

VOICELESS RHOTIC / RETROFLEX CONSONANTS AS AN AREAL FEATURE OF THE ATLANTIC LANGUAGES

John T. M. Merrill

Princeton University
j.merrill@princeton.edu

Abstract: Voiceless rhotic or retroflex obstruents like [t̥ʂ] or [t̥ʄ] are rare in Africa, but the existence of a single such sound is relatively common among the Atlantic languages within Niger-Congo. The sound is not a shared retention across languages, but was repeatedly innovated through areal pressure. The sound can furthermore be reconstructed for multiple Atlantic protolanguages, and as such seems to have existed in the area for some time. This paper catalogues the occurrence of this rare sound in Atlantic languages as documented in existing sources, and suggests that due to its persistence in the area through time, it is relatively likely that the most recent common ancestor of the Atlantic languages contained a sound similar to [t̥] or [t̥ʂ], supporting existing proposals which reconstruct such a sound.

Key words: Atlantic languages, Niger-Congo, historical linguistics, voiceless rhotics, retroflex consonants, areal effects

1. Introduction

A small but significant number of the Atlantic languages within Niger-Congo¹ possess a consonant sound which is otherwise rare within the family and in Africa more broadly. This sound is not phonetically the same in all of these languages, but can broadly be described as a voiceless rhotic or retroflex obstruent like [t̥, t̥ʂ, ʂ, t̥ʂ], henceforth “t̥ʂ”. In none of these languages does this consonant have a voiced counterpart, and in fact very few have more than a single retroflex consonant. In addition to the modern languages which possess “t̥ʂ”, such a sound can be reconstructed for a number of Atlantic protolanguages, though it has developed to other sounds in the daughter languages. Crucially, the source of this single voiceless rhotic/retroflex consonant is not the same in each language or protolanguage — i.e. it does not descend from the same segment in the most recent common ancestor of these languages — rather, the presence of this consonant is an areal feature. Though the sound is seemingly rather prone to change, the density of languages in the Atlantic area has allowed for it to be found in a sufficient number of languages throughout time, such that it has remained an available target of sound change. An important implication of this areal tendency is that any protolanguage spoken in the Atlantic area has a higher-than-usual chance of possessing a voiceless rhotic/retroflex consonant. Notably – while this paper will not argue explicitly for this position – Proto-Niger-Congo itself may be one such language.

This paper is organized as follows. §2 defines the sounds under consideration, and establishes that they are rare in the world, especially in Africa and within Niger-Congo. §3 catalogues the modern Atlantic languages which possess this sound, and the existing descriptions of the sound in each language. §4 discusses the Atlantic protolanguages for which “t̥ʂ” can be reconstructed, and the arguments for reconstructing it.

¹ I make no assumptions about the more controversial inclusions in the family. As such, “Atlantic-Congo” as delineated in Glottolog (Hammarström et al. 2021) could be substituted throughout.

§5 discusses the historical mechanisms by which this rare sound has persisted in the area, drawing parallels to similar situations elsewhere in Niger-Congo, and briefly addresses the implications for the reconstruction of Atlantic subgroups and Niger-Congo more broadly. §6 concludes.

2. Voiceless rhotic/retroflex consonants in the world and in Africa

The terms “rhotic” and “retroflex” are both somewhat phonetically imprecise. Rhotics include various trills, taps, approximants, and sometimes other sounds. Here I consider only coronal rhotics, i.e. sounds transcribed with the IPA symbols [r, ɾ, ɽ], or similar. Retroflex consonants are similarly difficult to define. Hamann (2003: 32) writes that “variation in the class of retroflexes seems to be larger than in any other articulatory class, which makes it difficult to find common properties that hold for all retroflexes.” Crucially, the degree to which the tongue tip truly curls back varies considerably between languages, and a more reliable articulatory property may be the lowering of the tongue middle. Some sounds traditionally described as retroflex may in fact be laminal retracted sounds, as argued for example by Ladefoged & Maddieson (1996: 155) for Polish. In practical terms, I take as “rhotic” or “retroflex” any sound which has been described in the literature with either term, and/or is transcribed with a phonetic symbol reserved for these sounds. The joint category “rhotic/retroflex” is not arbitrary. Coronal rhotics and retroflex sounds are acoustically similar, having a very low third formant (F3). Rhotics trigger the retroflexion of adjacent coronals in many languages, synchronically and historically (Hamann 2003: 83–87). In Atlantic languages, a sound will sometimes be transcribed with a symbol for a retroflex in one source, and a rhotic in another, e.g. Bayot [t̪~ɽ] and Pepel [ʂ~ɽs].

Table 1 presents a list of the sounds in IPA transcription which I consider to be voiceless rhotic/retroflex consonants. Certain other sounds like [ɻ] and [t̪ʰ] are not considered, as these are even rarer, and

not reported for African languages. After each sound is shown the number of phoneme inventories² which contain it in the PHOIBLE database in: all languages, African languages, and Atlantic-Congo languages (the highest-level genetic unit containing the Atlantic languages that PHOIBLE recognizes).

