Chapter 2

Sorting out Proto-Bantu *j

Jeffrey Wills
Ukrainian Catholic University

The most problematic of the consonants that Meeussen reconstructed for Proto-Bantu (PB) phonology is *j, for which Guthrie used both *j and *y. Earlier generations had also sometimes omitted either in favour of vowel-initial roots. Recent progress in establishing a solid family tree of the Bantu languages allows the evidence to be re-evaluated based on phylogenetic significance, especially with the help of more data from the North-Western Bantu branches. It has long been recognised that Meeussen's *j has various outcomes throughout the Bantu area based on phonological or morphological environments. The primary method of this chapter is to sort out the evidence for PB *j into different phonological and morphological environments, and then consider possible scenarios for reconstruction of those categories. In most roots with initial *j, there is no support for a PB stop and an initial vowel or glide should be reconstructed. That includes common verbs like *(y)àd 'spread' and *(y)ɔ́m 'be dry', and nouns like *ícò 'eye' or *ɔ́bà 'sun'. Most modern reflexes in /z/ or /j/ are the result of developments at morpheme boundaries after the PB stage. Both *ny and *nj/nz are reconstructed as distinct phonemes.

1 Introduction

In his *Bantu Grammatical Reconstructions*, Meeussen (1967: 83) put forth the following Proto-Bantu (PB) reconstructions for simple consonants (with a parallel series of pre-nasalised versions of each stop):

m	n	ɲ	
b	d	j	g
p	t	c	k

Jeffrey Wills. 2022. Sorting out Proto-Bantu *j. In Koen Bostoen, Gilles-Maurice de Schryver, Rozenn Guérois & Sara Pacchiarotti (eds.), *On reconstructing Proto-Bantu grammar*, 59–101. Berlin: Language Science Press. DOI: 10.5281/zenodo.7575817

The most problematic of the consonants was **j*, which had been in flux for a century, and Meeussen noted that one might just as well use the notation "/z/ or /y/ instead of /j/". A generation later, Schadeberg (2003: 146–147) described the continuing uncertainty: "Guthrie (1967–71) distinguishes initial **j* from **y*, but BLR2 (Coupez et al. 1998) recognises only **j* to the exclusion of vowel-initial stems. I regard the two as allophonic but the question needs re-evaluation." Increasingly, there have been doubts about **j* in some lexemes and an inclination to return to at least some vowel-initial stems. This chapter goes further in that direction to argue for reconstructing vowel-initial roots more extensively in PB. After an introduction on the history of the scholarship and some methodological issues, the currently reconstructed **j* is systematically examined in the relevant phonological and morphological environments.

1.1 History of the problem

Why did early Bantu scholars reconstruct **j* in the first place? The topic was mostly handled in handbooks like Meinhof (1899; 1910) or Homburger (1913), or in discussions of individual languages and words. The stems which are today reconstructed with **j* in BLR3 (Bastin et al. 2002) were variously listed by Meinhof et al. (1932: 187–196) with three symbols: **ɣ*, **y̲*, and **ø*.[1] For example, **ɣala* 'spread out', **mu-ɣaka* 'year', **ɣîno* 'tooth', **y̲anî* 'leaf', **y̲oɣû* 'elephant', and **ato* 'boat, canoe'. Meinhof's effort to identify which root had which phoneme was complicated by his significant reliance on South Bantu languages where **g* > *ø* is widespread. Homburger sorted out some of these problems but reconstructed PB forms with only initial palatals or velars without much explanation, although her lists of reflexes gave evidence for some vowel-initial roots.

