#06
décembre 2020

LLA

Linguistique et Langues Africaines

9 782359 353259

ISSN 2429-2230
ISBN 978-2-35935-325-9 **20 €**

Lambert-Lucas

LLA

Linguistique
et Langues
Africaines

#06

décembre 2020

Comptes rendus / Book reviews

En couverture : Préparation du maïs concassé (Salineiro, Île de Santiago, Cap-Vert, *ca* 2005), photo Nicolas Quint

Design : Jean-Irénée Cuin

Production - diffusion : Éditions Lambert-Lucas

LLA

Linguistique et Langues Africaines

International Journal edited by
Revue internationale éditée par le
Llacan
(UMR 8135 CNRS / Inalco)

 # 6 - 2020

Diffusion

Éditions Lambert-Lucas
4 rue d'Isly
F-87000 LIMOGES
Tél. (+33) (0)5 55 77 12 36
(+33) (0)6 07 41 04 25
(+33) (0)6 88 29 04 14

Revue en accès libre

http://www.lambert-lucas.com/collection/linguistique-et-langues-africaines-revue-du-llacan

Commande d'exemplaires

contact@lambert-lucas.com

ISSN 2429-2230
ISBN 978-2-35935-325-9

Contact
llafrique@cnrs.fr

Envoi d'ouvrages pour compte rendu
Revue LLA
CNRS/Llacan BP 8
7 rue Guy-Môquet
F-94801 VILLEJUIF CEDEX

Site web
http://llacan.vjf.cnrs.fr/lla

Consignes de soumission
http://llacan.vjf.cnrs.fr/lla/apropos.html

Sommaire / Contents

Comptes rendus / Book reviews

1

Proto-Berber phonological reconstruction: An update

Maarten Kossmann
Universiteit Leiden

Abstract

Over the last decades, our insights in the phonological history of Berber and the reconstruction of its earlier stages greatly evolved. This is thanks to an emergent discussion and to new data on a number of languages that are crucial to reconstructing Proto-Berber, most importantly the works by Catherine Taine-Cheikh on Zenaga. In this article, I will provide an overview of the results and challenges in the reconstruction of Proto-Berber phonology.

Keywords

Afroasiatic, Amazigh, Berber, phonological reconstruction, Proto-Berber

Résumé

Au cours de ces dernières décennies, les études consacrées à la diachronie de la phonologie des langues berbères ont fait de nets progrès et ce pour deux raisons principales : (i) le développement de débats fructueux et constructifs sur la question et (ii) la publication de nouvelles données sur des langues revêtant une importance cruciale dans la reconstruction du proto-berbère, au premier rang desquelles le zénaga de Mauritanie, illustré notamment par les travaux de Catherine Taine-Cheikh. Dans cet article, je propose une synthèse des résultats déjà obtenus et fais le point sur les défis qui restent à relever pour mieux rendre compte de la phonologie reconstituée du proto-berbère.

Mots clés

afro-asiatique, amazigh, berbère, phonologie en diachronie, proto-berbère

1. Introduction

Historical reconstruction of Berber phonology[1] is still a niche activity in Berber studies. Emphasis has always been on the synchronic description of the different varieties, and on the comparison of their structures. While the latter often has a historical angle (as witnessed, among many others, in Galand 2010), it hardly ever takes the point of view of classical comparative-historical studies. As a result, the major method for reconstructing earlier stages of languages, viz. the systematic comparison of word forms in different languages according to the neogrammarian principles, has hardly any tradition in the field of Berber studies. The reasons behind this may partly be historical and ideological; one may think of the "societal relevance" of language description, rather than comparison, to colonial administration, and the post-independence ideology that all Berber varieties are in principle the same. However, an important reason also lies in the scientific positions taken by its most influential researchers. Especially the late Lionel Galand, while deeply interested in historical linguistics, always remained skeptical to the concept of proto-languages, and thus to the neogrammarian model and its methods. As a result, the systematic reconstruction of Berber phonology never really took off in the main center of Berber studies, Paris, thereby putting Berber in an entirely different situation than other Afroasiatic language groups, such as Semitic, Chadic, and Cushitic. In fact, one finds oneself in the paradoxical situation that one of the lexically and grammatically best described language groups in Africa lacks a serious tradition of historical phonology.

This is not to say that phonological reconstruction has been entirely absent from the field of Berber studies. This is to a large extent due to the work undertaken by one single scholar, Karl-G. Prasse, who, from the late 1950s onwards, undertook a massive endeavor of Berber reconstruction, based on Tuareg. His *A propos de l'origine de* h *touareg (tahaggart)* (1969) provides us with an analysis of several of the main problems in Berber historical phonology. His later *magnum opus*, the *Manuel de grammaire*

1. I would like to thank Marijn van Putten and Evgenia Gutova, as well as two anonymous referees, for their comments on an earlier draft of this article. Of course, all responsibility for errors and flaws in the argument remains with the author.

touarègue (tăhăggart) (1972-1974) is a bit different, as it is a historical reconstruction couched in the form of a synchronic description. While Prasse made use of Berber data from outside Tuareg, this was not the main focus of his study, and one remarks, for example, that he never analyzed the reflexes of his reconstructed laryngeals in northern Berber.

Thirty years later, I tried to fill in some of these gaps in my *Essai sur la phonologie du proto-berbère* (Kossmann 1999). This work presents full sets of cognates in all Berber languages concerning a number of phonological problems and proposes some reconstructions. Around the same time, a number of other scholars started to work on the same subject, esp. Catherine Taine-Cheikh (cf. already Taine-Cheikh 1999), Marijn van Putten, Lameen Souag, and Cécile Lux. While phonological reconstruction is still relatively marginal in Berber studies, it seems that there is at least something akin to scientific discussion going on nowadays.

Over the last twenty years, our knowledge of Berber has greatly expanded, especially because of the description of a number of hitherto largely unknown varieties. This includes varieties from Morocco (Ghomara, Mourigh 2015), Libya (Awjila, van Putten 2014a, a sophisticated reanalysis of earlier materials; Zwara, Mitchell 2009, the edition of a manuscript from the 1950s), and Egypt (Siwa, Naumann 2012; Souag 2013; Schiattarella 2016). Most importantly, two highly original Berber languages have now received extensive documentation: Zenaga in Mauritania, due to the long series of publications by Catherine Taine-Cheikh, and the closely related Tetserret in Niger as described in Lux (2013). Zenaga has proven to be key in our understanding of proto-Berber (Prasse 2011). As a result, our ideas of proto-Berber phonology have changed considerably since the turn of the millennium. Therefore, I think it is useful to provide the reader with an update on the problems and results of Berber historical phonology.

In this article, I will take the basic methods and historical assumptions of Kossmann (1999) as my point of departure. This means, among others, that evidence from Guanche—the extinct language of the Canary Islands—and the language of the antique Libyco-Berber inscriptions fall outside of its scope. I think that the state of our knowledge of these extinct languages makes it more profitable to use proto-Berber reconstructions for their analysis than the other way around. Similarly, no effort will be made to introduce or evaluate the analysis within the larger framework of Afroasiatic reconstructions. While I will try to give due reference to reconstructions made on the basis of Berber-internal data, I will remain

silent on those reconstructions that are mainly based on comparisons with other Afroasiatic languages.[2] In my view, the time span between a tentatively reconstructed Afroasiatic proto-language and Proto-Berber is too large to allow for a top-down approach, which simply feeds Afroasiatic reconstructions and comparisons into Berber. The term Proto-Berber will be used in the sense of the relatively unitary stage from which the modern dialectal diversity of Berber developed. There are many caveats to this position (see Galand 2010: 14 for a critique; and Múrcia Sànchez 2011: II/359-360 for an alternative scenario), but I doubt that the problems encountered in Berber are much bigger than, for example, those in Proto-Semitic or Proto-Germanic.

Berber languages have undergone enormous influence from other languages. Among the languages for which this influence can be traced, Punic, Latin, and Arabic have had impact on virtually all of Berber. Punic and Latin loans are in most cases phonologically and morphologically indistinguishable from Berber etyma. They can only be traced using our knowledge of these languages and studying the semantic fields to which they belong. Most Arabic loans, on the other hand, are easily set apart on phonological and morphological grounds. There is, however, a small stock of early loans, mainly expressing basic Islamic concepts, that is indistinguishable from native Berber forms on formal grounds. For more information on loanwords in Berber, including references to the large amount of literature on this subject, one may consult Kossmann (2013).

2. Labial consonants

Proto-Berber probably only had two oral labial consonants, which will be reconstructed here as *β and *f.[3] The reconstruction of *f is

2. One of the anonymous referees points to the fact that all researchers that have worked on comparative Berber also have knowledge of at least some other Afroasiatic languages and reconstructions. No doubt, being informed about other branches influences the directions and decisions taken, consciously or unconsciously. While acknowledging this influence, I still think there is an important distinction between analyses based on data from Berber only (however biased the analysis may be) and those explicitly using other language families in their argumentation.

3. Standard IPA symbols will be used in the transcriptions, with a number of exceptions: *š* = IPA [ʃ]; *y* = IPA [j]; *ž* = IPA [ʒ]; *ǧ* = IPA [ʤ]; *č* = IPA [ʧ]. Signs with a dot underneath designate pharyngealized consonants (e.g. *ḍ* = [dˁ]). Length is expressed by doubling the symbol. For Zenaga, the broad phonetic transcriptions by Catherine Taine-Cheikh have been maintained with vowels but adapted for consonants. The short (or central) vowels of Ghadames and Tuareg are marked by means of the *breve* sign. Of course, the phonetics of reconstructed consonants are difficult to establish, and, in spite of the use of IPA symbols, they should be taken as approximations.

entirely unproblematic: cognates with f in all word positions abound in modern Berber languages. All modern Berber languages have the same pronunciation, a voiceless labiodental fricative. The geminate counterpart of *f is *ff (for a large number of words including *f, see Naït-Zerrad 2002).

The situation with the consonant represented here as *β is different, and it has been the subject of one of the few longstanding debates in Berber historical phonology. While Beguinot (1924) and Kossmann (1999) propose a reconstruction as a bilabial consonant, Prasse (1969; 2011) and Acosta Armas (2017) prefer a reconstruction as a glottal fricative *h*.

*β is pronounced as a labial fricative in a few Libyan varieties: Ghadames [β], Awjila [v]. In these varieties, it occurs in all positions of the word.

In the other varieties, *β has different reflexes depending on its place in the word. When followed immediately by a consonant, the reflex of *β is regularly *b*. Elsewhere, a weakened reflex is found, which, depending on the language and the context, is *h* (in Tuareg), *ww ~ gg*^{w} (intervocalically in Tashelhiyt and Central Moroccan Berber) or the complete loss of the consonant (elsewhere) (see Kossmann 1999: 108-109 for details). In Zenaga, *β is lost when not immediately followed by a consonant, but its previous presence is still clear from vowel lengthening (Kossmann 2001a). In pre-consonantal position, it has become *w*, cf. Zenaga Aorist *äwðər* 'to mention' as compared to Ouargla *əbder* 'id.' and Ghadames *ăβðər* 'id.'. There are, however, a number of unexplained cases where Zenaga has *b*, e.g. *oʔbih* 'smoke', comparable to forms in other languages that point to *β, such as Ghadames *oβu* (for more cognates, see Kossmann 1999: 100).

While the complete loss of *β is widespread when not immediately followed by a consonant, it has had different vocalization effects according to the dialect. This shows that the loss of *β happened independently in these varieties.

The identification of pre-consonantal *b* in modern Berber with the other reflexes of *β (such as Tuareg *h*) is based on a number of observations. One is of a structural type: The non-*b* reflexes of *β seem to be in complementary distribution to *b*. Another structural argument is the fact that non-labial reflexes of *β still trigger labial dissimilation of *m* into *n* in some prefixes (for details, see Kossmann 1999: 131). A further argument is of a comparative nature: in Ghadames and Awjila, the same reflex is found for both *b* and non-*b* reflexes elsewhere. Finally, language-internal paradigmatic variation between *b* and non-*b* reflexes of *β show that they were originally one

single segment. There are abundant examples of pre-consonantal *b* varying with non-*b* reflexes in intervocalic position, e.g. the Aorist-Imperfective pair in Tashelhiyt Aorist *bzg*, Imperfective *azzg* (from *əβăzzăg) 'to be wet' (for a recent addition to our data concerning this variation, see Mourigh 2015: 152 on Ghomara).

An interesting detail in the history of *β is that it has assimilated to *f* in most varieties when immediately preceding a voiceless consonant. The assimilatory nature of these forms is also shown by variation between preconsonantal *f* and non-*b* reflexes in intervocalic position, e.g. Tashelhiyt Aorist *fk* (< *ăβkʸəʔ), Imperfective *akka* (< *əβăkkʸăʔ) 'to give'.

There is little, if any, evidence for a geminate counterpart to *β. In fact, in cases where one would morphologically expect to have geminated *β, this seems to have been prevented by means of metathesis. Thus modern instantiations of *β such as Mali Tuareg Aorist *ălh*, Imperfective *hall* 'to weep' (Heath 2006: 206) are most easily understood as coming from *ălβəʔ – *əβăllăʔ, in which the Imperfective form has metathesized *β.

In Kossmann (1999), some evidence was provided that would point to a third bilabial consonant *b. This comes from a small number of instances of *b* (instead of β) in Ghadames, and from a small number of Berber words that have a reasonably wide distribution and *b* in intervocalic or word-final position. At present, I think that the evidence for this reconstruction is not sufficient for proposing a separate proto-phoneme. The few unexpected Ghadames forms may well be inter-dialectal borrowings (cf. Souag 2017), while at least some of the more wide-spread forms with *b* may in fact be *Wanderwörter*, i.e. words with a non-Berber origin that spread over the Berber-speaking territory in post-proto-Berber times. I assume this could be the case of the terms for faba bean (*abaw* and variants) and for the term for pigeon (*edăber*, *itbir*, *adbir*…).

The original pronunciation of *β is difficult to pin down. The fact that *β became *f* when followed by a voiceless consonant is good evidence that it was originally a voiced consonant, as otherwise the assimilation would make no sense. It is more difficult to decide whether it was a fricative or a stop in proto-Berber times. On the one hand, the evidence from Ghadames and Awjila suggests a fricative pronunciation, a reconstruction which would concur well with the large-scale weakening of the consonant in the other varieties. On the other hand, one could also envisage a reconstruction where the proto-Berber pronunciation was *[b], a pronunciation preserved in most languages when followed by a consonant. In this case, one would have to assume that all languages underwent further weakening, including

Ghadames, Awjila and Zenaga, which would have carried the weakening over to all positions of the word. Of course it would also be possible to take an intermediate position, assuming that *β had two allophones, *[b] and *[β] depending on its position in the word, and that in Ghadames, Awjila and Zenaga the continuant pronunciation spread to all positions.

While the correspondences in the Berber languages make both *[β] and *[b] (or a combination) plausible reconstructions, there is some evidence that rather points to a fricative realization. In the first place, the voice assimilation *β > *f* is more easily understood if *β was a fricative (assuming that *f* was a fricative at that time, see below). In the second place, the apparent impossibility of having geminated *ββ makes much more sense if the consonant were quite weak (like a fricative or an approximant) than if it had been a stop. As shown below, restrictions on gemination seem to have applied to a number of other consonants, too, viz. *ʔ* and possibly *h, w, y*.

3. Dental/Alveolar stops

The system of dental/alveolar stops consists of three elements, for which I will use the symbols *d, *t and *ḍ.

The reconstruction of *d is straightforward. If one abstracts away from occasional assimilations, the consonant is well preserved in all Berber languages as a voiced consonant. In languages which have undergone the large-scale weakening process called spirantization in Berber studies, it is an interdental fricative, while in the other languages it is a dental or alveolar stop. Its geminate counterpart is *dd* everywhere. For cognates including *d, see Naït-Zerrad (1999).

The situation with *ḍ is slightly more complicated. Like with *d, *ḍ seems to have been preserved in all contexts in all Berber languages. However, there exists a rather erratically distributed dialectal variation between *ḍ* and *ṭ* for this consonant (*ṭ* is found in Ghomara, Dadès, Ayt Warayn in Morocco; in eastern Kabylia in Algeria, in Jebel Nefusa and Awjila in Libya, and in Siwa in Egypt), which makes it hard to decide which pronunciation should be considered original. In languages with spirantization and a voiced reflex of *ḍ, *ḍ* is pronounced as a pharyngealized interdental fricative, while in those spirantizing varieties that have a voiceless reflex of *ḍ, *ṭ* is never spirantized. Its geminate counterpart is *ṭṭ* virtually everywhere; in a few languages where it is normally *ḍḍ* (e.g. Ghadames), this is an analogical regularization based on the non-geminated pronunciation, as shown by the existence of some exceptional forms with *ṭṭ*. An important exception to this is Zenaga, which has *ḍḍ* throughout (Taine-Cheikh 2001-2003: 20ff.). It is therefore uncertain if the

voiceless pronunciation *ṭṭ* of the geminate really goes back all the way to Proto-Berber.[4]

Depending on whether one has to do with a spirantizing or a non-spirantizing variety, the reflex of *t is either a voiceless interdental fricative (which in a few languages can be weakened further to *h* or even be lost, e.g. Chaouia) or a dental or alveolar stop (which can have assibilated realizations, e.g. Figuig). Its geminate counterpart is *tt* (which, depending on the variety, may be assibilated). In contrast to the other two consonants discussed in this paragraph, *t presents a number of intricate problems (see already Marcy 1936: 51).

In Berber, there exists an interesting variation between forms with and without *t*. There are two main well-attested and highly salient contexts where this is found.

In the first place, there exists variation between *t* and *e* in verbs ending in *t*. This is a very common situation in Tuareg, which has a word-building suffix *-ăt* of unclear semantics (an "augment" in the terminology of Heath 2005), which becomes *e* when followed by a suffix, e.g. *i-ffurr<u>ăt</u>* 'he flew away', *əffurr<u>ă</u>-ɣ* 'I flew away' (Heath 2005: 295). The "augment" is also found in some residual forms in other Berber languages, but it does not have the same allomorphy there as found in Tuareg. In addition, there are a couple of short verbs ending in *t* that have similar variation between *t* and *e* or ø, esp. **ămmət* 'to die' and **ăwət* 'to hit'. In Tuareg, as well as in a few other languages, one finds forms such as the following. (Perfective forms; examples from Prasse *et al.* 2003: 563 and Kossmann 1997: 144):

Table 1 — Paradigmatic variation between t and zero in the verb 'to die'

	Tuareg (Ayer)	Figuig	
3SG:M	*y-əṃṃut*	*i-mmut*	'he died'
3PL:M	*ămṃe-n*	*mma-n*	'they died'

Elsewhere, the forms with final *t* are found all over the paradigm, probably representing paradigmatic leveling by analogy; one may note that in Zenaga the forms without *t* have been generalized (cf. Kossmann 2018). While one could assume that the "augment" has some intricate morphological history that would explain the presence of the variation in a non-phonological way, this would not work with the verb 'to die', as the widely attested nominal derivation **ta-măttan-t* 'death' shows that *t* is part of the stem.

In the second place, many Berber languages show variation between *t* and **e* in Direct Object pronouns (cf. Marcy 1936; Brugnatelli 1993; Kossmann 1997b). Thus, in Mali Tuareg one has two major

4. I owe this point to one of the anonymous referees.

allomorphs of the third person Direct Object clitics, one showing *t(ă)*, the other showing *e* (Heath 2005: 604-605), see Table 2.

Table 2 — Two series of Direct Object pronouns in Mali Tuareg

	I	II
3SG:M	*t(t)*	*e*
3SG:F	*tăt*	*et*

Similar forms are found all over Berber, and it seems reasonable to consider the variation a feature of Proto-Berber. It should be noted that Kossmann (1997b) proposes to reconstruct two entirely different sets of third person direct object clitics, which would not be related etymologically. In view of the wide-spread variation between *t* and *e* in the final-*t* verbs, I think a phonological explanation is to be preferred. It is, however, far from clear what the exact conditioning of the phonological change would have been.

One also wonders whether the frequent dropping of the verbal person prefix *t-* (2nd persons and 3SG:F) found in many Tuareg dialects is somehow related to the variations described above.

The fact that different cases of *t* – *e* variation were found, which do not seem to have any morphological connection, strongly suggest that they reflect a phonological process, presumably one where **t* was weakened under some specific conditions. Unfortunately, it is difficult to specify these phonetic conditions. The contexts where the *t* and *e* forms of the Direct Object clitics are used are different from language to language, and it is therefore difficult, if not impossible, to decide in which context the *t*-less forms were originally used. The only condition that is relatively clear is that concerning the *t*-final verbs, where word-final *t* is preserved, while intervocalic *t* is dropped.

It should be noted that positing a (somehow conditioned) loss of *t* in a very ancient stage of Berber would open up one more highly interesting etymological possibility (see Vycichl 1992: 259). It might be possible to analyze the nominal suffix of the feminine plural *-en* as being composed of the feminine suffix *t* followed by the plural suffix *-ăn* (Table 3).

Table 3 — A very tentative scenario for the development of the nominal plural suffixes

	old	new
m:sg	ø	ø
f:sg	*-t*	*-t*
m:pl	*-ăn*	*-ăn*
f:pl	*-t-ăn*	*-en*

All in all, the situation with **t* is complex as at present no clear conditioning for its variation with *ø* and *e* has been defined. There is little reason to consider the sound change as post-dating Proto-Berber. If the (quite adventurous) idea that the nominal suffix F:PL *-en* goes back to *t-ăn is correct, one can safely assume that the sound change had already taken place by Proto-Berber times, as *-en* and its cognates are attested in all Berber languages without exception.

4. Sibilants

In modern Berber languages, there are two sets of sibilants: *z*, *s*, *ẓ* and *ž*, *š*. As shown in Kossmann (1999), there is very little evidence that would suggest that this division goes back to Proto-Berber times (see already Basset 1952: 6 for a similar observation). In fact, with the exception of some dissimilatory occurrences of *ž* (Kossmann 1999: 228), there are hardly any forms with non-geminate *ž* and *š* that have sufficient attestation to reconstructed with confidence. The few cases of widespread *žž* and *šš* may go back to *zy and *sy, respectively (see Kossmann 1999: 229 for some details). As a result, we can only reconstruct one single set of sibilants, which will be represented here as *z, *s, *ẓ.

The original pronunciation of these sibilants is difficult to establish. The large majority of Berber languages have alveolar pronunciations ("sifflantes" in French linguistics). However, in Awjila, Tetserret and Zenaga different reflexes are found. As these languages all present interesting archaisms (as well as innovations) that set them apart from other Berber languages, this fact should be taken very seriously. In Awjila and Zenaga, there is quite some variation as to the reflexes of *s and *z. In Zenaga, two phonemes correspond to *z: one which is mostly pronounced *z̄* [5] (sometimes *ž*), and one which is mostly pronounced *θ* (sometimes *z*) (Taine-Cheikh 2001-2003: 23-28).[6] As far as I can see, there is no congruence in the distribution of the variant reflexes in Zenaga, and Awjila, and, for the time being, I consider them as unexplained rather than considering them as evidence for the existence of two sets of Proto-Berber sibilants. Table 4 presents the most common reflexes of the sibilants in these two languages (cf. Taine-Cheikh 2001-2003; 2008: lxxiii; Lux 2013: 132-136; van Putten 2014a; 2014b).

---

5. *z̄* stands for a non-strident pronunciation of *ž*.

6. Taine-Cheikh (2001-2003) considers *θ/z* the regular counterpart of geminate *zz*, and *z̄ž* the counterpart of *žž*. While this makes perfect sense in a synchronic framework, I do not see much evidence that *žž* corresponds to *žž* or *zz* elsewhere in Berber. Most examples of *žž* seem to reflect or contain attested or reconstructible *y(y)* in the other Berber languages.

Table 4 — Common reflexes of the sibilants

	Zenaga	Tetserret	Awjila
*s	š	š	š ~ s
*z	z̄/ž ~ θ/z	ž	ž ~ z
*ẓ	θ̣/ẓ[7]	ṣ	ẓ
*ss	ss	ss	šš ~ ss
*zz	zz	zz	žž ~ zz
*ẓẓ	ẓẓ	ṣṣ	ẓẓ

In addition to this, one can mention the well-known development of *z in Tuareg, which became [z] ~ [ž] in Niger, [š] in Mali, and [h] in Algeria.

5. The velar series

One of the few new ideas in Kossmann (1999) was the proposal to distinguish two sets of velar stops, one set tentatively reconstructed as *kʸ, *gʸ, while the other set was reconstructed as *k, *g. The main argument behind this reconstruction is the presence of two distinct sets of cognates in the so-called Zenatic varieties. In these varieties, *kʸ and *gʸ would have become palatal sibilants (*š* and *ž*, respectively), while *k and *g are represented by velar stops (or developments thereof). In all other varieties, the two sets would have merged. In languages with spirantization, *k* has become *ç* or *š*, while *g* has become *j* or *y*. As a result, in some Zenatic dialects with spirantization the reflexes of *kʸ (> š) and of *k (> ç > š) have merged for non-geminates. In many other Zenatic dialects, the distinction is well maintained. The distinction proposed in Kossmann (1999) also pertains to geminate *kkʸ (> *čč* > *šš* in Zenatic) / *kk (> *kk*) and *ggʸ (> *ǧǧ* > *žž* in Zenatic) / *gg (> *gg*).

While there is little doubt about the existence of these two cognate sets, their interpretation as reflecting two different phonemic sets is not without caveats (cf. the discussion in Kossmann 1999: 169ff.). In the first place, the distribution of *kʸ and *gʸ as shown by the Zenatic correspondences is partly conditioned: Unsurprisingly, the palatalized set is found before *i*. Moreover, the palatalized series is not attested when followed by a consonant or by *u*.

One possible interpretation, suggested and rejected in Kossmann (1999), would be to assume that the palatalization in *kʸ and *gʸ was the effect of a following palatal vowel. This would concur with the two distributional restrictions. It would imply a reconstruction with three

7. Except for a few exceptional cases, *θ̣* and *ẓ* are in complementary distribution (Taine-Cheikh 2001-2003: 30).

different short vowels, *ă*, *ĭ*, and *ŭ*, which is possible, but not certain (see below). As mentioned in Kossmann (1999), such a reconstruction would work for many forms, but produces problems elsewhere. In the first place, there are a number of verbs which have the same stem structure, but different reflexes, for example Tashelhiyt *agʷr* 'to be more', Zenatic *ažər 'id.' versus Tashelhiyt *agʷl* 'to hang', Zenatic *agəl 'id.'. One could venture to explain such forms by assuming different vocalization types in Proto-Berber (i.e. *agŭl vs. *agĭr), but such an analysis is not unproblematic.

Such ways around are more difficult in the case of the opposition between *k(k)ʸ and *g(g)ʸ in the position between two plain vowels, the second of which is *a*. In such phonetic contexts, there is no obvious place for a conditioning factor. It should be noted, however, that some of the words with the palatalized series under this condition can be shown to contain an ancient laryngeal consonant on the basis of Zenaga data (Table 5).

Table 5 — Tentative alternative interpretations of words with *gʸ, *kʸ

1999 reconstruction		Zenaga	alternative interpretation
*(a)gʸa	'bucket'	*äʔgäh*	*a-ʔĭga(h)
*(a)kʸal	'land'	*aʔgäy*	*a-ʔĭkal

Kossmann (1999) also tackles the challenging problems of variation between *y and *kʸ. According to this study, a number of phonetic rules would have been at work in Proto-Berber that changed *kʸ to *y under some circumstances: The most consequential condition is when *kʸ is preceded by a plain vowel and followed by schwa. Of course, this can easily be converted into an analysis without a palatalized velar phoneme, as the schwa in question could be *ĭ, while the cases where Kossmann (1999) has *Vkə or *Vgə, the relevant short vowel would have been *ŭ. It is not entirely clear that the variation *y ~ *kʸ was already present in Proto-Berber (cf. van Putten 2014a for counterevidence from Awjila).

All in all, it is not clear whether the reconstruction of two series of velars as proposed in Kossmann (1999) is absolutely necessary. In a large number of cases, an analysis where the palatalized velars are in fact phonetically conditioned by an adjacent *i or *ĭ seems to work out quite well (as long as one is willing to reconstruct *ĭ of course). It would, however, be necessary to go through all the evidence—especially in the light of the Zenaga data now available—in order to see whether this solution really works better than the one proposed in 1999.

6. The uvular consonant

Most Berber languages have a voiced uvular or back-velar fricative *ʁ* or *ɣ* which corresponds to a geminate voiceless uvular stop *qq*. The corresponding proto-phoneme will be represented here as *ɣ. The main exceptions are Zenaga and Tetserret, where the normal cognate of *ɣ* is a glottal stop or zero, respectively. As shown in Taine-Cheikh (2004) and in Kossmann (2001b; 2012), Zenaga /ʔ/ presents a merger of two different phonemes, *ɣ and *ʔ (on the latter, see below); in the phonological system of Zenaga the reflexes of these consonants behave largely the same way (see below under /ʔ/ for details). Cases where Zenaga has *ɣ* are rare, and they sometimes seem to go back to ancient velar stops, and otherwise may be borrowings from other Berber languages (Souag 2017). In Tetserret, which lacks the glottal stop of Zenaga, *ɣ has been lost entirely, although loans from Tuareg brought it back into the phonemic system (Lux 2013: 127ff.). Finally, in Awjila, *ɣ has the reflex [q] in the vicinity of a pharyngealized consonant /ṭ/ or /ẓ/ (van Putten 2014b).

The geminate form of *ɣ is *qq* everywhere except in Tetserret. In Zenaga, there is evidence for a form *qq* corresponding to *qq* elsewhere in forms such as *aqqiy* 'to look' (cf. for example Beni Iznasen *qqəl* 'to look'; Taine-Cheikh 2001-2003: 34). In a number of varieties, the relationship between *ɣ* and *qq* is no longer used paradigmatically (single *ɣ* corresponding to geminate *ɣɣ*, as in Awjila and Ghadames), but, like in Zenaga, there is ample evidence from irregular and non-derived forms that these varieties originally also had *qq*.

Generally speaking, the reflexes of *ɣ are fairly consistent within Berber. There are, however, a couple of problems that indicate that the situation may originally have been more complex than it looks now.

In the first place, there are about ten lexemes in which one finds dialectal and/or paradigmatic variation between *ɣ and *gʸ. In most of these lexemes, *ɣ/*gʸ stands in the vicinity of /r/ (5 out of 10) or /ẓ/ (2 out of 10); among the three other cases, only one (*e-sămăɣ/gʸ 'slave') can be considered convincing (for details, see Kossmann 1999: 212-216). There does not seem to be a major dialectal conditioning to having forms with *ɣ or forms with *gʸ; every word form has its own dialectal distribution, and in a number of cases forms with *ɣ and with *gʸ are found in one single variety, e.g. Tashelhiyt *rɣ* 'to burn', *ti-rg-in* 'embers'. In the second place, there are a number of nouns where *ɣ corresponds to *š* in Ghadames and to *ẓ* in Tuareg (Vycichl 1990; Kossmann 1999: 216-218). There may be some correlation with the presence of /i/ in these words, and one remarks that three out of five items have /r/ as a root consonant. No convincing explanation for these two phenomena has been put forward yet.

A further element in the analysis of *ɣ is its relationship to the voiceless uvular (or back-velar) fricative, i.e. *x*. Kossmann (1999: 236ff.) attributes wide-spread cases of *x* in modern Berber to two sources: voice assimilations to a following voiceless consonant, and final devoicing. The latter process mainly affects grammatical elements. This may be due to the fact that lexical stems tend to have paradigmatic variation between forms where *ɣ would be in final position and forms where it would be followed by a suffix or a clitic; *ɣ* may therefore have been analogically reintroduced in lexical items. According to this analysis, the Zenaga correspondent of *x* would be *k* rather than a glottal stop. While the idea of devoicing of *ɣ in non-Mauritanian Berber has remained unchallenged in the scholarly community, the rule *[x] > k in Zenaga was rejected by the main specialist in this language, Catherine Taine-Cheikh (2004; 2005). The discussion was continued in Kossmann (2006), which points to the existence of paradigmatic variation between *ʔ* and *k* in Zenaga.

The original pronunciation of *ɣ is difficult to establish. A sound change *ɣ > ʔ is unexpected and, more importantly, the plosive realization of the geminate is suggestive of an ancient plosive realization in the non-geminated consonant. Inspired by the geminate form and by Arabic dialects where *q has become [ʔ], a reconstruction *[q] would make perfect sense. This reconstruction is problematic once one accepts the idea that in Zenaga *ɣ became [ʔ], while a devoiced variant of *ɣ became [k], as the consonant would have been voiceless from the outset. One way of solving this problem is to posit an original pronunciation as a voiced uvular plosive rather than a voiceless one, i.e. *[ɢ], a reconstruction which I think is phonetically as plausible as *[q].

7. The glottal series

Already in 1969, Karl-G. Prasse proposed a reconstruction of Berber involving a glottal series. In his 1969 study of historical Ahaggar Tuareg phonology, he distinguished several different elements *h*. His h^2 and h^3 correspond to the non-pre-consonantal cases of *β in our analysis.[8] His h^1, on the other hand, is an element reconstructed entirely on the basis of an internal reconstruction of Tuareg morphology, and especially on the idea that the major morphological verb type should be analyzed as consisting of triconsonantal roots (see for the advanced analysis of this internal reconstruction Prasse 1972-1974). This latter argument is in itself questionable—why should

8. h^2 stands for those *h* that are attested in Ahaggar Tuareg and do not go back to *z; h^3 stands for those *h* that are not present in Ahaggar Tuareg, but appear in corresponding words in Ghadames and/or Mali Tuareg (Prasse 1969: 5).

one assume that Proto-Berber (or its ancestor) did not have more diverse verbal root shapes?—but of course not necessarily wrong. Kossmann (1999) did not discuss glottal consonants, as his reconstructions did not aim to pass the threshold of Proto-Berber (defined as the starting point of differentiation), while, at that time, Prasse's analyses seemed to go far beyond that point in time.

7.1 The glottal series: *ʔ

Our new data on Zenaga have dramatically changed the picture concerning the glottal series. The studies by Catherine Taine-Cheikh clearly show that Zenaga preserves at least one ancient glottal consonant, /ʔ/. This consonant has a specific distribution: in word-final position it is not realized (though different from underlying final vowels, see 6.2); and in non-final position it can only appear in coda position, preceding a consonant. Zenaga /ʔ/ has two etymological backgrounds. In the first place, it is the regular correspondent to *ɣ. In the second place, it occurs in a large number of words which have no corresponding consonant elsewhere in Berber. This situation brought Taine-Cheikh (2004) and Kossmann (2001b) to the conclusion that the Zenaga glottal stop should be reconstructed into Proto-Berber. The most important argument comes from verbs with a final glottal stop in Zenaga. It can be shown that these verbs correspond to verbs belonging to a specific type of biconsonantal roots in other Berber languages, that have different morphology from other biconsonantal verbs (Destaing 1920; Kossmann 1994). In addition, it was shown that in Ghadames cognates of verbs with an initial glottal stop in Zenaga have a specific morphological shape not found in other verbs (Kossmann 2001b). As a result, there can be no doubt that the Zenaga glottal stop is the reflex of an ancient Berber consonant—symbolized here as *ʔ. The fact that different non-Mauritanian languages have different vocalic reflexes of *ʔ (see Kossmann 2001b and van Putten 2015 for some details) corroborates the idea that the consonant was part of the Proto-Berber system and was lost independently everywhere except in Zenaga.

While the presence of *ʔ may be considered well-established now (see also Prasse 2011), its reflexes in non-Mauritanian Berber present numerous problems, some of which were solved by van Putten (2015), while others still stand out. Moreover, while no doubt the large majority of glottal stops in Zenaga go back to *ʔ and *ɣ, one cannot rule out that some of them have a different background. There exists a large amount of lexical and paradigmatic variation between forms with and without glottal stop in Zenaga (cf. for example Taine-Cheikh 2006 for plural formations), and as long as these variations are not

well understood historically, one should not too easily equal any Zenaga /ʔ/ (if not < *ɣ) with *ʔ.

On the other hand, the absence of a glottal stop in Zenaga in a certain word cannot automatically be taken as proof of its absence in Proto-Berber (or its predecessor). Thus one remarks that Zenaga data (and reflexes elsewhere) show that there are many reconstructible triconsonantal verbs of the types *ʔCC and *CCʔ. There is, however, not a single verb that can be reconstructed as *CʔC on the basis of Zenaga. This opens the way for an analysis where some biconsonantal verbs (i.e. those with the structure vCvC in Zenaga) in fact go back to *CʔC verbs, a possibility already envisaged by Prasse (1972-1974). This would imply that *ʔ was lost under some conditions in Zenaga.

There is no indication that *ʔ could be geminated.

The original pronunciation of *ʔ is difficult to establish. Of course, as Zenaga is the only variety that preserves the consonant, a reconstruction as *[ʔ] is logical. On the other hand, there is little reason to assume that Zenaga is particularly archaic in its phonetics, and evidence from one single language is hardly persuasive.

Souag (2011) points to an intriguing fact with *ʔ. Not only is there no indication for a geminate counterpart to *ʔ, verb morphology suggests that the presence of *ʔ prevented a following consonant from geminating. Thus in the Imperfective of triconsonantal verbs, normally the second consonant is geminated, e.g. Aorist *ăfrəs 'to divide', Imperfective *əfărrăs. In *ʔ-initial verbs, however, this gemination does not take place, and instead a prefix is used—a device otherwise used with longer verbs and with plain-vowel-initial and geminate-initial verbs. Thus one has the verb Aorist *ăʔkər 'to steal', Imperfective *ətt-ăʔkăr instead of **əʔăkkăr. This is the situation in all Berber languages, including Zenaga. It shows that *ʔ was different from normal consonants, and that maybe the present distribution of /ʔ/ in Zenaga—word-internally only in pre-consonantal position—is much older than one might have expected.

7.2 The glottal series: *hʔ

In an overview article that Prasse wrote after the relevance of Zenaga became clear (Prasse 2011), he proposed the existence of two different glottal elements (in addition to the *h* that I reconstruct as *β here), which one could symbolize as *ʔ and *h. The second element would cover the cases of his *h^{1} that cannot be explained by the Zenaga glottal stop. This would explain different pairs of verbs such as found in Ghadames (Table 6).

Table 6 — Reconstructions of Ghadames verb types following Prasse (2011)

	‘to dress’	‘to sneeze’	reconstruction
Aorist 3SG:M	*y-ăls*	*i-nzu*	*y-ălsəʔ – *y-ănzəh
Perfective 3SG:M	*i-lso*	*i-nza*	*y-əlsăʔ – *y-ənzăh
	‘to rub’	‘to hang’	reconstruction
Aorist 3SG:M	*y-oməs*	*y-agəl*	*y-ăʔməs – *y-ăhgəl
Perfective 3SG:M	*y-omăs*	*y-ugăl*	*y-əʔmăs – *y-əhgăl

The idea that there could be a second ancient glottal consonant around has also been advocated by Taine-Cheikh (2004: 186-187) on the basis of Zenaga verbs ending in *h*. Kossmann (2001b), on the other hand, analyzes word-final *h* in Zenaga as a phonetic off-glide, due to the presence of a word-final vowel. While I think this makes sense on a synchronic level, it is of course very well possible that, diachronically, word-final *h* in Zenaga perpetuates an old consonant that used to have a broader distribution. One should emphasize, however, that there is no opposition between (non-glottal-stop) *Vh#* and (non-glottal-stop) *V#*, so extending the Zenaga distribution to Proto-Berber would mean that there were no vowel-final forms originally (of course one cannot rule out the possibility that in Zenaga word-final *h* has several origins, Taine-Cheikh 2004, Prasse 2011).

There is no major dialectal variation in the reflexes of this putative *h (although large-scale analogical reformations make them sometimes difficult to establish, cf. Kossmann 1994), and it seems that its loss had already taken place in proto-Berber times. This means that the only reason to reconstruct this consonant is morphological simplicity: by positing *h in addition to *ʔ, one can explain most (if not all) biconsonantal verb roots as originally triconsonantal; otherwise, one would have to posit root types including a long vowel position. While this is an alluring take on the matter, it should be emphasized that it is hardly necessary that proto-Berber (or earlier stages) only had consonantal root elements.