Table 1

Occurrence of voiceless rhotic/retroflex sounds in PHOIBLE

sound	world (3020)	Africa (885)	Atlantic-Congo (513)
ʈ	481	18	1 (Bijogo)
ʈʂ	106	2 (Bench)	1 (Ronga)
ʂ	198	5	2 (Eggon, Shall-Zwall)
ɽ	36	1 (Maasai)	0
ɽ̥	8	0	0
ɽ̥̄	4	0	0
ʈr	3	3	0

At present, PHOIBLE is the largest database of sound inventories. Nonetheless, well under half of the world's languages are represented, and there are inevitably inconsistencies in how it adapts the descriptions of various sounds in the primary sources to a unified IPA transcription. As such, it would be unwise to draw any very specific conclusions about the distribution of these sounds — indeed the areal effect described in this paper does not emerge at all due to undersampling. However, it is safe to conclude that voiceless rhotic/retroflex sounds are very rare in Africa, and with the exception of [ʈ] are quite rare in the world.

Globally, there are a number of areas in which voiceless retroflex sounds are common. The most notable are the Indian subcontinent (extending somewhat to the north and east) and Australia. In both areas,

² PHOIBLE often lists multiple phoneme inventories for the same language based on different sources. Sometimes these represent different dialects, but not always. The 3020 inventories correspond to 2186 languages. There are 718 African languages, and 423 from Atlantic-Congo.

languages generally have a rich inventory of retroflex consonants, and there is nothing unique phonologically about the retroflex place of articulation when compared with other places of articulation. In contrast, Atlantic languages generally have at most a single retroflex consonant.

In Africa, the only area with a significant number of retroflex consonants is the Kordofanian area. The genetic status of the Kordofanian languages is the subject of much uncertainty, but it is generally accepted that there is no particularly close genetic connection between all of the languages in the area, even if they are ultimately related (see e.g. Güldemann 2018: 223–231). As such, the presence of retroflex consonants in almost all of these languages is all but certainly an areal effect. Retroflex consonants are also found in some languages to the south, including Central Sudanic languages like Moru (/ɽ/ and /ɽʁ/), and the Omotic language Bench (/ʂ, z, ʂ, ʂʁ/). Following the global trend, Kordofanian languages generally have multiple contrastive retroflex consonants, in most languages including voiced /d/. It is notable that the Kordofanian languages do not include fricatives or affricates at the retroflex place of articulation, in contrast with the Atlantic area in which the single retroflex consonant is more often an affricate or continuant.

The last relevant area within Africa does not emerge in the PHOIBLE database. A number of Edoid languages of Nigeria are described as having a voiceless rhotic segment. These languages, as well as others in the area, are more generally rich in rhotic segments, often contrasting a trill [r] with an approximant [ɹ]. The most notable language with a voiceless rhotic is Edo (aka Bini), with a voiceless trill /ɽ/ (Amayo 1976; Elugbe 1989). Elugbe reports that other than Bini, only Urhobo has a voiceless rhotic among Edoid languages. PHOIBLE does include an entry for Bini based on other sources that are not specifically focused on the language, but does not list /ɽ/ as a phoneme. The only Nigerian language that PHOIBLE lists with a voiceless rhotic is Degema, in which [ɽ] is given as an allophone of /ɽ/. Overall, voiceless rhotics are seemingly rare even among Edoid languages, and are unlikely to be historically relevant to the Atlantic consonants.

3. Voiceless rhotic/retroflex consonants in modern Atlantic languages

The Atlantic languages are a geographical grouping of Niger-Congo languages spoken near the coast between the Senegal River in the north and northern Liberia in the south. Whether these languages or various subsets of them form a valid subgroup or subgroups within Niger-Congo has been the subject of much debate. The uncontroversial genetic groupings of Atlantic languages are listed in Table 2, along with notes on the presence of voiceless rhotic/retroflex consonants. The precise number of languages in each group depends on rather subjective decisions about language vs. dialect boundaries.

Table 2

Established Niger-Congo subgroups in the Atlantic area

subgroup		# of langs.	“t/ɽ” in modern languages	in protolang
Cangin		3~5	–	*ɽ > s, l
Fula-Sereer		2	–	
Wolof		1	–	
Bainunk-Kobiana-Kasanga		5~8	–	*ɽ > r, h, l
Biafada-Pajade		2	–	*ɽ > r, s
Tenda		3	–	*ɽ > r, s
Bak	Joola	Other Joola	~10	3 langs: t/tʂ (lenited /t/)
		Bayot	2~4	Kugere tʂ, Edamme ⟨rh⟩
	Manjak cluster		~3	tʂ~tʂ
	Balanta		1~2	–
Bijogo		1~3	tʂ~tʂ	Proto-Bak: *t(ʂ) > l, t, s *t > l, l, t
Sua (Mansoanka)		1	–	
Nalu		1	–	
Rio Nuñez		2	–	
Limba		1~2	“tʂ”	
Gola		1	–	
Mel		~10	Temne “tʂ”	

Among the Atlantic languages, at least seven are clearly described as having a voiceless rhotic/retroflex consonant. Some others seem to possess such a sound, but the existing descriptions are not sufficiently clear. These Atlantic languages are listed in Table 3, along with the phonetic symbols used in various primary sources for the sound in question. Symbols in angled brackets are those used in an official orthography. The following sections will examine the descriptions of these sounds in more detail.