To clarify this situation, in 1954, André Coupez wrote the first article ever focused on the question of PB **j* – a mere 3-page note with wordlist. His explicit goal was to correct Meinhof as well as Bourquin (1923). He based his analysis on Yao P21 and Kongo H16, as the only well-attested languages which have regular 'positive reflexes' for **g* and **j* in verb-initial and intervocalic positions. His choice of those languages was unfortunate for elucidating **j* because they often introduce hiatus-fillers in those positions. Coupez concluded that at an

[1] Approximate orthographic comparisons are: ɣ (Meinhof, Bourquin) ≈ g^w, g´ (Homburger) ≈ g (Greenberg, Guthrie, Meeussen, BLR); y̲ (Meinhof, Bourquin) ≈ g^w, g´ (Homburger) ≈ z (Greenberg) ≈ j, (j) (Meeussen) ≈ j, y (Guthrie) ≈ j (BLR); ṅg (Meinhof, Bourquin) ≈ ng' (Homburger) ≈ nj (Meeussen) ≈ nj, ny (Guthrie, BLR). But of course, these authors do not always reconstruct the same series in specific lexemes.

early stage *j had been lost at the beginning of many nouns,[2] and there simply were not regularly any verbal stems with initial vowels. He stated his support for Homburger's "conclusions" and gave a wordlist with *g and *j, without any vowel-initial PB nouns or verbs.[3]

At that same time, Malcolm Guthrie (1953) had revived a three-way distinction. In addition to *g and *j (e.g. *jàdà 'hunger'), he added *y* (e.g. *yúdù 'nose', *yímb 'sing'). But Meeussen & Tucker (1955: 170, 175–177) writing on Ganda (which has many *j*-initial verbs) rejected Guthrie's *y as not yet justified and affirmed the unitary *j of Coupez and Homburger, even adding initial *j to some of Bourquin's (1923) vowel-initial reconstructions for PB. In *Bantu Grammatical Reconstructions*, Meeussen (1967) generally followed Coupez with *j as the default (e.g. *jojo 'life', *jáka 'year', *júba 'sun'), but *ø was allowed to return at the beginning of some verbs (e.g. *ɨgad 'shut' and *ánik 'spread in the sun'). The parenthetical consonants in words like *(g)amb 'speak', *(j)ɟjib 'know', and *(b)óba 'fear' further signalled an openness to initial root vowels. This style was continued by Meeussen's (1969) *Bantu Lexical Reconstructions*, already with some changes in particular words, e.g. *(j)áka 'year', *jɟji-b 'know', *ɨcó 'your father'. Meeussen did not reconstruct PB semi-vowels but he noted their similarity to contexts with his parenthetical *(j).

Guthrie's large dataset (finally published in 1970) continued his approach from the 1950s. He could not confirm a unitary *j, so he used both PB *j and *y (often for the same lexeme) and thought it likely that "there was a mutation *Ɉ » *Y" and that "*G » *Y has to be postulated for most of the *g/*y pairs" (1967: 114). This allowed him to have consistent CV 'units' and roots with initial consonants.[4] Guthrie's idiosyncratic approach with multiple proto-forms made it a difficult path for others to follow – certainly for Meeussen (1973: 10) whose review of Guthrie included: "On the whole, it appears that there is no real ground for setting up *j and *y as two distinct correspondences."

BLR2, with a team led by Coupez, maintained his approach with *j everywhere (without parentheses). As BLR3 (Bastin et al. 2002) notes in the online

[2] Coupez (1954: 158): "Sans doute *j s'est-il amuï de bonne heure à l'initiale des thèmes nominaux: les thèmes nominaux qui nous sont attestés avec voyelle initiale seraient en réalité des thèmes en *j." It was also Coupez who introduced a rather vague sense of unspecified allophones (ibid. 157).

[3] Greenberg (1969: 430) followed this line, pointing out problems in Meinhof's correspondences: "Nor has Meinhof explained any of these deviations in the text of his work. It is now generally accepted that, as first suggested by Homburger 1913, there are two proto-phonemes involved, which are usually symbolized *g and *y."

[4] Guthrie (1967: 44, §42.11): "It is from these various unit features that the patterns are made up, and the principal ones involved prove to be C_1V_1, and C_2V_2 […]".

legend:[5] "Guthrie's **j* and **y* have been merged into **j*. The problems regarding **j/y/zero* are far from being resolved." Subsequently, two standards for reconstruction were in play (Guthrie and BLR), both without initial vowels except in functional morphemes. But scholars periodically pointed out the case for initial vowels in specific roots or specific groups.[6]

In recent years, an increase in knowledge about the North-Western languages has allowed major advances in our understanding of the Bantu family tree. This chapter has taken advantage of these developments to give greater weighting to data from zones A and B. The resulting analysis supports a substantial number of PB reconstructions with an initial vowel or glide rather than a unitary **j*.