8. Semivowels

The consonant *w* is a labiovelar approximant. Its geminate counterpart is *gg*^{w} in most languages. The consonant *y* is a palatal approximant. The question of its geminate counterpart will be studied below.

Many Berber languages have a large-scale merger of *w / *y with the plain vowels *u / *i, to the extent that it has been claimed for Tashelhiyt that semivowels and high vowels are positional variants of one single set of segments. The Tashelhiyt situation is extreme, but in many varieties *w and *y are quite labile and prone to changes. To

this date, no comprehensive analysis of the semivowels from a historical-comparative perspective has been undertaken, and the following remarks are to be taken as general impressions rather than well-established analyses. I will leave regular dialectal vocalizations of *w / *y out of the discussion and concentrate on the more problematic points.

8.1 The semivowel *w

The main context where *w is largely preserved in all languages is the intervocalic position with at least one plain vowel present. Thus, the verb 'to help' has the following reflexes of what must have been Aorist *awəs: Tashelhiyt *aws* 'to help', Ghadames *awəs* 'id.', Zenaga *äwuš* 'id.' (cf. Taine-Cheikh 2008: 556).

In word-initial position, its fate is more volatile. Even in languages that do not regularly have vocalization of *wə* to *u*, one sometimes finds forms with initial *u* instead of *wə*, e.g. Figuig *ul* 'negation' < *wăr or *wər as opposed to *wətna* 'sister' (< *wălătma). In addition, there is an unexplained variation between generally attested *u*-initial nouns and much rarer forms with *w*, such as Mali Tuareg *ulh*, general Berber *ul* 'heart' as opposed to Niger Tuareg *əwəl* and the pan-Berber word *ulli* 'small cattle' as opposed to Ghadames *wălli* (maybe rather *wəlli*) 'goat'. One may also compare forms such as Tashelhiyt *urɣ* 'gold' with adjectives like *a-wraɣ* 'yellow' in the same language.

In word-final position, *w* is not very frequent. It is found in a relatively small number of triconsonantal verbs, the most widespread being *arəw (Aorist form) 'to give birth'. In longer words, final *w may have been elided in final position after a plain vowel, as witnessed by widely attested forms such as Tashelhiyt *a-zgza* 'blue (M)', *ta-zgzaw-t* (F).

The consonant *w* is regularly found as the second consonant in a triconsonantal verb root, e.g. *ăswəʔ 'to drink'; *ănwəʔ 'to be cooked'; *ărwəs 'to resemble' (cf. also *əẓwu 'to traverse'). In Zenaga and Tetserret, *w has become *b* in a number of these verbs (Lux 2013: 141). The large majority (if not all) of these forms have a dental or alveolar consonant before *w.

In a reconstructible preconsonantal position, *w is extremely rare. It has long been understood that this is due to a process in which preconsonantal *w underwent full assimilation to a following consonant (cf. Prasse 1972: III/69 for a full analysis). Thus a form like *ăwfəɣ 'to go out' would have become *ăffəɣ*; traces of the ancient initial root consonant can be found in forms like the causative *ssufəɣ* (possibly < *ss-əwəfəɣ, van Putten p.c.) and in various verbal noun formations. While it is plausible that the large majority of geminate-initial verbs go back to forms with initial *w, this does not mean that

this is always the case. In the first place, there are indications that geminate-initial verb stems may also go back to forms with *w as their second consonant. The best evidence for this is found in forms such as Kabyle *əqqən* 'to tie' and its instrumental noun *asəɣwən* 'rope', which point to *ăɣwən.[9] In the second place, initial gemination is a word-building process in Berber, present in a fair number of verb classes (cf. Basset 1955). In some cases, this may provide a better explanation than the phonetic one. Thus, in a number of varieties, there are verbs with similar semantics that are differentiated by the presence or absence of initial gemination, e.g. Beni Iznasen *ərẓ* 'to break' *vs. rrəẓ* 'to be broken' and *ɣṛəs* 'to cut the throat, to slaughter' vs. *qqəṛs* 'to rip' (notes by the author). In such cases, it seems preferable to consider this an ancient difference in ablaut pattern, e.g. *ărẓəʔ vs. *ərrăẓăʔ and *ăɣrəs vs. *əqqărăs.

A special problem is posed by verbs with initial *gg*^{w}, like *ăggʷəd 'to jump, to fear'. *gg*^{w} is the geminate counterpart of *w* (cf. nominal forms such as Figuig *t-iwdi* 'fear'), which would lead to a reconstruction *ăwwəd, which, then, would be the only remaining geminate-initial verb in the language. I have no solution to this question.

It should be emphasized that the reconstruction of initial *w in these verb forms is based entirely on internal evidence. The proposed assimilation has taken place in all Berber languages in the same way, and therefore there are good reasons to assume that it had already taken place when Berber started to differentiate.

As mentioned above, the most common pronunciation of the geminate variant of *w is *gg*^{w}. In a number of languages labialization was lost, and *gg*^{w} merged with *gg* (e.g. Tuareg, Ghadames). In some other languages the pronunciation of geminate *w is *bb*^{w}, *bb* or *ḅḅ*. The best-known cases are (parts of) Kabyle, Zenaga and Tetserret. In the case of Kabyle, it makes sense to consider *bb*^{w} a later phonetic development of earlier *gg*^{w}, as this is well-attested in Kabyle dialects more to the east. There is no reason to posit an intermediate stage *gg*^{w} for Tetserret *bb* and Zenaga *ḅḅ/bb*. In Awjila, there is no trace of a plosive realization of geminate *w. This includes at least one word that does not have paradigmatic variation with non-geminate *w* in any Berber language, *ašəw(w)áša* 'this year'. As remarked by van Putten (p.c.) this strongly suggests that the situation in Awjila is old, as there would be no basis for an analogical substitution of earlier *gg*^{w} by *ww*. It may therefore be wise to reconstruct the Awjila situation of *w* vs. *ww* to the Proto-Berber stage and consider the plosive geminates to be post-Proto-Berber developments.

9. Note that this solution only works if we assume that earlier forms of *asəɣwən* had a vowel between *ɣ* and *w* (*a-səɣəwən or something similar).

8.2 The semivowel *y

The situation with the semivowel *y* is even more intricate than with *w. In word-initial position, *y* is well-attested in the 3SG:M verbal prefix *y-* and in some pronominal forms. Like with *w, vocalized variants are common and not always predictable.

It is not easy to establish the geminate counterpart to *y. In the large majority of Berber languages, there are no forms with morphological alternations which allow us to establish what geminate would correspond to *y. Most languages do have *yy*, which may develop into *gg* (esp. in Kabyle and Chaouia), but these forms mostly seem to be new, being either loans from Arabic or otherwise innovative. There are, however, a few indications that *y originally corresponded to *ğğ ([ʤː]) (Kossmann 1999: 232ff.).

There exists important dialectal variation between *y* and *k* when it is the last root consonant of a triconsonantal verb. While most languages have *-y* and only rarely *-k*, the inverse is true in Ghadames, Awjila and Zenaga. Most cognates of *-y#* verbs elsewhere have *-k* (> *g* in Zenaga) in these languages, e.g. Mali Tuareg *əẓməy* vs. Ghadames *ăẓmək*; Awjila *ẓmək*; Zenaga *aẓẓṃug* 'to sew'.

On the other hand, the number of verbs which have *-y* in Ghadames, Awjila and Zenaga is much smaller, although certainly not a marginal number. In Zenaga, there are also cases where the final *-y* of other varieties corresponds to the absence of a consonant, which is mostly—but not exclusively—the case when the verb also contains a velar consonant or *w* (Kossmann 2018). In Ghadames and Awjila, no conditioning factors have been found for the distribution of *k* and *y* in these verbs.

The background to this dialectal distribution is unknown. One possibility is that (part of?) the final *-y#* verbs in fact go back to *k or *kʸ, and that the situation in Ghadames, Awjila and Zenaga is original. If one assumes this, it is an open question whether there is any evidence for *-y# as opposed to *-k#. While all three languages with *-k#* also have *-y#* or *-ø* forms, the lexical distribution of these forms is different, and some forms with final *g* (< *k) in Zenaga correspond to forms without a velar in Ghadames or Awjila and vice versa, cf. Figuig *ɣrwəy*; Zenaga *ärwih*;[10] Ghadames *ărwək* 'to stir, to mix'; Figuig *mḍəy*; Zenaga *anḍug*; Awjila *ənṭi* 'to taste'.

Word-final *y* is normally lost in longer forms when following *a*, as shown by forms like Tashelhiyt *a-ɣrda* 'rat', *ta-ɣrdəy-t* 'mouse'.

There is little evidence for *y* in pre-consonantal position. As far as I know, no clear case of a *y*-initial triconsonantal verb exists, while

10. The final *h* in Zenaga *ärwih* is an automatic effect of the vowel being in word-final position, see section 7.2.

the evidence in nouns is relatively scarce. Kossmann (1999: 200-202) presents ten reconstructible forms that may contain pre-consonantal *y*, but in most cases it is impossible to decide whether *y* was really preconsonantal, or rather intervocalic, as they are mainly attested in varieties where short vowels are elided in open syllables. It is therefore very well possible that pre-consonantal *y underwent similar assimilations as pre-consonantal *w. However, in contrast to *w, there is no internal evidence that would corroborate this.

In post-consonantal position, the evidence for *y is also quite meagre, although a full analysis is still lacking. The best evidence seems to come from a number of assimilations. As shown in Kossmann (1999: 223ff. and 229ff.), it makes sense to analyze some well-attested cases of *šš* and *žž* as assimilated forms of *sy and *zy. While the evidence for *šš* < *sy is mainly circumstantial, the evidence for *žž* < *zy is relatively strong. In Ghadames and Awjila, *žž* corresponds to *zi*, which may very well go back to *zyə, e.g. Tashelhiyt *žži* 'to be healed', Ghadames *əzik* 'to be healed', apparently from *ăzyəy/k. Similarly, Ahaggar Tuareg *əhyəḍ* (< *ăzyəḍ) 'to have scabies' corresponds to Iwellemmeden Tuareg, Tarifiyt *əžžəḍ* 'id'. Moreover, the following Zenaga verbs confirm that *žž*-initial verbs are different from other verbs with an initial geminate. Normally, such verbs take the prefix *ətt-* in the Imperfective. In *žž*-initial verbs, one also finds forms with medial gemination (Taine-Cheikh 2001-2003: 57; 2008: 629) (Table 7).

Table 7 — Zenaga verbs with paradigmatic variation of žž *and* žVdd^{y}

	'to have scabies'	'to be healed'
Aorist	*äžžuḍ*	*äžžig*
Imperfective	*ižäddʸaḍ*	*əžäddʸäg*

Assuming that *ğğ* (> *ddʸ* in Zenaga) represents the regular geminated form of *y, these forms would attest to *y* as a medial consonant: *ăzyəḍ / *əzăyyaḍ and *ăzyəy / *əzăyyăy.

The history of the very rare pan-Berber consonant *čč* (mainly found in the verb 'to eat') is difficult to point down. Kossmann (2008) proposes that it represents *ty and reconstructs the relevant verb as *ătyəʔ 'to eat'. While the argumentation behind this reconstruction is largely circumstantial, it is plausible in the light of the other palatalizations due to a following *y. Maybe assimilated *y is also behind otherwise unexplained variations such as Tashelhiyt *kšm* vs. Siwa *kim* 'to enter' (< *ăkyəm?).

Kossmann's study of three Berber verbs *ăswəʔ 'to drink', *ătyəʔ 'to eat' and *ănwəʔ 'to cook' (Kossmann 2008) proposes that a number

of irregular morphological features can be understood if one assumes that *w and *y could not be geminated at a certain moment in the prehistory of Berber. If this is correct, the constraint must have applied only at some very early stage, as all comparative evidence points to the possibility of gemination of *w and *y.

9. Other sonorants

In addition to the consonants analyzed above, Proto-Berber had four more sonorants: *m, *n, *l, *r. Except for some assimilations and dissimilations, *m and *n are preserved as such in all Berber languages. The fate of *l and *r is more complicated on the dialectal level, but their reconstruction is unproblematic. There exists no doubt that these four consonants can be reconstructed as such into Proto-Berber.

10. Remarks on pharyngealization and on consonant weakening

There are two points of discussion that surpass the level of individual consonants and that will therefore be treated in a section of their own. The first one is pharyngealization, the second one is the question of consonant weakening in Proto-Berber times.

As shown above, two consonants have pharyngealized pronunciations in all modern Berber languages: *ḍ and *ẓ. The most widespread pronunciation is voiced, but there are important dialectal differences at this point. Thus, the pronunciation [ṭ] instead of [ḍ] is found in a large number of eastern dialects, and, sporadically, in Algeria and in Morocco. In Tuareg and in Zenaga/Tetserret, *ḍ is always voiced. The geminate counterpart, on the other hand, is consistently voiceless all over Berber, except in Zenaga. Voiceless pronunciations of *ẓ are only found in Zenaga and Tetserret (see section 4). In all other Berber languages, it is voiced, both as a nongeminate and as a geminate. It is therefore impossible to make a decision on the original voicing of *ḍ and *ẓ on the basis of the comparative evidence.

The uvular fricative (maybe originally a stop, see above) *ɣ is isolated within the Berber consonantal system, while the velar series lacks a pharyngealized counterpart. It thus makes sense, on a rather abstract level, to consider *ɣ (whatever its original pronunciation) as the pharyngealized counterpart to the velar consonants.

On the basis of Afroasiatic evidence, one may assume that at some stage in the linguistic history of Berber the pharyngealized consonants and *ɣ were glottalized (ejective or implosive) rather than pharyngealized. There is, however, no hard evidence for a glottalized pronunciation coming forth from internal evidence in Berber, and the change from glottalized to pharyngealized was probably already in place at the Proto-Berber stage.

The second question concerns weakening in the Proto-Berber consonant system. As shown above, in a number of articulation places all consonants are most easily reconstructed as fricatives rather than as plosives. This is the case of *β, *f and, to a lesser degree *ɣ. Among these, *β has no geminate counterpart, the geminate counterpart of *ɣ is always a plosive, while the geminated counterpart of *f is always a fricative.

As the Proto-Berber system seems to have stops in the alveolar and velar series, it would not be unexpected to have stops at the other articulation places, too. Comparative evidence with *ɣ indeed suggests that it was still a stop in Proto-Berber (see above), possibly *[ɢ]. While I think that a consistently fricative pronunciation of *β is the simplest way to explain the present situation, a reconstruction *b or one with allophonic variation between *[b] and *[β] is also viable. For *f, Berber-internal reconstructions do not provide any evidence for an ancient plosive realization.

One way to explain this situation is by assuming that in a not-too-far-away stage preceding Proto-Berber, all these consonants were plosives. They would have been subjected to a general process of lenition, which first targeted *[p] and, maybe somewhat later, *[b]. When Proto-Berber broke up, the tendency towards lenition did not stall, and in a similar fashion *[ɢ] was targeted everywhere, except in Zenaga/Tetserret, and restricted to the non-geminate consonant. Of course, such an explanation is conjectural, and predicated upon the assumption that the original pronunciation must have been a plosive.

In addition to these lenitions, which would have taken place just before or just after the break-up of Proto-Berber, there are also lenition processes that still apply in modern Berber languages. These processes are known as "spirantization" in the tradition of Berber linguistics and involve the lenition and fronting of bilabial, alveolar and velar stops. Spirantization does not affect geminates, and is therefore quite different from the lenition that would have changed *[p] to *f*, but similar to that that would have changed *[ɢ] to *ɣ*.

From the point of view of dialectal distribution, spirantization appears mainly in the northern part of Morocco, Algeria, and Tunisia, following a line that cuts across all other linguistic boundaries. It therefore looks very much like an innovation that spread at a moment when older splits were being smoothened out by convergence (Kossmann 1999: 21). There are, however, two caveats to this. In the first place, lenition targeting non-geminate alveolar stops is also found in Zenaga (cf. Taine-Cheikh 2001-2003 for details) and in Tetserret (Lux 2013). If one assumes that this process is historically related to the northern "spirantization" it would push back the lenition deep in time, as Zenaga and Tetserret are neither geographi-

cally, nor linguistically, anywhere close to the spirantizing dialects of the north. Of course, this assumption is by no means necessary, lenitions being common in the history of languages, and Zenaga involving many other types of lenition (e.g. voicing of *k and *l > *y*). The second caveat lies in the phonetic conditioning of northern Berber spirantization. In most of these languages, very similar conditions are found, basically lenition everywhere except in some consonant clusters. However, at the geographical extremes of the spirantization-affected region, much more restricted lenition is found, which only targets syllabic codas. This is found both in north-western Morocco (Ghomara, Mourigh 2017) and in Tunisia. Vycichl (1975) proposes that this is the ancient state of affairs and suggests that the other spirantizing dialects generalized the lenition, while the non-spirantizing dialects underwent a process of strengthening. As far as I know, this proposal—which had the sympathy of Lionel Galand—has never been exploited further (see Acosta Armas 2017 for an argumentation that spirantization played a role in Guanche).

11. Vowels

There exists no doubt that Proto-Berber had two types of vowels, which will be called “short vowels” and “plain vowels” here. Alternative terms would be “central vowels” vs. “peripheral vowels” (cf. Louali 1992; Louali-Raynal 2000) and “long vowels” or “full vowels” instead of “plain vowels”.

Only Ghadames and Tuareg maintain a clear contrast between these two series. In most other languages, the short vowels have collapsed into one single element, schwa. While schwa remains fully phonemic in a number of varieties (mainly in Libya and Siwa), it is hardly phonemic (when at all) in the Berber languages of Algeria and Morocco.

In Zenaga, a different process has eliminated the length distinction, and short and plain vowels have collapsed into a small system of high and low vowels (see below). The vowel system of Tetserret is very different from all other Berber languages and would need an extensive analysis in order to be understood historically.

11.1 The short vowel system

Tuareg and Ghadames unambiguously show the existence of a binary contrast among the short vowels between a higher and a lower vowel, symbolized here by *ə* and *ă*. While in Zenaga the short vowels have merged with the plain vowels, the height contrast is maintained.

The main question concerning the short vowel system is whether one should reconstruct one or two high vowels. Prasse (1972-1974)

opts for two short high vowels *ĭ and *ŭ, even though only one short high vowel is attested in Tuareg, the language on which his analysis is based. Kossmann (1999) reconstructs a system with two high vowels on the basis of a synchronic and comparative analysis of labialization in a number of Moroccan and Algerian varieties. This analysis clearly shows that velars and uvulars can only be labialized if they occur in the vicinity of a reconstructed short high vowel; in the vicinity of *ă, non-labialized forms are found. It is much less clear that labialization would provide evidence for a contrast between two different high vowels, as no clear partition between labialized and non-labialized consonants in the vicinity of ancient short high vowels can be discerned. An analysis according to which a single short high vowel phoneme would automatically be rounded in the vicinity of velars and uvulars (similar to modern Zenaga) would probably work for a large part of the evidence. Therefore I think, different from twenty years ago, that, while labialization is indeed related to the presence of an original high vowel, it is not necessarily indicative for a phonemic contrast between *ŭ and *ĭ.

Another potential piece of evidence for a contrast between *ŭ and *ĭ may come from Zenaga. As mentioned above, in this language the short and the plain systems have collapsed, maintaining a difference between low and high vowels. There is an enormous amount of allophonic variation, some of which is marginally phonemic (see Taine-Cheikh 2008: lxxiv for details). Most of the phonetic variants, however, can be understood on the basis of a two-vowel system with allophones conditioned by the surrounding consonants (cf. also Cohen & Taine-Cheikh 2000). In earlier works this led the present author to use "phonological" forms that glossed over the phonetic detail of Taine-Cheikh's phonetic transcriptions. This was an unlucky choice, and it led to a blind eye to at least one potentially important phenomenon: the existence of unconditioned forms with *u* (which was already clear from Taine-Cheikh 1999: 301). While [u] is the regular pronunciation of high vowels in certain phonetic contexts, such as pharyngealized consonants and velars, it also appears in a number of forms where unrounded allophones would have been expected, and is therefore undoubtedly phonemic. One of the possible implications of this oversight was that ancient contrasts among short high vowels may have been missed. In order to check this, I performed a preliminary survey of forms with unconditioned *u* in Zenaga, which yielded about 40 unambiguous cases. The results are highly interesting, although no evidence for *ŭ vs. *ĭ was found. It seems that unconditioned *u* is mainly found in two historical contexts. In the first place, it is found in the vicinity of *ʔ* when the latter goes

back to non-final *ɣ. It is normally not found in the vicinity of *ʔ* going back to *ʔ. Examples are *uʔran* 'feet (of animal)' < *i-ɣir-an; *yuʔṛa* 'he aborted', cf. Kabyle *əɣṛi* 'to have a miscarriage'; and *yuʔräš* 'he slaughtered' < *y-əɣrăs. The second group of nouns with unexpected *u* seem to perpetuate the Proto-Berber plain vowel *u, e.g. *uði* 'melted butter' (Tashelhiyt *udi*), *uri* 'gold' (Tashelhiyt *urɣ*), *uzzäy* 'iron' (cf. Tashelhiyt *uzzal*). There are also forms that have no cognates elsewhere, and forms that show other complications, but there is no indication that Zenaga unconditioned *u* could tell us anything about an ancient contrast between *ŭ and *ĭ.

Thus, neither the presence of labialized consonants in northern Berber, nor the presence of unconditioned *u* in Zenaga provide us with evidence that Proto-Berber had more than a contrast between *ə and *ă. This does of course not rule out that such a contrast did exist in reality, and we already saw that assuming such a contrast might solve the problem of the two velar series proposed in Kossmann (1999). For the time being, I prefer to remain agnostic about this question.

11.2 The plain vowel system

Most elements of the plain vowel system are unproblematic. There exists no doubt as to the presence of *a, *i and *u in Proto-Berber. As shown by van Putten (2015), some irregularities in vowel correspondences are in fact due to the vicissitudes in the reflexes of short vowels in combination with *ʔ, and therefore do not affect the basic system.

Ghadames and Tuareg have two additional vowels, *o* and *e*. As shown in Prasse (1990) for Tuareg and in Kossmann (2001b) and van Putten (2015) for Ghadames, *o* can be understood either as a product of vowel harmony (Tuareg), or as a reflex of *ăʔ (Ghadames). There is no reason to posit *o in Proto-Berber.

On the other hand, over the last few years, an important discussion has been undertaken by van Putten (2016) and Souag & van Putten (2016) about the history of *e* (see also Prasse 1990). This discussion is not only informed by the Tuareg and Ghadames forms, but also by additional data on Siwa, which has preserved *e* in a number of contexts. Moreover, while the normal cognate of Tuareg/Ghadames *e* is *i* in the other varieties, it seems to be *a* in Ghomara and in some Kabyle varieties (van Putten, p.c., referring to ongoing work together with Lameen Souag and Massinissa Garaoun).

Van Putten, expanding on a lead by Prasse, has shown that many cases of *e* (and its correspondents elsewhere) can be understood as the effect of assimilatory and dissimilatory processes, esp. a raising of

*a to *e when the next syllable has *ă (the rules are more complicated, see van Putten 2016; van Putten 2018). However, as the author admits himself, there are a number of elements that defy an interpretation like this, most importantly the nominal ending of the feminine plural *-en*, the ablaut marking in the negative perfective, and the noun *e-sen 'tooth'. Some of these cases may be understood from the elision of *t described above, but this would be unlikely for the negative perfective marker and, in view of Afroasiatic cognates, less desirable for *e-sen (cf. Semitic *sinn). Thus, as it stands, there is some good evidence for the reconstruction of *e in Proto-Berber, although one suspects that further investigations could succeed in eliminating these last few cases.[11]

As for its reconstruction into Proto-Berber, there exists little doubt that the allophone [e] of *a is a very old feature, as its reflexes are found all over Berber.

12. The accent

Berber languages of Morocco and Algeria (except Tuareg) do not seem to have a system of word accent. Instead, accentual features are governed by higher level prosodic organization, which one could call intonation. The situation is different in the Berber languages of Libya and Egypt, and in Tuareg. Heath (2005) presents us with the first full-scale analysis of accent in any Berber language. The recent outburst of studies on Siwa in Egypt has brought much data on accentual patterns in this language, although a full analysis is still outstanding. Van Putten (2014a) provides us with a well-argued overview of the system in Awjila based on older literature, while Mitchell (2009) gives due attention to accentual factors in Zwara.

There is therefore a lot of evidence to make a large-scale comparison of accentual systems possible, but no such endeavor has been undertaken yet. In one (important) detail, however, it has been shown that some quirks of Berber accentual systems can be reconstructed into proto-Berber. This concerns the accentuation of Aorist and Perfective aspectual forms. As shown by Brugnatelli (1986), a number of Libyan varieties have stem-initial accent in the Aorist of disyllabic verbs, while the accent is on the second syllable of such verbs in the Perfective. This distribution is not found in Tuareg, but it does occur in Tetserret, a language of Niger closely related to

11. One should note that there is no evidence for an ancient phoneme /e/ in Zenaga. However, at least some of the forms that have /e/ in other Berber varieties have a high vowel in Zenaga (e.g. in the negative Perfective). Thus, while there is no way to distinguish *e from *i in this language, the reflex of *e is clearly different from *a and *ă.

Zenaga (Lux 2013: 284ff.). In view of the unexpected distribution among the forms, and of its presence in languages that are both linguistically and geographically far apart, there is little doubt about the archaic nature of this phenomenon.

13. A reconstruction of the Proto-Berber phoneme system

The Proto-Berber phoneme system can be reconstructed as follows.

13.1 Consonants

Single consonants

		d	t	ḍ	gʸ?	kʸ?	g	k	ɢ	ʔ
β	f	z	s	ẓ						
m		n								
		r								
		l			y		w			

Geminate consonants

		dd	tt	ṭṭ[†]	ggʸ?	kkʸ?	gg	kk	qq
	ff	zz	ss	ẓẓ					
mm		nn							
		rr							
		ll			ğğ		ggʷ[‡]		

[†] In view of the Zenaga evidence, maybe rather *ḍḍ

[‡] In view of the Awjila evidence, maybe rather *ww

13.2 Vowels

Reconstruction 1 (assuming kʸ *and* gʸ *are positional variants of* k *and* g*)*

short vowels		
ĭ		ŭ
	ă	

plain vowels	
i	u
e	
a	

Reconstruction 2 (assuming kʸ *and* gʸ *are different phonemes from* k *and* g*)*

short vowels
ə
ă

plain vowels	
i	u
e	
a	

*e is probably an innovation, but this seems to have taken place before the dialectal differentiation in Berber.

References

Acosta Armas, Jonay. 2017. Notas sobre la aspiración en los guanchismos. *Revista de Filología de la Universidad de La Laguna* (35). 9-49.

Basset, André. 1952. *La langue berbère*. Oxford: Oxford University Press.

Basset, André. 1955. Sur le verbe berbère signifiant 'vivre'. In Johannes Lukas (ed.), *Afrikanistische Studien, Diedrich Westermann zum 80. Geburtstag gewidmet*, 45-50. Berlin: Akademie-Verlag.

Beguinot, Francesco. 1924. Sul trattamento delle consonanti *b, v, f* in berbero. *Rendiconti della R. Accademia Nazionale dei Lincei (classe di scienze morali, storique e filologiche)* 5(33). 186-199.

Brugnatelli, Vermondo. 1986. Alternanze accentuali e morfo-sintassi nominale nel berbero orientale. In *Contributi di Orientalistica, Glottologia e dialettologia (= Quaderni di ACME 7)*, 61-72. Milan: Cisalpino Goliardica.

Brugnatelli, Vermondo. 1993. Quelques particularités des pronoms en berbère du Nord. In Jeannine Drouin & Arlette Roth (eds.), *À la croisée des études libyco-berbères, Mélanges offerts à Paulette Galand-Pernet et Lionel Galand*, 229-245. Paris: Geuthner.

Cohen, David & Catherine Taine-Cheikh. 2000. À propos du zénaga. Vocalisme et morphologie verbale en berbère. *Bulletin de la Société de linguistique de Paris* 95(1). 267-320.

Destaing, Edmond. 1920. Note sur la conjugaison des verbes de formes C1eC2. *Mémoires de la Société de linguistique de Paris* 22. 139-148.

Galand, Lionel. 2010. *Regards sur le berbère*. Milan: Centro Studi Camito-Semitici.

Heath, Jeffrey. 2005. *A grammar of Tamashek (Tuareg of Mali)*. Berlin, New York: de Gruyter Mouton.

Heath, Jeffrey. 2006. *Dictionnaire tamachek (Mali) – anglais – français*. Paris: Karthala.

Kossmann, Maarten. 1994. La conjugaison des verbes CC à voyelle alternante en berbère. *Études et documents berbères* 12. 11-19.

Kossmann, Maarten. 1997a. *Grammaire du parler berbère de Figuig (Maroc oriental)*. Paris, Louvain: Peeters.

Kossmann, Maarten. 1997b. Le pronom d'objet direct de la troisième personne en berbère. In Alessandro Bausi & Mauro Tosco (eds.), *Afroasiatica Neapolitana*, 69-79. Naples: Istituto Universitario Orientale.

Kossmann, Maarten. 1999. *Essai sur la phonologie du proto-berbère*. Cologne: Rüdiger Köppe.

Kossmann, Maarten. 2001a. L'origine du vocalisme en zénaga de Mauritanie. In Dymitr Ibriszimow & Rainer Voßen (eds.), *Études berbères. 1. Bayreuth-Frankfurter Kolloquium zur Berberologie* (Frankfurter Afrikanistische Blätter 13), 83-95.

Kossmann, Maarten. 2001b. The origin of the glottal stop in Zenaga and its reflexes in the other Berber languages. *Afrika und Übersee* 84(1). 61-100.

Kossmann, Maarten. 2006. Remarks on the history of some Zenaga pronouns. In Dymitr Ibriszimow, Rainer Voßen & Harry Stroomer (eds.), *Études berbères III*, 167-174. Cologne: Rüdiger Köppe.

Kossmann, Maarten. 2008. Three irregular Berber verbs: 'eat', 'drink', 'be cooked, ripen'. In Alexander Lubotsky, Jos Schaeken & Jeroen Wiedenhof (eds.), *Evidence and counter-evidence. Essays in honour of F. Kortlandt*, 225-236. Amsterdam, New York: Rodopi.

Kossmann, Maarten. 2012. Some new etymologies for glottal-stop initial Zenaga Berber words. *Folia Orientalia* 49. 245-251.

Kossmann, Maarten. 2013. *The Arabic influence on Northern Berber*. Leiden, Boston: E.J. Brill.

Kossmann, Maarten. 2018. Zenaga reflexes of Berber final weak verbs. *Folia Orientalia* 55. 161-194.

Louali, Naïma. 1992. Le système vocalique touareg. *Pholia* 7. 83-115.

Louali-Raynal, Naïma. 2000. Vocalisme berbère et voyelles touarègues. In Salem Chaker & Andrzej Zaborski, *Études berbères et chamito-sémitiques. Mélanges offerts à Karl-G. Prasse*, 263-276. Paris, Louvain: Peeters.

Lux, Cécile. 2013. *Le Tetserret, langue berbère du Niger. Description phonétique, phonologique et morphologique dans une perspective comparative*. Cologne: Rüdiger Köppe.

Marcy, Georges. 1936. Note sur le pronom relatif-sujet et le pseudo-participe dans les parlers berbères. *Bulletin de la Société de linguistique de Paris* 37. 45-57.

Mitchell, Terence Frederick. 2009. *Zuaran Berber (Libya): Grammar and texts*. (Ed.) Harry Stroomer & Stanly Oomen. Cologne: Rüdiger Köppe.

Mourigh, Khalid. 2015. *A grammar of Ghomara Berber (North-West Morocco)*. Cologne: Rüdiger Köppe.

Mourigh, Khalid. 2017. La spirantisation en berbère ghmari (province de Chefchaouen). In Ángeles Vicente, Dominique Caubet & Amina Naciri Azzouz (eds.), *La région du Nord-Ouest marocain : parlers et pratiques sociales et culturelles*, 251-264. Zaragoza: Prensas de la Universidad de Zaragoza.

Múrcia Sànchez, Carles. 2011. *La llengua amaziga a l'antiguitat a partir de les fonts gregues i llatines*. 2 vols. Barcelona: Promocions i Publicacions Universitàries.

Naït-Zerrad, Kamal. 1999, 2002. *Dictionnaire des racines berbères (formes attestées). II. C – DƐN, III. Ḍ – GƐY*. Paris, Louvain: Peeters.

Naumann, Christfried. 2012. *Acoustically based phonemics of Siwi (Berber)*. Cologne: Rüdiger Köppe.

Prasse, Karl-G. 1969. *À propos de l'origine de* h *touareg (tahaggart)*. Copenhagen: Munksgaard.

Prasse, Karl-G. 1972. *Manuel de grammaire touarègue (tăhăggart)*. 3 vols. Copenhagen: Akademisk Forlag.

Prasse, Karl-G. 1990. New light on the origin of the Tuareg vowels *e* and *o*. In Hans G. Mukarovsky (ed.), *Proceedings of the Fifth International Hamito-Semitic Congress*, vol. 1, 163-170. Vienna: Afro-Pub.

Prasse, Karl-G. 2011. Bilan sur les laryngales du protoberbère. In Amina Mettouchi (ed.), *« Parcours berbères », Mélanges offerts à Paulette Galand-Pernet et Lionel Galand pour leur 90^{e} anniversaire*, 85-96. Cologne: Rüdiger Köppe.

Prasse, Karl-G., Ghoubeïd Alojaly & Ghabdouane Mohamed. 2003. *Dictionnaire touareg–français (Niger). Tămažəq–Tăfrănsist (Niger) Alqamus*. Copenhagen: Museum Tusculanum Press.

Putten, Marijn van. 2014a. *A grammar of Awjila (Libya). Based on Paradisi's work*. Cologne: Rüdiger Köppe.

Putten, Marijn van. 2014b. Some notes on the historical consonantism of Awjila. *Folia Orientalia* 51. 257-274.

Putten, Marijn van. 2015. Reflexes of the Proto-Berber glottal stop in Nefusa and Ghadames. *Wiener Zeitschrift für die Kunde des Morgenlandes* 105. 303-314.

Putten, Marijn van. 2016. The origin of front vowel nominal prefixes in Berber. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 166(1). 11-39.

Putten, Marijn van. 2018. Proto-Berber mid vowel harmony. *Nordic Journal of African Studies* 27(3). 32.

Schiattarella, Valentina. 2016. *Berber texts from Siwa (Egypt), including a grammatical sketch*. Cologne: Rüdiger Köppe.

Souag, Lameen. 2011. Why *h1 and *h2 were not valid onsets in late proto-Berber. *Jabal al-Lughat*. http://lughat.blogspot.com/2011/04/why-h1-and-h2-were-not-valid-onsets-in.html.

Souag, Lameen. 2013. *Berber and Arabic in Siwa (Egypt). A study in linguistic contact*. Cologne: Rüdiger Köppe.

Souag, Lameen. 2017. La diffusion en berbère : réconcilier les modèles. *Mémoires de la Société de linguistique de Paris, nouvelle série* XXIV. 83-107.

Souag, Lameen & Marijn van Putten. 2016. The origin of mid vowels in Siwi. *Studies in African Linguistics* 45. 189-208.

Taine-Cheikh, Catherine. 1999. Le zénaga de Mauritanie à la lumière du berbère commun. In Marcello Lamberti & Livia Tonelli (eds.), *Afroasiatica Tergestina*, 299-324. Padova: Unipress.

Taine-Cheikh, Catherine. 2001-2003. La corrélation de gémination consonantique en zénaga (berbère de Mauritanie). *Comptes rendus du GLECS* (séances 1998-2002) XXXIV. 5-66.

Taine-Cheikh, Catherine. 2004. Les verbes à finale laryngale en zénaga. In Kamal Naït-Zerrad, Rainer Voßen & Dymitr Ibriszimow (eds.), *Nouvelles études berbères. Le verbe et autres articles*, 171-190. Cologne: Rüdiger Köppe.

Taine-Cheikh, Catherine. 2005. Les marques de 1ère personne en berbère. Réflexions à partir du zénaga. In Alessandro Mengozzi (ed.), *Studi Afroasiatici. XI Incontro Italiano di Linguistica Camitosemitica*, 97-112. Milan: Franco Angeli.

Taine-Cheikh, Catherine. 2006. Alternances vocaliques et affixations dans la morphologie nominale du berbère : le pluriel en zénaga. In Dymitr Ibriszimow, Rainer Voßen & Harry Stroomer (eds.), *Études berbères III. Le nom, le pronom et autres articles*, 253-268. Cologne: Rüdiger Köppe.

Taine-Cheikh, Catherine. 2008. *Dictionnaire zénaga–français*. Cologne: Rüdiger Köppe.

Vycichl, Werner. 1975. Begadkefat im Berberischen. In James Bynon & Theodora Bynon (eds.), *Hamito-Semitica*, 315-317. The Hague, Paris: Mouton.

Vycichl, Werner. 1990. Die Palatalisiering von Q im Berberischen. *Rivista degli Studi Orientali* 63. 39-43.

Vycichl, Werner. 1992. Les terminaisons du pluriel externe des noms berbères. Étude sémantique et historique. In Jeannine Drouin & Arlette Roth (eds.), *À la croisée des études berbères. Mélanges offerts à Paulette Galand-Pernet et Lionel Galand*, 255-268. Paris: Geuthner.

2

Harmony and disharmony in Mbat (Jarawan Bantu) verbs

Christopher R. Green
Syracuse University

Abstract

This paper is the first to describe aspects of the vocalic phonology of Mbat, a Jarawan Bantu language. Mbat exhibits a series of vowel-consonant interactions in its verbs that sometimes yield height harmony between a stem and suffixal vowel. Via this stem-controlled phenomenon, high vowels (i, u, ɪ, ʊ) harmonize across a stem-final non-dorsal sonorant while the low vowel (a) harmonizes across any stem-final dorsal. Under other conditions, these verbs appear disharmonic for height. A sixth contrastive vowel, schwa (ə), does not actively participate in harmony. I show that these otherwise straightforward generalizations on harmony vs. blocking are sometimes obscured by alternations triggered by a preceding glide that affect the stem vowel itself. I offer an analysis using a feature geometric model of vowel height. I show that an approach based on well-motivated binary vocalic features like [open], [closed], and [ATR] offers a transparent account of most Mbat outcomes. There is at least one instance, however, where these features seem unintuitive relative to the phenomena being modeled. For the sake of comparison, I discuss a possible reanalysis based on abstract features. Such an approach is unencumbered by expected phonetic correlates of vocalic features and focuses instead on featural interactions. This approach aligns itself with more recent "substance-free" approaches to phonology which assume a model of phonological computation based on features whose phonetic implementation is downstream and language-specific.

Keywords

feature geometry, Jarawan Bantu, vowel features, vowel harmony

Résumé

Cet article est le premier à décrire certains aspects de la phonologie vocalique du mbat, une langue bantoue jarawan. Le mbat présente un ensemble d'interactions voyelle-consonne au sein des verbes qui génère parfois une harmonie de hauteur entre un radical et une voyelle suffixale. Via ce phénomène contrôlé par la racine, les voyelles hautes (i, u, ɪ, ʊ) s'harmonisent à travers une sonante non dorsale située en fin de radical, alors que la voyelle basse (a) harmonise à travers toute dorsale située dans cette même position ; dans toute autre condition, les verbes semblent disharmoniques pour la hauteur. Une sixième voyelle contrastive, le schwa (ə), ne participe pas activement à l'harmonie. Je montre que ces généralisations, de prime abord simples sur l'harmonie *vs* l'opacité, sont parfois obscurcies par des alternances déclenchées par une glissante antéposée affectant la voyelle du radical elle-même. Je propose une analyse élaborée à l'aide d'un modèle géométrique des traits de hauteur vocalique. Je montre qu'une approche fondée sur des traits vocaliques binaires bien motivés comme [ouvert], [fermé] et [ATR] rend compte de manière transparente de la plupart des formes de sortie du mbat, mais qu'il y a cependant au moins un cas pour lequel ces traits semblent peu intuitifs à l'égard des phénomènes modélisés. À titre de comparaison, je discute une nouvelle analyse potentielle construite à partir de traits abstraits. Une telle approche n'est pas bloquée par les corrélats phonétiques attendus des traits vocaliques et se concentre plutôt sur leurs interactions. Cette approche est en accord avec les approches « sans substance » de la phonologie les plus récentes, qui supposent un modèle de calcul phonologique fondé sur des traits dont l'implémentation phonétique se situe en aval et est spécifique à la langue.

