation *NC$_{[+voice]}$ > N happened in a geographically contiguous area and across subgroups, this NC cluster reduction is best understood as a contact-induced areal innovation (Thomason and Kaufman, 1988; Thomason, 2001, 2008), one that affected the different ancestors of those modern WCB languages that did not retain *mb, *nd, and *ŋg.

The innovation *NC$_{[+voice]}$ > N in the Lower Kasai area created convergence (homogenization) among speakers belonging to different subgroups that took part in this lenition as well as divergence (diversification) between the innovative Lower Kasai area and the conservative western groups Nzebi-Teke West and KLC. By the time convergence in the Lower Kasai area kicked in, the ancestors of those two subgroups must have been sufficiently removed from the homeland region not to be affected by it.

What is more, this first convergence pulse in the Lower Kasai area (phase 2 in Fig. 5) was not only characterized by the sound shifts *mb > m, *nd > n, and *ŋg > ŋ in C2 position, but possibly also by other contact-induced phonological innovations which crosscut genealogical subgroups in the WCB homeland area, such as final vowel loss (Daeleman, 1977; Rottland, 1977; Pacchiarotti and Bostoen, 2021b), umlaut (Bostoen and Koni Muluwa, 2014), and diphthongization (Koni Muluwa and Bostoen, 2012). Additional evidence to claim that these were contact-induced changes comes from the fact that this convergence zone is not fully contained within WCB. It also incorporates more distantly related but neighboring languages from the Lower Kasai region, especially CWB languages of Guthrie's C80 group such as Bushong C83, Lele C84, and Wonk C85, where these same phenomena are also attested to varying degrees (Vansina, 1959; Tete Wer Sey, 1975; Daeleman, 1977; Ngwamashi Kabandji-Bola Kamu, 1979, 1981; Koni Muluwa and Bostoen, 2015; Pacchiarotti and Bostoen, 2021b).

When it comes to Kwilu-Ngounie, after the contact-induced innovation $^*NC_{[+voice]}$ > N had initiated phonological differentiation between Nzebi-Teke West and KLC on one side and the rest on the other, further divergence happened on the Bateke Plateau (phase 3 in Fig. 5). Lexicon-based phylogenetics signals three neat subgroups: Nzebi-Teke West, Mbete, and Kwa-Kasai North.

Diachronic phonological evidence suggests that the convergence phase characterized among other phenomena by *ŋg > ŋ was followed by a subsequent development, namely the subsidiary ŋ > ∅ innovation in C2 position. It is hard to pinpoint when exactly ŋ > ∅ started. Given the relatively low frequency of ∅ reflexes in Mbete (Section 4.4), it likely happened when their most recent common ancestor had already split off (see Fig. 4). Subsequently, Proto-Mbete would have acquired zero reflexes through contact with languages where C2 ŋ > ∅ was systematic. Although ŋ > ∅ is much more pervasive in Kwa-Kasai North (Section 4.3) than in Mbete, assuming that Kwa-Kasai North shares a most recent common ancestor with languages where ŋ > ∅ is systematic (Section 4.2), as we have tried to argue in Fig. 4, is difficult to reconcile with the lexicon-based phylogeny.

Probably, this distinctive sound shift has genealogical diagnosticity only for paraphyletic grade 2 (see Fig. 2). Given its low support in Pacchiarotti et al. (2019), it does not conflict with lexical evidence to assume that its members—that is, Bibaana B70x, South Teke B70y, Bwala B70z, Tio Bali B75, Wuumu B78, Boma Nkuu B80x, South Boma B80y, Mfinu B83, and Mpuono B84—go back to a most recent common ancestor where ŋ > ∅ affected the entire lexicon. Based on this innovation, they all form what we call the Plateau subgroup (see phase 3 in Fig. 5). Whether this also holds for the better lexically supported paraphyletic grade 3 including North Teke B71, Ngungwel B72, Tsaayi B73a, West Teke B73X and B73Y, Eboo B74, Kukuya B77a, and Fumu B77b or whether these varieties systematized the full loss of *ŋg under heavy contact influence of neighboring

Plateau languages is hard to tell at this stage and would need the examination of other types of evidence, for example, morphological, such as the reshuffling of noun class systems (Hyman et al., 2019; Bollaert et al., 2021).

For the time being, we do consider languages not included in the phylogeny of Pacchiarotti et al. (2019; i.e., those shaded in gray in Appendix 1), because they were documented only later on (Kouarata et al., 2023), but where ŋ > ∅ is systematic (Section 4.2 and Map 3) to be part of the Plateau subgroup. Most of them are spoken on the Bateke Plateau—for example, Bua B70v and Kikimi B70r—while a few are in the Mai Ndombe—for example, Ŋiŋi B76b and Mosieno B76a. Taking into account the trade-driven mobility of Teke people (see Section 1), the latter two possibly represent recent arrivals in the Mai Ndombe region.

Even though the sound shift ŋ > ∅ has some diagnostic power in distinguishing a new Plateau subgroup, it was clearly also subject to contact-induced spread. Language contact and trade networks in the realm of the Teke kingdom led to the horizontal transmission of ŋ > ∅ as part of a broader process of dialectal diffusion with a pattern of centrality vs. peripherality (Trudgill, 2011) or core vs. periphery (Winford, 2003). As Map 3 shows, Plateau languages manifesting systematic total loss of *ŋg were the center of innovation and diffusion of change toward Mbete languages in the (north)west and Kwa-Kasai North languages in the (north)east. The contact-induced spread of the subsidiary innovation ŋ > ∅ in the chain *ŋg > ŋ > ∅ led to a second phase of convergence or increased homogenization among the language groups which underwent it (to different extents). Concomitantly, it also led to divergence between these varieties and those where /ŋ/ was never lost (see phase 4 in Fig. 5).

To conclude, it seems useful to refer to the distinction proposed by Babel et al. (2013) between clades exclusively defined by descent from a most recent common ancestor and taxa whose defining innovations are due to contact-induced diffusion between related languages.[22] Babel et al. (2013: 481–482) argue that if

---

[22] To refer to taxa whose shared features have sources other than inheritance from its members' last common ancestor, Babel et al. (2013: 446) use the term "apomorphic" which they admit to be at odds with how this term is commonly used in biological cladistics. In biology, apomorphy refers to a novel feature that is unique to a particular species and all its descendants and which can be used as a defining character for a species or group in phylogenetic terms (Martin and Hine, 2008). We therefore avoid using the term "apomorphic" as Babel et al. (2013) do.

> taxa can be formed when innovations diffuse across pre-existing dialect and even language boundaries, then there is no reason to assume that taxonomies based merely on shared innovations must also represent true cladistic structures. We suggest that the dynamics of language contact and linguistic diffusion can obscure or even erase clades.

In this article we have presented a prime example in support of this claim: the taxon characterized by the C2 ŋ > ∅ innovation was in all likelihood formed mostly by diffusion rather than descent. Attempts at positing cladistic structures based on this shared innovation are hard to reconcile with pre-existing lexicon-based cladistic structures.

In this article, we have offered evidence from Bantu languages in support of two claims set forth by Babel et al. (2013): (i) phonological innovation through contact-induced diffusion can obscure cladistic structure; and (ii) diffusion processes not only contribute to language homogenization but also to language diversification. Our new Plateau taxon within Kwilu-Ngounie is not exclusively defined by descent through the inheritance of C2 ŋ > ∅ from a most common recent ancestor but also by its horizontal transmission through space. The inclusion of varieties in this new Plateau subgroup largely depends on how much contact-induced change one allows in defining it.

Acknowledgments

This research was funded by an ERC Consolidator Grant no. 724275 within the Horizon 2020 Research and Innovation program granted to the last author. The first author acknowledges the support of FWO (Fonds Wetenschappelijk Onderzoek Vlaanderen) postdoctoral fellowship no. 12ZV721N and of the ERC Starting Grant no. 101116562 under the Horizon Europe Framework Programme. We are much indebted to two anonymous reviewers for their attentive reading and constructive critiques. One of the reviewers was invigoratingly critical of our first analysis of the data and pushed us to consider completely different scenarios. They made us see what the broader contribution of this article could be well beyond Bantu and brought to our attention many relevant works outside of Africa that we were not aware of. We are greatly indebted to their fastidious feedback, which dramatically improved this article.

References

Ansari Pour, Naser, Christopher A. Plaster, and Neil Bradman. 2013. Evidence from Y-chromosome analysis for a late exclusively eastern expansion of the Bantu-speaking people. *European Journal of Human Genetics* 21(4): 423–429.

Babel, Molly, Andrew Garrett, Michael J. Houser, and Maziar Toosarvandani. 2013. Descent and diffusion in language diversification: A study of Western Numic dialectology. *International Journal of American Linguistics* 79(4): 445–489.

Bastin, Yvonne. 1985. *Les relations sémantiques dans les langues bantoues*. Brussels: Académie royale des sciences d'outre-mer.

Bastin, Yvonne and Pascale Piron. 1999. Classifications lexicostatistiques: Bantou, bantou et bantoïde; De l'intérêt des "groupes flottants." In Jean-Marie Hombert and Larry M. Hyman (eds.), *Bantu Historical Linguistics: Theoretical and Empirical Perspectives*, 149–164. Stanford: CSLI.

Bastin, Yvonne, André Coupez, and Bernard de Halleux. 1983. Classification lexicostatistique des langues bantoues (214 relevés). *Bulletin des séances de l'Académie royale des sciences d'outre-mer* 37(2): 173–199.

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. Database. Tervuren: Royal Museum for Central Africa. http://linguistics.africamuseum.be/BLR3.html (accessed November 28, 2017).

Batibo, Herman B. 1998. A lexicostatistical survey of the Bantu languages of Botswana. *South African Journal of African Languages* 18(1): 22–28.

Bissila, Sylvie. 1991. *Description phonologique du ilaale (dialecte teke du Congo)*. Mémoire de diplôme d'études supérieures, Université Marien Ngouabi, Brazzaville.

Biton, Alexandre. 1969. *Dictionnaire ndumu-mbede-français et français-ndumu-mbede: Petite flore de la région de Franceville (Gabon); Grammaire ndumu-mbede*. Bar-le-Duc: Imprimerie St-Paul pour l'Archevêché de Libreville.

Bollaert, Flore, Sara Pacchiarotti, and Koen Bostoen. 2021. The noun class system of Bwala, an undocumented Teke language from the DRC (Bantu, B70z). *Nordic Journal of African Studies* 30(1): 1–40.

Boone, Olga. 1973. *Carte ethnique de la République du Zaïre: Quart sud-ouest*. Tervuren: Musée royal de l'Afrique centrale.

Bostoen, Koen. 2008. Bantu spirantization: Morphologization, lexicalization and historical classification. *Diachronica* 25(3): 299–356.

Bostoen, Koen. 2018. The Bantu expansion. In Thomas Spear (ed.), *Oxford Research Encyclopedia of African History*. Oxford: Oxford University Press. https://oxfordre

.com/africanhistory/view/10.1093/acrefore/9780190277734.001.0001/acrefore-9780190277734-e-191.

Bostoen, Koen. 2019. Reconstructing Proto-Bantu. In Mark Van de Velde, Koen Bostoen, Derek Nurse, and Gérard Philippson (eds.), *The Bantu Languages* (*Second Edition*), 308–334. Oxford: Routledge.