1.2 Sources, method, and terminology

Reconstructing the phonology of a proto-language at a stage over 4000 years before any record of its descendant languages has significant challenges, and in the case of Bantu there are not even many intermediate reconstructions of late branches. Accordingly, recourse must be made to the primary lexical data in over 400 modern languages (many only partly documented), and then applying a judicious method of sorting out idiosyncrasies, proposing an inevitably simplified starting point, and elucidating the principal developments. One must admire the immense progress made by the early scholars of Bantu, who had developed a respectable grammar and 800-root lexicon of PB by the 1920s. But that was initially based on only a couple dozen languages (eventually becoming over 50), of which some were very closely related and most were from the eastern and southern regions.

Guthrie The first thorough Bantu lexical survey, including substantial attention to North-Western Bantu languages, was the monumental work of Malcolm Guthrie (1967-71). It remains the largest set of comparative data, listing reflexes

[5]The online version of BLR3 is available from: https://www.africamuseum.be/en/research/discover/human_sciences/culture_society/blr (database last updated on November 6, 2005). Note that BLR3 uses the symbols i and ɪ instead of the pair i̧ and i used by Guthrie and Meeussen. Likewise BLR has u and ʊ instead of u̧ and u.

[6]For example, Creissels (1999: 304): "Tswana data clearly supports the reconstruction of two different types of initials corresponding roughly to Guthrie's **y* and **j*", and he felt that the observed reflexes of one type supported "the hypothesis of the (relatively) ancient absence of any initial consonant." Bostoen (2019: 311–312): "If one admits the existence of vowel-initial noun stems in PB, it is enough to reconstruct just **j* and not **y*." More fully in Teil-Dautrey (2004: 161–192). See also Bulkens (2009: 29–34, written 1997), Bostoen (2009: 115), Bostoen & Bastin (2016: 14–15).

of over 2000 "comparative series" of "Common Bantu" roots and stems from hundreds of languages across all zones, including a systematic sampling of 29 "test languages". In Guthrie's system, each "comparative series" (C.S.) is represented by a form with a prefixed asterisk (the usual mark for an artificial construct or reconstruction, although Guthrie is explicit that they are not reconstructions).[7] In addition to the five test languages from Guthrie's North-Western zones ABC, there are 70 other languages in those zones which he cites ten or more times. However, Guthrie does not identify sources or informants, which is a problem for determination of speech variety or verification of specific forms since later published sources do not always confirm his data. Nevertheless, Guthrie is currently the only dataset with reflexes for a large number of lexical items in a large number of Bantu languages. Unless otherwise noted, examples below come from his data and are cited using his orthography.

Grollemund Dataset The other lexical dataset to which I will sometimes refer is that accompanying Grollemund et al. (2015), collected from published sources and fieldwork for 409 Bantu and 15 Bantoid languages.[8] The resulting dataset is notable for its geographical range and depth, including 150 languages in the North-Western zones ABC. Unfortunately, it is limited to up to 100 basic lexical items (meanings), only a few of which concern PB **j*.[9]

Bantu Lexical Reconstructions (BLR) The most complete set of lexical reconstructions is provided by the *Bantu Lexical Reconstructions* database at the Royal Museum for Central Africa and is based on a century of work by various scholars. This online database (current version: BLR3) is not a reconstruction of PB but rather a toolkit of reconstructions of lexemes of various Bantu language

[7] Guthrie used the word "reconstruction" occasionally in 1967 regarding Meinhof's work but avoided it with regard to his own PB X "stems" or "items". Guthrie takes pains to explain that his "starred forms are in no sense reconstructions of presumed ancestor items" (Guthrie 1965: 43). Rather, they are just "symbolic representations" of "sets of recurrent patterns" (Guthrie 1967: 19, §23.11, 21, §24.11), which become fodder for a process of analysing and attributing related comparative series to PB lects.