Mots clés

géométrie des traits, harmonie vocalique, langue bantoue jarawan, traits vocaliques

---

1. Introduction

Linguists know fairly little about the grammar of the approximately 20 Jarawan (or Jarawan Bantu) languages, a cluster of language spoken across portions of Nigeria and Cameroon. This excludes what has been reported about their lexicon in an array of historical and

typological studies conducted over the last several decades (see, for example, Blench 2006; 2015; Gerhardt 1982; Grollemund 2012; Grollemund & Hombert 2012; Grollemund *et al.* 2015; Maddieson & Williamson 1975; Piron 1995; 1997; 1998; Shimizu 1983). These works illustrate that a sizable portion of the core Jarawan lexicon is comprised of Bantu cognates but that it is also influenced by contact with Chadic and Bantoid languages. Beyond these studies, there is only: (i) a Master's thesis (Yabilsu 1991) on one variety (Galamkya), (ii) one short descriptive paper (Gerhardt 1988) that deals with aspectual morphology in Jarawan verbs, and (iii) one additional paper (Green 2020) concerned with formalizing details of syllable structure in Mbat, the same Jarawan variety that is the focus of the current paper.

The description of Jarawan verbs in Gerhardt (1988) serves as an inspiration of sorts for the current paper. Gerhardt's paper discusses Perfective and Habitual "verbal extensions" in two Jarawan varieties, namely Jaar and Kantana. He illustrates patterns of phonologically-conditioned suffixal allomorphy where stem shape dictates the choice of one of three allomorphs of the Perfective and Habitual extensions. Though they are mentioned only in passing, Gerhardt's data show complex alternations that arise in both stem and suffixal vowels. Similar alternations are found in Mbat.

In this paper, I present data that I have collected on Mbat, a Jarawan language spoken by approximately 40,000 individuals (Eberhard *et al.* 2019), primarily in Bauchi State, Nigeria. Mbat verbs inflect for the same Perfective and Habitual extensions reported by Gerhardt and likewise exhibit phonologically-conditioned suffixal allomorphy based on stem shape. Mbat manifests similar, but not identical alternations to those seen in Gerhardt's Jaar in its stem and suffixal vowels. In Mbat, there are three realizations of the Perfective (*-m*, *-Vm*, *-ma*) and three corresponding realizations of the Habitual (*-n*, *-Vn*, *-na*). As I illustrate in Section 3, the first realization in each is limited to short CV stems while the third occurs with "heavy" CVVC and disyllabic stems. Of particular importance in this paper is the behavior of verbs with C(C)VC stems that select the -VC form of these extensions. My use of a placeholder V here is to represent the fact that the suffixal vowel has different surface realizations, the choice between which depends primarily on characteristics of the stem, namely the stem vowel and stem-final consonant.

Mbat's suffixal vowels are the target of stem-controlled alternations affecting vowel height. Under well-defined and predictable conditions, there is height harmony between a stem and suffixal vowel. The quality of stem vowels themselves is sometimes obscured by alternations triggered by a preceding glide. But, what I would argue is their basic form is apparent elsewhere. A stem vowel triggers

harmonization of a suffixal vowel, with stem vowels of different heights (high vs. low) doing so only across particular consonants. The low stem vowel [a] harmonizes a suffixal vowel only across a dorsal consonant, be it a stop or sonorant. A high stem vowel [i, ɪ, u, ʊ], on the other hand, harmonizes a suffixal vowel only across a non-dorsal sonorant. In the presence of a blocker (i.e., an incompatible intervening consonant), harmony does not occur. In blocking contexts, a word's stem and suffixal vowels are disharmonic for height. Further complicating matters is that derived high stem vowels do not trigger harmonization of a suffixal vowel, thus resulting in an apparent underapplication of the process.

Given the essentially undocumented status of Jarawan languages, the primary goal of this paper is to describe the patterns of vowel alternations that arise in Mbat. In doing so, I also aim to add to the literature on vowel harmony systems by discussing interactions between vocalic triggers and transparent consonants vs. blocker consonants. To model these data, I adopt an autosegmental approach that employs geometrically-arranged features. My analysis proposes that harmony is strictly local such that a stem and suffixal vowel can harmonize only when the stem vowel and stem-final consonant also share the same feature. This entails that the stem-final non-dorsal sonorants and stem-final dorsal consonants that facilitate harmony are associated with certain vocalic features. When the stem vowel and stem-final consonant do not share the requisite feature, harmony is blocked.

The approach I take to modeling Mbat's vocalic alternations and interactions between stem vowels and stem-final consonants assumes many, but not all of the hallmark characteristics of standard feature geometry (Clements 1985; 1991a; Halle 1995; Sagey 1986). I utilize a closed set of binary geometric features—[open], [closed], and [ATR]—organized under what I call the Vowel Manner (V-Manner) node. I motivate the arrangement of these features based on the ways that the segments that they comprise interact with and affect one another within an Mbat word. Odden (1991) has called a similar constituent the Vertical Movement node and instead employs the features [low], [high], and [ATR], respectively. I also appeal to a sub-constituent Aperture node composed of [closed] and [ATR] based on the fact that these often pattern together in operations involving feature spreading. I show that this standard approach to feature geometry offers a transparent means by which to represent most of the Mbat facts. However, the Mbat outcomes sometimes present an analytical challenge given standard assumptions concerning the cross-linguistic patterning of these features and their expected phonetic implementation.

As one potential means of overcoming this challenge, I briefly entertain an alternative analysis based on abstract features. Doing so allows one to focus on featural interactions that are unencumbered by expected or assumed phonetic correlates. Clements (1991b) employs such an approach in modeling vowel height alternations in several canonical Bantu languages. He does so based on a single feature, [open], arranged into several "registers". Such an analysis is reminiscent of more recent "substance-free" approaches to feature geometry like the Parallel Structures Model (PSM) (Morén 2003a; 2003b; 2006; 2007). The PSM posits that consonants and vowels are comprised of identical constellations of geometrically-organized phonological features whose relationship to one another and whose phonetic realization must be established language-specifically. I do not concern myself with broader conceptual issues related to the "substance-free" approaches to phonology, but see Blaho (2008) for an overview and critique of various proposals.

The remainder of this paper is organized as follows: Section 2 provides an introduction to various phonological and morphological properties of Mbat. These facts represent my understanding of Mbat grammar based on approximately 18 months of data collection, though the larger documentation project is still underway. Section 3 discusses suffixal patterns found in Mbat verbs inflected for the Perfective and Habitual, and illustrates the various vocalic alternations that arise depending on the particular combination of stem and suffixal vowels. Section 4 presents a feature geometric analysis of the Mbat alternations. Section 5 offers discussion, an alternative analysis of one particular complex matter that arises in the language, and concluding remarks.

2. Background on Mbat

Mbat (ISO 639-3: bau, Glottocode: bada1258) is also called Bada or Badanchi, though it is locally known as Jar. Mbat and the other approximately 20 Jarawan (or Jarawan Bantu) languages are not well-represented in the descriptive and theoretical linguistics literature. Historical linguists and typologists have been keenly interested in their lexicon and what it can reveal about the place of Jarawan languages alongside A-group Narrow Bantu vs. Southern Bantoid. From a grammatical perspective, even the most basic characteristics of these languages have yet to be reported, with just a few notable exceptions. Gerhardt (1988) discusses verbal extensions in "Jaar" and Kantana (ISO 639-3: mma, Glottocode: mama1272). It is unclear what variety is represented by Gerhardt's "Jaar", however, given that speakers of many varieties refer to their language as Jar or Jaar; see Maddieson & Williamson (1975). In addition, there is an unpublished

MA thesis on aspects of Galamkya (also assigned ISO 639-3: bau, Glottocode: bada1258) by Yabilsu (1991).[1] Based on lexical similarities alone, Mbat's closest sibling appears to be Duguri (ISO 639-3: dbm, Glottocode: dugu1249).[2]

The data in this paper were collected in person via direct elicitation at Syracuse University (Syracuse, NY) from a 31-year-old mother tongue speaker of Mbat who was raised in Tadnum village, Bogoro LGA, Bauchi State, Nigeria. She spent 29 years in Nigeria before leaving the country to pursue graduate education in the US. In addition to Mbat, she is a fluent L2 speaker of both Hausa and English. The data represent her speech, but we have confirmed them via phone with friends and members of her family back home in Nigeria. The findings that I present below are representative of approximately 18 months of collaboration with this speaker, though the larger documentation project is very much work in progress.

Mbat and its closest siblings are called Jarawan Bantu, but these languages share few immediately apparent grammatical similarities with "canonical" Bantu languages. Rather, in many ways, they seem to share more characteristics with Bantoid languages. For example, words other than borrowings are seldom more than two syllables in length, and their morphology has only a few instances of syntheticity. There is no functioning system of noun classes, and distinctions within the verbal system are aspectual, with little evidence of tense. For more discussion of grammatical comparisons between Bantu and Bantoid, see Hyman (2017; 2018).

Table 1 — Consonant phoneme inventory

	Labial	Alveolar	Post-Alv./ Palatal	Velar	Glottal
Plosive	p b	t d		k g	(ʔ)
Implosive	ɓ	ɗ			
Affricate			dʒ	kx gɣ	
Nasal	m	n	ɲ	ŋ	
Fricative	f (β)	s z	ʃ		
Approximant	w	l r	j		

1. I have been in contact with the author, and even she does not have an electronic copy of her thesis available to share, so I have been unable to evaluate its contributions.

2. I am grateful to Rebecca Grollemund for kindly computing cognate percentages from my Mbat data against other Jarawan, Bantu, and Southern Bantoid languages in her database. It is upon these percentages that I base this close connection between Mbat and Duguri.

This paper is concerned primarily with Mbat's vowels and their interaction with consonants. Based on the data that I have collected thus far, Mbat's inventory of consonant phonemes is as presented in Table 1. I have found that any of these consonants can appear in the onset of a CV or CVC syllable, though glottal stop /ʔ/ does so only in non-word-initial syllables. The bilabial fricative /β/ appears only in a few items in my database and may only be marginal.

The language has a three-way stem-initial contrast in labial and alveolar voiceless vs. voiced vs. implosive stops, but only voiceless stops are found word-finally. Stem-final consonants are particularly susceptible to alternation. For example, word-finally, coronals are either palatalized or retroflexed regardless of their manner (/n/ ~ [ɲ], /l/ ~ [l̪], /r/ ~ [ɽ], /t/ ~ [t̪]), and /k/ is realized [q]. When followed by a vowel-initial suffix, stem-final stops alternate with implosives: /p/ ~ [ɓ], /t/ ~ [ɗ], though /k/ alternates with the uvular stop [ɢ] or sometimes instead the uvular fricative [ʁ]. I assume that the affricates /kx/ and /gɣ/ listed above are phonemic based on their word-initial realization. Their intervocalic realization varies in ways that are yet unclear, sometimes with velar fricatives [x] and [ɣ], but other times with uvular fricatives [χ] and [ʁ]. Affricates, implosives, glides, and fricatives other than /s/ do not occur in syllable codas.

The only consonants that can occupy the second onset position of a CCVC syllable are approximants, specifically /j/, /w/, and /l/. Though the glides /j/ and /w/ often appear in a singleton onset, Green (2020) has shown that when they are the second member of a complex onset in a CCVC verb stem, they are involved in an unusual alternation (cf. further below). The data in this paper are presented phonetically unless otherwise indicated.

Most pertinent to the current paper are characteristics of the language's vowel system. Based on the Mbat data that I have collected thus far, the language appears to make a phonemic contrast between six vowels: /i, ɪ, u, ʊ, ə, a/.[3] These six contrastive vowels occur in what is arguably the language's basic word shape, namely in CVC stems (1).

(1) Words with CVC stems

a.	bíl̪	'follow'	b.	ɓīl̪	'give birth'
c.	kúm	'find'	d.	gús	'wash'
e.	zùp	'cover'	f.	ɓūq	'throw'
g.	pūt̪	'go out'	h.	ŋə̀m	'cry'
i.	ɓál̪	'count'	j.	kə̀s	'cut'
k.	jáp	'climb'	l.	ʃāq	'chew'
m.	ját̪	'fetch'	n.	kāŋ	'fry'

3. The vowel inventory presented here differs from that in Green (2020) in that it establishes contrastive /ə/. Green recognized the presence of [ə] in the Mbat inventory, but its phonemic status was unclear at the time. Based on data collected since that time, there is now ample evidence to distinguish between /ə/ and /a/.

A subset of these vowels, /i, u, a/, have a broader distribution in that they also occur in CV stems (2). *Cə* also occurs, but I have found it only in function words in my database. For example, the progressive auxiliary is *bə*, and the relativizers are *mə/ɓə*, though the latter are bimorphemic in that they inflect for singular vs. plural number, respectively.

(2) Words with CV stems

a.	pà	'give'	b.	ɓì	'him/her'
c.	tù	'pluck'	d.	gú	'millet'
e.	sá	'will (n.)'	f.	zù	'judge'
g.	mí	'me'	h.	nù	'drink'
i.	lì	'eat'	j.	ra	negative particle
k.	jú	'come'	l.	wù	'die'

The vowels that occur in CCVC stems are restricted (3). I have found no high vowels (tense or lax) in this frame, but rather lax mid vowels [ɛ] and [ɔ] appear here, as does the low vowel [a].

(3) Words with CCVC stems

a.	zwɔ̄ʈ	'sew'	b.	kjɛ̄ʈ	'sweep'
c.	vwɔ̄ʈ	'take forcefully'	d.	swɔ́ʈ	'hide'
e.	fwɔ̄ʈ	'peel'	f.	swɔ́p	'pour'
g.	twɔ̄p	'wash'	h.	mjɛ̄ʈ	'kill'
i.	ŋ̄gláp	'woman'	j.	tjáq	'trek'
k.	kwáʈ	'bowl'	l.	kwáŋ	'spear'
m.	jwàq	'snake'	n.	ljáŋ	'iron'
o.	mjáp	'spatula'	p.	pwàɭ	'cold'

The behavior of verbs like those in (3) is discussed at length in Green (2020). According to the analysis therein, verbs of this shape with mid vowels undergo unusual alternations when a vowel-initial suffix like the Perfective or Habitual is added to them. Upon suffixation, the stem-final consonant is syllabified into an onset. As a result, the stem glide is deleted, and the stem vowel ɛ/ɔ alternates with its high counterpart ɪ/ʊ. For example, [zwɔ̄ʈ] 'sew' becomes [zū.ràm] 'sewn' when the Perfective extension is added to it. An analogous alternation can be seen for [kjɛ̄ʈ] 'sweep' and [kīɗ-ám] 'swept'. Similarly shaped stems with [a] do not undergo either alternation: cf. [kwāq] 'drive' vs. [kwāɢ-ám] 'driven'. As Green discusses, both in these instances and elsewhere in Mbat, the mid vowels [ɛ] and [ɔ] appear only after glides. He argues that their alternation with [ɪ] and [ʊ], respectively, upon the loss of a preceding glide, suggests that [ɛ] and [ɔ] are allophones of /ɪ/ and /ʊ/. These "lax" mid vowels do not alternate with their "tense" counterparts (i.e., [e] and [o]). Indeed, the latter vowels are not found in Mbat's inventory. I will have more to say about this particular issue below.

The details of Mbat's tonal system are still a work in progress. At present, I have identified three main surface tone levels—High, Mid,

and Low—and have marked tone on the data in this paper accordingly. In addition, the Habitual suffix appears lowered after a stem Low tone. I transcribe this as a Super-Low tone, but I have not yet found evidence for Super-Low more broadly in the language. That said, there is some (albeit preliminary) evidence that the language makes a four-way underlying tonal contrast, with the addition of toneless tone bearing units. One fact in support of this observation is that nouns surfacing with High tone in isolation can be divided into two classes based on how they interact with surrounding High tones. One group is susceptible to spreading from an adjacent H tone while the other does not participate in such spreading. For the current paper, the two extensions to be discussed (Habitual and Perfective) behave differently from a tonal perspective, but this varies and appears not to bear on the vocalic alternations under consideration.

3. Alternations in Perfective and Habitual verbs

Mbat behaves in a manner similar to its cousins discussed in Gerhardt (1988) in that it exhibits phonologically-conditioned patterns of suffixal allomorphy that depend on stem shape. Of particular interest are verb stems that select a vowel-initial suffix. These allow us to witness vocalic phenomena that I argue are, at least in part, attributable to vowel height harmony. I begin this section by establishing the patterns of suffixal allomorphy seen in Mbat.

3.1 Suffixes without vowel alternation

Examples of the simplest C(G)V verb stems (where G stands for a glide) are in (4). These stems select *-m* for the Perfective and *-n* for the Habitual. Mbat verbs do not inflect for person, number, or gender. Person and number are encoded, where relevant, via pronouns or through number marking on nouns.

(4) Perfective and Habitual verbs with C(G)V stems

	Stem	Perfective	Habitual	
a.	pà	pà-m	pà-n	‘give’
b.	tù	tù-m	tù-n	‘pluck’
c.	jú	jú-m	jú-n	‘come’
d.	lì	lì-m	lì-n	‘eat’
e.	tā	tā-m	tā-n	‘sow seed’
f.	kwā	kwā-m	kwā-n	‘enter’
g.	swá	swá-m	swá-n	‘pierce’
h.	ɓwà	ɓwà-m	ɓwà-n	‘carve’

Verbs with C(G)VVC stems in (5) instead select *-ma* for the Perfective and *-na* for the Habitual. The same allomorphs are selected by the disyllabic verb stems in (6).

(5) Perfective and Habitual verbs with C(C)VVC

	Stem	Perfective	Habitual	
a.	māās	māās-má	māās-nā	‘ask’
b.	mííŋ	míín-má	míín-nā	‘blow nose’
c.	lààm	lààm-má	lààm-nā	‘cook’
d.	dìḭ̀l	dììl-má	dììl-nā	‘know’
e.	gɣūūm	gɣūūm-má	gɣūūm-nā	‘yawn’
f.	fwāā̰t	fwāāt-má	fwāāt-nā	‘vomit’
g.	ljāā̰t	ljāāt-má	ljāāt-nā	‘paste’
h.	tjáá̰q	tjáá̰q-má	tjáá̰q-nā	‘trek’

(6) Perfective and Habitual verbs with disyllabic stems

	Stem	Perfective	Habitual	
a.	lāɣā̰t	lāɣāt-má	lāɣāt-nā	‘taste’
b.	kxəməə̰t	kxəməət-má	kxəməət-nā	‘squeeze’
c.	lāŋās	lāŋās-má	lāŋās-nā	‘lick’
d.	gɣāgɣā̰t	gɣāgɣāt-má	gɣāgɣāt-nā	‘open’
e.	ɲūŋwā̰l	ɲūŋwāl-má	ɲūŋwāl-lā	‘write’

In the next sections, I describe various outcomes in Mbat verbs whose stem shapes are CVC or CGVC. Both stem shapes select a vowel-initial allomorph of the Perfective (*-Vm*) and Habitual (*-Vn*). I show that the surface quality of the suffixal vowel depends on both the quality of the stem vowel and the nature of the intervening stem-final consonant. The data are organized based on the quality of the vowel in the uninflected stem. I take these forms as basic given that stem vowels are sometimes susceptible to alternation upon inflection, though in predictable ways.[4]

3.2 Alternations with high vowel stems

The examples in (7) and (8) show verbs whose uninflected stem contains a high vowel. These stem vowels represent Mbat’s four contrastive high vowels, being either front or back, tense or lax. Beginning with (7), these verbs have stem and suffixal vowels of the same quality. They are unique in that their stem-final consonant is a non-dorsal sonorant, be it a nasal or liquid.

(7)

	Stem	Perfective	Habitual	
a.	nùm	nùm-ùm	nùm-ṵ̀n	‘bite’
b.	kúm	kúm-úm	kúm-ūn	‘find’
c.	bíl̰	bíl-ím	bíl-īn	‘follow’
d.	pūr̰	pūr-ùm	pūr-ùn	‘go out’
e.	sḭ̄ŋ	sīn -ìm	sīn-ìn	‘see’

4. There are no tonal alternations affecting the stem upon suffixation, though the tone associated with the suffixal vowel alternates depending on the tone of the stem vowel. These alternations differ in some instances between the two suffixes.

Verbs with a high stem vowel in (7) differ from those in (8) in that they select *-əm/-ən* suffixes. The factor corresponding to this difference is that their stem-final consonant is not a member of the aforementioned class of non-dorsal sonorants.

(8)		Stem	Perfective	Habitual	
	a.	zùp	zùɓ-əm	zùɓ-ən	‘cover’
	b.	gūp	gūɓ-əm	gūɓ-ən	‘close’
	c.	gús	gús-ám	gús-ān	‘wash’
	d.	līp	līɓ-əm	līɓ-ən	‘moisten’
	e.	nùq	nùɢ-əm	nùɢ-ən	‘sit’

The analysis that I present below proposes that the compatibility of stem-final sonorants to full height vowel harmony in (7) is due to them sharing the same height feature, [+closed], with high vowels. Consonants not specified [+closed] do not permit harmonization via this feature. There is no direct evidence that a feature like [ATR] is associated with stem-final consonants, as it does not intervene in the process. Both [+ATR] and [–ATR] vowels harmonize across a [+closed] stem-final consonant, as in (7). The propensity for high vowel stems to select *-əm/-ən* instead of *-am/-an* in (8) may also have a featural explanation.[5] I discuss this further in Section 4.2.

3.3 Alternations with [a]- and [ə]-stems

As was the case for verbs with high stem vowels, there are verbs with stem [a] that realize a harmonization of sorts with their suffixal vowel. The verbs in (9) show uninflected stems with [a] that select suffixal [a]. These contexts have in common that their stem-final consonant is a dorsal, be it either a nasal or oral stop.

(9)		Stem	Perfective	Habitual	
	a.	ɓāŋ	ɓāŋ-àm	ɓāŋ-àn	‘blow (wind)’
	b.	ʃāq	ʃāɢ-àm	ʃāɢ-àn	‘chew’
	c.	wàq	wàɢ-àm	wàɢ-àn	‘hear’
	d.	kāŋ	kāŋ-àm	kāŋ-àn	‘fry’

These can be directly compared to verbs in (10) whose uninflected stem contains [a] but where the stem vowel alternates to [ə] following suffixation. Unlike the verbs in (9), these have a stem-final consonant that is not a dorsal.

(10)		Stem	Perfective	Habitual	
	a.	dʒàm	dʒəm-àm	dʒəm-àn	‘stop’
	b.	kām	kəm-àm	kəm-àn	‘teach’
	c.	gɣāl̩	gɣəl-àm	gɣəl-àn	‘find’
	d.	ɓál̩	ɓól-ám	ɓól-ān	‘count’

5. One verb in my database, *ɓīŋ* ‘dance’, behaves differently when an alternating coda nasal is involved. Here, the Perfective is *ɓīn-àm*, and the Habitual is *ɓīn-àn*. This likely arises because harmony is predicated on the featural specification of the dorsal nasal rather than on that of the alternant.

The analysis that I present below entertains the possibility that the underlying quality of the suffix vowel is /a/. Based on this, the analysis proposes that stem-final dorsals in (9) facilitate harmony between stem and suffix [a] because they share the same height feature, [+open], with [a]. Stems with non-dorsal final consonants in (10), on the other hand, are not compatible with the creation of a single harmonic span across the word. What is interesting in the latter case is that, rather than maintaining two separate [+open] vowels in the same word, Mbat resorts to dissimilation of the stem vowel to [ə]. The same surface outcome arises in verbs like those in (11) whose uninflected stem contains [ə]. These verbs select *-am/-an* and do not undergo alternation. This outcome suggests that [ə] differs minimally from [a], perhaps in that it is [–open].[6]

(11)		Stem	Perfective	Habitual	
	a.	ŋə̀m	ŋə̀m-àm	ŋə̀m-ə̀n	‘cry’
	b.	kə̀s	kə̀s-àm	kə̀s-ə̀n	‘cut’
	c.	ɗə̀p	ɗə̀ɓ-àm	ɗə̀ɓ-ə̀n	‘pick/carry’

There are other alternations that affect a subset of *CaC* stems whose stem-initial consonant is a glide (12). These verbs have stem [a] when uninflected, however, the stem vowel surfaces either as [i] or [u] upon inflection by the Perfective and Habitual suffixes. Stem [i] is found after [j], while stem [u] is found after [w].

(12)		Stem	Perfective	Habitual	
	a.	wáɻ	wúl-ám	wúl-ān	‘laugh’
	b.	jàʈ	jìɗ-àm	jìɗ-ə̀n	‘love’
	c.	wàɻ	wùl-àm	wùl-ə̀n	‘quench’
	d.	jáʈ	jíɗ-ám	jíɗ-ān	‘fetch’

Given that these alternations arise only in inflected forms suggests that /a/ > [ə] dissimilation of the stem vowel, as in (10), occurs before other features are contributed by spreading from the onset. Important to the matter of interest in this paper is that the result may yield opacity. In (12a) and (12c), for example, one might otherwise expect harmonization of a high stem vowel across a stem-final liquid.

3.4 Alternations with mid vowel stems

The verbs in (13) are unique in that their stem shape is CGVC. As introduced above, CGVC is the only stem shape in which the mid vowels [ɛ] and [ɔ] occur. When the stem-final consonant of such a verb is resyllabified into an onset before a vowel-initial suffix like the Per-

6. A reviewer asks about the status of [ə] stem vowels before a dorsal. As similarly noted in Green (2020) concerning mid vowels, there are no instances of [ə] before a stem-final dorsal in my data. Thus, in the case of stem /a/ and /ə/, the contrast between them appears to be neutralized in this environment.

fective and Habitual, their pre-vocalic glide is lost, and the stem vowel raises to its high counterpart. Stem [ɔ] alternates to [ʊ], and stem [ɛ] alternates to [ɪ].

(13)		Stem	Perfective	Habitual	
	a.	zwɔ̰	zūr-àm	zūr-àn	'sew'
	b.	vwɔ̰	vūɗ-àm	vūɗ-àn	'take forcefully'
	c.	swɔ́	súr-ám	súr-ān	'hide'
	d.	twɔ̄p	tūɓ-àm	tūɓ-àn	'wash'
	e.	zwɔ̰	zūɗ-àm	zūɗ-àn	'lose'
	f.	kjɛ̰̄	kīɗ-àm	kīɗ-àn	'sweep'
	g.	mjɛ̰̄	mīr-àm	mīr-àn	'kill'

These verbs reveal that, in a way similar to the verbs in (12), derived high vowels cannot trigger harmony. For example, we might expect, based on comparison to (7), that Perfective and Habitual forms like (13a), (13c), and (13g) would result in full harmonization across a stem-final non-dorsal sonorant. However, **zūr-ùm*/*zūr-ùn*, **súr-úm*/*súr-ún*, etc. do not occur. Likewise, even across a non-harmonizing stem-final consonant, we might expect a verb like (13b) to select suffixal *-əm/-ən*, as did the verbs in (8). Again, however, forms like **vūɗ-ùm*/*ūɗ-ùn* do not occur. These outcomes show that: (i) derived high vowels cannot act as harmony triggers, and (ii) mid vowel stems select *-am/-an*. Thus, mid vowel stems pattern with verbs whose stems contain [a] and [ə].

3.5 Perfective and Habitual verb summary

The alternations seen above raise an important question concerning the underlying quality of the *-Vm/-Vn* suffixal vowel. In order to begin to address this, the following list summarizes the basic realizations of these suffixes. They are:

- *-im/-in*, *-ɪm/-ɪn*, *-um/-un*, *-ʊm/-ʊn* in harmonizing contexts after a high stem vowel of the same quality;
- *-əm/-ən* in non-harmonizing contexts after a high stem vowel of any quality;
- *-am/-an* elsewhere (including after stem vowels *ə*, *ɛ*, and *ɔ*).

Based on distribution alone, it might be reasonable to posit *-am/-an* as the basic realization of the Perfective and Habitual suffixes. This is supported by at least two other factors. First, *-am/-an* would be transparent variants of the *-ma/-na* suffixes selected by "heavy" stems in (5) and (6). The vowels of these -CV suffixes do not alternate. Second, *-am/-an* would similarly be aligned with what Gerhardt (1988) proposes for the Jarawan varieties in his study. He states that *-aC* suffixes are basic and have a wider distribution than other suffixal

variants. He describes *-əC* as an "unexpected" or "reduced" variant. While it is true that the *-əC* variants are less widely distributed in Mbat, the appearance of *-əC* is entirely predictable, rather than being unexpected.[7]

What should be clear is that non-*am*/*-an* suffixal variants are selected only by high vowel stems. One way to view this might be to assume that *-am*/*-an* is basic and that the suffixal vowel raises after high vowels: raising to *-əm*/*-ən* is partial in a non-harmonizing context and complete in a harmonizing context across a non-dorsal sonorant. One matter to be explored is how to account for such a possibility from a featural perspective. Another possibility would be to propose that the suffixal vowel is underspecified for height. Under such a view, high vowel stems would select either the "partial" or "full" variant depending on the nature of the intervening stem-final consonant. The *-am*/*-an* variants would represent the elsewhere condition.

Either of these approaches seem tenable and entail a certain set of reasonable assumptions. Based on the facts taken together, however, I will assume *-am*/*-an*. A third possibility suggested by an anonymous reviewer is to treat *-əm*/*-ən* as basic, as a means to highlight or to reinforce the harmonizing nature of the alternations. This might be possible but would entail height alternations in two directions. The approach that I suggest based on *-am*/*-an* would instead implicate a singular harmonic raising imperative after high vowel stems whose degree differs according to context. This third alternative might also struggle to explain why *-am*/*-an* is selected after stem *ə* rather than non-alternating *-əm*/*-ən*.

4. Applying feature geometry to Mbat's vowels

Feature geometry, like other autosegmental approaches to phonology, provides an elegant means by which to model interactions and dependencies that features associated with adjacent and even sometimes non-adjacent segments have on one another. One advantage particular to feature geometry is that features that function or move together can be analyzed as forming a constituent or node within the feature tree. As discussed above, Mbat's stem high vowels have the ability to harmonize a suffixal vowel only when they are followed by a stem-final non-dorsal sonorant. Such a consonant acts as transparent to feature spreading from the stem vowel onto the suffix. In an analogous way, the low stem vowel [a] and dorsal consonants have a bearing on one another such that in the presence of a stem-final

7. Another possible factor pertains specifically to the Perfective. The Perfective almost certainly derives from Proto-Bantu **mad* 'finish', whose cognate synchronically in Mbat is *mal* 'finish'. Both contain [a].

dorsal, stem [a] and suffixal [a] do not alternate. To model these and other effects, I assume that consonants of some types have height features akin to those uncontroversially ascribed to vowels. This is not a significant departure from other feature-based approaches which often propose that dorsal consonants, for example, are [high] vs. [low] or [front] vs. [back]. In other works, sonorants have also been described as associated with typical vocalic features like [ATR] (see, for example, Carnie 2002; Uchihara & Báez 2016). Geometric models like the Parallel Structures Model (PSM) of feature geometry (Morén 2003a; 2003b) assume that vowels and certain consonants make use of the same constellation of geometrically-organized vowel manner (i.e., height) features.

The analysis that I present below proposes that vocalic manner features associated with some consonants are key to understanding the alternations in Mbat's Perfective and Habitual verbs. The behavior of stem-final dorsal consonants suggests that their ability to permit low vowel harmony relates to a shared featural specification between them and low vowels. Likewise, the behavior of stem-final non-dorsal sonorants and their ability to permit high vowel harmony suggests that they share some featural specification with high vowels. In this section, I model Mbat's vowel system with three fairly uncontroversial binary features—[open], [closed], and [ATR]—that pertain to vowel manner, i.e., height. The features [open] and [closed] could easily be restated as [low] and [high], respectively. To differentiate vowel place, i.e., backness, I use the feature [dorsal]. Based on these features, and upon the patterning of vowels in the data above, I propose the feature specifications in Table 2 for Mbat's six contrastive vowels.

Table 2 — Featural specification of Mbat vowels

	V-Place	V-Manner		
	[dorsal]	[open]	[closed]	[ATR]
i	–	–	+	+
ɪ	–	–	+	–
u	+	–	+	+
ʊ	+	–	+	–
ə	–	–	–	
a	–	+	–	

Some preliminary observations can be made based on these proposed featural specifications for Mbat's six contrastive vowels that are predictive of other outcomes. First, based on how these vowels behave, [ATR] appears to be a subsidiary feature that patterns

together with [closed]. Only a [+closed] vowel can be specified [+ATR]. Geometrically-speaking, this implies that [closed] and [ATR] are resident within the same node or constituent. Such a finding is not unique cross-linguistically. I refer to the node containing [closed] and [ATR] as Aperture, though other terms like Height (Odden 1991) or Vertical Movement (Green & Hantgan 2019) have been proposed for such a constituent that might also be appropriate. Second, and more important, is that a vowel can be specified either as [+open] or [+closed], but there are no vowels that are positively specified for both features. The features [open] and [closed] are precisely those that I propose are involved in word level vowel harmony. There is evidence that I entertain below that the incompatibility of [+open] alongside [+closed] extends beyond individual segments. Such a combination of features appears to be avoided at the level of the word. While these feature specifications pertain to Mbat's contrastive vowels, there is more to consider about the properties of the language's mid vowels, [ɛ] and [ɔ], before offering more detail concerning their associated features. Recall that these vowels pattern as if they are allophones of /ɪ/ and /ʊ/, respectively.

The analysis below attributes low vowel/dorsal consonant interactions to Mbat's preference for a single [+open] span within a word. In instances where such a span is not possible, and two instances of the feature might instead appear within the same word, the result is disharmony. In the case of /a...a/ across a non-dorsal stem-final consonant, there would be two vowels specified [+open]. The language responds by dissimilating the first vowel to [–open], yielding [ə]. Analogously, I attribute the behavior of non-dorsal sonorants and their ability to permit high vowel harmony to their shared specification for V-Manner [+closed] with high vowels. Mbat prefers a single [+closed] span within a word, which is possible only across a non-dorsal sonorant.

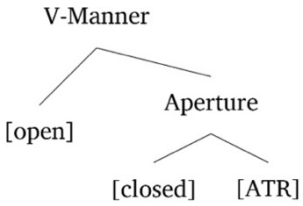


Figure 1 — Geometry of V-Manner features

Based on the behavior of Mbat's vowel and the proposed architectural relationships discussed thus far, I offer the basic geometric organization of Mbat's V-Manner features in Figure 1. For the sake of space and simplicity, I present only V-Manner trees below, as my

analysis treats V-Manner features as responsible for the harmony and dissimilation phenomena introduced above. Vowel roundedness/backness, as dictated by a place feature like [dorsal], does not play a major role in Mbat.

Also important to this geometric analysis is the featural specification of the Perfective/Habitual suffix. As suggested above, several pieces of evidence point toward underlying *-am/-an* and accordingly [+open, -closed] for the suffixal vowel. This is represented in Figure 2.

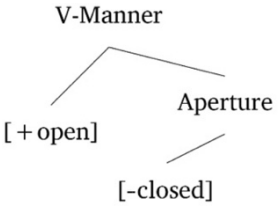


Figure 2 — Underlying form of the Perfective/Habitual suffixal vowel

With Mbat's basic vowel geometry proposed, I turn in the subsections below to modeling each of the alternations earlier discussed in Section 3.

4.1 Alternations involving [open]

Full harmony involving stem and suffixal [a] is possible within a word only across an intervening stem-final dorsal (e.g., *kāŋ* 'fry', *kāŋàm* 'fried'). Under an analysis where the suffixal vowel is underlyingly /a/, this state of affairs involves no overt alternation. I have shown, however, that the vowel of a *CaC* stem elsewhere dissimilates to schwa when such featural harmony cannot be achieved (e.g., *dʒàm* 'stop', *dʒàmàm* 'stopped'). I analyze low harmony between stem /a/ and suffixal /a/ as being possible because both are specified [+open], as is the intervening stem-final consonant.

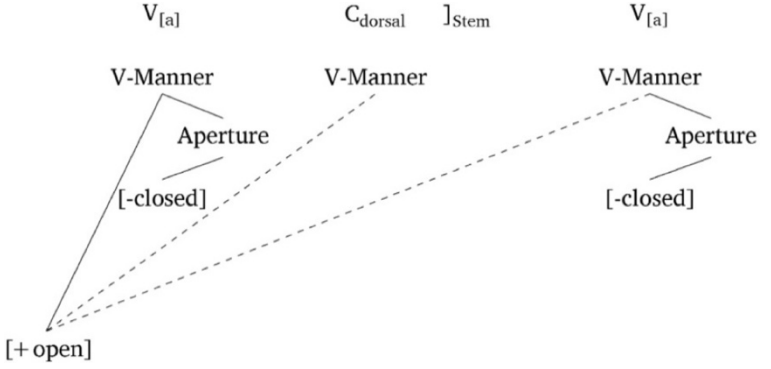


Figure 3 — [+open] conflation – e.g., kāŋ *'fry'* + am → kāŋ-àm *'fried'*

As illustrated in Fig. 3, I propose that Mbat avoids multiple instances of [+open] within a word via feature conflation. Such an outcome is reminiscent of what Cole & Trigo (1988) discuss for [tense] spreading in Menomini Height Harmony. That is, in Mbat, one condition on harmony is that adjacent features (in this case, [+open]) associated with the stem vowel and a stem-final consonant conflate. In the case of harmony with [a], this thereafter extends to the suffixal vowel, creating a single [+open] span across the word.

In related instances, as in (10), where a stem-final consonant is not specified [+open], conflation of the stem vowel's and suffixal vowel's [+open] features is not possible. As a result, Mbat avoids two independent instances of [+open] within the word by dissimilating the first to [–open]. Stem /a/ surfaces as [ə], being [–open, –closed]. This outcome is represented in Figure 4. It is unclear why Mbat opts for alternation in the stem vowel instead of the suffixal vowel under these conditions.[8]

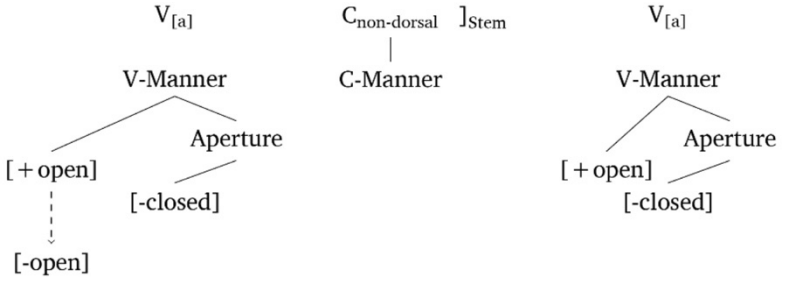


Figure 4 — [+open] dissimilation
– e.g., ɓál *'count'* + am → ɓál-àm *'counted'*

In those instances where the vowel in an uninflected stem is [ə], as in (11), there is no alternation when *-am*/*-an* is added to the stem. Compare, for example, verbs like *də̂p* 'pick/carry' and *də̂ɓ-àm* 'picked/carried'. Here, there is no issue of adjacent [+open] vowels, and thus no repair is necessary. Adjacent instances of [–closed] appear unproblematic.

In addition, there is the special case of "dissimilating" contexts, like those in (12), that involve an onset glide (e.g., *wál* 'laugh', *wúlám* 'laughed'). In Figure 5, rather than a simple dissimilation of the stem vowel [a] from [+open] → [–open] before the Perfective or Habitual suffix, the presence of an onset glide entails an additional alternation of the stem vowel to a high "tense" vowel. Such a vowel is specified

8. A reviewer suggests that there may be rhythmic preferences underlying this choice. It would be difficult to independently justify such a claim, however, given that the language elsewhere opts for alternations in the vowel of the same suffixes.