Table 3

Atlantic languages with voiceless rhotic/retroflex sounds

modern language	symbols used in sources
Joola Fulup (Ejamat)	ʂ, rʂ
Joola Buluf	ʈ
Joola Eegimaa (Banjal)	ʈ, ʈ̣
Bayot Kugere	ʈ, ʂ
Bayot Edamme	⟨rh⟩
Manjak	⟨ʈ⟩, ʈ̣, ts
Mankanya	⟨t⟩, t(ʂ), (n)tʂ, č ; ⟨ʂ⟩, ʂ, š
Pepel	⟨ʂ⟩, ʂ, rʂ
Bijogo	ʈ, tʂ, rʂ
Temne	⟨t⟩, ʈ, tr, ʈ (tš), ʈ̣
Limba	⟨t⟩, ʈ, ʈ̣

3.1. Bijogo

Bijogo, spoken in the Bissagos (or Bijagós) Islands of Guinea Bissau, has at least three major dialect areas: Kagbaaga (Bubaque), Kajoko (incl. Orango), and Kamona (incl. Caravela). These are different enough that they might be treated as separate languages. All three have a single retroflex consonant which corresponds straightforwardly across varieties. Wilson (1990) describes the sound as follows: “[Orango] tʂ is a slightly retroflex affricate. Caravela has a retroflex stop. Bubaque has a fricative rʂ.” Segerer (2002) identifies Bubaque /t/ as “une occlusive prépalatale rétroflexe sourde.”

3.2. Manjak cluster

Traditionally, three languages are identified within the “Manjak cluster” of languages: Manjak, Mankanya, and Pepel. In truth the dialectal situation is more complicated, as some dialects of Manjak are more closely related to Pepel than to other Manjak dialects. Doneux’s (1975) Manjak cluster family tree is shown in Figure 1.

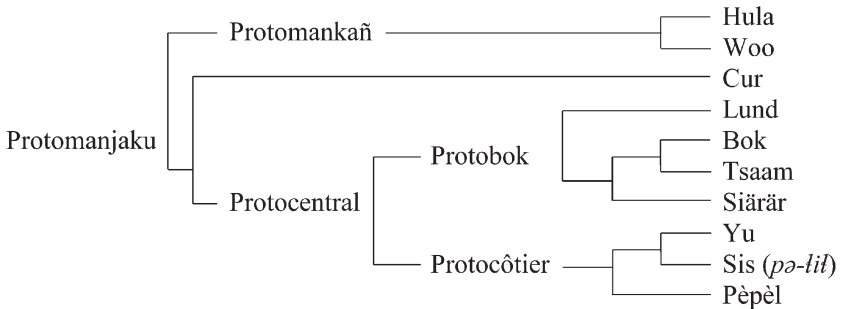


Figure 1. Manjak cluster family tree from Doneux (1975)

Doneux’s tree is of course subject to debate, but makes clear that the Manjak cluster cannot be divided neatly into the three languages traditionally presented.

A number of varieties are described as having a retroflex obstruent, though note that most dialects lack extensive documentation. I will refer to this sound as Proto-Manjak *tʂ. Table 4 shows the symbols used by Doneux (1975: 3–4) in various dialects for *tʂ and its prenasalized counterpart, along with symbols used by other authors where available.

For Bok, Karlik (1972) gives the sound as [tʂ]. Ndao (2013) gives the Pepel sound as a simple fricative [ʂ], while Doneux (1975: 4) describes it as [ɾs], an “affriquée dentale à attaque apicale vibrante sourde.” Gaved (2020: 25) gives Mankanya /t/ as a retroflex stop with a common word-final allophone [tʂ], while Trifkovic (1969: 15) gives it as [č], an alveo-palatal affricate. No source other than Doneux (1975) notes any phonetic difference when the sound is prenasalized.

Table 4

Reflexes of Proto-Manjak *tʂ

	*tʂ (Doneux)	*tʂ (Others)	*ntʂ (Doneux)	*ntʂ (Others)
Hula	tʃ	t~tʂ, ɕ	nc	nt~ntʂ, nɕ
Woo	ts		ntʂ	
Cur	h		nts	
Lund	t		nt	
Bok	ts	tɻ	nts	ntɻ
Yu	ts		nts	
Pepel	ɻs	ʂ	nts	nʂ

The official orthographies for Manjak and Mankanya use the symbol ⟨t⟩ for this sound, regardless of its realization, while Pepel uses ⟨s⟩. My impression from audio recordings of the Bible available online (Anonymous 2014; Jesus Film Project 1995–2022) is as follows: the Manjak speakers (dialect unspecified) have [tɻ] or [tʂ]; of the Mankanya speakers, some have a stop [t^(h)] with no affrication (seemingly Woo judging by other phonological features like [θ] instead of [s]), others have a strongly affricated [tʂ] (seemingly Hula, with [s] instead of [θ]), and at least one speaker shows frequent variation between [t] and [tʂ]; the Pepel speakers have [ʂ]. For dialects with an affricate realization, it seems likely that some have a laminal retracted alveolar, rather than an apical retroflex one — cf. American English /ɹ/, for which different speakers achieve the same acoustic effect with either a retroflex or “bunched/retracted” articulation. The development to [h] and [t] in Cur and Lund is interesting from a diachronic perspective, as it shows some of the ways in which the original retroflex affricate can develop historically.

In addition to its retroflex stop/affricate, Mankanya has a retroflex fricative [ʂ] (⟨s⟩ in the official orthography), which is equivalent to [ɻ], [l], or [s] in other Manjak cluster varieties. This makes Mankanya the only Atlantic language with more than one voiceless retroflex consonant (setting aside prenasalized consonants). Doneux reports that in the Woo

dialect, /θ/ (equivalent to Manjak-Pepel /c/) becomes [nʂ] when prenasalized. Doneux (1975: 4) gives [ʒ] (a “constrictive apicale rétractée”) as the Cur (Manjak) pronunciation of what is a voiced palatal stop/affricate /j/ in all other Manjak cluster varieties.