Bostoen, Koen and Gilles-Maurice de Schryver. 2018a. Langues et évolution linguistique dans le royaume et l'aire kongo. In Bernard Clist, Pierre de Maret, and Koen Bostoen (eds.), *Une archéologie des provinces septentrionales du royaume Kongo*, 51–55. Oxford: Archaeopress.

Bostoen, Koen and Gilles-Maurice de Schryver. 2018b. Seventeenth-century Kikongo is not the ancestor of present-day Kikongo. In Koen Bostoen and Inge Brinkman (eds.), *The Kongo Kingdom: The Origins, Dynamics and Cosmopolitan Culture of an African Polity*, 60–102. Cambridge: Cambridge University Press.

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.

Bostoen, Koen and Joseph Koni Muluwa. 2011. Vowel split in Hungan (Bantu H42, Kwilu, DRC): A contact-induced language-internal change. *Journal of Historical Linguistics* 1(2): 247–268.

Bostoen, Koen and Joseph Koni Muluwa. 2014. Umlaut in the Bantu B70/80 languages of the Kwilu (DRC). *Transactions of the Philological Society* 112(2): 209–230.

Bostoen, Koen, Bernard Clist, Charles Doumenge, Rebecca Grollemund, Jean-Marie Hombert, Joseph Koni Muluwa, and Jean Maley. 2015. Middle to late Holocene paleoclimatic change and the early Bantu expansion in the rain forests of western Central Africa. *Current Anthropology* 56(3): 354–384.

Bostoen, Koen, Jean-Pierre Donzo Bunza Yugia, Guy Noël Kouarata, Lorenzo Maselli, and Sara Pacchiarotti. 2023. Pre-Bantu substrate in Batwa Bantu languages of the Congo rainforest: A comparative study of nasal-oral stop cluster reduction. ICHL (International Conference of Historical Linguistics) 26, Heidelberg University, September 4.

Bostoen, Koen, Sara Pacchiarotti, and Wannes Hubau. Forthcoming. À la recherche des sentiers perdus du passé en forêt: Réflexions sur l'effondrement de la population en Afrique centrale il y a 1500 ans. *Proceedings of the Royal Academy for Overseas Sciences*. http://www.kaowarsom.be/documents/PDF%20BULLETIN/BOSTOEN.pdf (accessed February 17, 2024).

Burssens, Nico. 1999. *Dictionnaire français-buma*. Bandundu: Centre d'études ethnologiques de Bandundu.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14(3): 261–290.

Chacon, Thiago Costa and Johann-Mattis List. 2016. Improved computational models of sound change shed light on the history of the Tukanoan languages. *Journal of Language Relationship* 13(3–4): 177–204.

Coupez, André. 1956. Application de la lexicostatistique au mongo et au rwanda. *Aequatoria* 19: 85–87.

Coupez, André, Yvonne Bastin, and Evariste Mumba (eds.). 1998. Reconstructions lexicales bantoues 2 / Bantu Lexical Reconstructions 2. Electronic database. Tervuren: Musée royal de l'Afrique centrale.

Daeleman, Jan. 1977. A comparison of some Zone B languages in Bantu. *Africana Linguistica* 7: 93–144.

Dammann, Ernst. 1972. Das Meinhofsche Gesetz. *Afrika und Ubersee* 55: 242–244.

de Filippo, Cesare, Chiara Barbieri, Mark Whitten, Sununguko Wata Mpoloka, Ellen D. Gunnarsdóttir, Koen Bostoen, Terry Nyambe, Klaus Beyer, Henning Schreiber, Peter de Knijff, Donata Luiselli, Mark Stoneking, and Brigitte Pakendorf. 2011. Y-chromosomal variation in sub-Saharan Africa: Insights into the history of Niger-Congo groups. *Molecular Biology and Evolution* 28(3): 1255–1269.

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: 87–162.

Dunn, Michael. 2015. Language phylogenies. In Claire Bowern and Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, 190–211. London: Routledge.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2023. *Ethnologue: Languages of the World*. 26th edn. Dallas: SIL International. https://www.ethnologue.com.

Ellington, John Ernest. 1977. *Aspects of the Tiene language*. PhD thesis, University of Wisconsin, Madison.

François, Alexandre. 2014. Trees, waves and linkages: Models of language diversification. In Claire Bowern and Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, 161–189. London: Routledge.

Geisler, Hans and Johann-Mattis List. 2013. Do languages grow on trees? The tree metaphor in the history of linguistics. In Heiner Fangerau, Hans Geisler, Thorsten Halling, and William Martin (eds.), *Classification and Evolution in Biology, Linguistics and the History of Science: Concepts, Methods, Visualization*, 111–124. Stuttgart: Franz Steiner.

Goes, Heidi and Koen Bostoen. 2019. Progressive vowel height harmony in Proto-Kikongo and Proto-Bantu. *Journal of African Languages and Linguistics* 40(3): 1–51.

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 of the United States of America* 112(43): 13296–13301.

Gunnink, Hilde, Natalia Chousou-Polydouri, and Koen Bostoen. 2023. Divergence and contact in Southern Bantu language and population history: A new phylogeny in cross-disciplinary perspective. *Language Dynamics and Change* 13(1): 74–131.

Guthrie, Malcolm. 1971. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu languages*, vol. 2, *Bantu Prehistory, Inventory and Indexes*. London: Gregg International.

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

Heine, Bernd. 1973. Zur genetischen Gliederung der Bantu-Sprachen. *Afrika und Ubersee* 56(3): 164–185.

Heine, Bernd, Hans Hoff, and Rainer Vossen. 1977. Neuere Ergebnisse zur Territorialgeschichte der Bantu. In Wilhelm Möhlig, Franz Rottland, and Bernd Heine (eds.), *Zur Sprachgeschichte und Ethnohistorie in Afrika*, 57–72. Berlin: Dietrich Reimer.

Hinnebusch, Thomas J. 1999. Contact and lexicostatistics in comparative Bantu studies. In Jean-Marie Hombert and Larry M. Hyman (eds.), *Bantu Historical Linguistics: Theoretical and Empirical Perspectives*, 173–205. Stanford: CSLI.

Hinskens, Frans, Peter Auer, and Paul Kerswill. 2009. The study of dialect convergence and divergence: Conceptual and methodological considerations. In Peter Auer, Frans Hinskens, and Paul Kerswill (eds.), *Dialect Change: Convergence and Divergence In European Languages*, 1–48. Cambridge: Cambridge University Press.

Hochegger, Hermann. 1972. *Dictionnaire buma-français avec un aperçu grammatical*. Bandundu: Centre d'études ethnologiques de Bandundu.

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: Biological Sciences* 269(1493): 793–799.

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: MacDonald Institute for Archaeological Research.

Holden, Clare Janaki, Andrew Meade, and Mark Pagel. 2005. Comparison of maximum parsimony and Bayesian Bantu language trees. In Ruth Mace, Clare J. Holden, and Stephen Shennan (eds.), *The Evolution of Cultural Diversity: A Phylogenetic Approach*, 53–65. London: UCL Press.

Hombert, Jean-Marie. 1986. The development of nasalized vowels in the Teke language group. In K. Bogers, H. van der Hulst, and M. Mous (eds.), *The Phonological Representation of Suprasegmentals*, 359–379. Dordrecht: Mouton de Gruyter & Foris.

Hombert, Jean-Marie and Médard Mouélé. 1988. Éléments de phonologie diachronique du wanzi (langue bantu du Gabon, groupe B50). *Pholia* 3: 183–205.

Huelsenbeck, John P. and Bruce Rannala. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology* 53(6): 904–913.

Hyman, Larry. 2019. Segmental phonology. In Mark Van de Velde, Koen Bostoen, Derek Nurse, and Gérard Philippson (eds.), *The Bantu Languages* (*Second Edition*), 128–149. Oxford: Routledge.

Hyman, Larry M., Florian Lionnet, and Christophère Ngolele. 2019. Number and animacy in the Teke noun class system. In Samson Lotven, Silvina Bongiovanni, Phillip Weirich, Robert Botne, and Samuel Gyasi Obeng (eds.), *African Linguistics across the Disciplines: Selected Papers from the 48th Annual Conference on African Linguistics*, 89–102. Berlin: Language Science Press.

Impubi Mukwa, Mfum. 1987. *Le syntagme nominal en yans B85d: Parler de Mikingi Kiakia—Collectivité de Nkara/Niadi.* Travail de fin d'études, Institut supérieur pédagogique de Kikwit.

Jacques, Guillaume and Johann-Mattis List. 2019. Save the trees: Why we need tree models in linguistic reconstruction (and when we should apply them). *Journal of Historical Linguistics* 9(1): 128–167.

Jacquot, André. 1965. Précisions sur l'inventaire des langues teke du Congo. *Cahiers d'études africaines* 5(18): 335–340.

Kalyan, Siva and Alexandre François. 2019. When the waves meet the trees: A response to Jacques and List. *Journal of Historical Linguistics* 9(1): 168–177.

Kalyan, Siva, Alexandre François, and Harald Hammarström. 2019. Problems with, and alternatives to, the tree model in historical linguistics. *Journal of Historical Linguistics* 9(1): 1–8.

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.

Koni Muluwa, Joseph and Koen Bostoen. 2012. La diphtongaison dans les langues bantu B70–80 (Bandundu, RDC): Typologie et classification historique. *Africana Linguistica* 18: 355–386.

Koni Muluwa, Joseph and Koen Bostoen. 2015. *Lexique comparé des langues bantu du Kwilu* (*République démocratique du Congo*). Cologne: Rüdiger Köppe.

Koni Muluwa, Joseph and Koen Bostoen. 2019. Nsong (B85d). In Mark Van de Velde, Koen Bostoen, Derek Nurse, and Gérard Philippson (eds.), *The Bantu Languages* (*Second Edition*), 414–448. Oxford: Routledge.

Kouarata, Guy Noël. 2016. *Dictionnaire beembe-français*. Saint-Denis: Edilivre.

Kouarata, Guy Noël, Sara Pacchiarotti, and Koen Bostoen. 2023. Nouvelle contribution à l'inventaire et la description des parlers teke (bantou, B70–80). *Linguistique et langues africaines* 9(2): 1–33.

Kristensen, Ole, Don Webster, and Thelma Webster. 1984. Enquete dialectale teke. Unpublished manuscript, SIL, Brazzaville.

Labov, William. 1981. Resolving the Neogrammarian controversy. *Language* 57(2): 267–308.

Labov, William. 1999. *Principles of linguistic change: Internal factors*. Oxford: Blackwell.

Lane, Bryan David. 1989. *Toward Understanding Multilingualism among the Mbete of Northern Congo*. MA thesis, University of Texas, Arlington.

Lema, Gwete. 1978. *La statuaire dans la société teke: Morphologie et contexte culturel.* Thèse de doctorat, Katholieke Universiteit te Leuven.

Linton, Pauline. 2013a. Dictionnaire latege (Webonary). SIL. https://www.webonary.org/latege/?lang=en

Linton, Pauline. 2013b. *La négation en latege*. Léconi: Global Outreach Mission.

Loubelo, Fidèle. 1987. *Description phonologique du itsaangi, parler de Madouma-Mossendjo*. Mémoire de diplôme d'études supérieures, Université Marien Ngouabi, Libreville.