[+closed, +ATR], and also takes on the specification for backness from the preceding glide. The status of Mbat's onset glides as [+high] (here, [+closed]) is discussed in Green (2020). Notice that the two Aperture features could be said to spread together.

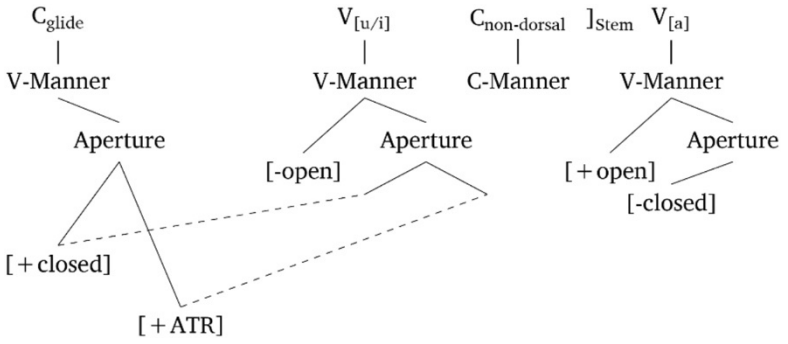


Figure 5 — Dissimilation with onset glide
– e.g., wá[*'laugh'* + am → wúl-ám *'laughed'*

I have shown in this section that Mbat adopts two strategies to avoid multiple instances of [+open] within a word, whether through feature conflation across a stem-final dorsal or via dissimilation of the stem vowel. I have also shown that stem vowels that are underlying [–open, –closed] do not undergo alternation under the same conditions. In doing so, I have suggested that positive-valued features are marked relative to their negative-valued counterparts. Last in this section I showed a special case of alternations involving onset glides. These alternations bring to light that there are some, albeit few, instances where two marked feature values, [+closed] and [+open], co-occur within the same Mbat word. In the next section, I illustrate that this is elsewhere avoided and that there are clear instances in which [+closed] dominates [+open] when a choice between retaining one or the other must be made.

4.2 Alternations involving [closed]

In this section, I turn to harmony involving high vowels, whether they are "tense" [i, u] or "lax" [ɪ, ʊ]. Such harmony was illustrated in (7) and is possible only across a stem-final non-dorsal sonorant (e.g., *kúm* 'find', *kúmúm* 'found'). I attribute this outcome to these stem-final consonants sharing the feature [+closed] with high vowels. For the sake of simplicity, I again set aside the issue of backness as it does not play a key role in the harmony process.

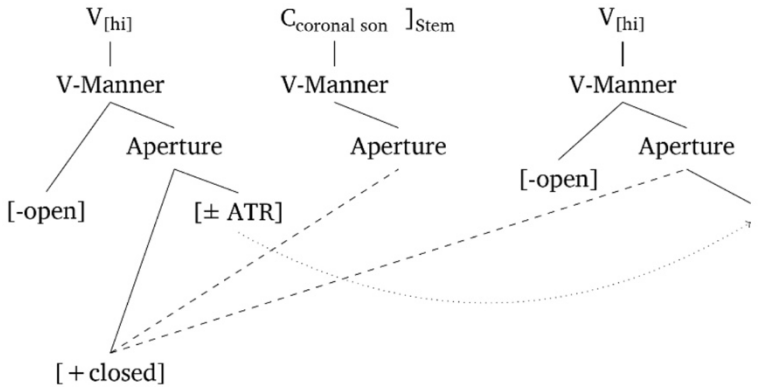


Figure 6 — High vowel harmony
– e.g., bíl *'follow'* + am → bíl-ím *'followed'*

Similar to what I propose in Section 4.1 for the role of [+open] in low harmony, a key factor in harmony involving high stem vowels is the ability for [+closed] to conflate with the same feature associated with the stem-final consonant. Doing so avoids two independent instances of [+closed] within the same word. What is different here, however, is that [+closed] thereafter spreads onto the suffix. In doing so, it replaces the suffix vowel's specification for [–closed] and also leads to a change in the value the suffix's [+open] to [–open]. I have suggested previously that a specification for *[+open, +closed] is avoided in Mbat. This might not be surprising given that the two gestures are contradictory. A general illustration of this outcome is in shown in Figure 6.

Though the [ATR] specification of the stem vowel also arises on the suffixal vowel, it is not clear if [ATR] spreads alongside [closed] (first to the stem-final consonant and then on to the suffixal vowel) or otherwise spreads independently, directly to the suffixal vowel. In this particular figure, I represent [+closed] as operating in a semi-independent way relative to [ATR]. I do this because there is no evidence that stem-final sonorants in Mbat are associated with the latter feature. As stated, it may be possible that both features spread together, or even that the Aperture node itself spreads, even though there is no overt effect on the intervening stem-final consonant. There is evidence elsewhere in Mbat, however, that although the two features often function together, they do not do so in all instances.

One of the more puzzling outcomes witnessed in Mbat Perfective/Habitual inflection concerns the alternation that occurs when a high stem vowel cannot harmonize a suffixal vowel across a non-dorsal sonorant, as in (8). In these instances, the surface quality of the

suffixal vowel is *-əm*/*-ən*. This is seen in verbs like *līp* 'moisten', whose Perfective counterpart is *līɓ-àm* 'moistened'. Under the view stated above that the suffixal vowel is underlyingly /a/, such an outcome involves the suffixal vowel undergoing an alternation from [+open] → [–open]. This occurs after a stem vowel also specified [–open]. Such an outcome is unusual in that there is no immediately apparent motivation for the alternation of the suffixal vowel to [–open] after a [–open] stem vowel. Considering the featural specification of these vowels more broadly, however, offers more insight to this outcome. Had alternation of the suffixal vowel not taken place, the result would be a form like **līɓ-àm*, where adjacent, "positive" specifications of [+closed] and [+open] would result on the stem and suffixal vowel, respectively, within the same word. We have not seen such an outcome, except in verbs like those in Figure 5 (cf. *wál* 'laugh' and *wúlám* 'laughed') where [+closed] spreads from an onset glide onto the stem vowel. We know that such spreading from the onset is late because even across a stem-final non-dorsal consonant, high harmonization of the suffixal vowel does not occur.

If I am correct that the spreading of [+closed] from an onset is exceptional, one possibility to explain the suffixal /a/ > [ə] alternation in *līp/līɓàm*-type verbs is that Mbat avoids positive values for opposing manner features within a word wherever possible. As such, and with [+closed] being dominant as elsewhere, the stem vowel [+closed] is maintained while the suffixal vowel alternates from [+open] → [–open].[9]

Compared to other alternations, another way to view this would be that it yields partial satisfaction of an imperative for vowel raising, achieved formally via the dominance of [+closed]. Under this view, total harmonization of the suffixal vowel is possible only with high stem vowels across a transparent consonant: this yields a single [+closed] span. Partial harmonization, leading to *-əm*/*-ən*, would instead be the result with a high stem vowel and a blocker consonant: there is no [+closed] span, but also no opposing [+open]. Elsewhere, when the stem vowel is not [+closed], the result is non-alternating *-am*/*-an*.

4.3 Opacity in stems with derived high vowels

I briefly discussed above that Mbat manifests opacity in some instances in that derived high vowels do not trigger vowel harmony. I illustrated this in (12) for verb stems with /a/ like *wál* 'laugh' whose

9. Starting with the assumption that the suffixal vowel is underlyingly /ə/ fares no better in that one would have to explain two alternations, one affecting the stem vowel and the other affecting the suffixal vowel, in *ɓál*/*ɓəlám*-type verbs.

stem vowel alternates to a high tense vowel under the influence of a preceding singleton onset glide. For example, the Perfective form of this verb is *wúl-ám* 'laughed'. One might otherwise expect vowels like [u] and [i] to trigger full harmonization of the suffixal vowel across a non-dorsal sonorant (e.g., **wúl-úm*), but this does not occur. The result is an apparent underapplication of height harmony.

A similar outcome obtains for CG[ɛ/ɔ]C verb stems like those in (13). Green (2020) has shown that Mbat manifests an unusual alternation in verb stems of this type such that they lose their glide upon the addition of a vowel-initial suffix like the Perfective *-am* or Habitual *-an* (e.g., *swɔ́ɣ* 'hide', *súrám* 'hidden'). Such inflection unsurprisingly results in the stem-final consonant being resyllabified into an onset due to a near-universal typological tendency for languages to avoid onsetless syllables wherever possible. That this resyllabification leads to the loss of the pre-vocalic glide is far more unusual. Green has argued that this outcome is due to an inherent connection between the second member of a complex onset and a singleton coda, as famously discussed by Kaye & Lowenstamm (1981) and formalized by Baertsch (2002). Important to the matters under consideration in the current paper, however, is that the loss of this pre-vocalic glide entails an alternation in quality of the stem vowel.

Following resyllabification, mid vowels raise to high but do not alternate in their "tenseness", i.e., [ɛ] → [ɪ] and [ɔ] → [ʊ]. Whereas stems with [ɪ] and [ʊ] otherwise have the ability to trigger harmony, these vowels cannot do so when they are derived. An example of such an outcome is seen in a verb like *swɔ́ɣ* 'hide' and its Perfective counterpart *súr-ám* 'hidden'. Here, stem [ɔ] alternates to [ʊ] upon the loss of the pre-vocalic glide, and the suffixal vowel surfaces [a]. There are other pairs in (13), like *pūɣ* 'go out' and *pūrʊ̀m* 'gone out', in which stem [ʊ] harmonizes the suffixal vowel across the same stem-final consonant. In this latter case, however, the triggering high vowel is not derived. Thus, in the case of 'hide' and verbs like it, a derived high vowel does not act as a harmony trigger. Harmony in these instances is opaquely blocked.

One matter pertaining to these alternations that I have not yet entertained in detail is the formal connection between [ɪ/ʊ] and [ɛ/ɔ] from a featural perspective alongside other vowels. As I have suggested, following Green (2020), [ɛ] and [ɔ] behave as if they are allophones of /ɪ/ and /ʊ/, respectively, that occur only after a glide. Their featural properties alongside those of Mbat's other vowels are discussed in the Section 5.

5. Discussion and concluding thoughts

The primary goal of this paper has been to begin to establish certain details of the Mbat vowel system through the lens of alternations that I analyze as being connected to vowel harmony. One challenge inherent in doing so has been that so little is known about the grammar of Jarawan languages more broadly. This means, therefore, that few comparisons can be made to related languages in order to determine whether vowel harmony of a similar or perhaps even of a more transparent type is a pervasive characteristic of this group. Absent the ability to make such comparisons, Mbat still presents an opportunity to consider factors that appear to motivate and govern interactions between stem vowels and adjacent consonants of different types and the downstream effect(s) that these interactions have on suffixal vowels. Another challenge, of course, is that the larger documentation project from which these data are taken is still in progress, meaning that it is yet unclear where else in the language that other instances of harmony, on vowel height or otherwise, may ultimately arise.

While it is clear that Mbat favors harmony in the quality of its stem and suffixal vowels wherever possible, I have suggested that even some instances of disharmony may still implicate an imperative towards raising after high stem vowels. Of course, I have also shown that this trend is sometimes masked by the outcome of other alternations. Overall, vowel harmony in Mbat is triggered by a stem vowel but is blocked across certain intervening stem-final consonants. I have argued that the fact that low harmony occurs only across dorsal consonants and high harmony occurs only across non-dorsal sonorants implicates features that are generally associated with vowel height being intimately involved in the process. There are also two instances in which harmony underapplies in morphologically-derived contexts.

I have attributed the imperative towards raising to the behavior and dominance of [+closed]. This feature is associated with high vowels, but also glides. Mbat prefers a single span of [+closed] across a word, but when this is not possible, other outcomes still suggest that [+closed] is dominant relative to [+open]. The analysis that I present contends that Mbat actively avoids the co-occurrence of *[+closed, +open]. This appears absolute within the same segment, but it is also apparent in vowel alternations from [+open] → [–open]. This avoidance is seen on adjacent vowels in *līp/līɓàm*-type verbs, though it is relaxed somewhat when [+closed] propagates from a syllable onset. This is seen in *wál/wúlám*-type verbs. Of course, although [+closed] appears dominant relative to [+open], it is clear that the language also

acts to create a single [+open] span wherever possible, or otherwise dissimilation occurs.

The generalizations stated above serve as a baseline for analyzing other phenomena in Mbat whose motivations are in some ways less clear cut. One such matter concerns the behavior of the CG[ɛ/ɔ]C stems discussed in Section 4.3, and particularly their alternation with [ɪ/ʊ]. The status of mid vowels is arguably marginal to the extent that they appear only after a glide. When a glide is lost, a mid vowel alternates to its high counterpart. Based partially on this behavior, Green (2020) analyzes mid vowels as being allophones of /ɪ/ and /ʊ/. He proposes that /ɪ, ʊ/ lower to [ɛ, ɔ] due to an OCP constraint against adjacent [+high] segments. Translated to the feature set employed in the current paper, where [+high] equates with [+closed], this resembles other cases where multiple instances of [+closed] are avoided. In a C_1C_2VC syllable, if C_2 and the nuclear vowel are both underlyingly [+closed], the vowel alternates to [–closed]: /zwʊr/ → [zwɔ̄ɽ] ‘sew’. When inflection for the Perfective/Habitual entails loss of the onset glide, there is no antagonistic [+closed]/[+closed] sequence to compel lowering: [zūràm] ‘sewn’.

If this featural analysis is correct, it raises a question about the featural specification of [ɛ/ɔ] relative to other vowels, and particularly to [ə]. Recall from Table 2, repeated here in Table 3 for convenience, that [ɪ, ʊ] are the [–ATR] counterparts to [i, u], all four of which are [+closed]. If in order to avoid an OCP violation on [+closed], mid vowels alternate to [–closed], one would need a means by which to disambiguate these vowels from [ə], which is also [–closed]. One reasonable possibility is that these vowels differ in their specification for [ATR]. Until now, of course, there has been no direct evidence to assume an [ATR] specification for [ə].

Table 3 — Featural specification of vowels

	V-Place	V-Manner		
	[dorsal]	[open]	[closed]	[ATR]
i	–	–	+	+
ɪ	–	–	+	–
u	+	–	+	+
ʊ	+	–	+	–
ə	–	–	–	?
a	–	+	–	

Based on the distribution and participation of mid vowels and [ə] in alternations within Mbat’s verb system, I would propose that their

behavior could be modeled using [ATR], provided that one step away from the usual assumption that [ɛ, ɔ] are "lax" and thereby [–ATR] vowels. That is, the phonological behavior of mid vowels in Mbat suggests that, despite their phonetic quality being [ɛ, ɔ], they pattern and are best analyzed featurally as [+ATR] relative to [ə]. This proposed featural distinction is shown in Table 4, and I will have more to say about it below. The low vowel [a] may be [–ATR], but there is no clear evidence to substantiate whether or not this feature has any role to play relative to this vowel.

Table 4 — Featural specification of vowels

	V-Place	V-Manner		
	[dorsal]	[open]	[closed]	[ATR]
i	–	–	+	+
ɪ	–	–	+	–
u	+	–	+	+
ʊ	+	–	+	–
ɛ	–	–	–	+
ɔ	+	–	–	+
ə	–	–	–	–
a	–	+	–	(–)

The status of mid vowels as [–closed] seems straightforward based on their alternation with [ɪ, ʊ] and their selection of *-am/-an* suffixes, but their [ATR] status in light of what I proposed just above is in need of further justification. If [ɛ, ɔ] are indeed allophones of /ɪ, ʊ/ but specified [+ATR], how do they receive this specification and what conditions their alternation? I would like to propose that one possibility might be that [+ATR] is contributed by spreading from the glide that precedes them.

There is an undeniable featural connection between glides and high vowels, in particular, with the substantive difference between them perhaps being that glides contain some consonantal manner feature. In traditional featural accounts, this would be attributed to glides being [–syllabic] relative to [+syllabic] vowels. This connection is discussed in Green (2020) and entailed in his analysis of stem /ɪ, ʊ/ alternation to [ɛ, ɔ] after a pre-nuclear glide. As introduced above, Green's earlier analysis refers to [+high], which correlates with [+closed] in the current paper. As such, a prenuclear glide triggers a [+closed] → [–closed] alternation on the nuclear vowel in CG[ɛ/ɔ]C stems. I augment this perspective here in proposing that Mbat's pre-nuclear glides are [+ATR] and that this feature is shared with a

following vowel. This would apply both in CG[ɛ/ɔ]C stems, but it also has an analog in *wál/wúlám*-type verbs.

If this is correct, the ability for [+ATR] to be associated with [ɛ] and [ɔ] (though marginally) would entail a slight modification to the proposition stated previously that only [+closed] segments in Mbat can be specified [+ATR]. That is, [+ATR] can be licensed on a [–closed] segment, but only if it has spread from a preceding segment. Such a result would also illustrate that [ATR] and [closed] do not always act or spread in concert with one another. To be clear, assigning [–ATR] to [ə] as a counterpart to mid vowels being [+ATR] (as opposed to being underspecified for the feature) makes no incorrect predictions.

The apparent patterning of Mbat's [ɛ] and [ɔ] as [+ATR] vowels highlights a conceptual issue inherent in standard assumptions about features and their phonetic implementation. Based on their phonetic quality alone, one might presume that these vowels are "lax". Mbat indeed makes an [± ATR] contrast, at least in its high vowels. As such, one might expect these mid vowels to be paired with "tense" [e] and [o] based on what occurs in other languages where such a contrast is found. But, of course, these vowels do not appear in Mbat's inventory. It is only upon a more detailed comparison of their behavior that their featural relationship to [ə] can be appreciated. One might equally be tempted to assume instead that [ə] could be the [+ATR] counterpart of [a]. Of course, this is not the only instance in Mbat where assumptions about phonetic implementation might present an analytical challenge. Let us consider, for example, the selection of Perfective/Habitual *-am*/*-an* vs. *-əm*/*-ən* in context of the featural representation of Mbat's vowel in Figure 7.

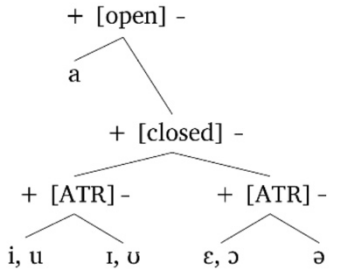


Figure 7 — Mbat vowels – standard features

Recall from earlier in this paper that *-əm*/*-ən* is selected by high vowels in non-harmonizing contexts. In harmonizing contexts, the suffixal vowel is identical in quality to that of the stem. Elsewhere, *-am*/*-an* is selected. The generalization based on the features in Figure 7 is that stems with a [+closed] vowel select a suffix with a [–open] vowel, *-əm*/*-ən*. Elsewhere, a suffix with a [+open] vowel

chooses *-am/-an*. All things being equal, the phonological connection between [+closed] and [–open] is not apparent given phonetic assumptions about these features. Why would a [+closed] vowel choose a suffix with a [–open] vowel? As an alternative, however, consider the representation in Figure 8 where features are instead assigned as abstract categories, with the focus being on how they function phonologically, rather than how they are phonetically implemented.

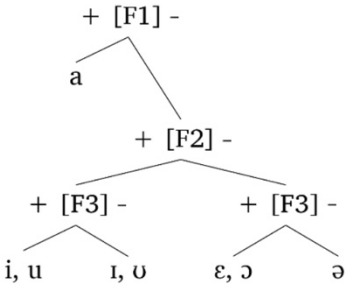


Figure 8 — Mbat vowels – abstract features

An analysis based on such abstract features is unencumbered by phonetic assumptions and is instead concerned only with phonological relationships and, more broadly, the computation entailed in the featural alternations. Surface outcomes are subsidiary. In light of this abstract approach, the choice of suffix that Mbat makes is more straightforward. Stems with a [+F2] vowel select a [–F1] suffix. The [+F1] suffix is selected elsewhere. Here, there is a simple dichotomy in the choice of suffix associated with one value or the other of a particular feature. The phonetic implementation of these featural specifications would be downstream and language-specific. A similar approach is taken by Clements (1991b) in modeling vowel height alternations in narrow Bantu languages and is reminiscent in some ways of substance-free approaches to phonology where phonology is responsible for computation and separate from phonetics. The suitability of such an approach to modeling Mbat's phonology overall must be left to future research.

In closing, this paper has accomplished the goal of bringing to light for the first time characteristics of the vocalic system of a nearly undescribed group of languages that have thus far been unavailable to the theoretical and descriptive linguistics community. This research sets the stage for further exploration not only of Mbat and its vowel system, but also the degree to which vowel harmony represents an areal feature of these languages. In addition, the analysis that I have presented explores the merits of a feature geometric approach to modeling both local and longer-distance vowel-consonant inter-

actions. I have shown that there are both transparent and opaque outcomes in Mbat that are nicely captured by such an approach. I also showed that while standard vowel features offer a means by which to analyze many of Mbat's alternations and interactions, there are some interactions whose transparency is improved by employing an analysis based instead on abstract features.

Acknowledgments

I am incredibly grateful to two anonymous reviewers and to the LLA editors for their helpful comments and suggestions which allowed me to improve upon my original submission. I am indebted to Milkatu Garba to sharing her language with me and to Mr. Burga Kunkadma and the Tadnum Traditional Council in Bauchi State, Nigeria for their continued support of my research on Jarawan languages. Thanks also to Gary Mullin for technical assistance with the preparation of figures for publication.

References

Baertsch, Karen. 2002. *An optimality theoretic approach to syllable structure: The split margin hierarchy*. Bloomington, IN: Indiana University, PhD dissertation.

Blaho, Sylvia. 2008. *The syntax of phonology: A radically substance-free approach*. Trømso: University of Trømso, PhD dissertation.

Blench, Roger. 2006. Jarawan Bantu: New data and its relation to Bantu. Kay Williamson Foundation, ms.

Blench, Roger. 2015. *The Bantoid language*. Oxford handbooks online. Oxford: Oxford University Press

Carnie, Andrew. 2002. A note on diphthongization before tense sonorants in Irish: An articulatory explanation. *Journal of Celtic Linguistics* 7. 129-148.

Clements, George N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225-252.

Clements, George N. 1991a. Place of articulation in consonants and vowels: A unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5. 77-123.

Clements, George N. 1991b. Vowel height assimilation in Bantu languages. In Kathleen Hubbard (ed.), *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society: Special session on African language structures*, 25-64. Berkeley, CA: Berkeley Linguistics Society.

Cole, Jennifer & Loren Trigo. 1988. Parasitic harmony. In Harry van der Hulst & Norval Smith (eds.), *Features, segmental structure and harmony processes (part II)*, 19-38. Dordrecht: Foris.

Eberhard, David M., Gary F. Simons & Charles D. Fennig. 2019. *Ethnologue: Languages of the world*. 22nd ed. Dallas, TX: SIL International.

Gerhardt, Ludwig. 1982. Jarawan Bantu: The mistaken identity of the Bantu who turned north. *Afrika und Übersee* 65(1). 75-95.

Gerhardt, Ludwig. 1988. A note on verbal extensions in Jarawan Bantu. *Journal of West African Languages* 18(2). 3-8.

Green, Christopher R. 2020 (forthcoming). On the link between onset clusters and codas in Mbat (Jarawan Bantu). *Natural Language & Linguistic Theory*. https://link.springer.com/article/10.1007/s11049-020-09469-9.

Green, Christopher R. & Abbie E. Hantgan. 2019. A feature geometric approach to Bondu-so vowel harmony. *Glossa* 4(1). 35.

Grollemund, Rebecca. 2012. *Nouvelles approches en classification : application aux langues bantu du nord-ouest*. Lyon: Université Lumière Lyon 2, PhD dissertation.

Grollemund, Rebecca, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti & Mark Pagel. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the NAS of the USA* 112(43). 13296-13301.

Grollemund, Rebecca & Jean-Marie Hombert. 2012. Sorting out the position of Bantu languages in Bantoid. Paper presented at the 7th World Congress on African Linguistics, Buea, 17-21 August 2012.

Halle, Morris. 1995. Feature geometry and feature spreading. *Linguistic Inquiry* 26(1). 1-46.

Hyman, Larry M. 2017. Multiple argument marking in Bantoid: From syntheticity to analyticity. In Walter Bisang & Andrej Malchukov (eds.), *Unity and diversity in grammaticalization scenarios*, 67-95. Berlin: Language Science Press.

Hyman, Larry M. 2018. Common Bantoid verb extensions. In John R. Watters (ed.), *East Benue-Congo: Nouns, pronouns, and verbs*, 173-198. Berlin: Language Science Press.

Kaye, Jonathan D. & Jean Lowenstamm. 1981. Syllable structure and markedness theory. In Adriana Belletti, Luciana Brandi & Luigi Rizzi (eds.), *Theory of markedness in generative grammar*, 287-315. Philadelphia: John Benjamins.

Maddieson, Ian & Kay Williamson. 1975. Jarawan Bantu. *African Languages* 1. 125-163.

Morén, Bruce. 2003a. Affricates, palatals, and iotization in Serbian: Representational solutions to longstanding puzzles. *Poljarnyj Vestnik* 6. 46-70.

Morén, Bruce. 2003b. The Parallel Structures Model of feature geometry. *Working Papers of the Cornell Phonetics Laboratory* 15. 194-270.

Morén, Bruce. 2006. Consonant-vowel interactions in Serbian: Features, representation, and constraint interaction. *Lingua* 116(8). 1198-1244.

Morén, Bruce. 2007. The division of labor between segment-internal structure and violable constraints. In Sylvia Blaho, Patrick Bye & Martin Krämer (eds.), *Freedom of analysis?*, 313-344. Berlin: Mouton de Gruyter.

Odden, David. 1991. Vowel geometry. *Phonology* 8(2). 261-289.

Piron, Pascale. 1995. Identification lexicostatistique des groupes bantoïdes stables. *Journal of West African Languages* 25(2). 3-39.

Piron, Pascale. 1997. *Classification interne du groupe bantoïde*. 2 vols. Munich: Lincom Europa.

Piron, Pascale. 1998. The internal classification of the Bantoid language group, with special focus on the relations between Bantu, Southern Bantoid and Northern Bantoid languages. In Ian Maddieson & Thomas J. Hinnebusch (eds.), *Language history and linguistic description in Africa*, 65-74. Trenton, NJ: Africa World Press.

Sagey, Elizabeth. 1986. *The representation of features and relations in nonlinear phonology*. Cambridge, MA: Massachusetts Institute of Technology, PhD dissertation.

Shimizu, Kiyoshi. 1983. Die Jarawan-Bantusprachen des Bundesstaates Bauchi, Nordnigeria. In Rainer Voßen & Ulrike Claudi (eds.), *Sprache, Geschichte und Kultur in Afrika*, 291-301. Hamburg: Helmut Buske.

Uchihara, Hiroto & Gabriela Pérez Báez. 2016. Fortis/lenis glides and vowels in Quiaviní Zapotec. *Glossa* 1. 1-24.

Yabilsu, J. Kulu. 1991. *The phonology of Jar Galamkya dialect*. Port Harcourt: University of Port Harcourt, Master's thesis.

3

Inheritance and contact in the genesis of Gisamba (Bantu, L12a, DRC): A diachronic phonological approach

Sifra Van Acker & Koen Bostoen
Ghent University

Abstract

Gisamba (ISO 639-3: smx) is a nearly undocumented and undescribed as well as highly endangered Bantu language spoken in the Kwilu and Kwango provinces of the Democratic Republic of the Congo (DRC). It belongs to the Kikongo Language Cluster (KLC), a discrete subclade of the West-Coastal Bantu (WCB) branch of the Bantu language family. Within the KLC, Gisamba forms a distinct subgroup called "Kikongoid" together with Kiyaka, Kisuku and Kihungan, which are also spoken in the Kwilu and Kwango provinces of the DRC. In this article, we show how both divergence as part of WCB and the KLC and convergence through contact with neighboring WCB and South-West Bantu (SWB) languages contributed to the genesis of Gisamba as spoken today. For this purpose, we provide a synchronic and diachronic account of the phonology of Gisamba. Data used in this article stem from original fieldwork which the first author conducted in 2017 in the village of Kimafu. Some of the diachronic sound changes confirm Gisamba's affiliation to WCB, the KLC and Kikongoid. Others show that Gisamba's synchronic phonology cannot be accounted for as being only the result of vertical transmission through inheritance, but must be the outcome of horizontal transmission through space. This is well in line with the fact that Gisamba is currently endangered and that, historically speaking, its speech communities have been scattered in the Kwilu and Kwango provinces of the DRC where they are surrounded by much larger WCB and SWB speech communities.

Keywords

genealogical classification, historical-comparative linguistics, language contact, Samba, West-Coastal Bantu

Résumé

Le gisamba (ISO 639-3 : smx) est une langue bantoue pratiquement non documentée ou décrite, particulièrement menacée, parlée dans les provinces du Kwilu et du Kwango en République démocratique du Congo (RDC). Elle appartient au groupe kikongo (GK), un sous-ensemble spécifique de la branche dite « bantou de la côte occidentale » (BCO) de la famille des langues bantoues. Au sein du GK, le gisamba forme un sous-groupe distinct appelé « kikongoïde » avec le kiyaka, le kisuku et le kihungan, également parlés dans les provinces du Kwilu et du Kwango en RDC. Dans cet article, nous montrons en quoi les divergences au sein du BCO et du GK aussi bien que les convergences issues des contacts avec les langues voisines du BCO et du bantou du sud-ouest (BSO) ont contribué à la genèse du gisamba tel qu'il est parlé aujourd'hui. Pour ce faire, nous proposons un traitement synchronique et diachronique de la phonologie du gisamba. Les données utilisées dans cet article proviennent d'un travail de terrain spécifique que le premier auteur a mené en 2017 dans le village de Kimafu. Certains des changements sonores diachroniques confirment l'affiliation du gisamba au BCO, au GK et au kikongoïde. D'autres montrent que la phonologie synchronique du gisamba ne peut être considérée comme étant uniquement due à une transmission verticale par héritage, mais qu'elle doit résulter d'une transmission horizontale à travers l'espace. Ceci s'accorde bien avec le fait que le gisamba est actuellement en danger et que, historiquement parlant, ses communautés de locuteurs ont été dispersées dans les provinces du Kwilu et du Kwango en RDC, où elles sont entourées par des communautés beaucoup plus importantes de locuteurs du BCO et du BSO.

Mots clés

bantou de la côte occidentale, classification généalogique, contact linguistique linguistique historico-comparative, samba

---

1. Introduction

In this article, we analyze the diachronic phonology of Gisamba, a poorly studied Bantu language from the Democratic Republic of the Congo. Such a historical-comparative study is relevant because of the striking mismatch between the referential and genealogical classification of the language.

In his first referential classification of the Bantu languages, Guthrie (1948: 54) clusters Samba L12 together with two of its closest neighbors, i.e. Pende L11 and Kwese L13, into Group 10 of Zone L, mainly based on the language's geography, but also certain assumedly shared phonological and morphosyntactic features. In a later update of this same classification, Guthrie (1971) includes Holu in the L10 group with the same L12 code as Samba. (2003; 2009) adds Sonde L101 and re-labels Samba and Holu as L12a and L12b, respectively, classifying them as two varieties of the same language.

From a genealogical point of view, however, the other L10 languages do not seem to be Gisamba's closest neighbors. Except for Kisonde, which was never included in a genealogical classification, all L10 languages other than Gisamba belong to South-Western Bantu (SWB) (Bastin *et al.* 1999; Grollemund *et al.* 2015; Vansina 1995). In contrast, Gisamba is part of the Kikongo Language Cluster (KLC), one of the sub-branches of West-Western or West-Coastal Bantu (WCB) (Bostoen & de Schryver 2018a; 2018b; de Schryver *et al.* 2015). Gisamba's closest relatives are Kiyaka H31, Kisuku H32 and Kihungan H42 (de Schryver *et al.* 2015) (cf. Figure 1). Together they constitute the so-called "Kikongoid" cluster, which de Schryver *et al.* (2015) consider to be the first split off from Proto-Kikongo, the most recent common ancestor of the KLC. Nonetheless, all genealogical classifications of Gisamba are quantitative approaches—either lexicostatistical (Vansina 1995; Bastin *et al.* 1999) or phylogenetic (de Schryver *et al.* 2015; Bostoen & de Schryver 2018a; 2018b)—to so-called "basic vocabulary", i.e. the Tervuren-92-list, a reduced Swadesh 100 word list (cf. Swadesh 1971: 283). Hence, alternative approaches to Gisamba's evolution and historical classification are important.

In this article, we develop such an alternative approach by analyzing Gisamba's diachronic phonology. We examine how both divergence and convergence contributed to the genesis of the language as it is spoken today. We show that certain of the phonological innovations which Gisamba underwent are shared with other languages of the KLC and must be inherited from a most recent common ancestor. Other changes are rather to be understood as contact-induced, i.e. as the result of relatively recent interactions with neighboring languages, both SWB of Guthrie's L10 group (+ Gimbala H41) and WCB languages of Guthrie's B80 group, such as Ensong B85d, Kinsamban B85F, Kimpiin B863, and Engong B864 (cf. Figure 1). Although our historical-comparative approach mainly focuses on sound change, we also consider some morphosyntactic features.

In Section 2, we present the little which is known on Gisamba and its speakers and we briefly assess how severely endangered it is. In Section 3, we provide a basic description of the synchronic phonology

of Gisamba. In Section 4, we analyze the main sound changes the language underwent through a systematic association of present-day Gisamba lexicon to Bantu lexical reconstructions (Bastin *et al.* 2002). In Section 5, we analyze which phonological innovations of Gisamba result from divergence, i.e. are shared with its closest relatives in the KLC, and which from convergence, i.e. are induced through contact with more distantly related languages, both SWB (i.e. L10 and Gimbala H41) and WCB (i.e. B80) (cf. Figure 1). Conclusions are presented in Section 6. The article also includes two appendices: a Gisamba-English wordlist and a short overview of the language's noun class system.

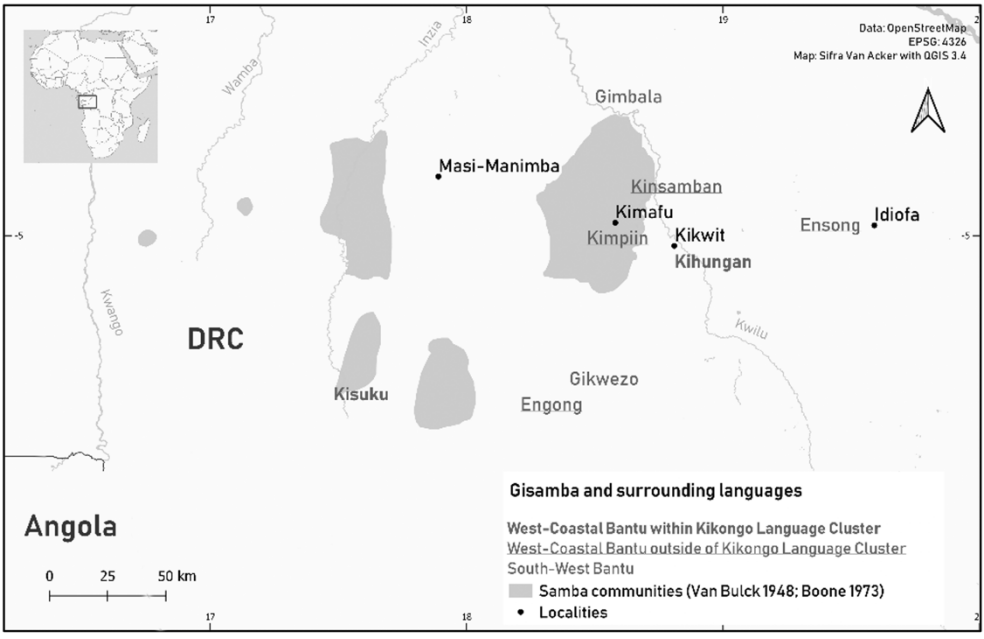


Figure 1 — Samba communities reported in the literature and surrounding languages subcategorized according to phylogenetic classification (Van Bulck 1948; Boone 1973)

2. Gisamba: Speakers, vitality and documentation

Gisamba (ISO 639-3: smx),[1] also called Kisamba, Samba, Tsamba or Tsaam (Eberhard *et al.* 2019), is a Bantu language spoken in the Kwilu and Kwango provinces of the Democratic Republic of the Congo (DRC). Little is known about the speakers of Gisamba, also referred to as

1. While Ethnologue 22 uses the ISO 639-3-code [smx] for Gisamba (Eberhard *et al.* 2019), the same code refers in Glottolog 4.2.1 (Hammarström *et al.* 2020) to Ntsambaan B85F. This is probably due to the confusion between both languages in Bastin *et al.* (1999), see fn. 2.

Basamba or the Samba people. As their villages are scattered among those of larger Bantu language groups, such as speakers of Gimbala [mdp], Kihungan [hum], Kiyaka [yaf] and Kisuku [sub], the geographical location of Gisamba speech communities is not well established (Boone 1973: 286-290; Felix 1987: 178; Plancquaert 1971: 13-14; Torday & Joyce 1907: 135).

Furthermore, not much scientific literature is available on the history and culture of Gisamba speakers, who tend to be mentioned only in passing. Boone (1973: 26) describes them as being specialized in palm oil production, while Plancquaert (1971: 13) and Felix (1987: 178) rather present them as being experts in blacksmithing. Their subsistence economy is reported to be largely dependent on the cultivation of crops, such as peanuts and cassava, and the exploitation of useful trees, such as the oil palm (Felix 1987: 178). During her 2017 fieldwork in Kimafu (DRC) (cf. Figure 1 and infra), the first author could indeed observe the continued importance of peanuts, cassava and oil palms for their subsistence along with that of other crops such as bananas and maize. The practice of blacksmithing was not observed in 2017. According to Leon Kimoko Babakala, the *chef de groupement* of the Samba people, iron working is nowadays mainly the prerogative of Hungan people.

Finally, it is not clear how many people still speak Gisamba. Felix (1987: 178) estimates the number of people identifying themselves as Samba (or *Musamba*, singular of *Basamba*) at 20,000. However, not all people who identify themselves today as Samba still speak Gisamba. Eberhard *et al.* (2019) give 4,200 speakers in 2002 as an estimate, and they consider the language to be threatened, which the 2017 fieldwork confirms. Currently, Gisamba's youngest speakers are over 40 years old and none of them use the language in every domain of life. Its use is mainly restricted to the homestead, especially for conversations among older people. Children are no longer raised in Gisamba. Instead, parents talk to them in vehicular Kikongo (also known as *Kikongo ya Leta*) or in Lingala. Gisamba speakers more generally rely on these two vehicular languages as their main means of communication, along with French in the context of administration and schooling.

Gisamba is not only endangered, it is also little documented and described. As far as we know, no published Gisamba data exist and no "grey" literature (i.e. unpublished theses, dissertations or manuscripts) is available either, apart from a list of basic vocabulary accessible on the website of the Royal Museum for Central Africa.[2] To the

2. See https://www.africamuseum.be/sites/default/files/media/docs/research/human-sciences/culture-society/lexico_bantu/B/B851_kitsamb.pdf (last consultation:

best of our knowledge, the first author's BA and MA theses are the first and only preliminary studies of the language (Van Acker 2016; 2018). The BA thesis is written in Dutch and based on data provided by the only native speaker she could identify in Belgium (Van Acker 2016). The MA thesis is written in English and based on fieldwork done in the village of Kimafu (4°57'34.7"S 18°34'38.9"E, Kwilu province, DRC) (Van Acker 2018). Data used in this article stem from fieldwork carried out in Kimafu in 2017 and consist of (i) a wordlist of 748 items (included in Appendix A) and (ii) a small text corpus consisting of both spontaneous discourse and phrases elicited through translation from French, all of which have been morphologically parsed and glossed. Two main language consultants were involved in the fieldwork: Leon Kimoko Babakala and Henry Nzimabu Kambombo Djopy; both men were born in 1964. Additional data were obtained from Valentine Salayumbu and Pasi Mibengo, two women born in 1938 and 1979, respectively. These four language consultants speak Gisamba in several domains of life, but mainly at home with elders, family members, and friends of the same generation. Different data collection techniques were used to obtain the most representative possible documentation of Gisamba. Several elicitation sessions targeting lexicon, the noun class system, the noun phrase and verbal conjugation (TAM) were conducted with Leon and Henry. Apart from elicitation, spontaneous language data were recorded, i.e. a conversation between Leon and Pasi and stories by both Valentine and Pasi. In total, there are about 21 hours and 15 minutes of audio recordings for the elicitation sessions and about 14 minutes of spontaneous discourse. All fieldwork data are stored on the BantUGent documentation server and can be accessed upon request to the authors.