3.3. Joola languages (except Bayot)

Some Joola languages have a voiceless retroflex obstruent which results from the lenition of earlier [t]. Multiple sources document such a sound in Joola Fulup (aka Ejamat). In Carlton & Rand’s (1994) comparative Joola wordlist, postvocalic [t] in most other varieties generally corresponds to [ʂ] in their Fulup lists from the towns of Efok and Kahem (but not Youtou). Wilson (2007) has [rʂ] for this sound in his short Fulup wordlist. Other sources do not indicate this lenition, including Barry’s (1987) comparative Joola wordlist, a recent translation of the Gospel of Luke (Anonymous 2010a), and the earliest Fulup wordlist from the late 17th century (published in D’Avezac 1845), which always spells ⟨t⟩. My impression from online audio recordings from different dialects (Global Recordings Network 1961–2006) is that there is a range between [t~t^h~ʂ~ʂ] across speakers — but consistent for each speaker — in line with the dialectal variation documented in written sources.

In Carlton & Rand’s Buluf list (closely related to Fonyi) from the town of Djimande, [t] appears where most other varieties have postvocalic [t]. Their other Buluf list (“North Mlomp”) has only [t]. Unfortunately Buluf is not represented in other sources.

Joola Eegimaa (a Banjäl dialect) is much better described than Fulup or Buluf, but unfortunately the phonetic nature of its lenited /t/ is not made clear in existing sources. All Eegimaa singleton stops have lenited allophones after a vowel. Both Sagna (2008) and M. Bassène (2012) note that /t/ is subject to this lenition. A. Bassène (2007) does not mention lenition of /t/, while explicitly listing the lenited allophones of all other stops. Sagna uses the symbol [t̚] for the lenited allophone, and M. Bassène uses [t̚], but neither comment on its realization. Given

the use of ⟨t̥⟩ in nearby Manjak and Mankanya for [t̥(ʂ)], this is likely what is intended for Eegimaa. In all Eegimaa sources, /t/ becomes /l/ root-finally before a vowel, and Proto-Joola *t becomes /l/ between any two vowels outside of root-initial position — this is distinct from the lenited allophone of the synchronic phoneme /t/.

3.4. Bayot (Joola)

The Bayot languages were the first to split off in the Joola family tree. They are Bayot Gubaare and Edamme/Arame in Guinea Bissau, and Bayot Kuxinge and Kugere in Senegal. Sources differ on dialect/language boundaries, but agree that at least two languages should be differentiated. In some sources “Bayot” refers specifically to Bayot Kugere. All Bayot varieties are poorly documented, with Diagne (2009) on Kugere being the most extensive source. Bayot varieties are also documented in Barry (1987) and Carlton & Rand’s (1994) wordlists.

In an inventory that is unusual for the area, Diagne lists five retroflex consonants for Kugere: [t̥, d̥, l̥, ʂ, z̥]. However [t̥] and [z̥] are given in only one word each, and [d̥] in three words, compared with [t] in 146 words, and [l] in 78. Pending corroboration from other sources, only /t/ and /l/ can be confidently taken as contrastive phonemes. Diagne’s [t̥] corresponds to [ɕ] in Carlton & Rand’s Bayot (Kugere) list, while Barry’s list has [t]. Where Kugere has [t̥~ɕ], Kuxinge has [ɕ] in both Barry and Carlton & Rand, Gubaare has [ɕ] in Barry, and most other Joola languages have /t/. A translation of the gospel of Luke in the Kulaje dialect of Bayot Edamme (Anonymous 2010b) spells the sound as ⟨rh⟩. Kugere does have a contrastive phoneme /t/, which corresponds in some words to /s/ in other Joola languages (including Kuxinge), and in other words to /t/. Kugere /l/ corresponds to Proto-Joola *ɭ, and appears as [ɕ] in both Barry’s and Carlton & Rand’s Bayot lists.

While Bayot appears at first to present a rather different situation from the other Joola languages with a voiceless retroflex obstruent, in fact the situation is essentially the same. The single Bayot voiceless

rhotic/retroflex consonant [t] or [ɽ] present in Kugere and Edamme is the historical lenition of Proto-Joola *t, just as in the other varieties, and corresponds to [t] in Joola languages that have not undergone the lenition. While Kugere does seem to have a second contrastive retroflex consonant /ɽ/ in at least one dialect (/ɽ/ in others), there is no good evidence for any more than these two.

3.5. Temne and Limba

The Mel group contains around 10 languages spoken in the southern part of the Atlantic area. Gola and Limba have at times been taken as part of a “Southern Atlantic” subgroup along with Mel (as in Sapir 1971), but they are now generally treated as isolates within Atlantic or Niger-Congo. The largest Mel language, Temne, has a single retroflex consonant, variously described as a stop or affricate. While this sound does not appear in other Mel languages, it seems to be found in Limba based on rather limited documentation.