[8] This dataset is available from: http://www.evolution.reading.ac.uk/DataSets.html. It is an expanded version of Grollemund (2012), a study of about 200 North-Western Bantu and Bantoid languages using a modified version of the wordlist for *Atlas Linguistique du Gabon* (ALGAB).

[9] Another useful dataset is that collected for Bastin et al. (1999), which has 93 meanings from 335 languages, but the Grollemund Dataset often includes the earlier dataset and has fuller zone coverage (with A10, G60, P10, as well as Jarawan). The earlier Bastin et al. (1999) dataset is available from: https://www.africamuseum.be/nl/research/discover/human_sciences/culture_society/lexicostatistic-study-bantu-languages.

groupings and historical stages with varying reliability. For each reconstruction, it provides no reflexes or mentions of specific languages, only zones based on the sources in its bibliography.[10] For our purposes, BLR3 lists over a thousand forms with **j* from various time depths, so our focus will be on its 183 "Main" entries with **j* in C_1 position and 25 more with **j* in C_2 position. I have usually also provided zone information, because only about 2/3 of the "Main" reconstructions have descendants in zones AB, and some which do are not labelled as "Main". Throughout, I will be using BLR3 reconstructions (which uses only **j*), although the comparative data discussed often comes from Guthrie, whose C.S. use **j* and **y*.

Reconstruction based on parsimony and the Bantu phylogenetic tree Historical reconstruction is based on parsimony (or economy). We propose ancestral states requiring the fewest independent changes needed to derive later reflexes. For this process, we must have languages structured by a reliable family tree, i.e. a phylogeny. One of the major fruits of the half-century since Meeussen's *Bantu Grammatical Reconstructions* is the determination of a basic family tree for the Bantu language group.[11]

In Grollemund et al. (2015: Fig. 1 and 2), the evolution of the Bantu languages is graphed in a consensus time tree and a map of migration routes. Although more refinement needs to be done at lower levels, the progressive "backbone" of the tree and major branches is statistically very solid. Node 1 on that tree is the Bantu common ancestor treated here as PB, and then a series of binary splits (repeated 7 to 12 times) leads to a detailed structure with over 400 terminal nodes (the modern languages). In theory, each split is the result of innovations distinctive to one branch or the other, and it is the accumulation of these innovations which marks the divergence from the ancestral language. But the quantity and quality

[10] BLR2's system of *fiabilité* 'reliability' had some advantages, but the BLR2 version is no longer supported and the current BLR3 has useful grouping and numerous corrections of details so it was used for this chapter. The history and method of BLR is described by Bostoen & Bastin (2016).

[11] "[F]rom a purely classificatory point of view, the various trees published over the last 15 years or so by and large agree in their results" (Philippson & Grollemund 2019: 347). Ideally, a family tree classifies languages based on all linguistic changes, both lexical and non-lexical, which are assessed in various ways. Since Bantu is a fairly recent family with much internal contact, lexical and non-lexical innovations sometimes give conflicting isoglosses. The most recent non-lexical analysis (Nurse & Philippson 2003), based on thirty phonological and morphological features, proposes several historical scenarios but does not propose a tree. Accordingly, this chapter follows the most recent and detailed tree based on lexical innovation, being Grollemund et al. (2015).

of innovations vary between splits, and the Bantu phylogenetic tree is scaled for time not divergence, so the *depth* (number of levels a clade is from node 1) is merely a useful approximation of how close or far the clade is to the *root* (the proto-language).