3. Synchronic phonology

In this section we provide a basic overview of synchronic Gisamba phonology: the inventory of vowel phonemes (§3.1), consonant phonemes (§3.2), syllable structure (§3.3), tone (§3.4), and the principal morphophonological rules (§3.5). High tones are marked with an acute accent <´>, while low accents are left unmarked. Falling tones are marked with a circumflex <^>. Aspirated consonants are followed with <h> in superscript.

November 17, 2020). This online word list, titled "Kitsamb" and wrongly labelled with the Guthrie B851 code, was collected in 1975 by Jan Daeleman (1922-2014) from a consultant called Kingunza. It was used for the lexicostatistical study by Bastin *et al.* (1999), where it was erroneously included as a variety of Tsambaan or Kisamban B85F (i.e. their dataset B85/7). After completion by Joseph Koni Muluwa, it was also used for the phylogenetic study by de Schryver *et al.* (2015), where it was included as a dataset representing Gisamba L12a.

3.1 Vowel inventory

Gisamba has five vowel phonemes which we represent as follows: *i e a o u*. The minimal pairs in (1) show that they are phonologically contrastive.

(1)	i/e	*bín* 'you (PL)'	vs.	*bén* 'breast'
		gubíl 'to forget'		*gubél* 'to be sick'
	i/u	*gilímb* 'symbol, sign, scar'		*gilúmb* 'day'
		guhín 'to push'		*guhún* 'to deceive'
	e/a	*gulénd* 'to be able to'		*guláand* 'to follow, continue'
		gudél 'to call, name'		*gudál* 'to visit, admire, watch'
	e/o	*gubél* 'to be sick'		*guból* 'to rot'
		gubéng 'to be red'		*gubóng* 'to get'
	a/o	*bál* 'liver'		*ból* 'two'
		láand 'follow, continue (IMP)'		*lónd* 'height'
	o/u	*guból* 'to rot'		*gubúl* 'to hit, break, peel'
		ngónd 'moon, month'		*ngúnd* 'field'

The mid vowels /e/ and /o/ can be phonetically realized as [e] or [ɛ] and [o] or [ɔ], respectively. The half-close and half-open mid vowels, both front and back, are in free variation. They have no straightforward phonological conditioning.

Vowel length is not phonologically contrastive in Gisamba. As in many Bantu languages (cf. Hyman 2019: 135), vowels are automatically long when they occur in CGV and VNC position. For instance, the verb stem *dwál* 'drive' is always realized phonetically as [dwá:l] and the verb stem *dámb* 'play' always as [dá:mb]. Their vowel length is entirely predictable and thus not noted. The only exception to this automatic lengthening occurs in word-final position. Vowels at the end of a word are never long, also not in CGV context. For instance, *mbwa* 'dog' is realized [mbwa] and not [mbwa:].

The three nouns in (2) are the only ones to manifest non-predictable vowel length. The long vowel here results from vowel concatenation, another common source of vowel length in Bantu (cf. Hyman 2019: 135). These exceptional cases of phonetically non-predictable long vowels result from the historical merger between a noun class prefix and a vowel-initial noun stem. The retention of the vowel length can be considered here as an instance of "archaic heterogeneity" (cf. Dimmendaal 2011: 99). Synchronically, they do not contrast with short vowels in Gisamba to distinguish lexical meaning. These few words with long vowels do not form minimal pairs with equivalent words having short vowels.

(2) *gʊ̀dʊ̀ ‘leg; hind leg’ > *gúúl/mííl* (*kʊ-/mi-) ‘leg, foot/legs, feet’
*jícò ‘eye’ *líís/míís* (li-/mi-) ‘eye/eyes’
*jínò ‘tooth’ *líín/míín* (li-/mi-) ‘tooth/teeth’

3.2 Consonant inventory

As shown in Table 1, Gisamba has 28 consonant phonemes including seven pre-nasalized ones. At least three of these, viz. the ones in parentheses, are to be considered marginal for reasons explained below.

Table 1 — Inventory of Gisamba consonant phonemes

	Bilabial	Labio-dental	Alveolar	Palatal	Velar	Glottal	Labio-velar
Plosive	p b		d		g		
Aspirated plosive	p^h		t^h		k^h		
Nasal	m		n				
Fricative		f v	s (z)	ʃ		h	
Affricate	pf		ts dz	tʃ dʒ			
Lateral approximant			l				
Approximant				y			w
Pre-nasalized plosive	(mp) mb		nd		ng		
Pre-nasalized fricative		(mv)					
Pre-nasalized affricate			nts	ntʃ			

We demonstrate in (3) the phonemic status of each of the consonants in Table 1 by means of at least one minimal pair.

(3) p/d *pondo* ‘millet’ vs. *dondo* ‘hills’
p^h/k^h *p^háb* ‘flatness’ *k^háb* ‘angriness’
b/g *gób* ‘glass, mirror, window’ *góg* ‘arm, hand’
g/k^h *gémb* ‘plantain’ *k^hémb(o)* ‘happy’
m/l *gumímb* ‘to lay down, sleep’ *gulímb* ‘to wrap up’
n/ng *gudún* ‘to refuse’ *gudúng* ‘to build, construct’
f/l *gufúd* ‘to pay’ *gulúd* ‘to drill’
v/d *guvúgís* ‘to mix’ *gudúgís* ‘to provoke’
s/ʃ *los* ‘rice’ *loʃ* ‘noise’
h/g *guhín* ‘to push’ *gugín* ‘to dance’
pf/d *gupfá* ‘to die’ *gudá* ‘to tell’
n/nd *ginán* ‘eight’ *gindánd* ‘cassava’

dz/l	*mudzób* 'idiot'	*mulób* 'fisherman'
tʃ/n	*mwítʃ* 'smoke'	*mwín* 'heat, warmth, sun'
t^h/b	*t^ho* 'source'	*bo* 'take'
dʒ/y	*gudʒím* 'to switch off, extinguish'	*guyím* 'to dry'
y/b	*yúlu* 'sky'	*búlu* 'disease'
y/k^h	*yáb* 'know'	*k^háb* 'angriness'
w/g	*guwónd* 'to be tired, weak'	*gugónd* 'to miss'
mb/nd	*mudámb* 'trap'	*mudánd* 'back, spine'
nd/nts	*ndóng* 'line'	*ntsóng* 'top'
ng/mb	*ngúnd* 'field'	*mbúnd* 'heart'
ntʃ/mb	*ntʃi* 1. 'fly', 2. 'country'	*mbi* 'fault'
nd/nts	*ndóng* 'line'	*ntsóng* 'top'
ngónd	'moon, month'	*ngónts* 'bat'

We consider the consonants in parentheses in Table 1 as marginal in that they are rather rare in the lexicon and cannot be contrasted to other phonemes as part of a minimal pair. However, they also cannot be considered as a free or conditioned allophone of one of the other phonemes. All attestations of these marginal phonemes in our lexicon are presented in (4). All of them are clearly borrowings from vehicular Kikongo as spoken in the Kwilu and Kwango provinces (cf. Swartenbroeckx 1973). Kikongo, in turn, borrowed them from European languages, such as French and Portuguese.

(4)	z	*gizúʃí* 'picture'	cf. Kikongo	*(ki)zídi*, *(ki)zízi*
		mizíg 'music'		*muzíki*, *musíki*
	mp	*límpa* 'bread'		*dímpà*
		múmpe 'priest'		*mumpé*
		mpa 'news'		*mpà*
		mpíl 'way'		*mpílà*
		mpév 'soul'		*mpêvé*
	mv	*mvud* 'answer'		*mvútu*

We do not consider the phoneme /p/ to be marginal because it is more frequent and does contrast with other phonemes, as illustrated in (3). However, it is probably also a loan phoneme. As shown in §4.2, the regular reflex of Proto-Bantu (PB) *p in Gisamba is /h/. As shown in (5), most Gisamba words featuring /p/ also have /p/ in vehicular Kikongo (cf. Swartenbroeckx 1973).

(5)	p	*pondo* 'millet'	cf. Kikongo	*mpòndo*
		mupép 'wind'		*mupépé*
		gípus 'skin'		*púsu*
		gupúlúmúg 'to fly'		*pùlumúka*
		lupáng 'fence'		*lupángu*
		pimb 'good, easy'		*pímbu*

podopod 'porridge'	*pòto-póto*
púlúpulu 'diarrhea'	*pùlu-púlu*
puluʃ 'police'	*pulúsi*
pwede 'silence'	*pwèdi*
pwis 'thirst'	*pwîsa*
páypáy 'papaya'	*pàpái*, *pàyi-páyi*, *pàipái*

Apart from the consonant phonemes listed in Table 1, Gisamba has several consonants that are allophones of one of these phonemes, either phonologically conditioned or free.

The alveolar tap [ɾ] can be considered a positional variant of the alveolar lateral approximant /l/. The two sounds are in complementary distribution: [l] never occurs in front of the close front vowel /i/; /l/ is always pronounced [ɾ] in this environment. Hence, *tsámbwali* 'seven' is pronounced [tsámbwaɾi], *tsúli* 'smell' as [tsúɾi]. Likewise, when followed by suffixes having an initial /i/, such as the causative suffix *-is*, a root-final /l/ is pronounced as [ɾ], e.g. *gól* 'enter' → *gólís* [góɾís] 'make enter, introduce'.

The voiceless plosives [t] and [k] and the voiced plosives [d] and [g] are realizations of the phonemes /d/ and /g/. Except when following a nasal, the voiced and voiceless allophones are freely interchangeable. In case of prenasalization, only voiced stops occur. Conversely, the aspirated voiceless stops /t^h/ and /k^h/ do not have voiced counterparts. Hence, *dúng* 'sew' can be realized either [dúŋ] or [túŋ]; *búd* 'give birth' either [búd] or [bút]. Likewise, *gal* 'charcoal' can be pronounced either [gal] or [kal]; *ngág* 'bag' either [ŋgág] or [ŋgák]. However, in post-nasal position, only the voiced variants are observed, i.e. /nd/ and /ng/. Hence, *bóngól* cannot be realized as **[bóŋkól]; *bénd* 'pull' cannot be pronounced as **[bént]. Exceptions have been noted in the following words, which are probably borrowings: *gandín/dundín* [gantín/duntín] 'bucket(s)', *gindéd* [gintét] 'Monday', *gindúndu* [gintúntu] 'flower', *ndwál* [ntwál] 'front', *ndúmb* [ntúmb] 'apostle', *ngén* [nkén] 'grain, seed'. Orthographically, the phoneme with the allophones [t] and [d] is always represented by the symbol of the voiced counterpart.

In word-final position, the phoneme /ng/ is mostly—but not always—realized as [ŋ], e.g. *dúng* 'construct' [dúŋ]. In word-initial or word-medial position, this allophonic alternation is seldom observed. As a consequence, when a word is realized with a word-final vowel, [ŋ] shifts again to [ŋg]. For instance, *líng* 'search' has been recorded as both [líŋ] and [líŋg], and also as [líŋga].

3.3 Syllable structure

Gisamba has both open and closed syllables. As illustrated in (6), open syllables take the following forms: CV, V, CGV. Syllable onsets are either zero, simple (C) or complex (CG). (Recall from §3.2 that prenasalized consonants are treated as phonemic units.)

(6)

CV	*ba* 'palm tree'
	nde 'if'
V	*e* 'and, with'
CGV	*tswí* 'ear'
	mbwa 'dog'

As listed in (7), closed syllables take the forms CVC and CGVC.

(7)

CVC	*bál* 'liver'
	dénʃ 'mirror'
	mbén 'enemy, rebel'
	ngónd 'moon, month'
CGVC	*byóts* 'all'
	mwámb 'yellow'
	ngwáʃ 'uncle'
	mbwang 'jaw'

All words in (6) and (7) are monosyllabic. When it comes to polysyllabic words, closed syllables usually only occur word-finally, while open syllables occur in all positions, as illustrated in (8).

(8)

CV.CV.CV	*bá.ga.ga* 'certain'
	gi.lá.ndí 'secretary'
CV.CV.CVC	*bá.lú.múg* 'turn, change, become'
CV.CVC	*má.ndob* 'fishhooks'

The two words in (9) are the only ones also having closed syllables in word-initial or word-medial position. The first one, which has already been identified as a loanword in (5), is also exceptional in that it has a glide as syllable coda. The second one is possibly also a loanword having its ultimate origin in French *exercice* or Portuguese *exercício*, but no equivalent could be traced in vehicular Kikongo.

(9)

CVC.CVC	*páy.páy* 'papaya'
CV.CVC.CVC	*gi.ngár.sis* 'exercise'

Closed syllables almost only occur in word-final position due to final vowel loss. Synchronic evidence for this process is provided by the words in (10), which have been found in the data both with and without final vowel. The fact that the vowel varies across words and is not necessarily a copy of the root vowel suggests that the final vowel does not result from a synchronic phonological rule such as default

final vowel insertion or vowel copying. The lexical reconstructions given in parentheses (cf. Bastin *et al.* 2002) indicate that when the final vowel surfaces it is indeed a reflex of the original vowel (except for the last example). There is no clear conditioning for the absence or presence of the final vowel, even if it seems to appear more often at the beginning of a sentence than at the end. In any case, if it surfaces, it entails the resyllabification of the coda, which becomes the onset of the second syllable.

(10)	CVC ~ CV.CV	*pʰug ~ pʰu.gu* ‘rat’ (*púkù)
		pʰíb ~ pʰí.ba ‘night’ (*pɪmpa)
		móʃ ~ mó.ʃi ‘one’ (*mòtí)
	CGVC ~ CGV.CV	*mwêtʃ ~ mwê.ʧi* ‘water’ (*gèdì)
	CGVC ~ CGV.CV	*tswéng ~ tswé.ngi* ‘little bird’
	CGVC ~ CGV.CV	*mwémb ~ mwé.mbu* ‘drug, medication’
	CVC ~ CV.CV	*mbúng ~ mbú.ngi* ‘fog’ (*bùngì)
	CV.CVC ~ CV.CV.CV	*mu.gád ~ mu.gá.di* ‘in, inside’ (*kàtí)
	CV.CVC ~ CV.CV.CV	*gi.dúmb ~ gi.dú.mbu* ‘war’ (*tùmbà)

The nouns in (11) are some of the few that have always been observed with a final vowel. This might be because all of them are possibly Kikongo loanwords.

(11)	CV.CV	*tʰá.lu* ‘number, price’	cf. Kikongo	*ntálu*
		yú.lu ‘sky’		*yúlu*
	CV.CV	*ndá.mbu* ‘bit’		*ndámbu*
	CGV.CV	*pwe.de* ‘silence’		*pwéti*, *pwèdi*

3.4 Tone

Gisamba has two level tone phonemes: high (H) and low (L). H is marked with an acute accent (´), low is not marked graphically. As the minimal pairs in (12) show, these tones are phonologically distinctive.

(12)	*bul* ‘village’	vs.	*búl* ‘wrong, evil’
	gudug ‘to appear’		*gudúg* ‘to exit’
	ndund ‘foam’		*ndúnd* ‘vegetable’

Gisamba also has a falling (F) contour tone, which is marked with a circumflex (^). No rising tones have been observed. Underlyingly, a falling tone can always be analyzed as the sequence of a high and a low tone resulting from the contraction of two vowels belonging to distinct morphemes, most commonly a noun prefix and a vowel-initial noun stem, e.g. *bú-ad* ‘canoe’ → *bwâd*; *mú-an* ‘child’ → *mwân*. As the minimal pair in (13) shows, a falling tone can contrast with a level tone, but this is only very rarely the case. Monosyllabic words can thus be either L, H or F.

(13) *lyang* 'mango' vs. *lyâng* 'branch'

Noun prefixes in Gisamba can either be H or L. Among disyllabic nouns consisting of a noun prefix and a monosyllabic noun stem, five tone patterns are observed in isolation: L-L, L-H, H-L, H-H, and L-F.

(14)

L-L	*mu-nun* 'old man'	
L-H	*bu-gáf* 'lie'	
H-L	*má-yemb* 'shoulders'	(SG *yemb*)
H-H	*má-gód* 'frogs'	(SG *gód*)
L-F	*bu-mwân* 'childhood'	

When consonant-final monosyllabic stems as the ones in (14) take a final vowel, the latter is most often low, but not always, as shown in (15).

(15)

L-LL	*mu-dangi* 'reader, lector'	
L-HL	*mu-dúngi* 'builder'	
H-LL	*má-dzandu* 'markets'	(SG *dzandu*)
L-HH	*bi-béndé* 'iron'	
L-LH	*ma-ndagá* 'speech'	

Tone patterns of noun stems which are minimally disyllabic are listed in (16). They are too rare in our data to distinguish recurrent patterns.

(16)

H-LL	*gí-fanus* 'monster'
(L-)HL	*yóngon/ma-yóngon* 'chameleon(s)'
L-HLL	*gi-sánunu* 'comb'
L-HH	*gi-gáláng* 'bridge'
(L-)HHL	*bágála/ba-bágála* 'man/men'
(L-)HLL	*góngoʃig/ba-góngoʃig* 'cricket(s)'
(L-)HHH	*yóngólól/ma-yóngólól* 'centipede(s)'
(L-)HLH	*dáladál/ma-dáladál* 'window(s)'
(L-)HHHL	*búlúmwáli/ma-búlúmwáli* 'cabinet(s)'
(L-)HLHL	*ndúngundúngu/ba-ndúngundúngu* 'moustache(s)'
(L-)LLHL	*gusumbóngi/ma-gusumbóngi* 'parrot(s)'

As noticed in the examples above and elsewhere in this article, several Gisamba nouns have a high tone on the prefix in citation form. Leftward high tone spreading is common in the KLC and often results in high-toned noun class prefixes (cf. Blanchon 1998; Daeleman & Meeussen 1983), also in Gisamba's closest relatives such as Kiyaka (Van den Eynde 1968: 13-17). More reliable tone data and a more systematic study would be needed to understand the synchronic and diachronic operation of these tonal processes in Gisamba. Moreover, as illustrated in (17), the tone pattern of a word in isolation is not necessarily the same as in context, which is also a well-

established feature of the KLC known as "tone case" (cf. Blanchon 1998; 1999; Daeleman & Meeussen 1983; Philippson & Boungou 1999). An in-depth tone study would also be needed to determine how such contextual tone patterns are conditioned in Gisamba.

(17) *magóndó mélí mugubol*
'The bananas are rotting.'

ma-góndo má- ílí mu-gu-bol
NP_6-banana SP_6-be NP_{18}-NP_{15}-rot
(cf. *magóndo* 'bananas')

3.5 Principal morphophonological rules

3.5.1 Post-nasal strengthening

Following a non-syllabic homorganic nasal prefix, such as the 1SG subject prefix, the 1SG object prefix or the plural noun class 10 prefix, several phonemes undergo a process of post-nasal strengthening. The fricative /s/ and the approximate /y/ affricate to /ts/, while the liquid /l/ strengthens to /d/, see (18).

(18) s → ts / N_ *sál* 'stay' vs. *ntsál* *N-sál* 'I stay.'
súmb 'buy' vs. *ntsumbi* *N-sumb-I* 'I have bought.'
y → ts / N_ *yámbúl* 'leave' vs. *ntsámbúl* *N-yámbúl* 'Leave me.'
yágúl 'talk' vs. *ntságúlí* *N-yágúl-í* 'I have talked.'
l → d / N_ *léf* 'borrow' vs. *ndéf* *N-léf* 'Borrow me.'
lulémb 'finger' vs. *ndémb* *N-lémb* 'fingers'

3.5.2 Aspiration

Gisamba has three voiceless aspirated stops, i.e. p^h, t^h, and k^h. They are usually the result of a morphophonological process and rare in monomorphemic words. They mostly occur in stem-initial position, but not exclusively, e.g. *edét^{h}i* 'before, first, already'. As evidenced by the morphophonological changes observed in (19), aspiration can be triggered by an underlying non-syllabic nasal, which becomes zero on the surface. In the case of alveolars and velars, only stops that are historically voiceless get aspirated.

(19) h → p^h / N_

/u-N-hég/ → *up^{h}ég* 'You give me.'
SP_{2SG}-OP_{1SG}-give

/ga-N-háng-il-i/ → *gap^{h}ángili* 'He has done for me.'
SP_1-OP_{1SG}-do-APPL-PFV

d ~ t → t^h / N_

/u-N-twal-i/ ~ /u-N-dwal-i/ → *út^{h}wali* 'You have brought me.'
SP_1-OP_{1SG}-bring-PFV

g ~ k → k^h / N_

/N-kán-i/ ~ /N-gán-i/ → *k^háni* 'I have visited.'
SP_1-visit-PFV

3.5.3 Palatalization

As discussed in §3.2, /l/ is realized as its positional variant [ɾ] when followed by the high front vowel /i/. Followed by the same vowel, /s/ and /ts/undergo palatalization, or more precisely post-alveolarization, to [ʃ] and [tʃ], respectively, as shown in (20). Before /i/ these palatal consonants can be considered as phonologically conditioned allophones of their non-palatal counterparts. However, there are other phonological contexts where the non-palatal and palatal consonants may contrast, cf. (3). That is why, in contrast to the alveolar tap [ɾ], we consider /ʃ/ and /tʃ/ to be phonemes in Gisamba.

(20) s → ʃ /_i

sá 'put'	vs.	[uʃi]	*u-s-i* 'You have put.'
dugis 'shine'		[odugiʃí]	*o-dugis-í* 'It shines.'
ts → tʃ /_i			
íts 'come'		[wítʃí]	*u-íts-í* 'You have come.'
bwáts 'sit'		[obwátʃí]	*o-bwáts-í* 'He is seated.'

3.5.4 Progressive vowel height harmony

Based on data from the BA thesis of the first author (Van Acker 2016), Goes & Bostoen (2019: 43) classify Gisamba among the KLC languages in which the high vowels of verb extensions (i.e. verbal derivational suffixes) do not undergo lowering to assimilate to mid root vowels. The data in (21), also presented in their study, do indeed testify to the absence of progressive vowel harmony.

(21)

kéng-id-il-a 'oversee, supervise, watch'	(*-il-il* = APPL-APPL)
lémb-ik-a 'appease'	(*-ik* = IMPO)
déf-ís-a 'lend'	(*-is* = CAUS)
t^hók-ís-a 'cook'	
sóng-ul-a 'sharpen'	(*-ul* = SEP.TR)
tób-ul-a 'pierce'	

During her 2017 fieldwork on the Kimafu variety described here, the first author did observe instances of progressive vowel harmony as in (22). It was only observed when both the root and extension vowel have either a front or a back vowel.

(22) *bóng-ól* ‘restore’ *-ul* = SEP.TR
sóg-ól ‘gather’
sól-ól ‘chat’
dzód-óg ‘jump’ *-ug* = SEP.INTR
yéb-él ‘clean, wash’ *-il* = APPL
démb-és ‘believe’ *-is* = CAUS
lémb-és ‘lighten up’
ség-és ‘sharpen’

The only exceptions to this conditioning observed in the 2017 data are presented in (23).

(23) *léf-is* ‘lend’ *-is*= CAUS
lémb-íg ‘appease’ *-ig* = IMPO

Gisamba potentially has a type of progressive vowel harmony that is attested in no other language of the KLC, i.e. only when the root and extension vowel have the same mid vowel. However, the available data are too sparse to make strong claims in this respect.

4. Diachronic phonology

In this section we analyze the main diachronic sound changes which Gisamba underwent by examining the present-day reflexes of Bantu lexical reconstructions (Bastin *et al.* 2002) in the language. The translation of the Gisamba reflex is only given when it differs from the meaning posited for the reconstruction.

4.1 Reflexes of PB vowels

4.1.1 7-to-5-vowel reduction

As discussed in §3.1, Gisamba has five vowel phonemes. As elsewhere in the KLC (Bostoen & Goes 2019), Gisamba reduced the PB 7-vowel system (cf. Meeussen 1967: 82) by merging the first two degrees of aperture, i.e. *i and *ɪ to /i/ and *u and *ʊ to /u/, as shown in (24).

(24) *i > i *jíkì ‘smoke’ > *mw-ítʃ*
*kíngó ‘neck’ *tʃingu*
*ɪ > i *jímb ‘sing’ *yímb*
*dímì ‘tongue’ *lú-lim*
*u > u *cùb ‘urinate’ *sub*
*bùmò ‘belly’ *dzúm*
*ʊ > u *bʊ̀ngì ‘fog’ *m-búng*[3]
*gʊ̀dʊ̀ ‘leg’ *gú-úl*

3. To facilitate comparison with reconstructed noun stems, we mark here the nasal of stem-initial prenasalized consonants as noun prefixes of cl. 9/10, which they certainly are from a diachronic point of view. Synchronically, they could be analyzed as being integrated into the noun stem.

As illustrated in (25), the PB vowels *e *a *o were mostly retained as is. The first two examples also show that PB root-internal vowel length got lost in Gisamba.

(25) *e > e *bèèdé 'knife' > *mbél*
*béénè 'breast' *bén*
*a > a *támbí 'footprint' *lu-dámbi*
*càbuk 'cross' *sábúg*
*o > o *dòng 'teach' *lóngi* 'teacher'
*bòd 'rot' *ból*

4.1.2 Final vowel loss or apocope

As discussed in §3.3, final vowel loss is the source of word-final closed syllables in Gisamba. The data in (26) show that all PB vowels are subject to this type of apocope.

(26) *cónì 'shame' > *n-tson*
*dédé 'cloth' *mu-lél*
*pácà 'twins' *has*
*bìndò 'dirt' *m-fínd*
*kúmú 'chief' *pfum*

4.1.3 Final vowel raising

If preserved, noun-final PB mid vowels are raised in Gisamba. Unlike in other languages of the KLC where final vowel raising only occurs when the preceding syllable also contains a mid vowel (cf. Meinhof & van Warmelo 1932: 168), this is not a dissimilatory process in Gisamba. As shown in (27), it took place regardless of the degree of aperture of the preceding vowel.

(27) *bògó 'buffalo' > *gí-bogu* 'hippopotamus'
*cémbò 'horn' *mu-sémbu* 'whistle'
*tádè 'stone' *dáli*
*kíngó 'neck' *tʃingu*
*pígò 'kidney' *lú-pʰigu/má-pʰigu*[4]

Exceptions to this regular diachronic sound change, as in (28), turn out to be loanwords, probably from vehicular Kikongo. As discussed

4. The stem-initial consonant of the reflex of *pígò is particular in two regards. First, given that it is followed by a PB high front vowel, one would expect a fricative or an affricate instead of a plosive, as the result of the common Bantu sound change known as spirantization. As Bostoen & Goes (2019) point out, *pígò also escaped spirantization in other languages of the KLC. Second, the aspiration of /p/ is unexpected and indicates an underlying nasal preceding *p (cf. §3.5.2). As the aspiration is observed in both the singular and the plural, we assume that the *N*-prefix of class 9 was integrated into the stem prior the reclassification of the noun into classes 11/6.

in §3.2, some of these words also manifest other phonological irregularities.

(28)	*góndo* 'banana'	cf.	*kòndò	Kikongo *dinkòndo*
	pondo 'millet'		*pòndó	Kikongo *mpòndo*
	bi-béndé 'iron'			Kikongo *dibèndé*

4.2 Reflexes of PB consonants

4.2.1 Unconditioned intervocalic reflexes of PB consonants

In this section, we discuss the unconditioned intervocalic regular reflexes of PB consonants. By unconditioned we mean not constrained by a specific phonological environment. Hence, they are the most common reflexes of PB consonants. PB consonants following nasals (§4.2.2) and/or close front/back vowels (§4.2.3) are not considered here.

As shown in (29), PB *b, *m and *n were retained in Gisamba.

(29)	*b > b	*càbʊk 'cross (river)' >	*sábúg*
		*dóbò 'fishhook'	*ndob*
		*bèng 'be red'	*béng*
		*bóngó 'knee'	*bong*
	*m > m	*dímì 'tongue'	*lú-lim*
		*dúmì 'male'	*mu-lúm* 'husband'
		*mòtí 'one'	*mó∫*
	*n > n	*càmbànò 'six'	*sambanu*
		*nénè 'big'	*néné*
		*nók 'rain'	*nog*
		*nú 'drink'	*nwa*

PB *p, *d, *c, and *j underwent lenition to /h/, /l/, /s/ and /y/ or ∅, respectively, as evidenced by the data in (30). As discussed in §3.2, /l/ is realized as [ɾ] in front of /i/.

(30)	*p > h	*páan 'give' >	*hán*
		*pácà 'twins'	*has*
		*pang 'act, make'	*hang*
		*pìkà 'slave'	*mu-híg*
	*d > l	*càmbʊ̀àdì 'seven'	*tsámbwali*
		*cʊ̀dì 'smell'	*tsúli*
		*jédò 'door'	*gy-él*
		*dó 'sleep'	*gí-lu*
	*c > s	*cónɪk 'write'	*sónig*
		*paaco 'locust sp.'	*p^{h}ásu*
		*jícò 'eye'	*líís*
		*kèc 'cut'	*gés*
		*cúmb 'buy'	*súmb*

*j > y, ∅	*jájì ‘yawn’	*mw-áy*
	*jíjab ‘know’	*yáb*
	*jánɪk ‘spread to dry’	*yáníg*
	*kájá ‘leaf’	*lú-gay*
	*jʊ̀dʊ̀ ‘leg’	*gú-úl*

As already discussed in §3.2 with the data in (5), Gisamba words having retained *p instead of shifting it to /h/, such as those in (31), are actually loanwords.

(31)	*pòndó ‘millet’	>	*pondo*
	*pèep ‘blow (as wind)’		*mu-pép* ‘wind’
	*pʊ̀cʊ̀ ‘skin’		*gí-pus*

We end this section with the reflexes of PB *t, *k, and *g. We treat them together, because they undergo the same voicing effect. As discussed in §3.2, [t] and [k] are free allophones of /d/ and /g/ in present-day Gisamba, except after nasals. As shown in (32), the voiced alveolar /d/ including its free unvoiced allophone [t] are reflexes of *t, while the velar stop /g/ and its free allophone [k] are reflexes of both *k and *g. While both *t and *k were retained before being optionally voiced, *g first underwent devoicing and merged with reflexes of *k before becoming accessible to optional (re)voicing. This velar merger is not unique to Gisamba, in fact this sound change occurred everywhere in WCB (Pacchiarotti & Bostoen 2020). Following insights from this article, we know that *g underwent devoicing and merged with reflexes of *k in Proto-WCB. We can therefore strongly assume that the optional (re-)voicing of *k to *g was a later innovation, and thus that /g/ is not a direct reflex of PB *g in Gisamba.

(32)	*t > d	*kàtí ‘inside’	>	*mu-gád*
		*tí ‘tree’		*mú-d*
		*tátʊ̀ ‘three’		*dádu*
		*bútʊk ‘come back’		*vúdúk*
		*játò ‘canoe’		*bú-ad*
	*k > g	*bàkàdà ‘man’		*bágála*
		*kàndá ‘letter’		*mu-gánd*
		*kímà ‘monkey’		*gim* ‘monkey sp.’
		*kídà ‘tail’		*mú-gil*
		*tòk ‘boil up’		*dóg* ‘boil’
		*kʊdɪ ‘bone’		*gí-guli*
	*g > g	*tég ‘sell’		*dég*
		*dʊ̀gú ‘sibling, friend’		*n-dúg* ‘friend’
		*gábán ‘divide up, share’		*gábán*
		*jògʊ̀ ‘elephant’		*n-tʃóg*
		*jògà ‘mushroom’		*b-óg*

In contrast to the reflexes of PB voiceless stops *t and *k, the reflex of *p is not subject to regular voicing. This indicates that Gisamba only started voicing voiceless stops after intervocalic *p had shifted to /h/, which does not provide the right input for this voicing rule. However, as explained in the following section, particularly in (36) in §4.2.2, de-aspirated word-final reflexes of *mp, i.e. /p/, can undergo voicing to /b/. They are just very rare. We also found two examples of the opposite effect, i.e. an evolution by analogy towards the optional voiceless realization of /b/ in word-final position, e.g. *jíb 'steal' > *yíp* or *yíb*, *gàb 'divide; give away; make present' > *gap* or *gab* 'share'. This indicates that the free allophony between voiced and voiceless alveolar (*d*/*t*) and velar stops (*g*/*k*) also occurs or has started to extend to labial stops (*b*/*p*).

4.2.2 Post-nasal reflexes of PB consonants

In this section, we give an overview of how PB pre-nasalized consonants evolved in Gisamba in both stem-initial (C1) and stem-internal (C2) position and in front of vowels that do not trigger specific sound changes, such as PB half-close front and back vowels. In §4.2.3 we consider sound changes triggered/conditioned by close front/back vowels.

The PB voiced pre-nasalized consonants *mb, *nd, and *ng are preserved in Gisamba, as shown in (33).

(33)	*mb > mb	*N-bèèdé 'knife'	>	*m-bél*
		*N-bòmbó 'nose'		*m-bómb*
		*N-kómb 'swipe'		*gómb*
		*N-tùmbà 'war'		*gi-dúmb*
	*nd > nd	*N-dáká 'tongue, language'		*n-dag* 'voice, word'
		*N-dìmbò 'birdlime'		*n-dímbu* 'glue'
		*N-dànd 'follow'		*land*
		*N-gàndá 'clan'		*gi-gánd* 'family'
	*ng > ng	*N-gàngà 'medicine-man'		*n-gáng*
		*N-gàndó 'crocodile'		*n-gánd*
		*N-tóng 'sew; build'		*dúng* [dúŋ] (cf. §3.2)
		*N-táng 'read; count'		*dáng* [dáŋ] (cf. §3.2)

Several of the examples in (33), just like those in (34), indicate that prenasalized voiced consonants are never reduced to simple nasals. This suggests that Gisamba did not undergo the common Bantu sound shift known as Meinhof's Rule (Dammann 1972; Meeussen 1962). It happens neither when a prenasalized voiced consonant is followed by another prenasalized voiced consonant in the next syllable ($*NC_{[+voiced]} > NC_{[+voiced]}/\_\_NC_{[+voiced]}$), nor when a simple N follows in the next syllable ($*NC_{[+voiced]} > NC_{[+voiced]}/\_\_N$).

(34)	*N-gòmà ‘drum’	>	*n-góm*	
	*N-gòndò ‘moon, month’		*n-gónd*	
	*N-bèng ‘be red’		*m-béng* ‘red’	
	*N-bùngì ‘fog’		*m-búng*	

As shown in (35), the PB voiceless pre-nasalized consonants *mp, *nt, and *nk undergo aspiration accompanied by the loss of the preceding nasal, which is a common process in Bantu (cf. Kerremans 1980).

(35)	*mp > p^h	*N-pangɪ ‘brother, sister’	>	*p^{h}áng* ‘sibling’
		*N-púkù ‘rat’		*p^{h}ug*
		*N-pémbé ‘white’		*p^{h}émb*
		*N-paaco ‘locust sp.’		*p^{h}ásu*
	*nt > t^h	*N-tàdò ‘number’		*t^{h}álu* ‘number, price’
		*N-to ‘river’		*t^{h}o* ‘source’
		*N-támbí ‘sole foot, footprint’		*t^{h}ámbi* ‘paws’
	*nk > k^h	*N-kómbò ‘goat’		*k^{h}omb*
		*N-kódá ‘snail’		*k^{h}ol*
		*N-kádì ‘bitter’		*k^{h}áli*
		*N-kámá ‘hundred’		*k^{h}ám*

In word-final position, especially when the final vowel is omitted, the aspiration is lost. As illustrated in (36), as a result, there is free variation between [t] and [d] cf. §4.2.1).

(36)	*mp > p, b	*N-pɪmpa ‘night’	>	*p^{h}íp, p^{h}íb*
	*nt > d	*káíntù ‘woman’		*mu-géd, mu-kéd, mu-gét, mu-két*
		*ntù ‘person’		*mú-t, mú-d*

Finally, the reflexes of PB *nc and *nj seem to have merged into /nts/ in Gisamba. In front of /ts/, the nasal can be dropped, but it usually does not, as shown in (37).

(37)	*nc > (n)ts	*N-cádá ‘feather’	>	*tsál*
		*N-càngò ‘news’		*n-tsángu*
		*N-conco ‘hinge/dowel-pin’		*mu-n-tsóntsó* ‘nail’
		*còɪk ‘hide, cover’		*gi-n-tswég* ‘secret’
	*nj > (n)ts	*N-jádá ‘nail, claw’		*gí-n-tsal*
		*N-jàdà ‘hunger’		*n-tsál*
		*N-jùngú ‘cooking pot’		*n-tsúng*
		*pànjʊd ‘scatter’		*hántsúl* ‘crush’

Before a front vowel, the reflexes of *nc and *nj undergo palatalization.

(38)	*nci/nji > ntʃi	*N-cí ‘ground, country’	>	*n-tʃí*
		*N-jìdà ‘path’		*n-tʃíl*

4.2.3 Spirantization

As most Bantu languages which underwent the 7-to-5-vowel merger, Gisamba also underwent the common sound shift known as "Bantu Spirantization". This is the mutation which PB plosives undergo in front of the close vowels *u and *i, mostly to a fricative or an affricate (Bostoen 2008; Schadeberg 1995).

As shown in (39), in front of *u, a full merger of places of articulation took place in favor of labiodental fricatives, as is the case elsewhere in the KLC (Bostoen & Goes 2019). Furthermore, we observe an advanced loss of the voicing distinction as the result of a common process known as "spirant devoicing" (cf. Labroussi 2000). Admittedly, the number of spirantized reflexes in our dataset is limited, but voiced spirants have only been observed among the reflexes of *bu and only in C1 position. This could indicate that the process is systematic in C2 and intermittent or ongoing in C1. The *pfu* reflexes of *ku can be accounted for by either pre-nasalization (e.g. *N-fum* > *pfum*) or a CGV sequence (e.g. *fu-a* > *fwa* > *pfa*).

(39)	*bu > f/vu	*búdà 'rain'	>	*m-ful*
		*bùá 'nine'		*gí-va*
		*bùd 'become numerous'		*vulu* 'many'
		*bútʊk 'come back'		*vúdúg*
	*du > fu	*dùmbì[5] 'corpse'		*m-fúmb*
		*dèdù 'beard'		*gi-léf*
	*gu > fu	*dàgù[6] 'wine; beer'		*ma-láf*
	*pu > fu	*púdò 'foam'		*má-fuluful*
		*pùt 'pay'		*fúd*
	*tu > fu	*túnd 'accuse'		*fúnd*
	*ku > fu	*kúkam 'kneel'		*fúgám*
		*kúmú 'chief'		*pfum* (*N-fum*)
		*kú 'die'		*pfã* (*fu-a*)

5. The reconstruction *dùmbì 'corpse' does not occur in BLR3 (Bastin *et al.* 2002). We propose this new reconstruction for Proto-West-Coastal Bantu on the basis of numerous attestations in languages from both the KLC and Guthrie's B50-80 languages. While its reflexes in the KLC systematically manifest Bantu Spirantization in C1 position (most often a voiced spirant), those in B50-80 do not; they have either /l/ or /d/. That is why we propose *d as C1. The noun stem is probably an agentive derivation (suffix *-i*) of *dùmb 'smell' (BLR 1258) with reported attestations in Guthrie's zones C D H M S (Bastin *et al.* 2002).

6. The reconstruction *dàgù 'palm wine' is not present in BLR3 (Bastin *et al.* 2002). We propose this new reconstruction for Proto-Kikongo in analogy with the existing reconstruction *dògù 'wine; beer' (BLR3 1108) with reported attestations in Guthrie's zones A B D K L R (Bastin *et al.* 2002). Apart from V1, these two reconstructions correspond in both form and meaning.