Maho, Jouni Filip. 2009. NUGL online: The online version of the new updated Guthrie list, a referential classification of the Bantu languages. June 4, 2009. https://goto.glocalnet.net/mahopapers/nuglonline.pdf (accessed December 13, 2010).

Marchal-Nasse, Colette. 1989. *De la phonologie à la morphologie du nzebi, langue bantoue (B52) du Gabon*. Thèse de doctorat, Université libre de Bruxelles.

Martin, Elizabeth and Robert Hine. 2008. *A Dictionary of Biology*. Oxford: Oxford University Press.

Maselli, Lorenzo, Véronique Delvaux, Jean-Pierre Donzo Bunza Yugia, Sara Pacchiarotti, and Koen Bostoen. 2023. Labial-velar stops in Sakata (Bantu C34): A preliminary phonetic and diachronic phonological account. *Studies in African Linguistics* 52(2): 125–147.

Matakumba Kanika, Mirel. 1980. *Structure phonologique et morphologique de la langue teke (B.75)*. Mémoire de graduat. Université nationale du Zaïre, campus de Lubumbashi.

Mayanga, Tayeye. 1985. *Grammaire yansi (Rép. du Zaïre)*. Bandundu: Centre d'études ethnologiques de Bandundu.

Meeussen, Achiel Emiel. 1962. Meinhof's Rule in Bantu. *African Language Studies* 3: 25–29.

Mickala-Manfoumbi, Roger. 1988. *Éléments de description du duma, langue bantu du Gabon (B51)*. Mémoire de licence spéciale, Université libre de Bruxelles.

Möhlig, Wilhelm J.G. 1977. Zur frühen Siedlungsgeschichte der Savannen-Bantu aus lauthistorischer Sicht. In Wilhelm Möhlig, Franz Rottland, and Bernd Heine (eds.), *Zur Sprachgeschichte und Ethnohistorie in Afrika*, 166–193. Berlin: D. Reimer.

Möhlig, Wilhelm J.G. 1979. The Bantu nucleus: Its conditional nature and its prehistorical significance. *Sprache und Geschichte in Afrika* 1: 109–141.

Möhlig, Wilhelm J.G. 1981. Stratification in the history of the Bantu languages. *Sprache und Geschichte in Afrika* 3: 251–317.

Motingea Mangulu, André. 2004. Divergences en dialectes tiene (bantou B.81): Origine disparate ou perte séparée d'identité linguistique? *Annales Aequatoria* 25: 151–202.

Mouandza, Jean-Daniel. 2001. *Éléments de description du iyaa* (*parler bantu du Congo-Brazzaville*). Thèse de doctorat, Université de Nice Sophia Antipolis.

Mouélé, Médard. 1997. *Étude synchronique et diachronique des parlers dúmá* (*groupe bantu B.50*). Thèse de doctorat, Université Lumière Lyon 2.

Ndouli, Guy Blaise. 2001. *Description phonologique du mbere: parler de Tsama*. Mémoire de maîtrise, Université Marien Ngouabi, Brazzaville.

Nguma, Wendo. 1986. *Dictionnaire français-yansi* (*Rép. du Zaïre*). Bandundu: CEEBA.

Ngwamashi Kabandji-Bola Kamu, Ernest. 1979. *Equisse morphologique du luhilel: Formes nominales et pronominales*. Travail de fin d'études, Institut pédagogique national, Université nationale du Zaïre, Kinshasa.

Ngwamashi Kabandji-Bola Kamu, Ernest. 1981. *Etude confrontative des formes verbales du français et du luhileel*. Mémoire de licence, Institut pédagogique national, Université nationale du Zaïre, Kinshasa.

Niama-Niama, Japhet. 2021. *Approche historico-comparative des langues bantu du Gabon: Vers de nouveaux embranchements phonologiques partagés entre les groupes B50-60-70*. Libreville: Université Omar Bongo.

Nsuka-Nkutsi, François. 1980. Eléments de description du punu. Lyon: Centre de recherches linguistiques et sémiologiques, Université Lumière Lyon 2.

Ntsiba Ngolo, Edouard Wilfrid, and Ruth Raharimanantsoa. 2021. *Mukaana wu a mankwumu ma bilogho mu Etyeghe ya mu Limpuru: Dictionnaire visuel téké du sud-ouest–français*. Brazzaville: SIL-Congo.

Nurse, Derek and Gérard Philippson. 1980. The Bantu languages of East Africa: A lexicostatistical survey. In Edgar C. Polomé and Charles Peter Hill (eds.), *Language in Tanzania*, 26–67. London: Oxford University Press, for the International African Institute.

Nurse, Derek and Gérard Philippson. 2003. Introduction. In Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*, 1–13. London: Routledge.

Okoudowa, Bruno. 2016. *Morfologia verbal do Lèmbáámá: Língua banta falada no Gabaõ* (*África Central*). Saarbrücken: Novas Edições Acadêmicas.

Pacchiarotti, Sara and Koen Bostoen. 2020. The Proto-West-Coastal Bantu velar merger. *Africana Linguistica* 26: 139–195.

Pacchiarotti, Sara and Koen Bostoen. 2021a. The evolution of the noun class system of Ngwi (West-Coastal Bantu, B861, DRC). *Language in Africa* 2(3): 11–67.

Pacchiarotti, Sara and Koen Bostoen. 2021b. Final vowel loss in Lower Kasai Bantu (DRC) as a contact-induced change. *Journal of Language Contact* 14: 437–474.

Pacchiarotti, Sara and Koen Bostoen. 2022. Erratic velars in West-Coastal Bantu: Explaining irregular sound change in Central Africa. *Journal of Historical Linguistics* 12(3): 381–445.

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.

Pacchiarotti, Sara, Lorenzo Maselli, and Koen Bostoen. 2021. The emergence of interior vowels and heterosyllabic vowel sequences in Ngwi (Bantu B861, DRC). *Papers in Historical Phonology* 6: 107–146.

Paulian, Christiane. 1994. Nasales et nasalisation en ŋgùŋgwèl, langue bantu du Congo. *Linguistique africaine* 13: 83–129.

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* (*Second Edition*), 335–354. Oxford: Routledge.

Phillips, Betty S. 1984. Word frequency and the actuation of sound change. *Language* 60(2): 320–342.

Phillips, Betty S. 2006. *Word frequency and lexical diffusion*. London: Palgrave Macmillan.

Phillips, Betty S. 2020. Spread across the lexicon: Frequency, borrowing, analogy, and homophones. In Richard D. Janda, Brian D. Joseph, and Barbara S. Vance (eds.), *The Handbook of Historical Linguistics*, 343–356. Hoboken, NJ: Wiley Blackwell.

Raharimanantsoa, Ruth. 2021. Dictionnaire téké des plateaux–français (Boma-Nzikou: ebo/nzu): Unkaana a andɔ Ityɔɔ–Limpuru. Brazzaville: SIL-Congo & ABDATEK (Association pour la Bible, le Développement et l'Alphabétisation en Téké).

Raharimanantsoa, Ruth and Wilfrid Edouard Ntsiba Ngolo. 2015. *Petit guide d'orthographe et de grammaire du téké du sud-ouest parlé dans le Pool, la Bouenza et la Lékoumou, en République du Congo*. Brazzaville: SIL-Congo.

Rexová, Katherina, Yvonne Bastin, and Daniel Frynta. 2006. Cladistic analysis of Bantu languages: A new tree based on combined lexical and grammatical data. *Naturwissenschaften* 93(4): 189–194.

Ringe, Donald, Tandy Warnow, and Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100(1): 59–129.

Rottland, Franz. 1977. Reflexes of Proto-Bantu Phonemes in Yanzi (B85). *Africana Linguistica* 7: 375–396.

Scarborough, Matthew. 2019. Cognacy and computational cladistics: Issues in determining lexical cognacy for Indo-European cladistic research. In Matthew Scarborough, Matilde Serangeli, and Thomas Olander (eds.), *Dispersals and Diversification: Linguistic and Archaeological Perspectives on the Early Stages of Indo-European*, 179–208. Leiden: Brill.

Schadeberg, Thilo C. 1995. Spirantization and the 7-to-5 vowel merger in Bantu. *Belgian Journal of Linguistics* 9: 71–84.

Schadeberg, Thilo C. 2003. Historical linguistics. In Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*, 143–163. London: Routledge.

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(7). https://doi.org/10.1126/sciadv.abd8352.

Stappers, Leo. 1986. *Boma: Eine Sprachskizze*. Hamburg: Helmut Buske.

Swartenbroeckx, Pierre 1948. *Dictionnaire kiyansi ou kiyei, langage des bayansi ou bayey de territoire de Banningville (district du Lac Léopold II) au Congo belge*. Brussels: Royal Museum for Central Africa.

Tadmor, Uri, Martin Haspelmath, and Bradley Taylor. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2): 226–246.

Tete Wer Sey, Ikiero K. 1975. *Éléments de grammaire wóngo (wónk) C85: Phonologie et morphologie*. Mémoire de licence, Université nationale du Zaïre, Lubumbashi.

Thomason, Sarah G. 2001. *Language Contact: An Introduction*. Washington, DC: Georgetown University Press.

Thomason, Sarah G. 2008. Social and linguistic factors as predictors of contact-induced change. *Journal of Language Contact* Thema 2: 42–56.

Thomason, Sarah G. and Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.

Tonnoir, René. 1970. *Giribuma: Contribution à l'histoire et à la petite histoire du Congo équatorial*. Tervuren: Musée royal de l'Afrique centrale.

Trudgill, Peter. 1986. *Dialects in Contact*. Oxford: Basil Blackwell.

Trudgill, Peter. 2011. *Sociolinguistic typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.

Van Acker, Sifra, Sara Pacchiarotti, and Koen Bostoen. 2024. The introduction of sugarcane in West-Central Africa: Insights from comparative Bantu word histories. *Nordic Journal of African Studies* 33(1): 17–45.

Vansina, Jan. 1959. *Esquisse de grammaire bushong*. Tervuren: Musée royal du Congo belge.

Vansina, Jan. 1966. *Kingdoms of the Savanna*. Madison: University of Wisconsin Press.

Vansina, Jan. 1973. *The Tio Kingdom of the Middle Congo, 1880–1892*. London: Oxford University Press, for the International African Institute.

Vansina, Jan. 1973–1974. Probing the past of the Lower Kwilu peoples (Zaire). *Paideuma* 19–20: 332–364.

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

Wang, William S.Y. 1969. Competing changes as a cause of residue. *Language* 45(1): 9–25.

Weinreich, Uriel. 1953. *Languages in Contact*. The Hague: Mouton.

Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Malden: Blackwell.

Yanovich, Igor, Silvia Ghirotto, Patricia Alexandra Silva Santos, and Andrea Benazzo. 2021. Computational inference in non-tree models of language diversification: The case of Bantu. Abralin ao Vivo—Linguists Online, February 10. https://aovivo.abralin.org/en/lives/igor-yanovich-2/.