The early stages of this phylogenetic tree can be visualised in Figure 1 with names of major branches and their relevant language zones.[12]

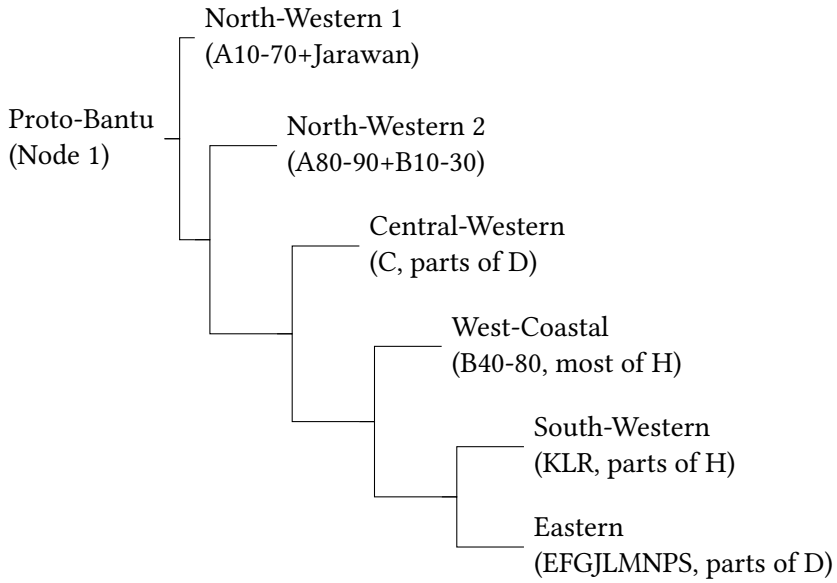

Figure 1: Simplified divisions of the Bantu phylogeny in Grollemund et al. (2015)

Our method is to work back from the modern languages, reconstructing ancestral forms for these major clades, and then proceeding to the nodes closer to the root. In general, we find that these major clades exhibit an internal unity in their reflexes of **j* that allows us to generalise at those levels, despite some inevitable innovations of a few languages among the dozens or hundreds in each branch.

Parsimony (the least number of changes) depends on the placement and distribution of the data in the structure of the tree. If an entire major branch has a distinctive form with **j* contrary to other branches, then it is possibly an innovation but possibly also a relic that escaped early changes in the other major

[12] From the detailed time tree in Grollemund et al. (2015: Fig. 1), I have collapsed three small neighbouring branches into North-Western 2 and three small neighbouring branches into South-Western.

branches, and so it must be studied seriously. But this is rarely the case in this chapter. Usually, the minority form (e.g. *zum* 'be dry' instead of the much more common *um*) is dominant in no branch, but rather is distributed across only a few languages in a few branches. So, it is likely to be a sign of innovations at later stages of Bantu development—for if the distinctive form belonged to PB, dozens and dozens of changes would be needed at multiple levels of the tree to account for the much greater number of languages lacking that distinctive form. We will often see that pattern of a few scattered innovations for **j*, indicating fairly recent developments. Of course, it is always theoretically possible that the scattered minority forms preserve an archaic heterogeneity, but then instead of a regional concentration we would require a concentration of the minority forms according to some original allophony or allomorphy.

When the evidence from both the North-Western branches is in agreement, it has great weight because these branches dominate the first splits in the tree. So, strong evidence from the North-Western branches and some currency in other major branches will make a good case for a PB reconstruction. On the other hand, the great majority of documented Bantu languages are in the Eastern branch, a clade which is several levels deep, and any reconstruction at that level must be reconciled with South-Western and West-Coastal (also called West-Western) before it can be given consideration for reconstruction at a higher level. In certain cases, a lexeme is not attested in all major branches, but any reconstruction at one or two levels below PB will be considered to be 'early' Bantu, i.e. early enough to be proposed as a candidate for PB but obviously not confirmable as such without support in some other way.

There are, however, two issues that must always be considered along with the phylogenetic approach: contact phenomena and directionality or naturalness of a sound change.

Contact phenomena across branches, which can create changes that are not independent innovations. This is particularly a concern in the North-Western regions of the Bantu domain where dozens of small languages belonging to different branches are geographically adjacent. So, although the North-Western clades have a privileged place in the phylogenetic tree, it is important to support reconstructions in those clades with at least some Bantu branches that are far enough away to discount an areal feature or borrowing of lexemes. Likewise, evidence beyond Narrow Bantu can support PB reconstructions, so relevant Bantoid data from Guthrie and the Grollemund Dataset will be cited.