Temne has two contrastive consonants written as ⟨t⟩ and ⟨th⟩ (a dental stop) in the official orthography. An early description of Temne is given by Schlenker (1864), who uses a digraph ⟨tr⟩ for the first of these, along with the following comment: “If r follows t, and is to be pronounced together with t, as is often the case both at the beginning and at the end of a word; the r is pronounced rather faint much like tr in *true* or in *truth*.” Thomas (1916a) gives a more concrete description, identifying the same sound as follows: “ɽ (tʂ), retroflex t, often heard as tr, or č.” Yillah (1992) uses the symbol ⟨ɽ⟩, identified as an “apical retroflex stop.” Wilson (2007: 154) writes, “/t/ or /ts/ is alveolar, somewhat retroflex and affricated at times.” From these descriptions it is clear that the sound is a voiceless retroflex obstruent, though whether it is a stop or affricate is not so clear. Based on Wilson’s comment it seems that the precise articulation varies, perhaps by dialect or speaker. Interestingly, Ladefoged & Maddieson (1996: 42) describe the sound differently, though they mention it only in passing. For them, the Temne sound is IPA [ɽ], a laminal alveolar stop, contrasting with [t],

the apical dental stop. This description is based on the authors' own observations, and it is unclear what accounts for the difference with other sources. The discrepancy once again calls to mind the retroflex~retracted variation for American English /ɹ/.

Limba is much more poorly documented than Temne, but seems to have the same distinction between a retroflex stop/affricate and a dental stop. Clark (1922) is the most detailed (albeit enigmatic) description of the first sound, identified as ⟨t⟩: “t and c are often used interchangeably, [t] having a smothered sound.” Thomas (1916b) has wordlists for five Limba dialects, and ⟨t⟩ commonly appears in the “Sella” and “Safroko” dialects. Sometimes other dialects have ⟨č⟩ where the first two have ⟨t⟩. It must be noted that in Thomas' own words, he had “no knowledge of the languages recorded” outside of Temne, so the Limba transcriptions cannot be assumed to be particularly accurate. Nonetheless, being familiar with the Temne sound, it is telling that Thomas used the same symbol ⟨t⟩ in transcribing some Limba dialects. Ladefoged & Maddieson (1996: 42) report the exact same distinction in Limba as in Temne, i.e. [t̠] vs. [t̠̠]. While the precise pronunciation of the sound in question is not all that clear from existing descriptions, they seem to be in agreement that the Limba sound is effectively the same as that found in Temne.

There is no convincing evidence for a voiceless rhotic/retroflex consonant in other Mel languages, though some are not well documented. Thomas (1916b) does use the symbol ⟨t⟩ for one morpheme each in Bullom (plural class prefix *tə-*) and Kisi (perfective *ki~ʔi*), but this consonant is not reported in dedicated descriptions of these languages. Wilson (1962) finds that Temne ⟨t⟩ corresponds to [ts] or [c/ʃ] in other Mel languages.

4. Voiceless rhotic/retroflex consonants in Atlantic protolanguages

Multiple Atlantic protolanguages can be reconstructed with a voiceless rhotic/retroflex consonant based on sound correspondences between

modern languages. Here I present my own thoughts on the matter, summarizing arguments made in Merrill (2018) and (2021). Table 5 presents the relevant sounds which can be reconstructed. Because the precise phonetic realization of these segments cannot be determined with any certainty, in most cases I simply use the symbol *ɾ without intending any very specific value.

Table 5

Atlantic protolanguages with a voiceless rhotic/retroflex sound

	proto-sound	modern reflexes
Proto-Cangin	*ɾ	Noon-Saafi [s], Ndut-Paloor [l]
Proto-Bainunk-Kobiana-Kasanga	*ɾ *s (= [ʂ]?)	Bai. [l], Ko. [h], Ka. [r] Bai. [r], KK [s]
Proto-Biafada-Pajade	*ɾ	Bi. [r], Pa. [s]
Proto-Tenda	*ɾ	Konyagi [r], Bassari-Bedik [s]
Proto-Bak	*t(ʂ) *ɿ	Joola *[h], Manjak *[t], Balanta [s] Joola *[l], Manjak *[h], Balanta [t]

The sound reconstructed as *ɾ in the first four groups almost certainly comes from an earlier **t (it corresponds to [t] in e.g. Wolof, Bijogo, and Proto-Bantu), and in the modern languages becomes [t] or [tt] when subject to fortition (this phenomenon is absent in Cangin). The value [ɾ] or similar seems likely in each case, though other reconstructed values are of course possible. In Proto-Cangin in particular, the value is difficult to determine — Pozdniakov & Segerer (2017) suggest [h]. In Bainunk-Kobiana-Kasanga, there is additional evidence for a phonetic value corresponding to a “tɿ” sound. In borrowings from (Proto-)Manjak, BKK *ɾ corresponds to Proto-Manjak *tʂ ((t) in Manjak orthography), for example: Proto-BKK *-ɾax (Bai. Gubëeher *-lax*, Ko. *-hah*) vs. Manj. *-ɿak* ‘forge’; Proto-KK *-boox (Ko. *-booh*, Ka. *-boor*) vs. Manj. *-buax* ‘fish with line.’

Proto-BKK *s corresponds to Bijogo [tʂ~t~ʂ] in cognates, and given its realization [r] in Bainunk may have been retroflex — however

rhotacism of [s] to [r] is not uncommon crosslinguistically. When subject to fortition in Kobia-Kasanga it becomes [c(c)].