Table 2 summarizes the outcomes of Bantu Spirantization in front of *u.

*Table 2 — Gisamba reflexes of PB stops in front of PB *u*[7]

PB		*pu	*tu	*ku	*bu	*du	*gu
Gisamba	C1	fu	fu	fu	vu; fu	fu; ¦vu¦	¦fu¦
	C2	¦f(u)¦	¦f(u)¦	¦f(u)¦	¦f(u)¦	f(u)	f(u)

There is a gap in the data available to examine Bantu Spirantization in front of *i. However, the examples presented in (40) show that the merger of places of articulation is only partial in this environment. As is also the case in many other languages of the KLC (Bostoen & Goes 2019), the reflexes of *bi and *pi have a place of articulation that is distinct from those of *di, *gi, *ti, and *ki. The latter have all become palatal, whereas in most other KLC varieties they have merged to become alveolar fricatives. In Gisamba, Bantu Spirantization in front of *i was thus accompanied or followed by palatalization (or rather post-alveolarization). As before *u, spirant devoicing is very advanced. The available reflexes of *di manifest a positional conditioning in that voicing is maintained in C1 and lost in C2. However, some of the few reflexes available for *bi and *gi did undergo devoicing in C1. These data seem to indicate that spirant devoicing is systematic in C2 and intermittent or ongoing in C1 in front of *i. It is striking that the mode of articulation of *ti reflexes are not only distinct from those of *ki and *gi, but also from *di. While all reflexes of *t in front of close *i are fricative, i.e. /ʃ/, *d, *k and *g all have affricate reflexes, i.e. *tʃ* and/or *dʒ*.

(40) *bi > fi *bìndò ‘dirt’ > *m-findu*
*búì ‘white hair’ *lú-m-fi*
*di > dʒi *dìık ‘bury’ *dʒíg*
*ding[8] ‘live’ *lu-dʒíng* ‘life’
*di > tʃi *bàdì ‘tomorrow’ *m-bátʃ(í)*

7. Forms in between ¦ ¦ are assumed but not attested in our dataset (see §5.1 for why we do not propose ¦vu¦ as a reflex of *gu).

8. The reconstruction *ding ‘live’ is not included in BLR3 (Bastin *et al.* 2002). We propose this new reconstruction for Proto-Kikongo based on numerous attestations in all subgroups of the KLC. All reflexes manifest Bantu Spirantization (BS) and have /z/ as C1. The root had probably already undergone BS in Proto-Kikongo. We need evidence from outside the KLC to confirm the reconstruction of *d as C1. It could also be *g, as PB *di and *gi merged in Proto-Kikongo (Bostoen & Goes 2019). If the C1 is indeed *d, it needs to be examined to what extent this root is related to *dìng ‘desire’ (BLR 1061) with reported attestations in Guthrie’s zones A B C L (Bastin *et al.* 2002). In the KLC, the root only occurs with the meaning ‘live’, not ‘desire’.

	*jàdí ‘oil’		*m-átʃ* ‘fat’
	*jàdí ‘lightning’		*n-tsatʃ*
	*gèdì ‘stream’		*mw-êtʃ(i)* ‘water’
	*kádí ‘woman, wife’		*mu-gátʃ* ‘wife’
	*bídì ‘fish’		cf. *m-bitʃ ó mwêtʃ* ‘fish, lit. animal of water’
	*dòòdí ‘dream’	≻	*n-dótʃ*
*gi > tʃi	*dògì ‘witch’		*mu-lótʃ(i)*
	*gì ‘fly (n.)’		*n-tʃí*
*pi > fi	*píp ‘suck’		*fí*
	*píná ‘pus’		*má-fin*
	*pínık ‘cover’		cf. *finam* ‘close’
*ti > ʃi	*tím ‘dig’		*ʃím*
	*tíkà ‘cold season, night’		*ma-ʃig* ‘evening’
	*tíngà ‘bow-string;		*mu-ʃíng* ‘rope; tendon; vein’
	*títú ‘forest’		*mú-ʃid*
	*mòtí ‘one’		*móʃ*
*ki > tʃi	*kíngó ‘neck’		*tʃingu*
	*jíkí ‘smoke’		*mw-ítʃ*

Table 3 summarizes the outcomes of Bantu Spirantization in front of *i.

*Table 3 — Gisamba reflexes of PB stops in front of PB *i*[9]

PB		*pi	*ti	*ki	*bi	*di	*gi
Gisamba	C1	fi	ʃi	tʃi	fi; ¦vi¦	dʒi; ¦tʃi¦	tʃi
	C2	¦f(i)¦	ʃ(i)	tʃ(i)	¦f(i)¦	tʃ(i)	tʃ(i)

5. Inheritance versus contact in the genesis of Gisamba phonology

In this section, we provide a short historical assessment of the sound changes that Gisamba underwent. It turns out that not all of them can be accounted for by vertical transmission through time or inheritance. Some are unmistakably the result of horizontal transmission through space or contact-induced change. We are very well aware that, as one of the reviewers pointed out, none of the single features discussed below are unique to Gisamba and its closest relatives and/or neighbors. They are also found often elsewhere in Bantu. Nonetheless, the sum of the shared innovations can still be taken as valid evidence for either inheritance or contact, especially if one considers them in the light of existing genealogical classifications and if one reckons their specific geographic distribution.

9. Forms in between ¦ ¦ are assumed but not attested in our dataset (see §5.1 for why we do not propose ¦dʒi¦ as a reflex of *gi).

5.1 Inherited phonological innovations

As presented in §1, Gisamba is part of the Kikongo Language Cluster (KLC), a discrete subclade of WCB. Within the KLC, Gisamba forms a distinct subgroup called 'Kikongoid' together with Kiyaka, Kisuku and Kihungan, which are spoken in the Kwilu and Kwango provinces of the DRC. As for the sound changes discussed in §4, we consider the following as being innovations shared with other Kikongoid languages specifically or with the KLC and WCB more generally, even if none of them is unique to WCB:

- PB velar stop merger — WCB
- *d-lenition — WCB/KLC
- 7-to-5-vowel reduction — KLC
- *p-lenition — Kikongoid
- Spirant devoicing — Kikongoid

The intervocalic reflexes of PB *g and *k have become indistinguishable in Gisamba. As Pacchiarotti & Bostoen (2020) demonstrate, this merger of PB velar stops occurred at the Proto-WCB stage. Within South-West Bantu (SWB), to which several languages neighboring Gisamba belong, this merger is not omnipresent and cannot be reconstructed to the most recent common ancestor. For instance, the merger occurred in Gipende (Ntitenguha 1984), Gimbala (Ndolo 1972) and Gikwezo (Forges 1983), but not in Cokwe K11 (Kanyamibwa 1982), Kiholu L12b (Kashika Katanga 1990), Ruwund L53 (Mulindabigwi 1981) and zone L more generally (Kabange Mukala 2009). The fact that it occurs specifically in the SWB language immediately bordering WCB is possibly contact-induced.

Gisamba has reduced the PB 7-vowel system to one consisting of only 5 vowel phonemes. This is a phonological innovation which it shares with all other languages of the KLC. Kihungan is the only language within the KLC to have 7 vowel phonemes, but these do not directly reflect the 7 vowels of PB; they result from a vowel split that happened after the 7-to-5-vowel reduction (Bostoen & Koni Muluwa 2011). Although this 7-to-5-vowel reduction is shared by all KLC languages, Bostoen & Goes (2019) argue, for reasons on which we cannot elaborate here, that this innovation does not go back to Proto-Kikongo, the most recent common ancestor of the KLC. In any event, it is still too widespread to consider it as an innovation which Gisamba inherited from the ancestor and shares with the other Kikongoid languages only. The same holds for the intervocalic lenition of PB *d to /l/, which is widespread in the KLC, in WCB and in Bantu more generally (cf. Hyman 2019: 142).

As for the lenition of PB *p to /h/, Gisamba shares this innovation with all its Kikongoid relatives, as shown, for instance, by the reflexes of PB *páan 'give', i.e. Kiyaka *hááná* (Ruttenberg 2000: 40), Kisuku *hana* (Piper 1977: 17), Kihungan *hán* (Kasuku-Kongini 1984: 33), or *páng 'make, act', i.e. Kiyaka *hángá* (Ruttenberg 2000: 41), Kisuku *hanga* (Kifindi 1997: 53), Kihungan *háŋ* (Koni Muluwa & Bostoen 2015: 93). The *p > *h* shift is also attested elsewhere inside and outside of Bantu. Within the KLC, it occurs most notably in several North Kikongo varieties (cf. Nguimbi-Mabiala 1999). However, elsewhere in the KLC, other types of *p-lenition occur, i.e. *p > *β* > *v*, which is most prominent in South and West Kikongo, *p > *ɣ*, which is characteristic for East Kikongo (cf. Bostoen *et al.* 2013). As a consequence, and within the broader context, *p > *h* can be considered as a phonological innovation which the Kikongoid languages, including Gisamba, inherited form their most recent common ancestor.

As Bostoen & Goes (2019) show, Bantu Spirantization in front of the PB close vowels *u and *i is an innovation that goes back to Proto-Kikongo. The same holds for the complete merger of places of articulation before *u and the partial one before *i, which accompanies Bantu Spirantization. In other words, this is a sound shift which Gisamba shares with the rest of the KLC and confirms the distinctive genealogical status of the KLC within WCB. In the remainder of WCB, Bantu Spirantization is irregular and sporadic (Pacchiarotti & Bostoen 2020). In neighboring SWB languages, such as Gimbala, Gipende and Gikwezo, it does occur regularly, but the patterns in terms of merger of places of articulation are quite distinct (Janson 2007: 111). However, spirant devoicing as observed in Gisamba, is not omnipresent in the KLC. It is completely absent in South and Central Kikongo and irregularly attested in West and North Kikongo. East Kikongo and Kikongoid are the only subgroups where this innovation is more or less regular and where it is likely to have occurred in the ancestor language of each of the subgroups (Bostoen & Goes 2019). In other words, Gisamba probably also inherited this innovation from the most recent common ancestor which it shares with Kiyaka, Kisuku and Kihungan. The fact that voiced spirants are still irregularly observed in Gisamba is possibly due to contact with surrounding SWB languages, such as Gimbala, Gipende and Gikwezo, where spirant devoicing did not take place (Janson 2007: 111). As we discuss in §5.2, the palatalization and affrication of spirants is in all likelihood also a feature which Gisamba adopted through contact with these languages.

5.2 Contact-induced phonological innovations

Not all diachronic phonological innovations that contributed to the genesis of Gisamba's synchronic phonology can be accounted for by vertical transmission through inheritance. Three of them are clearly contact-induced for at least two reasons, i.e. the fact that they are rare or not attested within the KLC and the fact that they are unsystematic in Gisamba itself.

- Final vowel loss B85-B87
- Voicing of voiceless stops SWB
- Spirant palatalization and/or affrication SWB

Final vowel loss is irregular in Gisamba in that some words never occur with a final vowel, others occasionally do so and yet others never do. Moreover, the phenomenon manifests inter-speaker variation. Apart from Gisamba, the Kikongoid language Kihungan is the only other one within the KLC where word-final vowels are commonly lost (Pacchiarotti & Bostoen forthcoming). It is absent from certain varieties and in those where it does occur, several words are not affected as in Gisamba (Bostoen & Koni Muluwa 2011: 254-255). Final vowel loss is systematic, however, in several neighboring WCB B80 languages (Daeleman 1977; Koni Muluwa 2010; Pacchiarotti & Bostoen forthcoming; Rottland 1977). It seems to be an areal feature, as it also occurs in languages of Guthrie's groups B70 (Guthrie 1960) and C80 (Grégoire 2003) as well as in certain SWB languages, such as Kanyok L32 (Mukash Kalel 1982) and Ruwund L53 (Nash 1992). However, it is absent from the SWB languages in the immediate vicinity of Gisamba, i.e. Gimbala (Ndolo 1972), Gipende (Gusimana 1972) and Gikwezo (Forges 1983). In the case of both Kihungan and Gisamba, we therefore consider it to be interference from neighboring B80 languages, such as Kimpiin B863 and Ensong B85d.

As discussed in §4.2.1, the reflexes of *t, *k, and *g are optionally voiced. This means that *t is realized as either [t] or [d] in present-day Gisamba, and *k and *g, which merged into *k in Proto-WCB lost (Pacchiarotti & Bostoen 2020), as either [k] or [g]. The fact that this voicing effect is optional and that the unvoiced and voiced allophones are in free variation suggests that this is not a deeply rooted sound shift, but rather a contact-induced interference. This assumption is borne out by the fact that such voicing is unattested everywhere else in the KLC, even in Gisamba's closest Kikongoid relatives, and also in most of the remainder of WCB, especially in Gisamba's B80 neighbors. Voicing is attested, however, in Gisamba's closest SWB neighbors, i.e. Gimbala, Gipende and Gikwezo. As the noun prefix of each of these languages suggests, *k has shifted to /g/. As for the Samba language,

both Gisamba and Kisamba occur. The voicing of *t and *k is only systematic in Gimbala and Gikwezo, and does not occur in Gipende, as shown in (41) with comparative data taken from Koni Muluwa & Bostoen (2015). The voicing of PB voiceless stops is also manifested in the reflexes of the noun prefixes of classes 7, 12, 13, and 15, which are *gi-*, *ga-*, *du-* and *gu-*, respectively, in these three SWB languages. Gisamba, on the other hand, has both the voiced and unvoiced equivalents for these class prefixes. Reflexes of *g are not considered in (41), because *g and *k did not merge in Proto-SWB. Their reflexes are still distinct in several SWB languages, such as Gipende and Gikwezo, e.g. *dòg 'bewitch' > Gipende *gu-loa*, Gimbala *gu-loga*, Gikwezo *gu-lowa*.

(41)	*jíkʊt 'be satiated'	>	Gipende	*gu-kúta*
			Gimbala	*gu-guda*
			Gikwezo	*gu-gúdá*
	*táp 'draw water'	>	Gipende	*gú-thaya*
			Gimbala	*gu-dáya*
			Gikwezo	*gu-daha*
	*tád 'look'	>	Gipende	*gu-tala*
			Gimbala	*gu-daala*
			Gikwezo	*gu-daala*
	*jókà 'snake'	>	Gipende	*nyógá*
			Gimbala	*nyoga*
			Gikwezo	*nyôga*
	*kún 'plant'	>	Gipende	*gu-kuna*
			Gimbala	*gu-gúna*
			Gikwezo	*gu-guna*

A last, probably contact-induced, phonological feature of Gisamba is the affrication of spirants in front of *i. As discussed in §4.2.3, the reflexes of *ki (> *tʃi*), *di (> *dʒi*, possibly also *tʃi*) and *gi (> *tʃi*) not only undergo palatalization before the high front vowel, they also affricate, unlike the reflexes of *ti (> *ʃi*), which only palatalize. The palatalization of these spirants is not unattested in Kikongoid. It also occurs optionally in Kiyaka (Van den Eynde 1968: 6). However, spirant affrication does not occur in Kkongoid, at least not when the spirant is not preceded by a nasal. Affricated palatal spirants do occur, however, in both Gipende and Gikwezo, while Gimbala has spirants that are neither palatalized nor affricated (Janson 2007: 111). Remarkably, affrication is only observed with voiced spirants in Gipende and Gikwezo, but not with voiceless ones and not for the reflexes of *di in Gikwezo. In other words, although the palatalization and affrication patterns of spirants are not exactly identical to those in Gi-pende and

Gikwezo, contact with those SWB languages might explain why the phonetic outcomes of Bantu Spirantization in Gisamba are distinct from those in its closest Kikongoid relatives. The fact that Gipende and Gikwezo only affricate voiced spirants might also explain why reflexes of *ti did not undergo affrication in Gisamba. The voicing of *ki reflexes could then be considered as a kind of hypercorrection.

Another striking contact-induced feature of Gisamba is double negation marking. Although it is not phonological, we wish to discuss this property here to show that the impact of neighboring languages on Gisamba has been substantial. As shown in (42)-(45), unlike what is commonly the case in Bantu, Gisamba does not have verbal negation markers. Negation is marked by means of independent words, i.e. *lo*, which is also used to answer 'no', and by a word that consists of the locative pronominal prefix of class 17 *gu-* and a possessive stem. The latter is always co-referential with the subject of the clause. In (42), *gobá* is built on the possessive stem *oba* 'their' of class 2, which co-refers just like the verb itself to the subject of the clause belonging to class 2, i.e. *batsútsú bámá bóló* 'my two chickens'. In (43), *gwándí* contains the possessive stem *andi* 'his, her, its' of class 1, which is co-referential here with the subject, i.e. 'Anne'. In (44), the *am* 'my' in *gwám* refers to the 1SG subject. Very often *lo* and the class 17 possessive mark negation conjointly as in (42) and (43). However, each of them can also be the sole negator as in (44) and (45). Most often they are pre-verbal, but they can also occur after the verb, as in (44).

(42) *Batsútsú bámá bóló lo gobá bálumbu meg.*

ba-N-sútsú	ba-ámá	ba-óló	lo	gobá
NP_2-NP_9-chicken	PP_2-myNP_2-two		NEG	NEG_{3PL}
ba-a-lumb-u	ma-eg			
SP_2-PRS-lay-FV.PRS	NP_6-egg			

'My two chickens do not lay eggs.'

(43) *Anná gélí mugusónígá mugándá wándí yábántsá lo gwándí gamanis*

Anná	gá-a-ílí	mu-gu-sóníg-á		mu-gándá	u-ándí
Anne	SP_1-PRS-be	NP_{18}-NP_{15}-write-FV		NP_3-book	PP_3-her
é-a-bánts-á		lo	gwándí	ga-man-is	
SP_{1SG}-PRS-think-PRS		NEG	NEG_{3SG}	SP_1-finish-CAUS	

'Anne is writing her book, but I think she will not finish it.'

(44) *émání gwám gél mugánd*

é-mán-í	gwám	gél	mu-gánd
SP_{1SG}-finish-PFV	NEG_{1SG}	yet	NP_3-book

'I have not finished the book yet.'

(45) *ntse ló wélí e lujidʒú mu dangu maʃig*

ntse	ló	ú-a-ílí	e	lu-dʒidú
2SG	NEG	SP_{2SG}-PRS-be	with	NP_{11}-rudeness
mu	∅-dangu	ma-ʃig		
NP_{18}	NP_{5a}-hour	NP_{6}-evening		

'You are not impolite tonight.'

Gisamba's Kikongoid relatives Kiyaka and Kisuku have none of these two negation markers. They combine verbal negation marking with the post-verbal marker *ku/ko* (cf. Piper 1977; Van den Eynde 1968). The negative marker *lo* does occur in Kihungan, the other Kikongoid relative of Gisamba. It also tends to be pre-verbal in that language, but unlike in Gisamba, it combines with negative prefixes on the verb (Takizala 1974: 13). Moreover, *lo* is also attested as a post-verbal negative marker in Ensong (B85d) (Koni Muluwa & Bostoen 2019: 441-442) as well as in Kinsamban (B85F) and Engong (B864) (Devos & van der Auwera 2013: 226-228; Koni Muluwa 2017: 233). It also occurs in Gisamba's SWB neighboring languages Gipende, Gimbala and Gikwezo (Devos & van der Auwera 2013: 226-228). What is more, the latter two also have the class 17 possessive negation marker co-referential with the subject. The use of a class 17 locative possessive pronoun to express negation has also been observed in other Bantu languages of the wider region, such as Kiyombe H16c (West Kikongo), Kizombo H16h (South Kikongo), Kiholu L12 (SWB), Ruwund L53 (SWB) and Kanincin L53 (SWB) (Devos & van der Auwera 2013). Nevertheless, Gikwezo and Gimbala are the only ones where it occurs in pre-verbal position and combines with *lo*, as shown in (46) and (47) with examples also cited in Devos & van der Auwera (2013). Just like in Gisamba, there is no negation marking on the verb itself. In other words, Gisamba shares exactly the same negation pattern with Gikwezo and Gimbala, except that Gikwezo can have triple negation marking, which does not seem to be the case in Gimbala and Gisamba.

(46) Gikwezo (Forges 1983: 216)

lo gwâmi ngaswěgá ídondo lo

lo	gwâmi	nga-swěg-á	ídondo	lo
NEG	NEG	1SG.PST-hide-PRF	9.meat	NEG

'I have not hidden the meat.'

(47) Gimbala (Ndolo 1972: 77)

lo gomi i-hosh-idi

lo	gomi	i-hosh-idi
NEG	NEG	1SG-talk-PRF

'I haven't spoken.'

Gisamba not only shares several phonological features with the neighboring SWB languages Gimbala and Gikwezo, but also the morphology and syntax of a negation pattern that is unattested in the KLC and fairly unusual within Bantu more generally. This indicates that contact-induced change had an impact on the genesis of Gisamba as we know it today.

6. Conclusions

We have provided here a synchronic and diachronic account of the phonology of Gisamba, a nearly undocumented, undescribed, and highly endangered Bantu language spoken in the Kwilu and Kwango provinces of the DRC.

With regard to the sound system reconstructed for PB, we have identified several phonological innovations which are in line with the phylogenetic classification of Gisamba established by means of basic vocabulary (Bostoen & de Schryver 2018a; 2018b; de Schryver *et al.* 2015). The merger of PB velar stops *g and *k adds to the hypothesis that Gisamba is part of WCB, as this innovation took place in their most recent common ancestor. Surrounding SWB languages did not undergo this merger. The reduction of the PB 7-vowel system to one of 5 vowel phonemes confirms that Gisamba is part of the KLC within the wider WCB branch, as most other WCB languages did keep the 7 vowels of Proto-Bantu or enlarged their vowel phoneme inventory. Finally, the lenition of *p to /h/ and spirant devoicing are the two sound shifts corroborating that Gisamba is indeed more closely related to the other members of the Kikongoid subgroup, i.e. Kiyaka, Kisuku and Kihungan, than to any other language within the KLC. These different sound changes reflect the role that divergence and vertical transmission of innovation through time played in the genesis of Gisamba as we know it today.

The synchronic phonology of Gisamba cannot be accounted for by divergence alone. Contact-induced change also had a considerable impact on the shaping of its sound system, especially through interaction with neighboring WCB languages of Guthrie's B80 group as well as with adjacent SWB languages, especially Gimbala, Gipende and Gikwezo. The B80 languages contributed to the frequent but not entirely systematic loss of word-final vowels. Contact with the SWB languages is at the origin of the irregular voicing of voiceless stops and the palatalization and affrication of spirants. The horizontal transmission of features from Gimbala and Gikwezo is not limited to phonology, but can also be observed in the morphosyntax as illustrated with pre-verbal double negation marking in §5.2. This indicates that interference from neighboring languages in Gisamba

was considerable and must have involved a certain degree of bilingualism. With reference to the analytic framework for contact-induced change of Van Coetsem (1988), the impact of contact on the genesis of Gisamba is too profound to be solely explained as "borrowing" under "recipient language agentivity". In the course of its history, Gisamba was clearly also subject to "imposition" under "source language agentivity". Either Gisamba speakers were well versed in their neighboring languages or many speakers of these languages shifted to Gisamba so that some of their features transformed the outlook of Gisamba itself. This is well in line with the little historical data available on Gisamba and its speech communities. First of all, speakers of Gisamba have been scattered through the Kwilu and Kwango provinces of the DRC and constituted minority groups within larger speech communities speaking other languages. That is why not all varieties of Gisamba exhibit the same contact-induced changes. The variety which the first author studied as part of her BA research, for instance, is spoken in an area where Kisuku is more prevalent and does not manifest the voicing of voiceless stops which is so characteristic of the variety studied here. Secondly, Gisamba is highly endangered. None of its speakers are monolingual and for none of them Gisamba is still the main means of communication. It is well-known that heavily endangered languages are more easily impacted by contact-induced change.

Acknowledgments

We wish to express our deep gratitude to Sara Pacchiarotti (BantU-Gent), Sebastian Dom (University of Gothenburg), Nadine Grimm (University of Rochester), two anonymous reviewers and the editors for their very useful feedback on earlier drafts of this article. Finally, we thank Sara Pacchiarotti for her help with developing the map in Figure 1. The usual disclaimers apply. We also thank the Special Research Fund of Ghent University as well as the European Research Foundation for research funding, more specifically through a Starting Grant (n° 284126) for the KongoKing project (2012-2016) and a Consolidator's Grant (n° 724275) for the BantuFirst project (2018-2022).

Abbreviations

APPL	applicative
BLR3	Bantu lexical reconstruction 3
C	consonant
C1	first stem consonant
C2	second stem consonant
CAUS	causative

DRC Democratic Republic of the Congo
F falling (tone)
FV final vowel
G glide
H high (tone)
INTR intransitive
IMP imperative
IMPO impositive
KLC Kikongo Language Cluster
L low (tone)
N noun, nasal
NEG negation
NP noun prefix
OP object prefix
PB Proto-Bantu
PFV perfective
PL plural
PP pronominal prefix
PRF perfect
PRS present
PST past
SEP separative
SG singular
SP subject prefix
SWB South-West Bantu
TR transitive
V vowel
V1 first stem vowel
V2 second stem vowel
WCB West-Coastal Bantu

References

Bastin, Yvonne, André Coupez & Michael Mann. 1999. *Continuity and divergence in the Bantu languages: Perspectives from a lexico-statistic study*. Tervuren: Royal Museum for Central Africa.

Bastin, Yvonne, André Coupez, Evariste Mumba & Thilo C. Schadeberg (eds.). 2002. *Bantu lexical reconstructions 3*. Tervuren: Royal Museum for Central Africa. Online database: http://linguistics.africamuseum.be/BLR3.html.

Blanchon, Jean Alain. 1998. Semantic/pragmatic conditions on the tonology of the Kongo noun-phrase: A diachronic hypothesis. In Larry M. Hyman & Charles W. Kisseberth (eds.), *Theoretical aspects of Bantu tone*, 1-32. Stanford: Stanford University, Center for the Study of Language and Information.

Blanchon, Jean Alain. 1999. “Tone cases” in Bantu group B.40. In Jean Alain Blanchon & Denis Creissels (eds.), *Issues in Bantu tonology*, 37-82. Cologne: Rüdiger Köppe.

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. 299-356.

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

Bostoen, Koen, Odjas Ndonda Tshiyayi & Gilles-Maurice de Schryver. 2013. On the origin of the royal Kongo title ngangula. *Africana Linguistica* 19. 53-83.

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

Bostoen, Koen & Gilles-Maurice de Schryver. 2018b. Seventeenth-century Kikongo is not the ancestor of present-day Kikongo. In Koen Bostoen & Inge Brinkman (eds.), *The Kongo kingdom: The origins, dynamics and cosmopolitan culture of an African polity*, 60-102. Cambridge: Cambridge University Press.

Daeleman, Jan. 1977. A comparison of some zone-B languages in Bantu. *Africana Linguistica* 7. 93-144.

Daeleman, Jan & Achiel Emiel Meeussen. 1983. The tonology of the noun in Kongo (Ntandu). *Africana Linguistica* 9. 137-148.

Dammann, Ernst. 1972. Das Meinhofsche Gesetz. *Afrika und Übersee* 55. 242-244.

Devos, Maud & Johan van der Auwera. 2013. Jespersen cycles in Bantu: Double and triple negation. *Journal of African Languages and Linguistics* 34. 205-274.

Dimmendaal, Gerrit. 2011. *Historical linguistics and the comparative study of African languages*. Amsterdam, Philadelphia: John Benjamins.

Eberhard, David M., Gary F. Simons & Charles D. Fennig. 2019. *Ethnologue: Languages of the world*. 22nd ed. Dallas, TX: SIL International.

Felix, Marc Leo. 1987. *100 peoples of Zaire and their sculpture: The handbook for collectors, scholars and students*. Brussels: Zaire Basin Art History Research Foundation.

Forges, Germaine. 1983. *Phonologie et morphologie du kwezo*. Tervuren: Musée royal de l'Afrique centrale.

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

Grégoire, Claire. 2003. The Bantu languages of the forest. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 349-370. London, New York: Routledge.

Grollemund, Rebecca, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti & 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.

Gusimana, Barthélemy. 1972. *Dictionnaire pende-français*. Bandundu: CEEBA.

Guthrie, Malcolm. 1948. *The classification of the Bantu languages*. London, New York: Oxford University Press for the International African Institute.

Guthrie, Malcolm. 1960. Teke radical structure and Common Bantu. *African Language Studies* 1. 1-15.

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, Robert Forkel, Martin Haspelmath & Sebastian Bank (eds.). 2020. *Glottolog 4.2.1*. Jena: Max Planck Institute for the Science of Human History. http://glottolog.org.

Hyman, Larry. 2019. Segmental phonology. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 128-149. 2nd ed. Oxford: Routledge.

Janson, Tore. 2007. Bantu spirantisation as an areal change. *Africana Linguistica* 13. 79-116.

Kabange Mukala, Ernest. 2009. *Eléments de transition linguistique entre les zones L et M du bantu : parenté et affinité*. Lubumbashi: Université de Lubumbashi, PhD dissertation.

Kanyamibwa, Melchior. 1982. *Confrontation historique du lexique du protobantu avec le cokwe (zone K), le sanga (zone L) et le rwanda (zone J)*. Ruhengeri: Université nationale du Rwanda, BA thesis.

Kashika Katanga. 1990. *Phonologie historique de la langue holu*. Lubumbashi: Institut supérieur pédagogique de Lubumbashi, Diploma thesis.

Kasuku-Kongini. 1984. *Langue kihungana : phonologie et systématique*. Kikwit: Institut supérieur pédagogique de Kikwit, Diploma thesis.

Kerremans, Richard. 1980. Nasale suivie de consonne sourde en proto-bantu. *Africana Linguistica* 8. 159-198.

Kifindi, Bunkheti. 1997. *Recherches en grammaire du suku : langue bantu (H.32) de la vallée du Kwango (Angola-Kongo Zaïre)*. Paris: Paris III, Université Sorbonne Nouvelle, PhD dissertation.

Koni Muluwa, Joseph. 2010. *Plantes, animaux et champignons en langues bantu. Étude comparée de phytonymes, zoonymes et myconymes en nsong, ngong, mpiin, mbuun et hungan (Bandundu, RD Congo)*. Brussels: Université libre de Bruxelles, PhD dissertation.

Koni Muluwa, Joseph. 2017. Negation in Ngong, Nsong and Nsamban: Preliminary study on morphosyntactic variation in three closely related B80 languages. Paper presented at the SOAS-BantUGent Workshop on Parametric Approaches to Morphosyntactic Variation in Bantu, Ghent University, March 31, 2017.

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

Koni Muluwa, Joseph & Koen Bostoen. 2019. Nsong (B85d). In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 414-448. 2nd ed. Oxford: Routledge.

Labroussi, Catherine. 2000. Spirant-devoicing and genetic classification in eastern Bantu. In H. Ekkehard Wolff & Orin D. Gensler (eds.), *Proceedings of the 2nd World Congress of African Linguistics, Leipzig 1997*, 323-358. Cologne: Rüdiger Köppe.

Maho, Jouni Filip. 2003. A classification of the Bantu languages: An update of Guthrie's referential system. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 639-651. London, New York: Routledge.

Maho, Jouni Filip. 2009. *NUGL online: The online version of the new updated Guthrie list, a referential classification of the Bantu languages*. https://brill.com/fileasset/downloads_products/35125_Bantu-New-updated-Guthrie-List.pdf.

Meeussen, Achiel Emiel. 1962. Meinhof's rule in Bantu. *African Language Studies* 3. 25-29.

Meeussen, Achiel Emiel. 1967. Bantu grammatical reconstructions. *Africana Linguistica* 3. 79-121.

Meinhof, Carl & Nicolaas Jacobus van Warmelo (eds.). 1932. *Introduction to the phonology of the Bantu languages being the English version of "Grundriss einer Lautlehre der Bantusprachen" by Carl Meinhof, translated, revised and enlarged in collaboration with the author and Dr. Alice Werner by N. J. Van Warmelo*. Berlin: Dietrich Reimer/Ernst Vohsen.

Mukash Kalel, Timothée. 1982. *Le kanyok, langue bantoue du Zaïre : phonologie, morphologie, syntagmatique*. Paris: Université Sorbonne Nouvelle, PhD dissertation.

Mulindabigwi, Emmanuel. 1981. *Confrontation historique de la phonologie et du lexique rund avec le protobantu et deux langues contemporaines : sanga et rwanda*. Butare: Université nationale du Rwanda, BA thesis.

Nash, Jay Arthur. 1992. *Aspects of Ruwund Grammar*. Urbana-Champaign, IL: University of Illinois at Urbana-Champaign, PhD dissertation.

Ndolo, Pius. 1972. *Essai sur la tonalité et la flexion verbale du gimbala Tervuren: Musée royal de l'Afrique centrale*. Tervuren: Musée royal de l'Afrique centrale.

Nguimbi-Mabiala, Jean-Noël. 1999. *Phonologie comparative et historique du koongo*. Lyon: Université Lumière Lyon 2, PhD dissertation.

Ntitenguha, François-Xavier. 1984. *Confrontation historique de la phonologie et du lexique pende avec le protobantu et deux langues bantoues contemporaines : rwanda et sanga*. Butare: Université nationale du Rwanda, BA thesis.

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

Pacchiarotti, Sara & Koen Bostoen. Forthcoming. Final vowel loss in Lower Kasai Bantu (DRC) as a contact-induced change. *Journal of Language Contact*.

Philippson, Gérard & Pierre Boungou. 1999. Éléments de tonologie beembe (H.11). In Jean Alain Blanchon & Denis Creissels (eds.), *Issues in Bantu tonology*, 83-107. Cologne: Rüdiger Köppe.

Piper, Klaus. 1977. *Elemente des Suku. Zur Phonologie und Morphologie einer Bantusprache*. Leiden: Rijksuniversiteit Leiden, PhD dissertation.

Plancquaert, Michel. 1971. *Les Yaka. Essai d'histoire*. Tervuren: Musée royal de l'Afrique centrale.

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

Ruttenberg, Piet. 2000. *Lexique yaka-francais, français-yaka*. Munich: Lincom Europa.

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

Schryver, Gilles-Maurice de, Rebecca Grollemund, Simon Branford & Koen Bostoen. 2015. Introducing a state-of-the-art phylogenetic classification of the Kikongo language cluster. *Africana Linguistica* 21. 87-162.

Swadesh, Morris. 1971. *The origin and diversification of language*. Chicago: Aldine.

Swartenbroeckx, Pierre. 1973. *Dictionnaire kikongo et kituba-français*. Bandundu: Ceeba.

Takizala, Alexis. 1974. *Studies in the grammar of Kihungan*. San Diego, CA: University of California, PhD dissertation.

Torday, Emil & T. Athol Joyce. 1907. On the ethnology of the south-western Congo Free State. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 37. 133-156.

Van Acker, Sifra. 2016. *De fonologie van het Kitsamba (Bantoe, L12a, DR Congo) vanuit een synchroon en diachroon perspectief*. Ghent: Ghent University, BA thesis.

Van Acker, Sifra. 2018. *A pilot study of Kisamba (Bantu, L12a, DRC): Phonology, morphology, wordlist and some texts*. Ghent: Ghent University, MA thesis.

Van Bulck, Gaston. 1948. *Les recherches linguistiques au Congo belge : résultats, acquis, nouvelles enquêtes à entreprendre*. Brussels: Librairie Falk - G. van Campenhout.

Van Coetsem, Frans. 1988. *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris.

Van den Eynde, Karel. 1968. *Éléments de grammaire yaka : phonologie et morphologie flexionnelle*. Kinshasa: Université Lovanium, Publications Universitaires.

Vansina, Jan. 1995. New linguistic evidence and the Bantu expansion. *Journal of African History* 36. 173-195.