A.1 Appendix 1: Inventory of Kwilu-Ngounie languages included in this study

In the table below, we list the 39 Kwilu-Ngounie languages on which our historical-comparative study focuses. We provide the glossonym followed by the reference place where the doculect is spoken in parentheses, referential code, country, geocoordinates for the reference place, and sources where the comparative data originate. Our study includes original fieldwork data collected in the DRC and Congo in 2021 and 2022 (for more information, see Kouarata et al., 2023), mainly from language varieties that had not previously been inventoried. These have a referential code that is unattested in referential lists of the Bantu languages (Guthrie, 1971; Maho, 2009; Hammarström, 2019). These new codes correspond to those proposed in Pacchiarotti et al. (2019) or abide by the principles exposed in that study (see Section 2).

Language varieties shaded in gray do not feature in the phylogenetic classification of Pacchiarotti et al. (2019). Based on their location and the preliminary data available to us, we tentatively consider them to belong to the Kwilu-Ngounie branch of WCB. As mentioned in Section 2, not all potentially relevant WCB varieties included in the phylogeny of Pacchiarotti et al. (2019) are included in this study due to lack of lexical data on the sound change investigated in the present article.

	Language	Code	Country	Geocoordinates		Source
				Lat.	Long.	
1.	Wanzi (Mayela)	B501	Gabon	−0.9	12.78	Hombert and Mouélé, 1988; Mouélé, 1997; Niama-Niama, 2021
2.	Duma (Lastoursville)	B51	Gabon	−0.8	12.6	Mickala-Manfoumbi, 1988; Mouélé, 1997; Niama-Niama, 2021
3.	Nzebi	B52	Gabon	−1.9	11.9	Marchal-Nasse, 1989; Mouélé, 1997; Niama-Niama, 2021
4.	Tsaangi (Lekoko)	B53	Gabon	−2.98	12.74	Loubelo, 1987; Mouélé, 1997; Niama-Niama, 2021
5.	Mbete (Tsama)	B61	Congo	−0.53	14.58	Ndouli, 2001
6.	Mbaama (Bidoua)	B62	Congo	−3.47	13.45	Own fieldwork 2021
7.	Nduumo (Yéyé)	B63	Gabon	−1.49	13.4	Biton, 1969
8.	Tua (Tua)	B70p	DRC	−3.63	16.61	Own fieldwork 2022
9.	Tiimi (Bokala)	B70q	DRC	−3.12	17.06	Own fieldwork 2021
10.	Kikimi (Nganambo)	B70r	DRC	−3.31	17.34	Own fieldwork 2021
11.	Kaan (Fankana)	B70s	DRC	−3.7	17.48	Own fieldwork 2021
12.	Nzali (Ibali)	B70t	DRC	−3.67	16.02	Own fieldwork 2021

(*cont.*)

	Language	Code	Country	Geocoordinates		Source
				Lat.	Long.	
13.	Tswaara (Nkana)	B70u	DRC	−3.9	15.92	Own fieldwork 2021
14.	Bua (Mbakana)	B70v	DRC	−4.44	16.19	Own fieldwork 2021
15.	Buu (Yuo)	B70w	DRC	−4.71	16.06	Own fieldwork 2021
16.	Nzinii (Ossélé)	B71bX	Congo	−1.41	15.34	Own fieldwork 2022
17.	Latege (Okoyo)	B71bY	Congo	−1.40	14.76	Own fieldwork 2022
18.	Latege (Léconi)	B71bZ	Gabon	−1.58	14.25	Linton, 2013a
19.	Ngungwel (Imporo)	B72a	Congo	−1.63	15.88	Own fieldwork 2021
20.	Laali (Mayeye)	B73bZ	Congo	−3.68	13.63	Bissila, 1991
21.	Laali (Kendi)	B73bW	Congo	−3.87	13.43	Own fieldwork 2022
22.	Yaa (Bihoua)	B73c	Congo	−3.81	13.32	Mouandza, 2001
23.	Tyee (Tsiaki)	B73d	Congo	−3.73	14.4	Own fieldwork 2022; Raharimanantsoa and Ntsiba Ngolo, 2015; Ntsiba Ngolo and Raharimanantsoa, 2021
24.	Eboo (Ngo)	B74	Congo	−2.5	15.6	Own fieldwork 2022; Raharimanantsoa, 2021
25.	Tio Bali (Mpila)	B75	Congo	−4.08	14.66	Own fieldwork 2022
26.	Ŋiŋi (Tshumbiri)	B76b	DRC	−2.63	16.24	Own fieldwork 2021
27.	Kukuya (Lékana)	B77a	Congo	−2.3	14.57	Own fieldwork 2022
28.	Fumu (Kintele)	B77b	Congo	−4.15	15.34	Own fieldwork 2022
29.	Wuumu (Kinkole)	B78X	DRC	−4.33	15.51	Own fieldwork 2022
30.	Wuumu (Odziba)	B78V	Congo	−3.57	15.51	Own fieldwork 2022
31.	Boma Yumu (Mondai)	B80z	DRC	−3.35	17.49	Hochegger, 1972; Burssens, 1999
32.	Boma Nkuu (Camp-Bankuu)	B80x	DRC	−3.42	17.26	Own fieldwork 2022
33.	South Boma (Boku)	B80y	DRC	−3.89	16.63	Own fieldwork 2022
34.	Tiene (Mansele)	B81	DRC	−2.08	16.45	Ellington, 1977
35.	North Boma (Bopaka)	B82X	DRC	−2.49	17.36	Own fieldwork 2022; Stappers, 1986
36.	North Boma (Inongo)	B82W	DRC	−1.93	18.28	Own fieldwork 2021
37.	Mfinu (Bingibingi)	B83	DRC	−4.31	15.54	Own fieldwork 2022
38.	West Yans (Mukonkie)	B85a	DRC	−3.48	17.29	Swartenbroeckx, 1948
39.	East Yans (Niadi)	B85b	DRC	−5.31	18.89	Nguma, 1986

A.2 Appendix 2: Combined table with 39 Kwilu-Ngounie varieties and corresponding percentages of reflexes of *ŋg (Tables 1–4 in main text)

Change	Variety	No. of reflexes	Reflex					
			ŋg		ŋ		∅	
			No.	%	No.	%	No.	%
Retention of C2 *ŋg	B501	42	42	100%	0	–	0	–
	B51	34	34	100%	0	–	0	–
	B52	45	45	100%	0	–	0	–
	B53	44	44	100%	0	–	0	–
	B73bZ	84	47	56%	36	43%	1	1%
	B73bW	82	58	71%	22	27%	3	2%
	B73c	46	44	94%	2	6%	0	–
Systematic loss of C2 *ŋg	B70p	32	1	1%	0	–	31	99%
	B70q	30	1	4%	0	–	29	96%
	B70r	24	0	–	0	–	24	100%
	B70s	28	7	25%	0	–	21	75%
	B70t	25	1	4%	0	–	24	96%
	B70u	37	2	5%	0	–	35	95%
	B70v	31	0	–	0	–	31	100%
	B70w	32	1	4%	0	–	31	96%
	B71bX	34	5	15%	0		29	85%
	B71bY	65	17	26%	0	–	48	74%
	B72a	29	0	–	0	–	29	100%
	B73d	72	3	4%	10	14%	59	82%
	B74	93	8	9%	0	–	85	91%
	B75	34	3	9%	0	–	31	91%
	B76b	34	0	–	0	–	34	100%
	B77a	28	0	–	0	–	28	100%
	B77b	28	0	–	1	4%	27	96%
	B78V	23	1	4%	0	–	22	96%
	B78X	26	0	–	2	8%	24	92%
	B80x	28	2	5%	0	–	26	95%
	B80y	25	1	4%	0	–	24	96%

(*cont.*)

Change	Variety	No. of reflexes	Reflex					
			ŋg		ŋ		∅	
			No.	%	No.	%	No.	%
	B83	35	1	4%	0	–	34	96%
	B85b	48	11	23%	0	–	37	77%
Pervasive loss of C2 *ŋg	B80z	55	0	–	30	55%	25	45%
	B81	35	2	6%	8	22%	25	71%
	B82X	38	0	–	11	29%	27	71%
	B82W	37	2	–	9	26%	26	74%
Buffer zone	B61	46	15	33%	8	17%	23	50%
	B62	44	15	34%	10	23%	19	43%
	B63	60	18	30%	17	28%	25	42%
	B71bZ	74	29	38%	21	29%	24	33%

A.3 Appendix 3: Most widespread BLR reconstructions featuring C2 *ŋg in the West-Coastal-Bantu-speaking area

1. ***báŋgá** ‘jaw’ (BLR 108) > B501 *báŋgá*, B51 *mùbááŋgá/mìbááŋgá*, B52 *báŋgá*, B53 *báŋgá*, B62 *báá/ábáá*, B70p *lìbáá/màbáá*, B70q *lèbàà/màbàà*, B70r *lɛ̀bá/mábá*, B70s *lébàà/mábàà*, B70t *líbàà/mábàà*, B70u *líbàà/mábàà*, B70v *múbáá ntsì*, B70w *múbàà*, B71bX *báá/àbáá*, B71bY *bá/àbá*, B71bZ *bá/àbá*, B72a *báá/àbáá*, B73bW *bìbááàŋá*, B73bZ *bááŋà/bɛ́bááŋa*, B73d *ébáá*, B74 *báá*, B75 *báá/ã́báá*, B76b *bòbáá/màbáá*, B77a *màbá*, B77b *bá/màbá*, B78V *líbáá/ã́báá*, B80x *bòbíà*, B80y *lìbáá/màbáá*, B80z *mbà*, B81 *mòbáá*, B82W *eːbá*, B82X *èbáá/mbáá*, B83 *móbàà/míbàà*, B85a *mbaa*, B85b *mbaa*
2. ***bèŋg-à** ‘be red’ (BLR 151) > B501 *ábɛ̀ːŋgɛ́*, B51 *líbɛ́ŋgù*, B52 *bɛ̀ɛ̀ŋgɛ̀*, B53 *bɛ̀ɛ̀ŋgɛ̀*, B61 *bí*, B62 *kébíɛ̀*, B63 *abyeŋe*, B70p *òbyɛ́ɛ́*, B70q *òbìĩ*, B70t *óbíɛ̀*, B70t *óbíɛ̀*, B70u *óbìɛ̀*, B70v *óbíɛ̀*, B70w *úbíɛ̀*, B71bX *lɛ̀bɛ̀ɛ̀*, B71bY *byɛ̀*, B71bZ *byɛ̀*, B72a *bɛ̀*, B73bW *òbééŋgè*, B73bZ *ɔ́bɛ́ɛ́ŋɛ̀*, B73c (*ù*)*-bɛ́ɛ́ŋgɛ̀*, B73d *obyɛ*, B74 *bìɛ̀*, B75 *byɛ̀*, B76b *òbìɛ̀*, B77a *bè*, B77b *úbìè*, B78V *úbìɛ̀*, B78X *òbyɛ̀*, B80x *kòbɨ́*, B80y *ìbvyɛ́*, B80z *òbɛ̀*, B81 *òbɛ̀ɛ̀*, B82W *bè.é*, B83 *óbìɛ̀*