The two Proto-Bak sounds require further comment. The first, $*t(\text{ʃ})$, corresponds to Bijogo [tʃ~tʃ̣] in cognates. Based on its modern reflexes, its phonetic value is difficult to determine. The second, $*ṭ$, is once again the lenition of an earlier $**t$ (again corresponding to [t] in e.g. Wolof, Bijogo, and Proto-Bantu). As with $*t(\text{ʃ})$, it is difficult to know its pronunciation in Proto-Bak. It is likely that at least one of these sounds was a voiceless rhotic/retroflex obstruent, and it is possible that one of them was [(t)ʃ].

Throughout Atlantic, the most common origin for “ $ṭ$ ” is the lenition of an earlier [t]. In addition to the proto-sounds discussed above, recall that this sound in modern Joola languages always comes from Proto-Joola $*t$, and Manjak $*ṭ$ comes from Proto-Bak $*t$ (equivalent to Proto-Joola $*t$). This lenition change must have operated independently many times in many different languages. For Bijogo [t~tʃ̣] (as well as the corresponding Proto-Bak sound $*t(\text{ʃ})$), there is no evidence of lenition from an earlier [t]. This sound corresponds to [s] or [ʃ] in other Atlantic groups, and to Proto-Bantu $*c$. I do not currently have any hypothesis about the origin of “ $ṭ$ ” in Temne or Limba.

5. Implications

§2 established that voiceless rhotic/retroflex obstruents are rare sounds in Africa, and especially in Atlantic-Congo. Nonetheless, such sounds are found in a number of modern Atlantic languages which are not closely related. Not only are these sounds found in modern languages, but they can be reconstructed for earlier stages of a number of Atlantic languages. In general, only a single such sound is found in any given language, with Mankanya being the only modern language with two. No language has a corresponding voiced retroflex obstruent. Based on the development of the reconstructed sounds, as well as the sound correspondences between Manjak cluster varieties, it is clear that “ $ṭ$ ” is not a particularly stable sound, and tends to develop to other, more

cross-linguistically common sounds like [s, r, t, h]. The sound “ $\{t/\}^{\text{r}}$ ” is thus a globally rare sound which is prone to loss, but is nonetheless persistent within a specific linguistic area (Atlantic) throughout time. For convenience, I will refer to such a sound as a PERSISTENT RARE SOUND. This phenomenon is encountered elsewhere in the world, and supports the idea that speakers of one language are influenced by the sounds of other languages spoken nearby, especially in situations of multilingualism. As such, when a sound exists within a linguistic area in at least one language, a sound change that gives rise to that same sound in another language is much more likely than if no nearby language contains the sound. Presumably, the more languages in the area contain the sound (and the more widely spoken those languages are), the more likely the sound is to arise in a new language. An important requirement for the existence of a persistent rare sound in an area is that there be a sufficient number of languages spoken, such that despite the sound’s tendency to be lost, at least a few languages will use the sound at any one point in time. If only a few languages are spoken in the area, a time will inevitably arise when the sound is absent in all languages, and it will no longer be available as a target of sound change. At this point the sound is unlikely to arise again, being subject to the same restrictions that make it rare globally.

I will present two other examples of persistent rare sounds in Niger-Congo to exemplify the phenomenon. The first example is rather small-scale, and in fact is found in the Atlantic area: [ɬ] in Joola and Manjak varieties. The sound [ɬ] is overall rare in the world, being found in only 5% of languages in PHOIBLE’s sample. In Africa it is common in Chadic languages and some S zone Bantu languages, but is otherwise rare. It is thus notable that [ɬ] is found in a number of languages in a small area near the coast in southern Senegal and northern Guinea Bissau. These are all Joola or Manjak cluster varieties. The Joola varieties with [ɬ] are at least: Kasa (all dialects), Huluf, Husuy, Bliss, Keeraak (aka Her), at least some Fulup dialects, Buluf, at least one Fonyi dialect (documented in Weiss 1940, so perhaps now > [l]), and all Bayot varieties. Within the Manjak cluster, at least Pepel and Pexixe

(*pə-lil*) have the sound. While these are all languages of the Bak group, [ɬ] does not have a shared origin across all languages. Proto-Joola *ɬ is the source of [ɬ] in most Joola varieties, but Bayot Gubaare has additional tokens of [ɬ] from Proto-Joola *t. For [ɬ] in the Manjak cluster I assume a value of *[ɬ] for Proto-Manjak, which becomes [s, l, ʂ] in other Manjak cluster varieties. Alternately, the sound was *[ʂ] or *[s], and developed to [ɬ] at Doneux's *protocôtier* node. Crucially, *ɬ in Proto-Joola and Proto-Manjak do not correspond. Proto-Joola *ɬ corresponds to Proto-Manjak *t, and Proto-Manjak *ɬ corresponds to Proto-Joola *l. As such, the sound cannot have a shared origin in both languages. Rather, one language must have originally innovated the sound, and the other developed the same sound from a different source under areal pressure. Including Gubaare [ɬ] from *t, there are three "independent" origins for the sound in the area.