Appendix A: Gisamba-English wordlist

The wordlist consists of 748 entries and is alphabetically ordered. The entries (and their tones) are based on elicitation in isolation. They are followed by the part of speech between brackets, along with the singular and plural classes when this is applicable. If we do not mention noun classes, it is because we do not know them. Most often it concerns prefixless nouns for which we have neither a plural nor concord patterns. The verbs are not written as infinitives, but as verb bases, because this results in a wordlist which is easier to search alphabetically. The last column represents the Bantu lexical reconstructions (Bastin *et al.* 2002) that can be linked to the present-day reflex in question. The index number of each Bantu lexical reconstruction is added, so one may find it easily in BLR3. If there is no index number given, then the reconstruction in question does not occur in BLR3 and is tentatively proposed based on comparative West-Coastal Bantu data collected during the KongoKing (2012-2016) and BantuFirst (2018-2022) projects. A translation of the reconstruction is given when it does not correspond with the translation of the reflex in Gisamba.

adá (adv.) ‘even’
ágag (adv.) ‘again, still’
ba/mába (n. 5a/6) ‘palm tree’ – *bá (1)
bádígíl (v.) ‘hold back a fart’
bág (v.) ‘tear’ – *bàag (55)
bágaga (adj.) ‘certain’
bágála/babágála (n. 1a/2) ‘man’ – *bàgàdà (56)
bál/mabál (n. 5a/6) ‘liver’ – *bàdì (9093)
balégag (adv.) ‘maybe’
balínéd (n. 2) ‘glasses’
bálúmúg (v.) ‘turn, change, become’
bámbígís (v.) ‘paste’ – *bàmb (74)
bánts (v.) ‘reason, think’ – *banj
báw (pron.) ‘they’
bél (v.) ‘be sick’ – *béed (119)
bélúg (v.) ‘be healed’ – *béeduk (7892)
bén/mabén (n. 5a/6) ‘breast’ – *béénè (147)
bénd (v.) ‘pull’ – *bend
béng (v.) ‘be red’ – *bèng (151)
bibéndé (n. 8) ‘iron’
bidʒó (adj.) ‘little’
bíl (v.) ‘forget’ – cf. *jíbɪdɪd (3327)
bímbíl (v.) ‘inflate’ – *bímb (240)
bín (pron.) ‘you (PL)’
bo (v.) ‘take’
bóg (n. 14) ‘mushroom’ – *jokʊ
bógón (v.) ‘grind, crush, break’
bógóg (v.) ‘break, collapse’
ból (num.) ‘two’ – *jòdè (3523)
ból (v.) ‘rot’ – *bòd (253)
bóng (v.) ‘get’ – *bong
bong/mábong (n. 5a/6) ‘knee’ – *bóngó (275)
bóngól (v.) ‘restore’
bótʃín (v.) ‘run, flee’
bótʃ (n. 14) ‘honey’ – *jíkì
bóvulu (num.) ‘a lot’ – *bùd (367)
búd (v.) ‘give birth’ – *bʊ́t (346)
búd/mád (n. 14/6) ‘gun’ – *táà (9207) ‘bow’
búd bo mbá/mabúd ma mbá (n. 14/6) ‘firearm’
budzób (n. 14) ‘idiocy’
búg lí mígand/mabúg má mígand (n. 5a/6) ‘book’ – cf. *kàndá (1706) ‘letter’
bugé (n. 14) ‘slowness’
bugóg (n. 14) ‘heaviness’
bug (n. 14) ‘fufu’
bugáf (n. 14) ‘lie’
búl (adj.) ‘wrong, evil’
búl (v.) ‘hit, break, peel’ – *bʊ́d (297)

búl musémbu (v.) ‘whistle’ – *cémbò (535) ‘horn’
bul/mábul (n. 5a/6) ‘village’ – *dá (781)
bulé (n. 14) ‘blue’
bulémfu (n. 14) ‘obedience’
búli/mabúli (n. 5a/6) ‘testicle’
bulímbu/ndímbu (n. 14/10) ‘glue’ – *dìmbò (985) ‘birdlime’
bulótʃi (n. 14) ‘witchcraft’ – *dògì (1104)
búlu (n. 14) ‘disease’
bulu mu mundél/mábulu ma mindél (n. 5a/6) ‘town’
búlúmwáli/mabúlúmwáli (n. 5a/6) ‘cabinet’
bulus/mábulus (n. 5a/6) ‘t-shirt’
bumwân (n. 14) ‘childhood’ – *jánà (3203) ‘child’
bwa (v.) ‘fall, collapse’ – *bù (281)
bwád (v.) ‘carry, wear’ – *búat (283)
bwád (v.) ‘flash’
bwád mulél (v.) ‘dress (lit. wear clothing)’ – *búat (283) ‘wear’
bwâd/mabwâd (n. 5a/6) ‘canoe’ – *játò (3252)
bwátʃ (v.) ‘sit’ – *buat (4746)
bwés (v.) ‘add’
bwíd (n. 14) ‘blackness’
bwón (adv.) ‘so’
dá (v.) ‘tell’ – *tá (2710) ‘call, name’
dá (v.) ‘launch’ – *tá (2708)
dá búgaf (v.) ‘lie (lit. tell a lie)’
dá lwíʃ (v.) ‘sneeze’
dád/badád (n. 1a/2) ‘father, parent’ – *tààtá (2806)
dád múk^{h}ed/badád bá bák^{h}ed (n. 1a/2) ‘aunt’ – *kádí (1674) ‘woman, wife’
dád ó mbud/badád bá bámbud (n. 1a/2) ‘grandfather’ – *bútò (350) ‘relative’
dádu (num.) ‘three’ – *tátù (2811)
dál (v.) ‘visit, admire, watch’ – *tád (2718) ‘look, look at’
dáladál/madáladál (n. 5a/6) ‘window’
dáli/madáli (n. 5a/6) ‘stone’ – *tádè (2726)
dam/mádam (n. 5a/6) ‘cheek’ – *támà (2744)
dámb (v.) ‘play’ – *támb (2757)
dámbúl (v.) ‘walk’ – *támbud (2755)
dánd (v.) ‘slim, become thin’ – *tànd (2770) ‘spread, tr.; extend, tr.’
dáng (v.) ‘read, count’ – *táng (2786)
dángidángi/madángidángi (n. 5a/6) ‘thigh’
dangu/mádangu (n. 5a/6) ‘hour, sun’ – *tángó (2797)
dangu lyáli (adv.) ‘now (lit. hour this)’
dangu lyóts (adv.) ‘constantly (lit. hour all)’
dánín (v.) ‘monitor, supervise’
dánu (num.) ‘five’ – *táànò (2768)
déb-débé (adj.) ‘delicateness’
dégéd (v.) ‘tremble’
dég (v.) ‘sell’ – *tég (2824)
dél (v.) ‘call, name’ – *tá (2710)

démbés (v.) ‘accept, believe’
dend (v.) ‘underline’
dénʃ/madénʃ (n. 5a/6) ‘mirror’
dií (prep.) ‘until, up to’
dín (v.) ‘flee’ – *tíɪn (2899) ‘fear; run away’
dín (adj.) ‘fast’ – *tíɪnʊ ‘speed’
dód (n.) ‘clay’ – *tótò ‘soil’
dóg (v.) ‘boil up’ – *tòk (2967)
dómbóg (v.) ‘increase’
dúd mwân (v.) ‘raise a child’
dud/mádud (n. 5a/6) ‘cloud’ – *tù (3093)
dúg (v.) ‘exit, appear’ – *tʊ́ʊk (3052) ‘come from’
dúgís (v.) ‘provoke’
dulu/mádulu (n. 5a/6) ‘chest’ – *túdò (3044)
dúlúg (v.) ‘come down’ – *túdʊk (4648)
dúlúl (v.) ‘put, deposit’ – *tʊ́ʊd (3038)
dún (v.) ‘refuse’ – *tún (3065) ‘deny’
dún gísal (v.) ‘strike’
dúng (v.) ‘build, construct, sew’ – *tʊ́ng (3081)
dwál (v.) ‘drive’
dzandu/mádzandu (n. 5a/6) ‘market’ – *jánjʊ (8599)
dzél (n. 5a) ‘sand’ – cf. *jédù (9501) ‘clear ground’
dzódóg (v.) ‘jump’
dzól (v.) ‘want’ – *jod ‘want, desire, love’
dzu (v.) ‘kill’ – *du
dzúm/madzúm (n. 5a/6) ‘belly’ – *dumʊ
dzwa (v.) ‘get, receive’
ʤid/mádʤid (n. 5a/6) ‘knot’
ʤíg (v.) ‘bury’ – *dìɪk (1044)
ʤig/mádʤig (n. 5a/6) ‘room’
ʤím (v.) ‘switch off’ – *dím (1046) ‘extinguish’
ʤin/mádʤin (n. 5a/6) ‘name’ – *jínà (3464)
ʤíng (v.) ‘live’ – *ding
ʤúng (v.) ‘surround’ – cf. *dʊnga ‘ring, bracelet’
e (conj./prep.) ‘and/with’
é (v.) ‘go’
ebún (adv.) ‘then’
ebúnú (adv.) ‘so’
ebwál, elónd (adv.) ‘on’
éé (intj.) ‘yes’
enda lud (v.) ‘walk’ – *gènd (1362)
éntswal (v.) ‘press’
edét^{h}i (ord.) ‘before, first, already’
fáng/mafáng (n. 5a/6) ‘pipe’
fáníg (v.) ‘compare’ – cf. *púan (2670) ‘resemble each other’
féd (v.) ‘have to’
fĩ (v.) ‘suck’ – *píp (2583)

fílímíg (v.) ‘cover’
finam (adj.) ‘close’ – *pìn (2572) ‘press, squeeze’
fófól/mafófól (n. 5a/6) ‘match’
fúb (n.) ‘ashes’
fúd (v.) ‘pay’ – *pùt (2694)
fúgám (v.) ‘kneel’ – *kúkam (2111)
fúlúmún (v.) ‘breathe’ – cf. *púdì (2675) ‘breath’
fúnd (v.) ‘accuse’ – *túnd (3122)
fúnʃ/mafúnʃ (n. 5a/6) ‘storm’
fwám (v.) ‘be able’
gábán (v.) ‘distribute, divide up’ – *gàban (8823)
gád (v.) ‘peel’ – *kat (9520) ‘cut’
gadug (v.) ‘leave’ – *katʊk ‘fly away, leave’
gag (adj.) ‘other’
gág (adv.) ‘only, inevitably’
gágu/magágu (n. 5a/6) ‘monkey’ – *gàgì (9337)
gál (v.) ‘be’ – *jìkad (3441)
gál e (v.) ‘have, own’ – *jìkad (3441)
gál e masóngo (v.) ‘be right’
gál e mfun (v.) ‘require’
gál e ntsal (v.) ‘need’
gál mu mbímb (v.) ‘be close’
gál ngol (v.) ‘be strong’ – *jìkad (3441) & *gòdò (1419) ‘strength’
gal/mágal (n. 5a/6) ‘charcoal’ – *kádà (1662)
gám (v.) ‘embrace, catch, hold, keep’ – *kám (1689)
gámís (v.) ‘pinch’
gán/magán (n. 5a/6) ‘mouth’ – *nùà (4709)
gandín/dundín (n. 12/13) ‘bucket’
gáng (v.) ‘link, attach, knot, close’ – *gàng (1331) ‘tie up’
gáng bítʃo (v.) ‘wink’ – *jícò (3405) ‘eye’
ganga mong/dondo myong (n. 12/13) ‘hill’
gangéng/bingéng (n. 12/8) ‘sparrowhawk’
gángúl (v.) ‘open’ – *kangʊd (8779)
gap (v.) ‘share’ – *gàb (1274) ‘divide; give away; make present’
gás (v.) ‘swim’ – *kac (9563)
gásál (v.) ‘deforest’ – *cád (404) ‘work’
gátsu/magátsu (n. 5a/6) ‘kola nut’ – *kacʊ
ge/meg (n. 5a/6) ‘egg’ – *gí (1368)
gél (adv.) ‘yet’
gémínín (v.) ‘hold, keep’
génínígˊ (v.) ‘leave open’
gés (v.) ‘cut’ – *kèc (1752)
gés bídʒini bídʒini (v.) ‘cut into pieces’
gibá/bibá (n. 7/8) ‘plague’
gíba/bíba (n. 7/8) ‘pineapple’
gíbag/bíbag (n. 7/8) ‘wall’ – cf. *bák (66) ‘build’
gibéd/bibéd (n. 7/8) ‘savannah, plain’

gíbeg/bíbeg (n. 7/8) ‘lip’
gíbogu/bíbogu (n. 7/8) ‘hippopotamus’ – *bògó (258) ‘buffalo (*Syncerus caffer*)’
gíbolol (n. 7) ‘bentness’
gibóng/bibóng (n. 7/8) ‘horn’ – *bongo (6767)
gibóng/bibóng (n. 7/8) ‘termite hill’
gíbulu/bíbulu (n. 7/8) ‘jackal’ – *búdú (309)
gidád (n. 7) ‘right’ – cf. *tààtá (2806) ‘father, my father’
gidádu (n. 7) ‘Wednesday’ – *tátù (2811) ‘three’
gidán (n. 7) ‘Friday’ – *táànò (2768) ‘five’
gidés/bidés (n. 7/8) ‘level’
gídi/bídi (n. 7/8) ‘chair’ – *tí (2884) ‘stool’
gídin/bídin (n. 7/8) ‘piece’ – *tínì (5531)
gído/bído (n. 7/8) ‘injury, wound’
gidúmb/gidúmb (n. 7/8) ‘war’ – *tùmbà (3117)
gídʒin/bídʒin (n. 7/8) ‘shortness’
gidʒól (n. 7) ‘Tuesday’ – *jòdè (3523) ‘two’
gífanus/bífanus (n. 7/8) ‘monster’
gigáláng/bibáláng (n. 7/8) ‘bridge’
gigánd/bigánd (n. 7/8) ‘family’ – *gàndá (1324) ‘house; village; clan’
gigés (n. 7/8) ‘strength’
gíguli/bíguli (n. 7/8) ‘bone’ – *kʊdɪ (4673)
gígum/bígum (n. 7/8) ‘cause, condition’
gihán/bihán (n. 7/8) ‘place’
gihúg/bihúg (n. 7/8) ‘plague’ – cf. *pùká (2638) ‘insect; bee; ant; caterpillar’
gigug/bigug (n. 7/8) ‘kitchen’ – *kʊkʊ
giláb gyé méd/biláb byé méd (n. 7/8) ‘pencil’
gilándí/bilándí (n. 7/8) ‘secretary’
giléf/biléf (n. 7/8) ‘beard’ – *dèdù (897)
gilímb/bilímb (n. 7/8) ‘symbol, sign, scar’ – *dìmbò (983)
gílu (n. 7) ‘sleep’ – *dó
gilúmb/bilúmb (n. 7/8) ‘day’ – *dʊmbʊ
gilúmbu gi pimbu (intj.) ‘good day’ – cf. *dʊmbʊ
gím gé ntso/bím byé ntso (n. 7/8) ‘furniture’ – *mà (2139) ‘thing’ & *jó (1600) ‘house’
gím/bím (n. 7/8) ‘thing’ – *mà (2139)
gim/bágim (n. 1a/2) ‘monkey sp.’ – *kímà (1798)
gimám (n. 7) ‘left’ – cf. *máá (2140) ‘my mother; mother’
gimfăm/bimfăm (n. 7/8) ‘heritage’
gímfug/bímfug (n. 7/8) ‘floor’
gimfúg/bimfúg (n. 7/8) ‘totality’
gímweng/bímweng (n. 7/8) ‘mosquito’
gín (v.) ‘dance’ – *kín (1805)
ginán (num.) ‘eight’ – *nàinài (6434)
gindánd/bindánd (n. 7/8) ‘cassava’
gindéd (n. 7) ‘Monday’
gíndug/bíndug (n. 7/8) ‘friendship’ – *dùgú (1175) ‘brother or sister (same sex); relative; friend’

gíng (v.) ‘wait’ – cf. *kíng (1812) ‘meet on path’?
gíngan/bíngan (n. 7/8) ‘adage’ – *gànò (1318) ‘tale, proverb’
gingársis/bingársis (n. 7/8) ‘exercise’
gingéng/bingéng (n. 7/8) ‘sparrowhawk’
gínguma/bínguma (n. 7/8) ‘pottery clay’
gípagal/bípagal (n. 7/8) ‘ant’
gísamba (n. 7) ‘Samba language’
gíntsal/bíntsal (n. 7/8) ‘nail, claw, toe’ – *jádà (1558)
gintsánts/bintsánts (n. 7/8) ‘box’
gintsúmb/bintsúmb (n. 7/8) ‘star’
gintswég/bintswég (n. 7/8) ‘secret’
gindúndu/bindúndu (n. 7/8) ‘flower’ – *túndà (4541)
gíp (v.) ‘travel’
gípap/bípap (n. 7/8) ‘cold’ – cf. *pépò (2478) ‘wind; cold’
gíped/biped (n. 7/8) ‘squash’
gípus/bípus (n. 7/8) ‘bark, skin’ – *pʊ̀cʊ (2621)
gísalu/bísalu (n. 7/8) ‘matter, thing, work’ – cf. *cád (404) ‘work (v.)’
gisánunu/bisánunu (n. 7/8) ‘comb’ – *càn (441)
giség ó mugel (adj.) ‘green’
giséle/biséle (n. 7/8) ‘cassava leaves’
gíson/bíson (n. 7/8) ‘character, letter’ – cf. *cón (661) ‘draw a line, write’
gíson/bison (n. 7/8) ‘cousin’
giʃúu/biʃúu (n. 7/8) ‘(dry) season’ – *cìpò (582) ‘dry season’
gíva (num.) ‘nine’ – *bùá (360)
giválu (n. 7) ‘shadow’
giyá (n. 7) ‘Thursday’ – *nàì (3683) ‘four’
giyáb (v.) ‘feel, smell, sense’
gizúʃí/bizúʃí (n. 7/8) ‘picture’
gób/magób (n. 5a/6) ‘glass, mirror, window’
góg/myóg (n. 5/4) ‘arm, hand’ – *bókò (260)/*jókò (3541)
gólís (v.) ‘make enter, introduce’
góm elondro golonʃi (v.) ‘crucify’
gómb (v.) ‘clean, sweep, brush’ – *kómb (1919)
gónd (v.) ‘miss’
góndan (v.) ‘miss’
góndo/magóndo (n. 5a/6) ‘banana’ – *kòndò (1939)
góngís (v.) ‘increase’
góngoʃig/bagóngoʃig (n. 1a/2) ‘cricket’
gónts (adj.) ‘each’
gu (adv.) ‘with’
gúlu (adj.) ‘old’ – *kúdú (2003)
gúlúd/magúlúd (n. 5a/6) ‘can’
gúlúmúg (v.) ‘flow, descend’
gúm (v.) ‘reach, arrive’ – cf. *kúm (2112) ‘come from’
gúm péw (v.) ‘become cold’
gúmi (num.) ‘ten’ – *kúmì (2027)
gúmi e móʃ (num.) ‘eleven’ – *kʊ́mì (2027) ‘ten’ & *mòtí (2212) ‘one’

gún (adv.) ‘where, there’
gún (v.) ‘plant’ – *kún (2041)
gusumbóngi/magusumbóngi (n. 5a/6) ‘parrot’ – *kùcù (1993)
gúúl/mííl (n. 5/4) ‘leg, foot’ – *gùdù (1490)
gwáng (v.) ‘scratch’
gwé (adj.) ‘how much/many’
gwél (v.) ‘marry’ – *kóɪd (7250)
gwíd (v.) ‘paint’
gyámf/magyámf (n. 5a/6) ‘bridge’
gyél/byél (n. 7/8) ‘door’ – *bédò (135)
gyés (adj.) ‘happy’
hán, hég (v.) ‘serve, give’ – *páan (2345)
háng (v.) ‘make, do, work’ – *páng (2397)
hántsúl (v.) ‘crush’
has/máhas (n. 5a/6) ‘twin’ – *pácà (2348)
hénd (v.) ‘lick’ – *pend
hín (v.) ‘push’ – cf. *pìn (2572) ‘press, squeeze (especially with the fingers)’
humun (v.) ‘blow, breathe’ – *púʊmʊd (2648)
hún (v.) ‘deceive’ – *pʊn
húw (v.) ‘improve’
íts (v.) ‘come’ – *jìj (3425)
íya (num.) ‘four’ – *nàì (3683)
gabíni/dubíni (n. 12/13) ‘toilet’
gafe (n. 12) ‘coffee’
gántʃ (conj.) ‘but’ – *kancɪ (9384)
gapágál/dupágál (n. 12/13) ‘bat’
gáy (n.) ‘grass’ – cf. *kájá (1736) ‘leaf; tobacco leaf; tobacco’
gémb/magémb (n. 5a/6) ‘plantain’
k^{h}áb (n. 9) ‘angriness, sharpness, meanness’
k^{h}ág/bak^{h}ág (n. 1N/2+) ‘ancestor’ – cf. *kààká (1685) ‘grandparent’
k^{h}áli (n. 9) ‘bitterness’ – *kádì (1667) ‘bitter, sour; sharp; fierce’
k^{h}ám (num.) ‘hundred’ – *kámá (1695)
k^{h}émbo (adj.) ‘happy’
k^{h}énd (n. 9) ‘pity, sadness’
k^{h}í (pron.) ‘what’ – *í (6432)
k^{h}í dangu (pron.) ‘when (lit. what hour)’
k^{h}ímudíndu (pron.) ‘how’
k^{h}intʃi/mák^{h}intʃi (n. 9/6+) ‘party’
k^{h}índe (pron.) ‘which’
k^{h}ol/bák^{h}ol (n. 1N/2+) ‘snail’ – *kódá (1880)
k^{h}omb/bák^{h}omb (n. 1N/2+) ‘goat’ – *kómbò (1926)
k^{h}óngi/bak^{h}óngi (n. 1N/2+) ‘eyebrow’ – *kongɪ
k^{h}óy/bak^{h}óy (n. 1N/2+) ‘lion, leopard’ – *gòì (7154)
k^{h}ud/bák^{h}ud (n. 1N/2+) ‘crab’
lamb (v.) ‘cook’ – *dámb (842)
lámíg (v.) ‘hang, hook’ – *dàmɪk (830) ‘stick (tr.)’
lánd (v.) ‘follow, continue’ – *dànd (853)

léf (v.) ‘borrow’ – cf. *dìp (1001) ‘pay; compensate’
léfis (v.) ‘lend’ – cf. *dìp (1001) ‘pay; compensate’
lél (adv.) ‘today’ – *dèèdó (896)
lél (v.) ‘say’
lélug (v.) ‘float’ – *déd (887)
lélumug (v.) ‘jump’
lémbés (v.) ‘light up’
lémbíg (v.) ‘soothe, appease’
lénd (v.) ‘be able’
léy (adj.) ‘long’ – *dàì (3705)
líín/míín (n. 5/6) ‘tooth’ – *jínò (3472)
líís/míís (n. 5/6) ‘eye’ – *jícò (3405)
líl (v.) ‘cry, sob’ – *dìd (959)
lílíʃ/ malílíʃ (n. 5/6) ‘window’
lilóng/malóng (n. 5/6) ‘course’ – *dòngà (1128) ‘river; valley; channel’
límb (v.) ‘wrap up’
limem/mamem (n. 5/6) ‘sheep’ – *méémé (2166)
límpa/mámpa (n. 5/6) ‘bread’
líng (v.) ‘search, request’ – *díng (997)
lís (v.) ‘feed, nourish’ – *díici (5871)
lisól/masól (n. 5/6) ‘(hi)story, dialogue, conversation’
lo (intj.) ‘no; not (NEG)’
lónd (n.) ‘height’
lóng (v.) ‘teach’ – *dòng (1124)
lóng/malóng (n. 11/6) ‘plate’ – *dòngà (1131)
lóngi/balóngi (n. 1a/2) ‘teacher’ – *dòng (1124) ‘teach’
los (n. 11) ‘rice’ – *joco (7364)
loʃ (n. 11) ‘noise’
lúd (v.) ‘pass, walk, drill’ – *dùt (1227)
lúd/malúd (n. 5a/6) ‘body’ – *dútù (1228)
ludámbi/madámbi (n. 11/6) ‘footprint’ – *támbí (2761)
ludámbi/t^{h}ámbi (n. 11/10) ‘paw’ – *támbí (2761) ‘sole of foot, footprint’
ludáng/madáng (n. 11/6) ‘branch’
ludónd (n. 11) ‘generosity, love’ – *tónd (2980) ‘desire, search for’
ludʒidu (n. 11) ‘rudeness, impoliteness’
ludʒíng (n. 11) ‘life’ – *ding ‘live’
lufúd/mafúd (n. 11/6) ‘present, gift’
lufúd/ malufúd (n. 5a/6) ‘destiny’
lúgay/k^{h}ay (n. 11/10) ‘leaf’ – *kájá (1736)
lugóg lé ludʒidu (n. 11) ‘rude person’
lulémb/ndémb (n. 11/10) ‘finger’ – *démbó (923)
lúlim/málim (n. 11/6) ‘tongue’ – *dímì (973)
lumb (v.) ‘lay (an egg)’
lúmba/mba (n. 11/10) ‘palm nut’ – *bá (1) ‘oil-palm’
lúmfi/bámfi (n. 11/2) ‘white hair’ – *búì (364)
lumíng (n. 11) ‘Sunday’
lung (v.) ‘accomplish’ – *dòng (1204) ‘put straight, right’

lungáng/bangáng (n. 11/2) ‘thorn’
lúngub/ngub (n. 11/10) ‘peanut’ – *gʊba
lungwéni (n. 11) ‘chameleon’
lúnts (v.) ‘bite’ – *dʊnj
lupáng/mapáng (n. 11/6) ‘enclosure, plot, parcel, fence’
lupáy/mapáy (n. 11/6) ‘game bag’
lúp^{h}igu/máp^{h}igu (n. 11/6) ‘kidney’ – *pígò (2568)
lútsugi/tsugi (n. 11/10) ‘hair’ – *cùkí (715)
lwál (v.) ‘be injured’ – *dúad (1153)
lyá (v.) ‘eat’ – *dí (944)
lyang/mang (n. 5/6) ‘mango’
lyâng/mâng (n. 5/6) ‘branch’
lyóng/malyóng (n. 5a/6) ‘spear’ – *gòngá (1448)
mád (v.) ‘climb’
mádondo (intj.) ‘thank you’
mádumad (n. 6) ‘tomatoes’
máfin (n. 6) ‘pus’ – *píná (2574)
máfululul (n. 6) ‘foam’ – *púdò (2677)
magángu (n. 6) ‘lover’
magáʃín/bamagáʃín (n. 1a/2) ‘shop’
mágay (n. 6) ‘tobacco’ – *kájá (1736)
magúmá gíva (num.) ‘ninety’
mágwel (n. 6) ‘dowry’ – *kóɪd (7250) ‘marry’
maláf (n. 6) ‘drink, beverage, beer’ – *dagu; cf. *dògù (1108)
maláfu ma mba (n. 6) ‘palm wine’ – *dagu; cf. *dògù (1108) & *bá (1) ‘oil-palm’
málya (n. 6) ‘food’ – *díà (5850)
mám mú mbud/bamám bá bámbud (n. 1a/2) ‘grandmother’ – *mààmá (2146)
mám/bamám (n. 1a/2) ‘mother’ – *máá (2140)
mámbu (n. 6) ‘situation, problem’ – *jàmbò (3200) ‘affair’
man (v.) ‘finish’
mandagá (n. 6) ‘speech’ – *dáká (820) ‘tongue; language; jaw’
mántsogand/bámántsogand (n. 1a/2) ‘student’
mánu (n. 6) ‘earth’ – cf. *mani (8241) ‘stone’
masámb/bamasámb (n. 1a/2) ‘garden’ – cf. *càmbú (8467) ‘field; landed property’
masúb (n. 6) ‘urine’ – *cùb (753) ‘urinate’
maʃig (n. 6) ‘evening’ – *tíkà (2913)
maʃiga ma pimbu (intj.) ‘good evening’ – cf. *tíkà (2913) ‘evening’
máʃin (n. 6) ‘low’
maʃíʃ (n. 6) ‘corn, maize’
máts (v.) ‘throw’
mátʃ (n. 6) ‘fat’ – *gàdí (1300) ‘oil-palm; nut of oil-palm’
mátʃ má mba (n. 6) ‘palm oil’ – *gàdí (1300) ‘oil-palm; nut of oil-palm’ & *bá (1) ‘oil-palm’
mátʃ ma mbitʃi (n. 6) ‘animal fat’ – *gàdí (1300) ‘oil-palm; nut of oil-palm’ & *bídì (6135) ‘fish’
mavang (n. 6) ‘disorder’
mayél (n. 6) ‘intelligence’ – *jédà (3276) ‘wisdom’

mayón (n. 6) ‘yesterday’ – *gòdò (1420)
mbad/bámbad (n. 1N/2+) ‘tortoise’
mbátʃ (n. 9) ‘tomorrow’ – *bàdì (44)
mbál/mambál (n. 9/6+) ‘potato’
mbál/mambál (n. 9/6+) ‘time’
mbél/mambél (n. 9/6+) ‘knife’ – *bèèdé (124)
mbén/bambén (n. 1N/2+) ‘enemy, rebel’ – *méni
mbéng (n. 9) ‘red’ – *bèng (151)
mbi (v.) ‘hate’
mbí/bambí (n. 9/2+) ‘badness’ – *bîî (5841) ‘bad’
mbi/mámbi (n. 9/6+) ‘fault’ – *bîî (5841) ‘bad’
mbitʃ/bámbitʃ (n. 1N/2+) ‘animal’ – *bídì (6135)
mbitʃ ó mwêtʃ/bámbitʃ bá mwêtʃ (n. 1N/2+) ‘fish (lit. animal of water)’ – *bídì (6135) & *gèdì (1351) ‘stream’
mbímb (n. 9) ‘side’
mbómb/mambómb (n. 9/6+) ‘nose’ – *bòmbó (265)
mbud/bámbud (n. 1N/2+) ‘old person’
mbúk^{h}éd (n. 9) ‘cleanliness’
mbúnd/mambúnd (n. 9/6+) ‘heart’ – *bundʊ
mbúng (n.) ‘fog’ – *bùngì (4455)
mbung/mámbung (n. 9/6+) ‘harbor’ – *búngò (341) ‘beach; shore’
mbúnts/mambúnts (n. 9/6+) ‘face, forehead’ – *cú (680)
mbwa/bámbwa (n. 1N/2+) ‘dog’ – *búà (282)
mbwang/bámbwang (n. 9/2+) ‘jaw’ – *bángá (108)
mbwé gága (adv.) ‘often’
mé, meníní (pron.) ‘I, me’
még (v.) ‘try’ – *mìg (2180)
mén (v.) ‘germinate’
méng (n. 6) ‘blood’ – *jíngà (3483)
més/bamés (n. 1a/2) ‘table’
mfăm/bamfăm (n. 1N/2+) ‘boss’
mfĭmb (n. 9) ‘completeness’
mfĭnd (n. 9) ‘dirt’
mfugu/mámfugu (n. 9/6+) ‘debt, borrowing’ – *puka
mful/mámful (n. 9/6+) ‘year, rain’ – *búdà (368)
mfúmb/bamfúmb (n. 9/2+) ‘corpse’ – *dùmbi
mfun (n. 9) ‘necessity’
míligi (n. 4) ‘milk’
mímb (v.) ‘lay down, sleep’
mímb (v.) ‘live’
míʃí, gúm mú mfĭnd (v.) ‘smear, make dirty’
mizíg (n. 4) ‘music’
mól (adj.) ‘weak’
mon (v.) ‘see, show’ – *món (2206) ‘see’
móng/myóng (n. 3/4) ‘mountain’ – *gongʊ
móʃ (num.) ‘one’ – *mòtí (2212)
moʃi (adj.) ‘some, a few’ – *mòtí (2212) ‘one’

mpa (n. 9) 'novelty' – *pía (2495) 'new'
mpév/bampév (n. 1N/2+) 'soul'
mpíl/bampíl (n. 9/2+) 'way'
mu gimfúg (adv.) 'together'
mubágabág/mibágabág (n. 3/4) 'wing'
múd ó gísal/bád bá gísal (n. 1/2) 'worker, laborer' – cf. *ntù (3005) 'person' & *cádá (408) 'work'
múd/bád (n. 1/2) 'person' – *ntù (3005)
múd/míd (n. 3/4) 'tree' – *tí (2881)
múd/míd (n. 3/4) 'head' – *túè (3023)
mudámb/midámb (n. 3/4) 'trap' – *támb (2759) 'set trap'
mudánd/midánd (n. 3/4) 'back, spine'
mudangi/midangi (n. 3/4) 'reader, lector' – *táng (2786) 'read, count'
mudég/badég (n. 1/2) 'vendor, seller' – *tég (2824) 'sell'
múdu gágin/bádu bágin (n. 1/2) 'dancer' – *kín (1805) 'dance; play; gamble'
múdu wa lánd/bádu ba lánd (n. 1/2) 'disciple, follower' – cf. *ntù (3005) 'person' & *dànd (853) 'follow'
mudumb/midumb (n. 3/4) 'truck, car'
mudúngi/badúngi (n. 1/2) 'builder' – *túng (3081) 'build'
mudzágíʃ/badzágíʃ (n. 1/2) 'blacksmith'
mudzómb/badzómb (n. 1/2) 'hunter'
mugánd/migánd (n. 3/4) 'letter, note' – *kàndá (1706)
mugánd o ntsámb/migánd mi ntsámb (n. 3/4) 'Bible (lit. book of God)' – cf. *jàmbé (3196) 'God'
mugáʃ/migáʃ (n. 3/4) 'root' – *gangi
mugátʃ/bagátʃ (n. 1/2) 'wife' – *kádí (1674)
*mugég (*n. 1) 'small person' – *kéèkéè (7983)
múgil/mígil (n. 3/4) 'tail' – *kídà (1793)
mugó felem/bagó máfelem (n. 1/2) 'farmer'
mugógya/bagógya (n. 1/2) 'owner'
mugúb/migúb (n. 3/4) 'painting, color' – *gúbù (4550)
mugúmb/migúmb (n. 3/4) 'belly button' – *kʊmba
mugúng/migúng (n. 3/4) 'song' – cf. *gùngà (1514) 'bell; cuphorn'
mugúʃ/migúʃ (n. 3/4) 'fart'
múgwen/bágwen (n. 1/2) 'comparable person'
muhíg/bahíg (n. 1/2) 'slave' – *pìkà (2513)
mugád (prep.) 'in, inside' – *kàtí (1732)
mugéd/bagéd (n. 1/2) 'woman' – *káíntù (9300)
mugíl/migíl (n. 3/4) 'road' – *kɪdɪ (5961) 'path'
mulamb/milamb (n. 3/4) 'pants'
mulél/milél (n. 3/4) 'clothing, fabric' – *dédé (892)
mulíngi/milíngi (n. 3/4) 'fruit'
mulób/balób (n. 1/2) 'fisherman' – *dobi (6875)
mulómbi/balómbi (n. 1/2) 'applicant' – *dómb (1112) 'ask for'
mulong/milong (n. 3/4) 'fish trap'
mulótʃ/balótʃ (n. 1/2) 'wizard, witch' – *dògì (7089)
múlum/bálum (n. 1/2) 'husband' – *dúmì (1188)

mulyá/milyá (n. 3/4) ‘intestine’ – *díà (5850) ‘food’
mumín/mimín (n. 3/4) ‘throat’ – cf. *mìn (2190) ‘swallow’
mumóng/myóng (n. 3/4) ‘top’ – *gongʊ ‘mountain’
múmpe/mímpe (n. 3/4) ‘priest’
mundánd/mindánd (n. 3/4) ‘caterpillar’
munéng/minéng (n. 3/4) ‘ring’
múng (n.) ‘salt’ – *jùngúá (3641)
mungáng/mingáng (n. 3/4) ‘doctor, nurse’ – *gàngà (1332) ‘medicine-man’
muntsénts/bantsénts (n. 1/2) ‘foreigner, outsider’ – *jenja ‘stranger, visitor’
muntsóntsó/mintsóntsó (n. 3/4) ‘nail’
munun/banun (n. 1/2) ‘old man’ – *nùnù (2314)
mupép (n. 3) ‘wind’ – cf. *pèep (2469) ‘blow (as wind)’;
cf. *pépò (2478) ‘wind; cold’
mupép ó ngól/mipép myé ngól (n. 3/4) ‘storm’
– cf. *pépò (2478) ‘wind; cold’ & *gòdò (1419) ‘strength’
mupfĭ mísu/bapfĭ mísu (n. 1/2) ‘blind person’
músamba/básamba (n. 1/2) ‘Samba person’
musáng/misáng (n. 3/4) ‘bead’ – *cángá (478)
músob/mísob (n. 3/4) ‘worm’
musúg o pimbu (intj.) ‘good morning’
musug/misug (n. 3/4) ‘morning’ – *cúgù (761) ‘day of 24 hours’
musúní/misúní (n. 3/4) ‘meat’ – *cùnì (724)
múʃid/míʃid (n. 3/4) ‘forest’ – *títú (2948)
muʃíng/miʃíng (n. 3/4) ‘rope’ – *tíngà (2941) ‘bow-string; tendon, vein’
muʃíngíléd/miʃíngíléd (n. 3/4) ‘bike’
muwangu (n. 3) ‘playground’
mvud/mámvud (n. 9/6+) ‘answer’
mwámb (n. 3) ‘yellow’
mwân/bán (n. 1/2) ‘child’ – *jánà (3203)
mwân mugéd/bán ba bagéd (n. 1/2) ‘girl’ – *jánà (3203) ‘child’
& *káíntù (9300) ‘woman’
mwân o bágál/bán ba babágál (n. 1/2) ‘boy’ – *jánà (3203) ‘child’
& *bàgàdà (56) ‘man, male’
mwáy/myáy (n. 3/4) ‘yawn’ – *jájɪ (3254)
mwélu/myélu (n. 3/4) ‘entrance’ – *jédò (3286) ‘door’
mwémb/myémb (n. 3/4) ‘drug, medication’
mwêtʃ (n. 3) ‘water’ – *gèdì (1351) ‘stream’
mwêtʃ o nén (n. 3) ‘sea’
mwín (n. 3) ‘heat, warmth, sun’ – *jínyà (3480) ‘day, daylight’
mwínd/mínd (n. 3/4) ‘lamp’ – *jendʊ (7606)
mwítʃ (n. 3) ‘smoke’ – *jíkì (3442)
ná (pron.) ‘who’ – *nai (3682)
ndag/mándag (n. 9/6+) ‘voice, word’ – *dáká (822)
ndámbu (n. 9) ‘bit’ – *dàmbú (847) ‘tribute’
ndé (conj.) ‘if’
ndé, ndeníní (pron.) ‘he/she’
ndob/mándob (n. 9/6+) ‘fishhook’ – *dóbò (1093)

ndól/mandól (n. 9/6+) 'punishment'
ndóng/mandóng (n. 9/6+) 'line' – *dòng (1119) 'arrange'
ndóngámán (v.) 'be up'
ndótʃ (v.) 'dream' – cf. *dóot (1145)
ndótʃ/mandótʃ (n. 9/6+) 'dream' – *dòòdí (1098)
ndóy/bandóy (n. 1N/2+) 'homonym' – *dùkì (1180)
ndúg/bandúg (n. 1N/2+) 'friend' – *dùgú (1175)
ndúmb/bandúmb (n. 1N/2+) 'apostle'
ndund (n. 9) 'foam'
ndúnd/mandúnd (n. 9/6+) 'vegetable' – *dʊnda
ndung (n. 9) 'pepper' – *dúngú (1223)
ndúngundúngu/bandúngundúngu (n. 9/2+) 'moustache'
ndwál (n. 9) 'front'
ndʒíg bámbud (n. 9) 'tradition'
néné (n.) 'bigness' – *nénè (2255) 'big'
ngág/bangág (n. 9/2+) 'bag'
ngán (v.) 'bark'
ngánd/bangánd (n. 1N/2+) 'crocodile' – *gàndú (1326)
ngáng/bangáng (n. 1N/2+) 'healer' – *gàngà (1332) 'medicine-man'
ngáy/mangáy (n. 1N/6+) 'cat'
ngén/mangén (n. 9/6+) 'grain, seed'
ngól (n. 9) 'force, serious, energy' – *gòdò (1419) 'strength'
ngóm/mangóm (n. 9/6+) 'drum' – *gòmà (1429)
ngónd/mangónd (n. 9/6+) 'moon, month' – *gòndà (1444)
ngónd o mful (n. 9) 'rainy season'
ngóy (n. 9) 'behind'
ngul/bángul (n. 1N/2+) 'pig' – *gùdú (1493)
nguli/bánguli (n. 1N/2+) 'mother' – *gʊdɪ
ngúnd/mangúnd (n. 9/6+) 'field' – *gùndà (1509) 'forest, garden'
ngúng/mangúng (n. 9/6+) 'bell' – *gùngà (1514)
ngúnts/bangúnts (n. 1N/2+) 'young boy'
ngwáʃ/bangwáʃ (n. 1N/2+) 'uncle'
níg (v.) 'scrub' – *nik 'stamp, smash, grind'
nog (v.) 'rain' – *nók (2286)
nótʃ/banótʃ (n. 1a/2) 'bee' – *jíkì (3350)
ntság/mantság (n. 9/6+) 'game' – cf. *càkan (427) 'play'
ntsal (n. 9) 'wing' – cf. *cádá (406) 'feather'
ntsál (n. 9) 'hunger' – *jàdà (1555)
ntsál á mwétʃ (n. 9) 'thirst (lit. hunger for water)' – *jàdà (1555) & *gèdì (1351) 'stream'
ntsáli/bantsáli (n. 1N/2+) 'follower'
ntsámb (n. 1N) 'divinity, God' – *jàmbé (3196)
ntsángu/bantsángu (n. 9/2+) 'novelty' – *càngò (479) 'news'
ntsatʃ/bántsatʃ (n. 9/2+) 'lightning' – *jàdí (1561)
ntsé, ntseníní (pron.) 'you (SG) '
ntsém/mantsém (n. 9/6+) 'light'
ntsó/mantsó (n. 9/6+) 'house' – *jó (1600)