3. ***bìŋg-à** ‘chase; chase away; go after’ (BLR 213) > B51 *mbìŋgù* ‘hunt (n.),’ B61 *ŋò-byɛ́*, B62 *kébìà/mbìà* ‘hunt (n.),’ B63 *mbi* ‘hunt (n.),’ B70p *mbvìù/bàmbvìù* ‘hunt (n.),’ B70q *mbìì/màmbìì* ‘hunt (n.),’ B70r *óbìà*, B70s *íbíí* ‘hunt (n.),’ B70t *mbìù/bámbìù* ‘hunt (n.),’ B70u *mbìù/màmbìù* ‘hunt (n.),’ B70v *mbìì/mámbìì* ‘hunt (n.),’ B70w *mbíî/màmbíî* ‘hunt (n.),’ B71bX *mbìì* ‘hunt (n.),’ B71bY *òbìì*, B71bZ *bìà*, B72a *mbìì/àmbìì* ‘hunt (n.),’ B73bW *óbíŋɛ̀*, B73bZ *ɔ́byɛ̀ŋɛ̀*, B73c *úbíŋgì*, B73d *óbíî*, B73d *mbìì* ‘hunt (n.),’ B74 *mbìì* ‘hunt (n.),’ B74 *bìà*, B75 *mbìù* ‘hunt (n.),’ B75 *úbìà*, B76b *mbìù/màmbìù* ‘hunt (n.),’ B77a *mbìà* ‘hunt (n.),’ B78V *úbìì*, B78X *óbíí*, B80x *mbìì* ‘hunt (n.),’ B80z *òbìà*, B83 *óbìì*, B85a *bii*, B85b *obie*
4. ***bìŋgá** ‘green pigeon’ (BLR 216) > B52 *mbèŋgà*, B53 *mbèèŋgá*, B62 *mbéèŋgá/ámbéèŋgá*, B63 *mbiŋga*, B73bW *bíîŋgɛ́*, B73bZ *mbyɛ́ɛ̀ŋgɛ̀/bámbyɛ́ɛ̀ŋgɛ̀*, B73c *mbíîŋgí/bámbíîŋgí*, B73d *mbìí*, B74 *mbìá*, B81 *kèbèŋá*, B82W *mbjé*, B82X *mbyà/mbyà*
5. ***bòŋgó** ‘brain’ (BLR 274) > B70r *mbvùó/bámbvùó*, B71bX *àbwɛ̀ɛ̀*, B71bY *òbvùɔ́*, B71bZ *bòí*, B72a *òbwɛ̄̀ɛ̄́ɛ̄̀*, B73d *bwɛ̄̀ɛ̄́*, B74 *búɔ́/àbúɔ́*, B82W *bɔɔ́ŋ*, B82X *bɔ̀ŋɔ̀/bɔ̀ŋɔ̀*
6. ***bóŋgó** ‘knee’ (BLR 275) > B501 *lìbɔ́ːŋgɔ́/màbɔ́ːŋgɔ́*, B51 *lìbɔ́ːŋgɔ́/màbɔ́ːŋgɔ́*, B52 *bɔ̀ŋgɔ̀*, B53 *bɔ́ɔ́ŋgɔ́*, B61 *vúwoː́*, B62 *vúɔ́/ávúɔ́*, B63 *abwo*, B70p *búɔ́/mábúɔ́*, B70q *bɔ́ɔ̄́*, B70r *bɔ́ɔ́/mábɔ́ɔ́*, B70s *bùù/mábùù*, B70t *bùɔ̀/mábùɔ̀*, B70u *bùɔ̀/màbùɔ̀*, B70v *bùɔ̀/mábùɔ̀*, B70w *bùɔ̀/mábùɔ̀*, B70z *búò/mábúò*, B71bX *búɔ́/àbúɔ́*, B71bY *bùlú bvúɔ́*, B71bZ *kúnàbwóŋó/àkúnàbwóŋó*, B72a *búó/àbúó*, B73bW *bóòŋgó*, B73bZ *bɔ́ɔ̀ŋɔ̀/mábɔ́ɔ̀ŋɔ̀*, B73c *bɔ̋ɔ̋ŋgɔ́/mábɔ̋ɔ̋ŋgɔ́*, B73c *mbɔ̋ŋíɲì/mámbɔ̋ŋíɲì* ‘elbow,’ B73d *bwɔ́ɔ́*, B74 *búɔ́*, B75 *búɔ́/ā̀búɔ́*, B76b *búɔ́/màbúɔ́*, B77a *búɔ̀/ábúɔ̀*, B77b *bwɔ́/màbwɔ́*, B78V *búɔ́/ā̀búɔ́*, B78X *bɔ́ɔ̄́/mábɔ̀ɔ̄̀*, B80x *búɔ̀/màbúɔ̀*, B80y *bùɔ̀/mábùɔ̀*, B81 *è-bɔ́ɔ́*, B82X *ìbɔ́ɔ́/màbɔ́ɔ́*, B83 *bwɔ́/mábwɔ́*, B85a *boo/maboo*, B85b *bɔɔ*
7. ***búŋg-à** ‘gather up’ (BLR 339)[23] > B63 *gibvunaŋa*, B71bY *lèbvùrùà*, B73bW *óvúút*, B73d *óbvúɣúlù*, B74 *ùbvɨ̀ɔ̀*
8. ***bʊ́ŋgò** ‘beach; shore’ (BLR 341) > B63 *obuŋgu*, B74 *ntsínà-bùù*, B80z *lèbóŋg*, B81 *èvwúò*, B85a *mbwuu*, B85b *mbwo*

---

[23] The BLR 339 protoform is actually reconstructed as *bʊ́ŋg with *ʊ as V1 instead of *u. However, all reflexes of this protoform presented in this appendix undergo a sound change known as Bantu Spirantization (Schadeberg, 1995; Bostoen, 2008), where a stop becomes a fricative or affricate when followed by the high Proto-Bantu vowels *i and *u. This is why we change V1 from *ʊ to *u. It is worth noting, however, that BLR3 also includes *búŋg ‘wrap up’ (BLR 384). It is likely that BLR 339 and BLR 384 are actually one and the same reconstruction. Metonymical extension (temporal and spatial) is a very common semantic change in Bantu (Bastin, 1985), i.e., gathering up objects can lead to them being wrapped up or placed together.

9. ***búŋg-à*** 'mix' (BLR 385) > B61 *vùŋwà*, B70p *ò-bvúɔ̀*, B71bY *lèbvùrùà*, B73bW *óvúútùl*, B73d *ɔ́bvúútò*, B74 *ùbvììnà*, B80z *òbwúnà*, B83 *óbvùù*, B85a *buün*
10. ***càŋgà*** 'island' (BLR 475) > B70q *èsàà/bèsàà*, B70r *kɛ̀sàī̀/bɛ̀sàī̀*, B70s *ìsàà/bìsàà*, B70u *íʃìɛ̀/bíʃìɛ̀*, B71bX *èsàŋgà*, B71bY *kèsàŋgà*, B75 *ísã̀ã̀/bísã̀ã̀*, B76b *èsàã̀/bèsàã̀*, B78X *èsã́ã́/bìsã́ã́*, B80x *kèsàŋgá/bèsàŋgá*, B81 *kèsàŋà*, B85a *esaa*, B85b *kesa*
11. ***càŋgò*** 'news' (BLR 479) > B63 *ntcya*, B70q *ntsìà/màntsìà*, B70r *ntsàã̀/bàntsàã̀*, B70s *ntsààŋg/bàntsààŋg*, B70u *ntsìà/màntsìà*, B70v *ntsìà*, B70w *ntsìà/màntsìà*, B71bX *ntsàà/àntsàà*, B71bY *ntsàà*, B71bZ *ntsàà/àntsàà*, B72a *ntsìà/àntsìà*, B73d *nsya*, B74 *ntsàà/àntsàà*, B76b *ntsìà/màntsìà*, B77a *ntsàà*, B77b *ntsà/màntsà*, B78X *ntsàã́/bàntsàã́*, B80x *ntsã̀ã̀*, B80y *ntsìà*, B81 *nsàà*, B82W *nsàː*, B82X *nsàà*, B83 *ntsyàà*, B85a *musaa*, B85b *musaa*
12. ***cáŋgʊ́*** 'millet, eleusine; maize, small seed' (BLR 486) > B53 *tsáːŋgù*, B63 *tca*, B70p *syá/màsyá*, B70t *ʃìà/máʃìà*, B70u *ìʃìà/máʃìà*, B70v *ʃìà/máʃìà*, B70w *ʃìà/máʃìà*, B71bX *sáá/àsáá*, B71bY *sáŋgí* 'millet,' B71bY *ntsàá* 'small grain(s),' B72a *síá/àsíá*, B73bW *másááŋá*, B73bZ *sááŋgà/másááŋgà*, B73c *sá/másá*, B73d *sya*, B74 *lìsàà*, B76b *sìá/màsìá*, B77a *lísáá*, B77b *sáá/màsáá*, B80y *ʃìà/máʃìà*, B80z *mà-ʃìà*, B81 *èsáá*, B82W *isá*, B82X *ìsáá/màsáá*, B83Z *má-syà*, B85a *masaa*
13. ***cìŋgà*** 'string; hair' (BLR 622) > B501 *mùsííŋgà/mìsííŋgà*, B51 *sììŋgà*, B61 *ó-ʃyà/è-ʃyà*, B63 *osya*, B70q *mòsìò/mìsìò*, B70r *mùʃà/mìʃà*, B70s *mùʃìì/mìʃìì*, B70t *múʃì/míʃì*, B70u *mùʃìì/mìʃìì*, B70v *múʃì/míʃì*, B70w *múʃì/míʃì*, B70z *mùsìì/ mìsìì*, B71bX *òsìà/èsìà*, B71bY *òsìà*, B72a *òsìī̀/àsìī̀*, B73bW *mósííŋgè*, B73c *músííŋgì/mísííŋgì*, B73d *músíî*, B74 *úʃìà/ìʃia*, B76b *mòsìɲá/mèsìɲá*, B77b *mùsìà/mìsìà*, B78X *múswɛ̀/míswɛ̀*, B80x *mòʃíá/mìʃíá*, B80y *mùʃìà/mìʃìà*, B80z *mù-ʃìà*, B81 *mò-sìà*, B85a *musii mwɛd*, B85b *muna musi*
14. ***còŋg-à*** 'show' (BLR 665) > B70p *òswɔ̀*, B70q *òsùò*, B70r *ɔ́swɛ̀*, B70s *ìsùù*, B70t *ósúɔ̀*, B70u *òswɔ̀*, B70v *ósúɔ̀*, B70w *úswɔ̀*, B72a *òʃúò*, B73bW *ósóóŋg*, B73d *oswo*, B74 *úfúɔ̀*, B75 *sùò*, B76b *òsùɔ̀*, B77b *úsùɔ̀*, B80x *kòʃùɔ́*, B80y *ìʃwɔ́*, B80z *óswɔ̀*, B81 *òsùɔ̀*, B82X *kòcɔ̀ː*, B83Z *súɔ̀*, B85a *swo*
15. ***còŋgè*** 'point' (BLR 674) > B501 *tsɔ̀ɔ̀ŋgí*, B53 *tsòòŋgì*, B63 *tcwoyi*, B70p *ntswɛ̃́*, B70r *lèsɔ̃́ĩ́/màsɔ̃́ĩ́*, B70t *ntsúɔ̀/mántsúɔ̀*, B70u *ntswóó/màntswóó*, B70v *ntsúɔ̀/mántsúɔ̀*, B70w *ntsúɔ́/mántsúɔ́*, B71bX *ntsùí*, B71bZ *ntʃɔ̀í /àntʃɔ̀í*, B72a *ntsyáá*(*l*)/*àntsyáá*(*l*), B73bW *ntswɛ́ɛ̀ŋgɛ́*, B73bZ *ɛ́sɛ́ɛ́ŋgɛ̀/bísɛ́ɛ́ŋgɛ̀*, B73d *nsweẽ*, B74 *ntsùɛ́*, B74 *ntsüɛ́*, B75 *tsúɛ́*, B77b *ntsúé*, B78X *ntsúɛ̀*, B80y *ntsùɛ̀/mántsùɛ̀*, B80z *kènswɛ́*, B82W *ntswé*, B85a *nswè*, B85b *nswɛ*
16. ***cʊ́ŋgé*** 'moon, month' (BLR 739) > B501 *cúːŋgí*, B51 *tsúúŋgí*, B52 *tsúúŋgí*, B53 *tsùùŋgì/màtsùùŋgì*, B70p *ntsúí/bàntsúí*, B70u *ntswì/bàntswì*, B70v

ntswìì/bántswìì, B70w *ntsùì/bántsùì*, B70z *ntswíì/ bántswíì*, B71bX *ntswíí*, B71bY *ntswíí*, B72a *ntsíí/àntsíí*, B73bW *tsűȕŋgɛ́*, B73c *tsűűŋgí/mátsűűŋgí*, B74 *ntswíí*, B75 *ntswíí/bantswíí*, B77a *ntswíí*, B77b *ncwii*, B78V *ntsȕȕ*, B78X *ntswìì*, B80y *ntswì*, B83 *ntswì/bántwì*