A wider-scale example involves fricative vowels in Bantoid languages of Cameroon. Fricative vowels are vowels produced with a very close, fricative-like constriction, often resembling a syllabic [v] or [z]. They are represented with various phonetic symbols in different sources, e.g. [ɺ, z̥, v̥]. Fricative vowels are rare globally, but occur for example in some Ryukyuan languages like Miyako, and Wu Chinese varieties. Faytak (2015) is the most comprehensive account of fricative vowels in Cameroon. He finds that the presence of fricative vowels is an areal feature centered around the Grassfields area. These sounds appear in multiple Bantoid subgroups (Mambiloid, Beboid, Ekoid-Mbe, multiple Grassfields subgroups, and A zone languages within Bantu proper), and geography is a much better predictor of whether a language will contain these sounds than genetic affiliation. Furthermore, fricative vowels in the area cannot be traced to a single source in a common ancestor, but rather continually arise in different languages from the raising of high vowels, often due to a chain shift involving the raising of upper mid or lower high vowels. These fricative vowels are subject to rapid loss and innovation, and in some languages like Kom, Faytak proposes that they arose very recently. The presence of fricative vowels in this area is important given that Proto-Bantu is generally agreed

to have been spoken there (or very nearby), and likely contained two fricative vowels (often called “super-high”): *j and *ɥ. These phonetic values are notably argued for by Maddieson (2003: 20) and Connell (2007). Other than a few A zone Bantu languages spoken in the Grassfields area, no modern Bantu language contains fricative vowels, but in many daughter languages the Proto-Bantu vowels had very pronounced effects on preceding consonants which are characteristic of the effects of attested fricative vowels. In summary, fricative vowels in Cameroon have the following properties: they are found in an area with a high density of languages, they have seemingly existed in the area for a long time, they are subject to rapid loss and renewal, and when languages move out of the area they inevitably lose the sounds.

Returning to “t̪ʰ” in Atlantic languages, this sound (or range or sounds) fits the profile of a persistent rare sound. It is cross-linguistically rare, and is subject to rapid loss and renewal, but due to the density of languages in the Atlantic area it is always present in at least a handful of languages at any point in time. Of course, there must be some point in the past when the sound first arose in the area, but this ultimate origin is perhaps impossible to identify. An important implication is that any protolanguage spoken in the Atlantic area would have a heightened chance of containing a sound “t̪ʰ”, just as the existence of fricative vowels in modern languages of the Grassfields area aligns with the reconstruction of fricative vowels to Proto-Bantu. This tendency would in theory apply even to the latest common ancestor of all Atlantic languages, given that this protolanguage was most likely spoken somewhere in the area where modern Atlantic languages are spoken. I know of two proposals which reconstruct a voiceless retroflex obstruent to such a protolanguage. Pozdniakov & Segerer (2017) use the term “Atlantic” for a subgroup containing all of the languages in Table 2 except Sua, Gola, Limba, and Mel, and reconstruct a segment *t̪ to their proposed Proto-Atlantic language. I myself do not believe that the (Northern) Atlantic languages form a valid subgroup within Niger-Congo, but rather represent multiple primary branches of the family. As such, I argue in Merrill (2021) that Proto-Niger-Congo itself was

spoken in the Atlantic area, and reconstruct a retroflex stop/affricate *ʈ (the source of Bijogo /t/ and the corresponding sounds in other groups mentioned at the end of §4). While these two proposals are not compatible with each other, both are consistent with the idea that voiceless rhotic/retroflex sounds have been common in the Atlantic area for some time.

6. Conclusion

Voiceless rhotic/retroflex obstruents are rare sounds in Africa, but are found in ten or so modern Atlantic languages, and can be reconstructed for multiple Atlantic protolanguages. The sound in each language does not have a shared origin, but rather is in most cases the result of a recurrent sound change affecting an earlier [t]. The presence of a voiceless rhotic/retroflex sound — and without a voiced counterpart — is thus an areal feature of Atlantic languages. This sound tends to be unstable, often developing to more common sounds like [s, r, h, t]. Nonetheless, due to the high density of languages in the area, the sound is able to be spread to new languages just as quickly as it disappears in others, and as such has survived over a seemingly lengthy span of time as an areal feature. I term such globally rare sounds which survive throughout time in a linguistic area *persistent rare sounds*. Drawing a parallel with Bantoid and Proto-Bantu fricative vowels — another example of persistent rare sounds — I argue that the common ancestor of the Atlantic languages quite possibly contained a sound “[ʈ],” as it was likely spoken in the Atlantic area. There is currently no consensus on what this latest common ancestor is, but whether Proto-Atlantic, Proto-Niger-Congo, or anything else, it is much more likely to have contained a voiceless rhotic/retroflex obstruent than would be expected based on the global prevalence of these sounds.

References

- Amayo, Moses Airen. 1976. *A generative phonology of Edo (Bini)*. Ibadan: University of Ibadan. (Doctoral dissertation.)