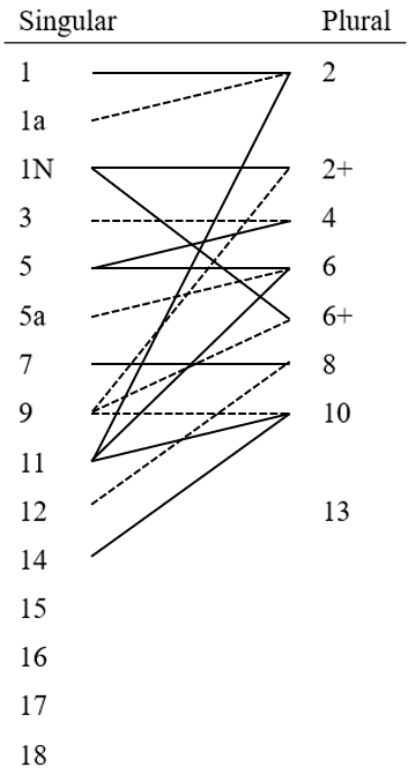
ntsó múngang/mantsó má míngang (n. 9/6+) 'hospital (lit. house of doctor)' – cf. *jó (1600) 'house' & cf. *gàngà (1334) 'medicine'
ntsógand/mantsógand (n. 9/6+) 'school, class' – cf. *jó (1600) 'house' & cf. *kàndá (1706) 'letter'
ntson/bántson (n. 9/2+) 'taboo, shame' – *cónì (664)
ntsóng (n. 9) 'top' – *còngè (674) 'point'
ntsóngil/mantsóngil (n. 9/6+) 'point' – *còngìdò (6827)
ntsug/mántsug (n. 9/6+) 'top' – *cúg (759) 'be finished, come to an end'
ntsug/bántsug (n. 9/2+) 'end' – *cúg (759) 'be finished, come to an end'
ntsúng/mantsúng (n. 1N/6+) 'termite'
ntsúng/mantsúng (n. 9/6+) 'cooking pot' – *jùngú (1632)
ntswálu (n. 9) 'immediacy'
ntʃi/mantʃi (n. 9/6+) 'country' – *cí (562)
ntʃi/mantʃi (n. 1N/6+) 'fly' – *gì (1389)
ntʃil (prep.) 'outside' – *jìdá (1594) 'path'
ntʃíl/mantʃíl (n. 9/6+) 'road, line' – *jìdá (1594)
ntʃímb (n. 9) 'money' – *címbí (617) 'iron, cowry'
ntʃóg/bantʃóg (n. 1N/2+) 'elephant' – *jògù (1606)
nún/banún (n. 1a/2) 'bird' – *jùnì (1627)
nung (v.) 'win' – *dʊng 'to win (a trial), to conquer'
nwa (v.) 'drink' – *nú (2342)
nwa mabén (v.) 'be breastfed' – *nú (2342) 'drink' & *béénè (147) 'breast'
nwán (v.) 'fight' – *dʊ̀an (1151)
nwánin (v.) 'rush'
nyámug (v.) 'get up, stand up'
nyámún (v.) 'raise, lift'
nyóng (v.) 'twist' – *níong (2275)
óts (adj.) 'all' – *ncò (627)
pá bumugu (v.) 'be flat'
pálág (n.) 'flat'
pang(í) dúgálag a yándi (n. 1N/2+) 'neighbor'
paʃ/mápaʃ (n. 5a/6) 'difficulty' – cf. *pàká (9599)
patʃ (adj.) 'hard, difficult' – cf. *pàká (9599) 'difficulty, contesting'
páypáy/mapáypáy (n. 5a/6) 'papaya'
pfá (v.) 'die' – *kú (2089)
pfum/bápfum (n. 1N/2+) 'chief' – *kúmú (2118)
pʰáb (n. 9) 'flatness'
pʰámb (n. 9) 'emptiness, unhappiness'
pʰándi (n. 9) 'situation'
pʰáng/bapʰáng (n. 1N/2+) 'sibling' – *pangɪ (8719)
pʰáng ó múgéd/bapʰáng bá bágéd (n. 1N/2+) 'sister' – *pangɪ (8719) & *káíntʊ̀ (9300) 'woman'
pʰángi ó bágál/bapʰáng bá bágál (n. 1N/2+) 'brother' – *pangɪ (8719) & *bàgàdà (56) 'man, male'
pʰásu/mapʰásu (1N/6+) 'locust' – *pààcò (9185)
pʰémb (n. 9) 'white' – *pémbé (2448)
p(ʰ)entsá (adv.) 'really'

pʰew (n. 9) ‘coldness’ – *píò (2558)
pʰíb (n. 9) ‘night’ – *pɪmpa (6015)
pʰíd (pron.) ‘we’
pʰós/mapʰós (n. 9/6+) ‘week, Saturday’
pʰu/mápu (n. 9/6+) ‘hat’
pʰug/bápʰug (n. 1N/2+) ‘rat’ – *púkù (2642)
pʰuntsu/bápʰuntsu (n. 9/2+) ‘arrow’
pʰúd (n. 9) ‘Europe’
pimb (adj.) ‘good, easy’
podopod (n.) ‘porridge’
pondo (n.) ‘millet’ – *pòndó (6702)
púlúmúg (v.) ‘fly’ – *pùdʊmʊk (4634) ‘go down; fall’
púlúpulu (n.) ‘diarrhea’
puluʃ/bápuluʃ (n. 1a/2) ‘police officer’
pus (v.) ‘drive’
pwede ‘silence’
pwis ‘thirst’
sá (v.) ‘put, contribute’ – *tá (2708)
sá tsongil (v.) ‘sharpen’ – *còng (670)
sábád/masábád (n. 5a/6) ‘shoe’
sábúg (v.) ‘cross’ – *càbʊk (394)
sal (v.) ‘stay, remain’ – *tígad (2911)
sálís (v.) ‘help’ – *cadɪc
sámb (v.) ‘pray’ – *càmb (8437) ‘meet’
sambanu (num.) ‘six’ – *cààmànò (433)
sámbu (conj.) ‘because’
sanún (v.) ‘comb’ – cf. *càn (441)
se (adv.) ‘also’
ségés (v.) ‘sharpen’ – *cekɪc
sél (v.) ‘weed’
sém (v.) ‘shine’
séy (v.) ‘laugh’ – *cèp (556)
ʃím (v.) ‘dig, grow’ – *tím (2918)
ʃímís (v.) ‘burn’
só ngon (v.) ‘snore’
sógól (v.) ‘gather’
sól (v.) ‘choose’ – *cóod (638)
sólól (v.) ‘chat’ – *codʊd
sónig (v.) ‘write (down), note’ – *cónɪk (662)
sub (v.) ‘urinate’ – *cùb (753)
súg (n. 5a) ‘room’ – cf. *cúkù (766) ‘night’
sugáli (n.) ‘sugar’
súgumúdu/masúgumúdu (n. 5a/6) ‘gorilla’
súmb (v.) ‘buy’ – *cúmb (720)
swám (v.) ‘hide’ – *còam (7215)
ʃamug (v.) ‘wake up’
tʰálátʃ/batʰálátʃ (n. 1N/2+) ‘snake’

tʰálu/matʰálu (n. 9/6+) 'price, number' – *tàdò (2736)
tʰo (n.) 'source' – cf. *to (7180) 'river'
tságád (v.) 'hunt' – *càk (420) 'drive; chase'
tsági/matsági (n. 9/6+) 'bag'
tsál/matsál (n. 9/6+) 'feather' – *cádá (406)
tsámbwali (num.) 'seven' – *càmbʊ̀àdì (440)
tsáts (v.) 'treat, nurse, raise' – *canc
tsáts mwân (v.) 'raise a child'
tseg/mátseg (n. 9/6+) 'bush' – *càká (423) 'thicket; bush-country'
tsítsa (adj.) 'small' – *cíicì (6426) 'ant sp. small; insect'
tsúg (v.) 'throw up' – cf. *dúk (1179) 'vomit'; cf. *dúk (1252) 'vomit'
tsúli/matsúli (n. 9/6+) 'smell' – *cʊ̀dì (704)
tsúng (adj.) 'open'
tsutsu/bátsutsu (n. 1N/2+) 'chicken' – *cúcú (698)
tswéng/batswéng (n. 1N/2+) 'little bird'
tswí/matswí (n. 9/6+) 'ear' – *túì (3030)
tʃa (prep.) 'under'
tʃin/bátʃin (n. 1a/2) 'louse' – *ná (2234)
tʃingu/mátʃingu (n. 5a/6) 'neck' – *kíngó (1845)
u (conj.) 'or'
váláʃ/maváláʃ (n. 5a/6) 'suitcase'
vés (v.) 'boo, hoot'
vúdug (v.) 'return' – *bútʊk (387)
vúdúl (v.) 'return (a thing), reply'
vúgáná (v.) 'reunite'
vúgís (v.) 'mix'
vulu (adj.) 'many' – *bùd (367)
válúg (v.) 'disappear'
wánán (v.) 'meet'
wónd (v.) 'be tired, be weak'
wul/máwul (n. 5a/6) 'hole, grave, tomb' – *bʊdʊ
yá (conj., prep.) 'as'
yá(ya) (adv.) 'here'
yáb (v.) 'know' – *jíjab (6207)
yágúl (v.) 'announce, say, talk' – *jàkʊd (3172) 'answer; speak'
yágúl lupfá (v.) 'announce, predict death'
yál (v.) 'rule' – *bíad (166)
yámbúl (v.) 'finish, let, exhibit'
yándíg (conj.) 'since' – *bádɪk (23)
yándíg (v.) 'begin' – *bádɪk (23)
yang/máyang (n. 5a/6) 'barrage, dam'
yanga/máyanga (n. 5a/6) 'pond, lake' – *jànjá (3221)
yángís (v.) 'bother'
yáníg (v.) 'dry' – *jánɪk (3206) 'spread to dry in the sun; to spread out'
yébél (v.) 'wash, clean'
yél (v.) 'fish'
yemb/máyemb (n. 5a/6) 'shoulder'

yílámán (v.) ‘improve’
yílíg (v.) ‘arrange, prepare, fix’
yím (v.) ‘dry (intr.) ‘
yímb (v.) ‘sing’ – *jímb (3361)
yímb/mayímb (n. 5a/6) ‘little sparrowhawk’
yín (v.) ‘push’
yíndúl (v.) ‘think’ – *jɪnduɗ
yíp (v.) ‘steal’ – *jíb (3387)
yóngólól/mayóngólól (n. 5a/6) ‘centipede’ – *góngòdó (1453)
yóngon/mayóngon (n. 5a/6) ‘chameleon’
yúg (v.) ‘hear, feel’ – *jʊg (3604)
yúg móy (v.) ‘be afraid’ – *jʊg (3604) ‘hear’ & *jógà (3528) ‘fear’
yúgú kʰábu (v.) ‘spoil’
yúl (v.) ‘fill, be full’ – *jíjɔdi (6203)
yúlu (n.) ‘sky’ – *gòdò (1486)
yúngúl (v.) ‘sift’ – *jʊ̀ngɔd (5046)
yúy/mayúy (n. 5a/6) ‘spider’

Appendix B: Gisamba noun class system



In order to improve readability, some lines have been dashed.

Figure 2 — Gisamba noun class system

Table 4 — Gisamba noun class system: Examples

Class	NP	Examples
1	mu-	*mulótʃ* ‘wizard, witch’; *múlum* ‘husband’
1a	∅-	*bágála* ‘man’
1N	N-[10]	*mbwa* ‘dog’; *ntʃi* ‘fly’
2	ba-	*balótʃ* ‘wizards, witches’; *bálum* ‘husbands’; *babágála* ‘men’; *bámfi* ‘white hair (PL)’
2+	ba-(N)-	*bámbwa* ‘dogs’; *bamfúmb* ‘corpses’
3	mu-	*mudumb* ‘truck, car’; *múmpe* ‘priest’
4	mi-	*midumb* ‘trucks, cars’; *mímpe* ‘priest’; *myóg* ‘arms, hands’
5	li-	*limem* ‘sheep’; *límpa* ‘bread’
5a	∅-	*gémb* ‘plantain’; *góg* ‘arm, hand’
6	ma-	*mamem* ‘sheep’; *mámpa* ‘bread’; *magámb* ‘plantains’; *madámbi* ‘footprints’
6+	ma-(N)-	*mantʃi* ‘flies’; *mambél* ‘knives’
7	gi-	*gigug* ‘kitchen’; *gído* ‘injury, wound’
8	bi-	*bigug* ‘kitchens’; *bído* ‘injuries, wounds’; *bingéng* ‘sparrowhawks’
9	N-[11]	*mbél* ‘knife’; *mfúmb* ‘corpse’; *ndímbu* ‘glue (SG)’
10	N-	*ndémb* ‘fingers’; *k^{h}ay* ‘leaves’
11	lu-	*ludámbi* ‘footprint’; *lúgay* ‘leaf’; *lúmfi* ‘white hair (SG)’
12	ga-	*gandín* ‘bucket’; *gangéng* ‘sparrowhawk’
13	du-	*dundín* ‘buckets’
14	bu-	*budzób* ‘idiocy’; *búlu* ‘disease’; *bulímbu* ‘glue (PL)’
15	gu-	*gubúd* ‘to give birth’
17	gu-	*gulónd* ‘above’
18	mu-	*mulónd* ‘above’

10. Although diachronically speaking the nasal of stem-initial prenasalized consonants in class 1N nouns is a reflex of the Proto-Bantu noun prefix of cl. 9/10, it can synchronically also be analyzed as being part of the noun stem. If so, class 1N nouns can be considered as prefixless, which implies that 1a and 1N merge.

11. Synchronically, the initial nasal of class 9/10 nouns can also be analyzed as part of the noun stem instead of as a prefix.

Comptes rendus / Book reviews

Michel Dieu, Louis Perrois & Henry Tourneux, *Dictionnaire encyclopédique koma-gɨ́mbē/français (monts Alantika, Nord-Cameroun), suivi d'un index français/koma-gɨ́mbē,* Naples, Università degli studi di Napoli "L'Orientale", 2016, 333 p.

par Gwenaëlle Fabre
Laboratoire ligérien de linguistique UMR7270
Université d'Orléans

La publication d'un dictionnaire d'une langue peu connue d'Afrique subsaharienne est toujours le résultat d'un travail long et minutieux, et ce d'autant plus quand il s'agit d'un dictionnaire encyclopédique, qui rassemble des informations riches et variées sur la structuration d'une société, sur ses activités et ses savoirs appréhendés à travers le prisme de sa langue.

Le *Dictionnaire encyclopédique koma-gɨ́mbē/français (monts Alantika, Nord-Cameroun)* est le fruit d'une coopération de terrain (une douzaine de séjours répartis sur six années) entre chercheurs de spécialités différentes qui se complètent dans l'analyse de la communauté koma : le linguiste Michel Dieu et l'anthropologue Louis Perrois. Le décès soudain de Michel Dieu en mai 1992 a mis en suspens l'édition de ce dictionnaire et Henry Tourneux a été chargé d'en revoir la partie linguistique quelques années plus tard, puis de rédiger la courte biographie de Michel Dieu présente en début d'opus. Cet ouvrage collectif est donc aussi un hommage à Michel Dieu et à la collaboration entre chercheurs de disciplines différentes, qui a permis de mieux connaître cette communauté montagnarde du Nord-Cameroun. Une profonde estime pour le travail de chacun et pour la communauté étudiée transparaît tout au long de l'ouvrage, à ce jour l'unique document scientifique paru sur la langue des Koma-Gɨ́mbē.

Cette langue appartient au groupe 4 de la branche Adamawa de la famille Niger-Congo selon la classification de Greenberg reprise dans Boyd (1989). S'appuyant sur une communication personnelle avec M. Dieu en 1984, Boyd (1989 : 183) indique « the Gɨmbe speaking Gɨmnɨme (1500 [speakers]) » et souligne la proximité du *gimnɨme* (ISO 639-3 : gmn, glottocode : gimn1238) avec les langues véré au sein de ce groupe 4. Kleinewillinghöfer (2015) et la classification dans la base Glottolog vont dans le même sens, plaçant cette langue dans un ensemble véré-gimmé.

Ce livre de 334 pages s'ouvre avec une introduction de seize pages qui retrace la genèse d'une collaboration et son objectif : une « enquête […] consacrée en parallèle aux cultures matérielles et à la

langue des Koma-Gɨmbē […] associant sur le terrain même, les approches ethnographique, linguistique et écologique » (p. 15), inscrite dans une démarche patrimoniale et documentaire plus vaste sur le Cameroun, réunissant des chercheurs de plusieurs disciplines.

La deuxième section recontextualise les premières mentions de la région et des habitants des monts Alantika, des « explorateurs » (Heinrich Barth en 1851, Eduard Robert Flegel en 1882, Siegfried Passarge en 1894, Leo Frobenius en 1913), puis des militaires (Max Moisel en 1913, capitaine Monteil en 1922, lieutenant Marchesseau en 1925, capitaine Coste en 1926) et des administrateurs coloniaux qui entreprennent une exploration plus systématique du pays koma vers 1925 (dont Henri Relly qui retrace en 1953 l'histoire du canton de Wangay dans son rapport de tournée et corrige quelques imprécisions des écrits antérieurs). Se référant aux travaux d'Eldridge Mohammadou (1979) sur le peuplement de la Haute-Bénoué, cette section se conclut par une discussion sur l'origine du terme « koma », sur les relations entre les différentes communautés de la région installées de part et d'autre de la frontière entre le Nigéria et le Cameroun (Koma, Tchamba-Leko, Véré, Bata) et sur la structuration interne de l'ensemble koma (Gimmé, Gɨmbē, Gewnu, Riitibé, Vomniyabé, Maribé et Véré) qui offre une diversité linguistique certaine.

La troisième et dernière section de l'introduction porte sur la géographie de la région. Elle mentionne aussi les travaux plus récents d'Edmond Dounias et manifeste une profonde connaissance du terrain et du terroir. Bénéficiant de la carte du début d'ouvrage, le lecteur découvre ici les extraordinaires paysages, les ressources, la faune et la flore, qui constituent l'environnement des Koma et tiennent une place importante dans leurs pratiques.

Dans la partie suivante en cinq pages sont rassemblées les notes ethnographiques concernant les Koma-Gɨmbē avec un exposé concis de la place que tiennent dans leur communauté les activités guerrières et agricoles, la sexualité et le mariage. La relation entre topographie et structuration sociale est soulignée : « les villages vont toujours de pair, l'un ancien en altitude où vivent les vieillards mais aussi ceux qui sont attachés à la tradition, avec femmes et enfants, et l'autre en piémont, dont les habitants sont plus ouverts au monde […] » (p. 31).

La bibliographie collige, notamment, les sources anciennes mentionnées ci-dessus. Viennent ensuite les notes linguistiques de M. Dieu, reprises par H. Tourneux, qui sont une esquisse phonologique (trois pages). Le système vocalique s'organise en neuf voyelles orales brèves et autant de voyelles orales longues, trois voyelles nasales brèves mais cinq voyelles nasales longues. Le système

consonantique organise les vingt et une consonnes en cinq ordres et huit séries. La présentation de la syllabe suggère que l'opposition voyelle longue/voyelle brève est neutralisée dans les syllabes fermées. Sur le plan tonal, les auteurs mentionnent trois tons ponctuels et deux tons modulés. À cela s'ajoute une liste des abréviations.

Si les circonstances particulières expliquent la brièveté de la partie linguistique, il est évident, à la lecture de l'ouvrage, que les connaissances des auteurs vont bien au-delà de cette esquisse. Le système de classes nominales productif (ce n'est pas le cas dans certaines langues Adamawa géographiquement et génétiquement proches, comme le tchamba leko) est évoqué, la typologie des verbes comprenant des verbes transitifs passifs aurait besoin d'être explicitée, de même que la forme du déterminé dans les nombreuses compositions nominales apparaissant dans le dictionnaire.

Le dictionnaire répertorie 2 500 entrées clairement présentées. Les vedettes, en gras, sont suivies de la catégorie grammaticale et d'une traduction en français auxquelles s'ajoutent, le cas échéant, des exemples d'emploi en composition, des commentaires linguistiques ou ethnologiques et l'identification des espèces zoologiques ou botaniques. Près de 378 visuels (dessins et photos indexés en fin d'ouvrage) illustrent les entrées et donnent au lecteur la représentation de réalités qui ne lui sont peut-être pas connues.

À titre d'exemple, la vedette *lálē/láʔē* (n. classe L/N) 'nom, mot' décrit dans le détail la cérémonie de dation du nom au nouveau-né, l'entrée *wārāgā/wārāgɨ̄bē* (n. classe Y/P) qui désigne une fête agraire est l'occasion d'évoquer la large place de la bière et les instruments de musique joués lors de cette festivité. Ou encore, le nom *kúnsā/kúunsɨ̄bē* (classe Y/P) 'fourreau de pipe, pipe' est suivi de sept termes composés relatifs aux parties de l'objet et aux différentes sortes de pipes fumées tant par les hommes que par les femmes (pipes masculines ou féminines, pipes réservées aux parents de jumeaux, pipe à usage thérapeutique comme remède contre la stérilité). Chacune est décrite en détail (matériaux et motifs) et six dessins illustrent cette entrée.

Le dictionnaire se termine par un index français/koma-gɨ́mbē qui permet un autre mode de navigation dans l'ouvrage (ex. cinq termes associés à 'classe d'âge', trois à 'maintenant').

Le volume précis et précieux permet ainsi d'apprécier la connaissance qu'ont les Koma-Gɨ́mbē de leur environnement, la variété des techniques qu'ils mettent en œuvre, la complexité de leur structuration sociale et des rites et coutumes associés.

On ne peut que saluer cet ouvrage, pour la densité et la précision des informations réunies en seulement six années de collecte. Et ce,

sur un terrain qui n'a jamais été facile d'accès, comme le souligne l'introduction. Vu la dramatique situation actuelle de la zone et les tristes circonstances dans lesquelles l'ouvrage a été terminé, on mesure combien il sera difficile de prolonger les travaux des auteurs ainsi que l'importance patrimoniale de ce volume pour la communauté scientifique, presque autant que pour les Koma eux-mêmes.

Références

Boyd, Raymond. 1989. Adamawa-Ubangui. In John Bendor-Samuel & Rhonda L. Hartell (eds.), *The Niger-Congo languages. A classification and description of Africa's largest language family*. Lanham, TX: University Press of America/Summer Institute of Linguistics.

Kleinewillinghöfer, Ulrich. 2015. Samba-Duru group. *Adamawa Language Projects*. Online: https://www.blogs.uni-mainz.de/fb07-adamawa/adamawa-languages/samba-duru-group.

Mohammadou, Eldridge. 1979. *Le peuplement de la Haute-Bénoué*. Garoua: CNRS-Onarest, ms.

Anthony Traill (ed. by Hirosi Nakagawa & Andy Chebanne), *A trilingual !Xóõ dictionary: !Xóõ] — English — Setswana*, Cologne, Rüdiger Köppe (Quellen zur Khoisan-Forschung / Research in Khoisan Studies 37), 2018, 318 p.

by Lee J. Pratchett
Goethe-Universität Frankfurt am Main / Humboldt-Universität zu Berlin

At a summer school for language description and documentation at the Leiden University in 2010, I placed a bid in a silent auction for Anthony Traill's (1994) *A !Xóõ dictionary* only to be outbid by the field phonetics instructor. Taking pity on the then still post-graduate about to embark on an adventure in "Khoisan" languages, the instructor withdrew his bid and I became the happy owner of a fine lexicographic work on one of the world's phonologically most complex languages. Well-thumbed and well-travelled, having made more than one journey to the Kalahari desert, it now stands alongside a yet more impressive work: a lexicon that is phonologically even more complex, with entries that are even richer in detail, and with a Setswana-!Xóõ section and an appendix full of flora and fauna terms, ideophones, and even a handful of profanities to top everything off.

Sadly, Anthony Traill passed away in 2007 before this dictionary project could be finished. In a preface to the dictionary, his wife, Jill Traill, reminds us that "of all [Tony's] unfinished work it was his dying wish to get this dictionary completed and published". It is particularly commendable that the editors have been able to see this wish through to completion, which now forms part of an important scientific legacy left to the !Xóõ and the linguistic community (see Güldemann and Nakagawa (2018) for an account of Anthony Traill's pioneering work in Khoisan linguistics and African linguistics). As far as possible, I will try to highlight the differences between the present dictionary and the original so that this review might serve those considering updating their Khoisan collection.

Just shy of thirty pages, the first section of the dictionary is tantamount to a grammatical sketch of the Lone Tree variety of East !Xóõ (!Xoon in other orthographies), of the Taa language complex. With approximately 2600 speakers in Botswana and Namibia, Taa is the last vital language of the Tuu language family. The sketch provides the necessary information for navigating the dictionary and teasing apart its many culturally and linguistically rich examples. Anyone coming across !Xóõ for the first time is well advised to consult this

section, if only to understand the particularities of the phonologically-based orthographic representation and ordering of entries. The subsection on general grammatical notes (p. 10-29) covers everything from pronouns and basic aspects of inflectional and derivational nominal morphology with a brief overview of the gender system, to verbal morphology, and introductory remarks on clauses and word order. This section is by and large identical in form to the original dictionary and the topics follow a functionalist description (e.g. what are referred to as adjectives in !Xóõ in the dictionary are in fact verbs). Interested readers are encouraged to consult complementary descriptions (see e.g. Güldemann 2013a; 2013b).

The subsection on general grammatical notes is followed by further phonetic definitions of the consonant and vowel contrasts and some phonetic rules (p. 30-36). It is worthwhile noting that the new dictionary maintains the phonologically-based linguistic orthography of the original. In a commentary found in both works (Traill 1994: 14 and p. 7 of the book under review), Traill himself recognises the arguments in favour of re-transcribing the dictionary with a user-friendly orthography (such as the one employed by the Ju|'hoan dictionary, cf. Dickens (1994)). A re-transcription entails not only considerable input from specialists, but consultation and consensus at community level. This aside, the language is still not fully described, effectively exemplified by the inclusion in the revised dictionary of five additional phonological contrasts discovered since the lexicon has been expanded. At this stage, the phonological representation remains the most faithful representation of what is phonologically the most complex language on earth.

The lexicon itself is divided into !Xóõ-English (p. 37-232), English-!Xóõ (p. 233-268) and Setswana-!Xóõ (p. 269-313). There is also a final appendix (p. 314-318), which brings together lexica for different semantic categories such as birds, which in the original dictionary were listed together unsuspectingly under the entry for 'bird' in the English-!Xóõ or distributed across the entire !Xóõ-English section. The appendix also brings together plant names, ideophones, smells, and curses in a user-friendly way. In the preface to the dictionary, Jill Traill reflects on her husband's "insatiable curiosity and an abiding passion for the !Xóõ people, their language and their culture". The !Xóõ lexicon is testimony to this. Let us take one example:

> **ǂhūn** 3 I **ǂhūm-tê** 2 I dim.: **gǂhùu-bâ** (sg.), **gǂhùu-Ɵ'ani** (pl.) 1. Shepherd's Tree *(Boscia albitrunca (Burch.) Gilg. Et Benedict var. Albitrunca)*. The fruit is eaten; the root has general medicinal value; the bark adds toxicity to arrow poison. *It is not used for*

> *firewood because its natural affinity is opposed to rain and burning it is believed to drive away the rain and to cause misfortune. This affinity may be harnessed deliberately to drive away unwanted incessant rain.* The diminutive is a metaphor for the Brown Hyena because its foliage resembles this hyena's shaggy coat. 2. Coffee. |**xāe ǂhūn** to play seesaw (cf. **ǂnōo īhì**). *Avoidance term **is !gà'n |èe**, lit. young girls, alluding to the soft foliage and lack of thorns (cf. **!náu ʘ'âni**).* (p. 160, my italics)

This entry is emblematic of the rich and informative descriptions that accompany many of the lexical entries. This entry shines some light on the morphophonological complexity of the language, particularly in the inflectional and derivational domains. *ǂhūn* is 'Shepherd's Tree', *ǂhūm-tê* is the same in the plural. The entry provides information on the agreement class pairing (3 I and 2 I) and thus the grammatical gender of the noun. The user is also provided with the singular and plural pairing for the diminutive forms, *gǂhùu-bâ* and *gǂhùu-ʘ'ani*, respectively. This is surely enough to entertain the specialist linguist for a while. The lexical item's description is nothing short of a journey into the world of the !Xóõ, and a bounty for cultural anthropologists, ethnobotanists, folklorists, and lexicologists alike. Above all else, in a fast changing world that in many respects scarcely resembles the one into which Anthony Traill stepped in 1969, it is a carefully curated record of an endangered way of life in the Kalahari made with and for the !Xóõ. I have italicised parts of the description which do not appear in the original dictionary. The entry itself in the original dictionary can be found under *ǂqhūn* with the agreement classes 3 I and 4 I (Traill 1994: 143), all illustrative of the attention to detail in the valuable additions and revisions to the original dictionary.

The Setswana-!Xóõ section, compiled by Botswanan linguist Andy Chebanne, is a welcomed addition that will hopefully make this lexicographic work accessible to a broader audience in Botswana, not least of all ethnic !Xóõ students and Tswana-speaking teachers. The Setswana-!Xóõ and English-!Xóõ sections follow the conventional alphabetical ordering for lexical entries and neither contain any of the rich details found in the !Xóõ-English section to which the reader is of course referred. The inclusion of the Setswana-!Xóõ section is a powerful symbol of the momentum in the documentation of Khoisan languages in southern Africa and Botswana in particular, and the strides made towards linguistic empowerment (see e.g. Batibo 2009), all of which was made possible by the significant achievements by Anthony Traill.

References

Batibo, Herman M. 2009. Language documentation as a strategy for the empowerment of the minority languages of Africa. In Masangu Matondo, Fiona McLaughlin & Eric Potsdam (eds.), *Selected proceedings of the 38th Annual Conference on African Linguistics: Linguistic Theory and African Language Documentation*, 193-203. Somerville, MA: Cascadilla Proceedings Project.

Dickens, Patrick. 1994. *English-Ju|'hoan Ju|'hoan-English dictionary*. (Quellen zur Khoisan-Forschung / Research in Khoisan Studies 8) Cologne: Rüdiger Köppe.

Güldemann, Tom. 2013a. Morphology: Taa (East !Xoon dialect). In Rainer Vossen (ed.), *The Khoesan languages* (Routledge Language Family series), 234-241. London: Routledge.

Güldemann, Tom. 2013b. Syntax: Taa (East !Xoon dialect). In Rainer Vossen (ed.), *The Khoesan languages* (Routledge Language Family series), 408-419. London: Routledge.

Güldemann, Tom & Hirosi Nakagawa. 2018. Anthony Traill and the holistic approach to Kalahari Basin sound design. In Tom Güldemann & Hirosi Nakagawa (eds.), *Kalahari Basin sound structure – in memory of Anthony T. Traill (1939-2007)* (Special Issue of *Africana Linguistica* 24), 1-29.

Traill, Anthony. 1994. *A !Xóõ dictionary* (Quellen zur Khoisan-Forschung / Research in Khoisan Studies 9). Cologne: Rüdiger Köppe.

Valentin VYDRIN, *Cours de grammaire bambara (ouvrage en réalité augmentée)*, Paris, Presses de l'Inalco, 2019, 597 p.

by Klaudia DOMBROWSKY-HAHN
University of Bayreuth

The *Cours de grammaire bambara* deals with the grammar of the Mande language Bambara (or Bamanankan), the most important national language spoken in Mali. The author characterizes the book as a compromise between a language textbook and a reference grammar. This review follows up on the question of whether this combination is advantageous.

The book is subdivided into 38 lessons, preceded by an introduction and followed by two appendices, which provide lists of glosses, grammatical morphemes and references, including printed and electronic resources. The book closes with an index of topics.

As expected from a textbook, the chapters are called "lessons", and each of them is designed for 90 to 120 minutes. They are arranged according to grammatical topics. In most cases, a lesson covers one topic (i.e. participles and converbs are the only topic of lesson 18, the infinitive is the only topic of lesson 31); exceptionally, several short topics are combined in one lesson (i.e. lesson 4 includes sections on the metrical foot, minor tone classes, identification clauses and personal pronouns). Every lesson discusses the grammatical topic it focuses on and illustrates it with examples taken from naturally produced texts. To justify the label "augmented reality", an application downloadable onto a smartphone provides the opportunity to listen to the recordings of the example sentences. The lessons end with suggestions for further reading and up to five exercises.

The first lesson introduces the geographical distribution and the sociolinguistic status of Bambara in Mali and discusses its external and internal classification. The grammar lessons set out very traditionally with a discussion of phonology and tonology. Other lessons are based either on a form or on a function. As an example, lesson 33 is dedicated to the subjunctive and the manifold constructions in which it is used, whereas lesson 34 focuses on the different ways to express conditional and temporal clauses.

The *Cours de grammaire bambara* is the first thorough grammar of the language since Dumestre's (2003) *Grammaire fondamentale du bambara*. It deals with all important grammatical topics of the language, taking into account the results of recent research not only on Bambara but on the cluster of Manding varieties as a whole. To mention but a few of them: the new interpretation of the tonal system

of Bambara based on the idea that next to syllables that carry a tone of their own there are toneless syllables; the introduction of the unit of foot; the categorization of the quotative *kó* among others as copula; the purely tonal distinction between a "visual" and a "non-visual" progressive, recalling different categories of evidentiality known in other languages.[1] The explanations are kept short and suppose some previous knowledge of the language, but the additional titles for further reading at the end of each lesson invite the reader to delve deeper into the respective topic. With respect to some subjects a parallel reading of Dumestre's (2003) and Vydrin's books may be beneficial.

The author follows the widely acknowledged scientific practice in language description by using almost exclusively attested examples from texts to illustrate the presented forms and functions. When compared to previous grammars, the presentation of example sentences is innovative in several respects. First, the innovative character relates to the type of transcription and translation including four levels. The first level presents the "surface realization" taking into account the tonal melody after application of tone rules, the vowel assimilation processes and some particularities of pronunciation, such as the production of certain instances of <s> as postalveolar fricative [ʃ], written as *sh*. The second level corresponds to the orthographic transcription plus the notation of lexical tone. It includes, among others, the tonal article in form of a floating Low and, in most examples, a separation of words into morphemes. The third level bears the glosses, the fourth is a free translation. The second innovation consists in providing audio files of all illustrating examples. These audio files were recorded with the help of several Bambara speakers in Bamako and in Paris, and are accessible by means of a smartphone application. Every sentence is repeated thrice by the same speaker. It is worth mentioning that the separation of the first two levels of transcription, and, especially, the marking of the Low tone article on the second level throughout all sentences makes the outstanding value of the book.

Students of linguistics may feel irritated by the author's idiosyncratic use of some technical terms. Deviating from the habitual usage, the author subsumes such grammatical morphemes as pronouns, postpositions, some determiners and some discourse particles under the term of "auxiliaries" and "semi-auxiliaries" (p. 554-560), next to predicate markers, which are the elements that convey tense, aspect,

1. The sentences meant to illustrate the "non-visual progressive" (p. 93-94) and the corresponding recordings are not considered to be grammatical by all speakers, though (S. Doumbia from Bamako, p.c.).

mood and polarity values of a verb. The designation "reflexive" is used not only for participants who are simultaneously initiator and endpoint of an event, but more widely for pronouns coreferential with the subject (p. 189 ff.). The terminology chosen for the description of Bambara tonology is likely to trigger misleading associations. In chapter 3, the author discusses the distinction between "dominant" and "recessive" syllables: only dominant syllables bear their own tone, while recessive syllables depend on the tone of the dominant syllable with which they constitute the tonal domain. The terminology of dominant and recessive syllables suggests a metaphorical reference to the meaning of these terms in genetics. In genetics, a recessive allele is that one of a pair of alleles that is masked by the activity of the second, when both are present on the same gene. But following the author's explanations there are not two tones on a recessive syllable, rather, it is toneless and its realization depends on the tone of a neighboring dominant syllable. Since nothing else is covered, as the author admits in another publication, they could have simply been called "no tone syllables" (Vydrin 2016: 85).

There seem to be some inconsistencies with respect to the description of constructions expressing resultative aspect. A state resulting from a preceding action, and sometimes even pure states which do not require a preceding action to occur,[2] are encoded by means of the resultative participle *V-len (-nen)* and the identifier *dòn* (neg. *tɛ́*) or the copula *bɛ́* (neg. *tɛ́*) otherwise found in locative and existential clauses, as in (1).

(1) Nê dén-` sà-len bɛ́
1SG.EMPH child-ART die-PTCP.RES COP[3]
'My child is dead [I am mourning it].' (8-28), p. 100

Following Idiatov (2000: 34), this construction is discussed as one of the aspects and moods, and its exponents figure in the list of aspectual and modal markers in lesson 8, entitled "verbal intransitive and transitive clauses, verbal conjugation". In contrast, in lesson 18, similar constructions are considered to be instances of a secondary predicate.[4] Although the participle does indeed function as a

2. For instance in sentence 8-29 cited by the author on p. 100: *jíriba jɔ̀-len bɛ dúkɛnɛ ` ná* 'There is a big tree in the yard', where the verb *jɔ̀* 'stand' does not imply an action.

3. Abbreviations: ART tonal article, COP copula in affirmative clause with nonverbal predicate; COP.NEG copula in negative clause with nonverbal predicate, PTCP.RES resultative participle; 1SG.EMPH first person singular emphatic pronoun.

4. The latter interpretation entails other misleading conclusions, e.g. in lesson 32 on the infinitive, where a similar construction is interpreted as a "semi-verbal or nonverbal predicate".

secondary predicate in other contexts, particularly in verbal clauses, this is not the case in sentences such as (2).

(2) Ní áw bɛ̀n-nen tɛ́ [...]
if 2PL agree-PTCP.RES COP.NEG
'If you do not agree [the family will be divided].' (18-11), p. 220

One of the major criteria of secondary predicates is their optionality. Thus, according to Schultze-Berndt & Himmelmann (2004: 65), when secondary predicates are omitted, the remaining string does not become ungrammatical and the relationship between the constituents is left unchanged. However, if the participle supposed to function as a secondary predicate is omitted in the mentioned example, the relationship between the remaining constituents and, as a result, the sense of the sentence (3) changes considerably. (3) is an identification clause, with its meaning changing to 'if you weren't; without you (lit.: if it is not you)'.

(3) Ní áw tɛ́
if 2PL COP.NEG
'If you weren't [without you] [the family would/will be divided].'
(18-11), p. 220

The *Cours de grammaire bambara* bears the following characteristics of a language learning textbook: it is divided into lessons, most of which span about 10 to 20 pages; it includes explanations of the topics discussed, and it ends with exercises. The amount of knowledge that is to be conveyed in 90 to 120 minutes is well measured. The grammatical phenomena are explained, albeit sparsely.

The assignments at the end of each lesson are adapted to the respective grammatical topic. Thus, the lists of minimal pairs at the end of lesson 2 on phonology are likely to be pronunciation exercises, even if there is no description of the task. In lessons 3 to 5 focusing on tone and in all the following lessons, the students are asked to note the tonal realization of sentences for which only the lexical tone is provided, by applying the tonal rules introduced in lesson 3. The bulk of the tasks are translations in both directions, but there are also more specific tasks such as the transformation of transitive into antipassive sentences (lesson 20), or searching for the meaning and the analysis of some particular constructions in the *Corpus Bambara de Référence* (lessons 22 through 24).

The usual efforts to ensure scientific correctness in language description (i.e. the exclusive use of attested examples) has prevented the illustrative examples from being arranged according to a difficulty level starting with the easier ones, and to avoid the occurrence

of difficult constructions before they are theoretically introduced. Thus, chapter 7 on locative clauses and simple postpositions mentions the use of the postposition *kɔ́* 'after' as final conjunction in temporal participial clauses and illustrates it with a complex sentence, before even simple clauses with a verbal predicate are introduced.

Furthermore, there are no vocabulary lists. This may be due to the close link of the textbook to the database *Corpus Bambara de Référence*, of which the author is the initiator and the person responsible (Vydrin *et al.* 2011-2020). Learners are apparently required to use the database dictionary to find unknown lexemes.

This book is strongly recommended for Bambara learners, albeit not for beginners. Not only the title but some of its features too suggest that it is intended for language learners: its organization into lessons including grammatical topics and exercises, the four-level representation of example sentences separating the levels of surface realization and the more abstract morpho-lexical level, and the accompanying audio recordings, which allow a sentence to be perceived in the visual and the auditive sensory modalities. The latter are valuable for the auditory perception of such difficult phenomena as the tonal and intonational realization of a sentence. However, since the exposition of the topics includes a lot of technical terms, learners' progression is taken into account only in the very beginning, and vocabulary lists are not provided, it does not seem suitable for beginners, at least not as a self-learning tool. Obviously, the intended audience of this book are rather (L1 or L2) speakers of the language: students who already have some skills in the language and want to learn more about the grammar of the language, such as for instance (future) teachers of Bambara, or students of linguistics interested in the language structure. In the sense that every person is engaged in lifelong language learning, this book is intended for learners on a high level.

As a reference grammar, the course is a very welcome addition, an extension and an update of Dumestre's (2003). One can safely state that this book is setting the standard for the grammatical analysis of Bambara. It would be definitely desirable to make it accessible in Mali, too.

References

Dumestre, Gérard. 2003. *Grammaire fondamentale du bambara.* Paris: Karthala.

Idiatov, Dmitry. 2000. Le sémantisme des marqueurs aspecto-temporels du bambara : une tentative d'analyse. *Mandenkan* 36. 1-59.

Schultze-Berndt, Eva & Nikolaus P. Himmelmann. 2004. Depictive secondary predicates in crosslinguistic perspective. *Linguistic Typology* 8(1). 59-131.

Vydrin, Valentin. 2016. Tonal inflection in Mande languages: The cases of Bamana and Dan-Gwɛɛtaa. In Enrique L. Palancar & Jean-Léo Léonard (eds.), *Tone and inflection: New facts and new perspectives*, 83-106. Berlin, Boston: De Gruyter Mouton.

Vydrin, Valentin, Kirill Maslinsky, Jean-Jacques Méric & Andrij Rovenchak (eds.). 2011-2020. *Corpus bambara de référence*. http://cormand.huma-num.fr.