17. ***díŋg-à** ‘search for; desire’ (BLR 997) > B53 *bàlíŋgìlì*, B61 *dyà*, B63 *ediŋi*, B70p *òdzìá*, B70q *òdìì*, B71bX *lèdzìà*, B71bY *lèdzìà*, B71bZ *dzìà*, B73d *ɔ́dzí*, B73d *ódzíì*, B74 *údzìà*, B75 *údzìà*, B76b *òdzìà*, B77a *kìdzìà*, B77b *údzí*, B78V *údzìà*, B78X *òdìì*, B80x *kòdìí*, B80z *òlìà*, B85a *lii*, B85b *olie*
18. ***díŋg-à** ‘turn round, wind round, wrap up’ (BLR 1062) > B70q *òlìĩ*, B70r *ɔ̀lɛ̀ɛ̃̏*, B70v *ódíì*, B70w *údíì*, B71bX *lèdzíà*, B71bY *lèdzɛ́nɛ́ɛ́*, B73bW *ódzíí*, B73bZ *ɔ́dzí*, B73d *ódzíɣì*, B74 *dzíà*, B75 *údzíà*, B77a *kílíɛ̃*, B77b *údzí*, B78X *òdìì*, B81 *òdía*, B82X *kòzìà*, B83 *ódzíì*
19. ***dòŋg-à** ‘speak; teach’ (BLR 1127) > B51 *lɔ̀ɔ̀ŋgɔ̀*, B63 *gilwoyo*, B70q *òlɔ̀ɔ̏*, B70s *ìlɔ́ŋgɔ́k*, B70t *ólúɔ̀*, B70u *ólúɔ̀*, B70v *ólúɔ̀*, B71bZ *lùò/àlùò*, B73bW *ndűȕŋgɛ́*, B73bZ *ndúúŋɛ̀/mándúúŋɛ̀*, B73c *ndóòŋgí/mándóòŋgí*, B73d *ólúɔ̀*, B74 *lùɔ̀*, B76b *òlùɛ̀*, B77a *kílúɔ̀*, B77b *úlúɔ̀* ‘learn,’ B77b *úlúɛ̀* ‘teach,’ B77b *mùlúɔ̀/mìlúɔ̀* ‘exercise,’ B80x *kòlɔ̏ɔ̏*, B80y *ìlúɛ̀*, B80z *ólɔ́ŋg* ‘teach,’ B80z *í-lwɔ̀* ‘advice,’ B82W *ìlːò*, B82X *ìlɔ̀ɔ̀/màlɔ̀ɔ̀*, B83 *ólűɔ̀*, B85b *olɔŋg*
20. ***dòŋgà** ‘plate, bowl’ (BLR 1131) > B70p *lóŋgá/màlóŋgá*, B70t *lùŋgá/màlùŋgá*, B70u *lóóŋgà/màlóóŋgà*, B73bZ *mɔ́lúúŋɔ̀/mɛlúúŋɔ̀*, B73d *ndɔ̋ɔ̏*, B78X *lɔ́ɔ́ŋ/màlɔ́ɔ́ŋ*, B80y *lùŋgá/màlùŋgá*, B80z *ílóŋg*, B83 *ídíɛ̀/mádíɛ̀*, B85b *lɔŋga ~ lelɔŋg*
21. ***dòŋgò** ‘line, row’ (BLR 1133) > B51 *lɔ̀ɔ̀ŋgɔ́*, B52 *lɔ̀ɔ̀ŋgɔ́*, B53 *lɔ̀ɔ̀ŋgɔ́*, B61 *òlɔ́ŋgɔ́*, B62 *ólúɔ̀/élúɔ̀*, B63 *olwo*, B70p *mùlúɔ́/mìlúɔ́*, B70q *mòlòò/mìlòò*, B70r *múlɔ́ŋg*, B70s *mùlùù/mìlùù*, B70u *múlúɔ̀/mílúɔ̀*, B70v *múlúɔ̀/mílúɔ̀*, B70w *múlúɔ̀/mílúɔ̀*, B71bX *òlùɔ̀/èlùɔ̀*, B71bY *ólùɔ̀*, B71bZ *òlùò/èlùò*, B72a *òlùò/ìlùò*, B73bW *mólóòŋgó*, B73c *múlɔ́ɔ̀ŋgɔ́/mílɔ́ɔ̀ŋgɔ́*, B73d *mólúɔ̀/mílúɔ̀*, B74 *ùlùɔ̀/ìlùɔ̀*, B75 *ű-lwɔ̀ɔ̀/í̋-lwɔ̀ɔ̀*, B76b *mòlùɔ̀/mèlùɔ̀*, B77a *úlúɔ́*, B77b *mùlúɔ́/mìlúɔ́*, B78V *űlùɔ̏/í̋lùɔ̏*, B78X *múlɔ́ɔ́ŋ/mílɔ́ɔ́ŋ*, B80y *mùlúɔ́/mìlúɔ́*, B81 *mòlɔ̀ɔ̀*, B82X *mùlɔ̀ɔ̀/mìlɔ̀ɔ̀*, B83 *múlwɔ́/mílwɔ́*, B85a *mulwo*, B85b *mulwɔ*
22. ***dòŋgò** ‘lineage, kinship, clan, tribe’ (BLR 1135) > B71bY *ólùɔ̀*, B73bW *múlóòŋgó/mílóòŋgó*, B73d *mólúɔ̀/mílúɔ̀*, B74 *ùlùɔ̀*, B80z *kélwɔ̀*, B85a *ndwo*, B85b *ndwɔ*
23. ***dʊ́ŋgʊ́** ‘shrub > red pepper; pepper’ (BLR 1223) > B501 *∅-ndúúŋgú/bà-ndúúŋgú*, B51 *ndúúŋgú*, B52 *ndúúŋgú*, B53 *∅-ndúúŋgú/bà-ndúúŋgú*, B61 *ndwó/andwó*, B62 *ndúó/àndúó*, B70u *ndùù/bándùù*, B70v *ndùù/bándùù*, B70w *ndùù/bàndùù*, B71bY *ndúú*, B71bZ *ndú/àndú ~ làndú*, B72a *ndű́ű́/àndű́ű́*, B73bW *ndúùŋgó*, B73bZ *ndúúŋɔ̀/bándúúŋɔ̀*, B73c *ndűűŋgú/bándűűndú*, B73d *ndúú*, B74 *nzúú/anzúú*, B75 *ndúú/bándúú*, B76b

lèlúú/bàndúú, B77a *àndzú*, B78V *mbvúú/bàmbvúú*, B78X *ndùù/bándùù*, B83 *ndùù/bándùù*