- Anonymous. 2010a. *Primeira edição do livro de Lucas na lingua Jola-Felupe de Suzana*. Dakar: Igreja Evangelica Projeto de Tradução da Guiné-Bissau.
- Anonymous. 2010b. *Primeira edição do livro de Lucas na lingua Jola-Bayote, dialeto de Kulaje*. Dakar: Igreja Evangelica Projeto de Tradução da Guiné-Bissau.
- Anonymous. 2014. *Ulibra wi Naşibaŋi* (Mankanya Bible). Wycliffe Bible Translators (text), Hosanna (audio). live.bible.is/bible/KNFWBT (Online resource, accessed 2022-01-22.)
- Barry, Abdoulaye. 1987. *The Joola Languages: Subgrouping and reconstruction*. London: University of London. (Doctoral dissertation.)
- Bassène, Alain-Christian. 2007. *Morphosyntaxe du Jóola Banjal: Langue atlantique du Sénégal*. Cologne: Köppe.
- Bassene, Mamadou. 2012. *Morphophonology of Joola Eegimaa*. Minneapolis: University of Minnesota. (Doctoral dissertation.)
- Carlton, Elizabeth M. & Rand, Sharon R.. 1994. Compilation de listes de mots Swadesh modifiées recueillis parmi les langues du groupe Diola, sud du Sénégal. *Journal of West African languages* 24. 85–119.
- Clarke, Mary Lane. 1922. *A Limba-English dictionary*. New York: Houghton.
- Connell, Bruce A. 2007. Mambila fricative vowels and Bantu spirantisation. *Africana linguistica* 13. 7–32.
- Diagne, Mbacké. 2009. *Bref aperçu grammatical et lexical du Bayot-Kugere*. Dakar: Éditions du livre universel.
- Doneux, Jean Léonce. 1975. *Lexique manjaku*. Dakar: Centre de Linguistique Appliquée de Dakar.
- Elugbe, Ben Ohiomamhe. 1989. *Comparative Edoid: Phonology and lexicon*. Port-Harcourt: University of Port Harcourt Press.
- Faytak, Matthew. 2015. High vowel fricativization as an areal feature of the northern Cameroon Grassfields. A paper presented at World Congress of African Linguistics 8, Kyoto.
- Gaved, Timothy. 2020. *A grammar of Mankanya*. Leiden: Leiden University. (Doctoral dissertation.)
- Global Recordings Network. 1961–2006. Jola-Felupe language. Online resource, accessed Jan. 22, 2022. <<https://globalrecordings.net/en/language/eja>>

- Güldemann, Tom. 2018. Historical linguistics and genealogical language classification in Africa. In Tom Güldemann (ed.), *African Languages and Linguistics*, 58-444. Berlin: DeGruyter Mouton.
- Hamann, Silke Renate. 2003. *The phonetics and phonology of retroflexes*. Doctoral dissertation, Utrecht University.
- Hammarström, Harald & Forkel, Robert & Haspelmath, Martin & Bank, Sebastian. 2021. *Glottolog* 4.5. Leipzig: Max Planck Institute for Evolutionary Anthropology. Online resource, accessed Jan. 22, 2022. <<https://glottolog.org/>>
- Jesus Film Project. 1995–2022. *Jesus* (film). Manjaku: <https://www.jesusfilm.org/watch/jesus.html/manjaku.html>; Papel: <https://www.jesusfilm.org/watch/jesus.html/papel.html>. Online resource, accessed 2022-01-22.
- Karlik, Jan. 1972. *A Manjako grammar with special reference to the nominal group*. London: University of London. (Doctoral dissertation.)
- Ladefoged, Peter & Maddieson, Ian. 1996. *The sounds of the World's languages*. Oxford: Blackwell.
- Maddieson, Ian. 2003. The sounds of the Bantu Languages. In Nurse, Derek & Philippson, Gérard (eds.), *The Bantu languages*, 15–41. London: Routledge.
- Merrill, John. 2018. *The historical origin of consonant mutation in the Atlantic languages*. Berkeley: University of California. (Doctoral dissertation.)
- Merrill, John. 2021. Atlantic groups as primary branches of Niger-Congo. A paper presented at “Westermann Workshop” in Humboldt University, Berlin, Nov. 6, 2021.
- Ndao, Dame. 2013. *Phonologie, morphologie et structures syntaxiques du Pepel*. Saarbrücken: Éditions universitaires européennes.
- Moran, Steven & McCloy, Daniel (eds.). 2019. *PHOIBLE* 2.0. Jena: Max Planck Institute for the Science of Human History. Online resource <https://phoible.org/>, accessed 2022-01-22.
- Pozdniakov, Konstantin & Segerer, Guillaume. 2017. A Genealogical classification of Atlantic languages. In Lüpke, Friederike (ed.), *The Oxford guide to the Atlantic languages of West Africa*. Oxford: Oxford University Press. (Draft.)

- Sagna, Serge. 2008. *Formal and semantic properties of the Gújjolaay Eegimaa (A.k.a Banjal) nominal classification system*. London: University of London. (Doctoral dissertation.)
- Sapir, J. David. 1971. West Atlantic: An inventory of the languages, their noun class systems and consonant alternations. *Current trends in linguistics* 7. 45–112.
- Schlenker, C. F. 1864. *Grammar of the Temne language*. London: Church Missionary Society.
- Segerer, Guillaume. 2002. *La langue Bijogo de Bubaque (Guinée-Bissau)*. Paris: Peeters Louvain.
- Thomas, Northcote W. 1916a. *Timne grammar and stories*. London: Harrison and Sons. (Anthropological Report of Sierra Leone, III.)
- Thomas, Northcote W. 1916b. *Specimens of languages from Sierra Leone*. London: Harrison & Sons.
- Trifkovič, Mirjana. 1969. *Le mancagne: Étude phonologique et morphologique*. Dakar: Institut Fondamental d’Afrique Noire, Université de Dakar.
- Wilson, William André Auquier. 1962. Temne, Landuma and the Baga languages. *Sierra Leone language review* 1. 27–38.
- Wilson, William André Auquier. 1990. *An outline of the Bijagó language*. Ms. 42pp.
- Wilson, William André Auquier. 2007. *Guinea Languages of the Atlantic group: Description and internal classification*. Frankfurt am Main: Peter Lang.
- Yillah, M. Sorie. 1992. *Temne phonology and morphology*. Ann Arbor: UMI. (Doctoral dissertation, City University of New York.)