24. ***gàŋgà** ‘medicine man’ (BLR 1332) > B501 *ŋgààŋgà*, B51 *ŋgààŋgà*, B52 *ŋgààŋgə̀*, B53 *ŋgààŋgà*, B61 *ŋgàː/aŋgàː*, B62 *ŋgáà /áŋgáà*, B70p *ŋgàá/bàŋgàá*, B70q *ŋgàà/bàŋgàà*, B70s *ŋgàá/bàŋgàá*, B70t *ŋgáà/báŋgáà*, B70u *ŋgáà/báŋgáà*, B70v *ŋgáà/báŋgáà*, B70w *ŋgáà/báŋgáà*, B71bX *ŋgàà/àŋgàà*, B71bY *ŋgàà*, B71bZ *ŋgà/àŋgà*, B72a *ŋgàà/àŋgàà*, B73bW *ŋgááŋg*, B73bZ *ŋgááŋà/bàŋgááŋà*, B73c *ŋgááŋgà/báŋgááŋgà*, B73d *ŋgáà/báŋgáà*, B74 *ŋgàà/aŋgàà*, B75 *ŋgàà/báŋgàà*, B76b *ŋgàà/bàŋgàà*, B77a *ŋgàà/àŋgàà*, B78V *ŋgàã/báŋgàã*, B78X *ŋgáà/bàŋgàà*, B80x *ŋgàà/bàgàà*, B80y *ŋgàá/bàŋgàá*, B80z *mùŋgáŋgà*, B81 *ŋgàà*, B82W *ŋgáː*, B82X *ŋgàà/bàŋgàà*, B83 *ŋgáà/báŋgáà*, B85a *ŋgaa*, B85b *ŋgaa mankier*
25. ***gòŋgá** ‘spear’ (BLR 1448) > B501 *kɔ̀ŋgɔ́*, B51 *kɔ̀ŋgɔ́*, B52 *lə̀kɔ́ɔ́ŋgɔ̀/màkɔ́ɔ́ŋgɔ̀*, B53 *kɔ̀ŋgɔ́*, B70s *lɛ́kɔ̀ŋgá/mákɔ̀ŋgá*, B71bY *yùɔ́*, B73bW *lìkɔ́ŋg/nkɔ́ŋg*, B73bZ *lɛ́kɔ́ɔ́ŋgɔ̀/nkɔ́ɔ́ŋgɔ̀*, B73d *yùɔ́*, B74 *jʊ́úɔ̀/àyʊ́úɔ̀*, B75 *yùɔ́/ã́yùɔ́*, B78X *líkɔ̏ɔ̏/mákɔ̏ɔ̏*, B80y *kɔ́ŋgɔ̀*, B81 *èkɔ̀ɔ́*, B82W *ìkːɔ̀ɔ́*
26. ***gòŋgò** ‘back, backbone’ (BLR 1450) > B501 *ŋgɔ̀ːŋgɔ̀/bàŋgɔ̀ːŋgɔ̀*, B51 *ŋgɔ̀ɔ̀ŋgɔ̀*, B52 *gɔ̀ɔ̀ŋgɔ̀*, B53 *gɔ̀ɔ̀ŋgɔ̀*, B61 *òŋgúwò/èŋgúwò*, B62 *óŋgwɔ́ɔ̀/éŋgwɔ́ɔ̀*, B63 *oŋgwo*, B71bX *òŋgùɔ̀/èŋgùɔ̀*, B71bY *òŋgùɔ̀*, B71bZ *èŋgwò*, B72a *òŋgùò/ìŋgùò*, B73bW *móŋgóóŋ*, B73bZ *mɔ́ŋgɔ́ɔ́ŋɔ̀/mɛ́ŋgɔ́ɔ́ŋɔ̀*, B73c *múŋgɔ́ɔ́ŋgɔ̀/míŋgɔ́ɔ́ŋgɔ̀*, B73d *múŋgùɔ̀/míŋgùɔ̀*, B74 *úŋgwɔ̀ɔ̀*, B75 *ṹŋgùɔ̀/ĩ́ŋgùɔ̀*, B76b *mòŋgùɔ̀/mèŋgùɔ̀*, B77a *ùmvùɔ̀*, B78V *ũ̀ŋgùɔ̀/ĩ́ŋgùɔ̀*, B78X *mùŋgɔ̋ɔ̋/mìŋgɔ̋ɔ̋*, B81 *mòŋgɔ̀ɔ̀*, B85a *ŋgwe*, B85b *ŋgwe*
27. ***góŋgòdó** ‘millipede’ (BLR 1453) > B53 *ŋgɔ̀ŋgùlɔ́*, B61 *ŋgɔ́nɔ́ː/aŋgɔ́nɔ́ː*, B62 *ŋgɔ́nɔ́ɔ́/áŋgɔ́nɔ́ɔ́*, B63 *ŋgonoŋo*, B70p *ŋgwɔ̀ɔ̀nɔ̀/bàŋgwɔ̀ɔ̀nɔ̀*, B70q *ŋgúŋgwɔ̀ɔ̀nɔ̀/bàŋgúŋgwɔ̀ɔ̀nɔ̀*, B70s *ŋgwòònò/bàŋgwòònò*, B70t *ŋgùɔ̀nàkùù/báŋgùɔ̀nàkùù*, B70u *ŋgwɔ́ɔ́n’ákúú/bàŋgwɔ́ɔ́n’ákúú*, B71bX *ŋgúŋgwɔ̀ɔ̀*, B71bY *ŋgɔ́ŋgɔ̀*, B72a *ŋgwɔ̏ɔ̋ɔ̏(n)/àŋgwɔ̏ɔ̋ɔ̏(n)*, B73bW *ŋɔ́ɔ̀ŋgúl*, B73bZ *ŋɔ́ŋɔ́nɔ̀/báŋɔ́ŋɔ́nɔ̀*, B73c *ŋgɔ́ɔ̀ŋgɔ́lɔ̀/báŋgɔ́ɔ̀ŋgɔ́lɔ̀*, B73d *ŋgɔ̀ŋgɔ́nɔ̀*, B74 *ndàláŋgùŋgɔ̀ɔ̀nɔ́*, B76b *ŋgɔ̀ɔ̀nɔ́/bàŋgɔ̀ɔ̀nɔ́*, B80x *ŋgùŋgwɔ̃̀/bàŋgùŋgwɔ̃̀*, B80y *ŋgwɔ̀ɔ̀nɔ̀/báŋgwɔ̀ɔ̀nɔ̀*, B82X *ŋgɔ́ɔ́/ŋgɔ́ɔ́*, B85a *ŋgoon*, B85b *ŋgɔkɔn*
28. ***gʊ̀ŋgà** ‘bell’ (BLR 1514) > B501 *ŋgòòŋà*, B61 *kɔ́ːŋgà/àkɔ́ːŋgà*, B62 *ŋgóòŋgá/áŋgóòŋgá*, B63 *o-kuŋgu*, B70s *ŋgùŋg/bàŋgùŋg*, B71bX *ŋgòŋgà*, B71bY *ŋgùnàà*, B73bW *ŋgúúŋg*, B73c *ŋgúúŋgù/máŋgúúŋgù*, B73d *ŋguũ*, B74 *ŋgùŋgà*, B75 *ŋgúúŋgà*, B80z *ŋgùŋgà*, B82X *ŋúŋà/ŋúŋà*, B83Z *ŋgɔ́ɔ̏*, B85a *ŋgwuu*, B85b *ŋguŋga*
29. ***jʊ̀ŋgʊ́** ‘cooking pot’ (BLR 1632) > B501 *nzúúŋgù/mànzúúŋgù*, B61 *ɲúŋgù*, B62 *ndzúò/ándzúò*, B70p *ndzúù/màndzúù*, B71bY *nzùŋgù*, B71bZ *ndzùŋgù/andzùŋgù ~ ndzǔ/andzǔ*, B73bW *nzúùŋgó*, B73bZ *nzúúŋgɔ̀/mánzúúŋgɔ̀*, B73c *nzúùŋgú/mánzúùŋgú*, B73d *nzùú/mánzùú*, B74 *ndzùù/*

àndzùù, B75 *ndzùú/ã́ndzùú*, B78V *mbvùù/ambvùù*, B80z *nzɔ́ŋg*, B81 *nzùú*, B82W *ndʒùó*, B82X *ɲɟùú/ɲɟùú*, B83 *ndzùù/mándzùù*, B85a *nzwuu*, B85b *nzuu*

30. ***gàŋg-à** ‘tie up; seize’ (BLR 1331) > B501 *ìkàːŋgà*, B51 *kààŋgà*, B70q *ókàà*, B71bY *lèkàŋgà*, B71bZ *kàŋgà*, B73bW *ókáá*, B73c *úkááŋgà*, B80x *kòkàà*, B80z *òkàŋg*, B81 *kàː*, B82W *kòkːá*, B82X *kòkáà*
31. ***kádaŋg-à** (BLR 1665)/***káŋg-à** (BLR 1719) ‘fry; roast’ > B61 *yáŋà* ~ *káŋglà*, B62 *kékááŋgà*, B63 *gikaŋga*, B70p *òkáá*, B70r *ɔ̀kɛ́ɛ̀*, B70s *ìkáá*, B71bY *lèkáà*, B71bY *lèyáà*, B71bZ *káà*, B73bW *ókááŋg* ~ *òyáŋg*, B73bZ *ɔ́kááŋà*, B73bZ *ɔ́yáŋgà*, B73d *ókáà*, B73d *óyáã̀*, B74 *úkáà*, B74 *yɔ̃́ɔ̃̀*, B75 *úkáá*, B77a *kìkáà*, B77b *úká*, B78V *úkáã̀* ~ *úyɔ̃́ɔ̃̀*, B78X *òkààlà*, B80z *okáŋg*, B81 *okáa*, B82W *kòkːá*, B83 *ókáá* ~ *ówɔ́ɔ̀*, B83Z *kàː*, B85a *kaa*, B85b *okaa*
32. ***káŋgà** ‘guinea fowl’ (BLR 1720) > B61 *ŋká*, B70p *nkáá/bànkáá*, B70q *ŋkyàá/bàŋkyàá*, B70r *ŋkáã̀/bàŋkáã̀*, B70s *ŋkàà/bàŋkàà*, B70u *ŋkáà/bàŋkáà*, B70v *ŋkáà/báŋkáà*, B70w *ŋkàà/báŋkàà*, B71bY *nkáà*, B71bZ *ŋká/àŋká*, B72a *ŋkáà/àŋkáà*, B73bW *nkááŋá*, B73bZ *kááŋgá/bákááŋgá*, B73d *nkáà/bánkáà*, B74 *nkáà/ànkáà*, B75 *nkáà/bánkáà*, B76b *ŋkáá/bàŋkáá*, B77a *nkáà*, B77b *nkáà/bànkáà*, B80x *ŋkyá/bàŋkyá*, B80z *nkáː*, B81 *nkáà*, B82W *èkːáŋà*, B82X *èkáŋà/ŋkáŋà*, B85a *nkaa*, B85b *nkaa*
33. ***kíŋgó** ‘neck; nape; voice’ (BLR 1845) > B501 *kíːŋgù/màkíːŋgù*, B51 *kííŋgú*, B52 *kíːŋgà/màkíːŋgà*, B53 *lìkííŋgù/màkííŋgù*, B61 *ŋkíː/aŋkíː*, B62 *nkíí/ánkíí*, B70r *lèkíí/màkíí*, B70t *ŋkìì/báŋkìì*, B70u *ŋkìì/màŋkìì*, B70v *ŋkìì/máŋkìì*, B70w *ŋkìì/máŋkìì*, B70z *ŋkíì/máŋkíì*, B71bX *nkíí/ànkíí*, B71bY *nkíí*, B71bZ *nkí/ànkí*, B72a *ŋkíí/àŋkíí*, B73bW *nkyéèŋé*, B73bZ *nkyɛ́ɛ̀ŋɛ́/mánkyɛ́ɛ̀ŋɛ́*, B73c *kĩ̋ĩ̋ŋgí/mákĩ̋ĩ̋ŋgí*, B73d *nkíí*, B74 *nkíí/ànkíí*, B75 *nkíú/ã̀-nkíú*, B76b *ntsíú/màntsíú*, B77a *nkíí*, B77b *nkíú/mànkíú*, B78V *nkìṹ/ã̀nkìṹ*, B78X *mkíí/mánkìì*, B80z *lè-kìɛ̀*, B81 *nkíì*, B82X *nkíò/nkíò*, B83 *nkí/mánkí*
34. ***táŋg-à** ‘read; count’ (BLR 2786) > B501 *ìtááŋgà*, B51 *ìtááŋgà*, B52 *náŋgɔ̀*, B62 *kétááŋgà*, B70p *òtsyɛ́ɛ́*, B70q *òtyà*, B70r *ɔ́tɛ́ɛ̀*, B70s *ìtàà*, B70u *ótɛ́ɔ̀*, B70w *útã́ã́*, B71bX *lɛ̀tsɛ́ɛ̀*, B71bY *lètààrà*, B72a *òtíò*, B73bW *ótááŋg*, B73bZ *ɔ́tááŋgà*, B73c *ú-tã̋ã̋ŋgà*, B73d *ɔ́táà*, B76b *òtàà*, B78X *òtã́ã́*, B80x *kòtɛ́ɛ́*, B81 *òtáŋà*, B82W *kòtːáŋgà*, B83 *ótã́ã́*, B85a *taa*, B85b *otaa*
35. ***tʊ́ŋg-à** ‘put through; thread on string; plait; sew; tie up; build; close (in)’ (BLR 3081) > B501 *ìtóːŋgà*, B51 *tóóŋgà*, B52 *tóóŋgà*, B61 *ŋòtwá*, B62 *kétúá*, B63 *hotwa*, B70p *òtùò*, B70q *òtùò*, B70r *ɔ̀tswáà*, B70s *ìtsyú* ‘build,’ B70s *ítsúú* ‘sew,’ B70u *ntòŋgà/bàntòŋgà* ‘needle,’ B70u *ótúà* ‘build,’ B70v *ótúù*, B70w *útúù*, B70z *útù*, B71bX *lètúà* ‘build,’ B71bX *òntòŋgá/èntòŋgá* ‘needle,’ B71bY *lètsúà* ‘build,’ B71bY *ntòŋgà* ‘needle,’ B71bZ *tsúà*, B73bW *ótúúŋg* ‘build, sew,’ B73bW *tswɛ́ɛ̀ŋgɛ́/mátswɛ́ɛ̀ŋgɛ́* ‘needle,’ B73bZ *ɔ́tú-úŋgɔ̀*, B73c *útṹ̋ṹ̋ŋgù*, B73d *ótúú*, B74 *twíà*, B75 *útswíà*, B76b *òtùà*, B77a

kìtsúà, B77b *útsyà*, B78V *útswíà*, B78X *òtùù*, B80x *kòtùá*, B81 *òtúà*, B82W *kòtːswá*, B82X *kòtʃúà*, B83 *ótúù*, B83Z *ótúù*, B85a *twuu*, B85b *otwo*

36. ***jóŋgò** 'bile' (BLR 3573) > B501 *nzùùŋgì*, B53 *nzùùŋgì*, B63 *ndjuŋgu*, B73bW *nzónyóóŋg*, B73bZ *ɲɔ́ɔ́ŋgɔ̀/máɲɔ́ɔ́ŋgɔ̀*, B74 *ŋwɩ̃́ɩ̃́/àŋwɩ̃́ɩ̃́*, B77b *nwé*, B81 *nyɔ́ɔ̀*, B82W *ɲɔ́*, B82X *ɲɔ̀ɔ̀/ɲɔ̀ɔ̀*
37. ***bʊ̀ŋgì** 'fog' (BLR 4455) > B61 *lemvwɔ́yì*, B70q *lèbwìì*, B70r *lévìì*, B70s *lìbùì*, B70t *líbwɩ̂ì*, B70u *lìbùì*, B70v *líbwɩ̂ì*, B70w *lìbwì*, B71bX *lèbwì*, B71bY *lèbvùì*, B71bZ *làbùì*, B72a *lèbùī̀*, B73bW *lébýýŋə̀*, B73bZ *lɛ́búúŋɛ̀*, B73d *libwi*, B74 *mbìì*, B76b *lèbùù*, B77a *íbìì*, B78X *lìbùí*, B80x *lèbwí*, B80y *lìbùì*, B82W *évːué*, B85a *mibuu*
38. ***cʊ̀ŋgʊ̀** 'elephant grass; sugarcane'[24] (BLR 5111) > B501 *mùsùŋgù/mìsùŋgù*, B51 *músúŋgù/mísúŋgù*, B52 *sùŋgù*, B53 *mùsùŋgù/mìsùŋgù*, B61 *ʃùù*, B62 *ósùù/ésùù*, B63 *o-cu/e-cu*, B70p *mùsùù/mìsùù*, B70t *músúù/mísúù*, B70u *mùʃòò/mìʃòò*, B70v *músúu/mísúù*, B70w *músù/mísù*, B71bX *òsùù/èsùù*, B71bY *òsùù*, B71bZ *òʃù/èʃù*, B72a *ʃù/àʃù*, B73bW *mósúŋgò*, B73bZ *mɔ́súŋgɔ̀/mɛ́súŋgɔ̀*, B73c *músúŋgù/mísúŋgù*, B73d *músʉ̃́ʉ̃̀/mísʉ̃́ʉ̃̀*, B74 *ùʃù ~ ùfùà*, B75 *ɔ̃̀fʉ̃̀ʉ̃̀/ɪ̃̀fʉ̃̀ʉ̃̀*, B76b *mòsùù/mèsùù*, B77b *mùsùà/mìsùà*, B78V *ʉ̃́ntsɛ́/ɪ̃́ntsɛ́*, B80y *mùʃúù/mìʃúù*, B82 *mùcûː/mìcû*, B82W *mòtʃú*, B82X *mùcùː/mìcùː*, B83 *músù/mísù*, B85a *muswuu/miswuu*, B85b *muswo/miswo*
39. ***cíŋgé** 'iron needle' (BLR 6279) > B62 *kétséì/étséì*, B73bW *éséèŋgé*, B73bZ *ɛ́sɛ́ɛ́ŋgɛ̀/bísɛ́ɛ́ŋgɛ̀*, B73d *ɛ́séè*, B74 *ntsùɛ́*, B75 *ísééŋgò/bísééŋgò*, B78V *ísɛ̃́ɛ̃́/bísɛ̃́ɛ̃́*
40. ***boŋgo** 'horn' (BLR 6767) > B70q *èbɔ̃̀ɔ̃̀/bèbɔ̃̀ɔ̃̀*, B70r *kɛ̀bɔ̀/bɛ̀bɔ̀*, B70s *ìbùù/bìbùù*, B70u *íbúɔ̀/bíbúɔ̀*, B70v *íbúɔ̀/bíbúɔ̀*, B70w *ìbùɔ̀/bìbùɔ̀*, B70z *íbúò/ bíbúò*, B80x *kèbùɔ̀/bèbùɔ̀*, B80z *kébwɔ̀*, B85a *ebo*, B85b *kebɔɔ*
41. ***cèŋgè** 'sand' (BLR 7719) > B501 *lìtsɛ̀ɛ̀ŋgɛ̀/màtsɛ̀ɛ̀ŋgɛ̀* 'sand, country,' B61 *ncyé*, B62 *ntsìɛ̀* 'earth, ground, soil,' B63 *tsye* 'ground,' B70u *ntsìɛ̀* 'earth, ground, soil,' B70v *ʃíɛ̀*, B70z *sìè*, B71bX *ntsìɛ̀*, B71bY *ntsìɛ̀*, B71bZ *ntsyè/àntsyè* 'soil, sand,' B72a *ntsìè* 'earth, ground, soil,' B73bW *lísyɛ́ɛ́ŋgɛ̀*, B73c *∅-tsyééŋgì*, B73d *nsìɛ̀*, B74 *ntsìì/àntsìì* 'earth, ground, soil,' B76b *ntsìè/màntsìè* 'earth, ground, soil,' B77a *ntsìɛ̀*, B77b *ntsìɛ̀* 'earth, ground, soil,' B78X *ntsìɛ̀/màntsìɛ̀* 'sand, ground,' B80x *ntsìé* 'earth, ground, soil,' B80z *nsìɛ̀* 'sand, ground,' B82W *ésːjè*

---

[24] Although BLR lists both elephant grass and sugarcane as meanings for *cʊ̀ngʊ̀, Van Acker et al. (2024) show that the original meaning of *cʊ̀ngʊ̀ is elephant grass (Pennisetum Purpureum), an autochthonous species which resembles morphologically the imported sugarcane (Saccharum). Due to physical resemblance, *cʊ̀ngʊ̀ was later used to refer to sugarcane in the languages of the region.

42. ***taŋg-à*** ‘flow; drip’ (BLR 8732) > B61 *táŋíní* ‘drop,’ B62 *táŋí/átáŋí* ‘drop,’ B63 *taŋi* ‘drop,’ B70p *lì-tɛ́/má-tɛ́* ‘drop,’ B71bZ *tàí/àtàí* ‘drop,’ B73d *ótáŋà*, B75 *tɛ̀ɛ́/ã́tɛ̀ɛ́* ‘drop,’ B77b *téé/màtéé* ‘drop’
43. ***tàŋgí*** ‘bedstead’ (BLR 8741) > B501 *tááŋgì/màtááŋgì*, B51 *tááŋgì/màtááŋgì*, B70r *ntɛ̀ɛ́/bántɛ̀ɛ́*, B70t *ntàà/mántàà*, B70v *ntàà/mántàà*, B72a *ètóó/ìtóó*, B75 *ítã́ã́/bítã́ã́*, B80x *kètàŋgɔ́/bètàŋgɔ́*, B81 *ntàá*, B83 *ítɔ̀ɔ̀/bítɔ̀ɔ̀*, B85a *ntea*, B85b *ntea*
44. ***cáŋgá*** ‘tear’ B52 *lə̀tsààŋgà/màtsààŋgà*, B61 *lèntyá:/àntyá:*, B63 *lìtʃà/àtʃà*, B70q *ntsìà/màntsìà*, B70r *ntsìá*, B70s *ntsìà*, B70t *ntsìà*, B70t *ntsìà*, B70v *ntsìà*, B70w *mántsìà*, B71bX *àntsáá*, B71bY *antsáá*, B72a *ntsíá/àntsíá*, B73bW *líntsáŋá*, B73c *lítsã́ŋgá/mátsã́ŋgá*, B73d *lítsã́ã́/mátsã́ã́*, B74 *lìntsá/àntsá*, B76b *ntsìà*, B77a *màntsá*, B78X *ntsyà/mántsyà*, B80x *ntsìà/màntsìà*, B80z *nsìà*, B85a *nsaa/mansaa*, B85b *lesaa/nsaa*
45. ***dʊ̀ŋgá*** ‘ring, bracelet’ > B63 *olwa*, B70p *mùlùú/mìlùú*, B70q *mòlùò/mìlùò*, B70t *múlùù/mílùù*, B70u *mólùà/mílùà*, B70v *múlùù/mílùù*, B70w *múlùù/mílùù*, B71bX *òlùà/èlùà*, B71bY *òlùá*, B72a *òlùù/ìlùù*, B73bW *mólúŋgò*, B73bZ *lúúŋɔ̀/málúúŋɔ̀*, B73c *múlúùŋgú/mílúùŋgú*, B73d *múlùú*, B74 *ùlùá/ìlùá*, B75 *ṹ-lùá/ĩ́-lùá*, B76b *mòlùá/màlùá*, B77a *úlúú/mílúú*, B78V *ṹlùṹ/ĩ́lùṹ*, B78X *múlùù/mílùù*, B80x *mòlúà/mìlúà*, B80y *múlùà/mílùà*, B83 *múlúù/mílúù*, B85a *mulwuu*, B85b *mulwo*
46. ***dàŋgì*** ‘bottle (made of calabash)’ > B63 *olaŋgu*, B71bY *òlàŋgì*, B71bZ *òlàŋgì/èlàŋgì*, B74 *ùlàŋgì*, B80z *mùlàŋg*, B81 *mòlàŋè*, B82X *mùláŋì/mìláŋì*, B83 *mólèè/mílèè*, B85b *mulaŋgi*
47. ***déŋgé*** ‘savanna, bush’ > B61 *ɔ̀díː/ɛ̀díː*, B70p *mù-líɛ́*, B70w *múlìɛ̀/mílìɛ̀*, B71bZ *kàlíɛ́/èlíɛ́*, B73bW *lɛ́ɛŋgɛ́*, B73bZ *lɛ́ɛ́ŋɛ̀/málɛ́ɛ́ŋɛ̀*, B73c *lɛ̃́ɛ̃́ŋgɛ́/málɛ̃́ɛ̃́ŋgɛ́*, B73d *elehe*, B74 *ùlíɛ́/àlíɛ́*, B77a *ùlìɛ̀*, B80z *mù-lìɛ̀*, B85b *mulie*
48. ***bʊ́ŋgʊd-à*** ‘breed (animals)’ > B63 *gibunaŋa*, B70p *ɔ̀búúlɔ̀*, B71bZ *búnùŋà*, B73bW *óbúŋúnù*, B73bZ *ɔ́búŋúnɔ̀*, B73d *óbwɔ́ɔɣ́ɔ̀*, B74 *bʊ́únɔ̀*, B77a *kìbúúna*, B80x *kòbwɔ́ɔ́r*, B82X *kòbúò*, B83 *óbúɔ̀*, B85a *bwol*