Directionality or *naturalness* of a sound change, which could lead us to prefer one variant over another. In the case of PB **j*, the most common reflexes are null (ø) or glides (*y*, *w*), but sometimes stops, fricatives or affricates (*j*, *z*, *ʒ*, *dʒ*) are seen.

Weakening (lenition) is the common direction for consonantal sound change, but the strengthening of glides is so common across languages of the world that it has also been argued to be the result of articulatory pressures. In fact, in the systematically compiled AlloPhon database, the strengthening of glides to fricatives is more common than the contrary (12 processes to 8).[13] Furthermore, the particular strengthening of palatal glides is attested in a dozen language families beyond the database. For example, the initial glide in the Latin month *Ianuarius* becomes the fricative /ʒ/ in French *janvier* and the affricate /dʒ/ in Italian *gennaio* but disappears in Spanish *enero*.[14] Cross-linguistically, the most common environments for palatal glide strengthening are at a word or syllable onset, and before a high and/or front vowel—both environments where PB **j* is most common.

In short, there is some basis for preferring the reconstruction of glides to fricatives, but our default will be to follow parsimony and the usual Comparative Method without assuming a strong natural direction for change one way or the other.

Zones Guthrie's coding of languages by letter and number, based mainly on geographical zones, has remained standard for identification. But with an increasingly solid family tree, Bantu historical linguists can now group data based on phylogenetic significance, with an emphasis on historical branches rather than geography. Accordingly, the symbol "+" here indicates additional zones, that is to say, "ABDE+" is a shorthand for zones A, B, D, E and some further letter(s). This abbreviation is used partly to save space but also to reduce reliance on Guthrie's geographical zones as meaningful indicators of a PB ancestry. A lexeme solidly attested in zones AB and E (or any Eastern zone) already implies the first eight or more branchings and 2000 years of geographical spread. An item only attested in zones ABDG is just as likely to have been present in PB as one found in all 16 zones ABCDEFGHJKLMNPRS, although evidence from multiple zones may improve the quality of certain features of the reconstruction or demonstrate the stability of a word in the lexicon.

[13] Bybee & Easterday (2019) describe the data collection and provide examples. For Romance and Basque examples, see Hualde (2011: 2232). For more on Spanish palatal fortition, see Baker & Wiltshire (2003). Meeussen & Tucker (1955: 174–175) noted that the development of Ganda JE15 *ggyá* 'new' < **hya* < *PB *pɪa* "exactly parallels" the glide hardening in Old Norse *tveggja* < Proto-Germanic **twa-jē* 'of two'. Ganda also has *-jjwa* < **hwa*. In modern German, the initial [*j*] in words like *Jahr* 'year' surfaces as an obstruent in various regional varieties, e.g. [ʒ] in the Mecklenburg dialect and [*ɟ*] in a variety of Thuringian (Hall 2014: 257–262).

[14] Likewise, in medial position, Latin *maior* 'greater' > Italian *maggiore*, and Latin *ego* > Vulgar Latin **eo* > Spanish [jo, dʒo]. For initial Indo-European **y* > Greek ζ [z, dz], see Sihler (1995: 187–190).

1.3 Outline

The primary method of this chapter is to "sort out" the evidence for PB **j*: first, into different phonological and morphological environments, and then into possible scenarios for reconstruction. Proposals (from BLR, Guthrie) for PB **j* and **y* will be tested using lexical data (from Guthrie, Grollemund, etc.), organised by a phylogenetic tree (from Grollemund et al. 2015).

Procedurally, let us begin by accepting the main reconstructions written with the symbol **j* in BLR3, and then try to elucidate what values they might have had. PB **i* and **n* tend to condition the evolution of subsequent consonants, so three environments can be distinguished:

Group 1: **j* not preceded by **i* or **n*

- Initial **j* in noun stems, e.g. **játò* 'canoe'
- Initial **j* in verb stems, e.g. **ját* 'split'
- Medial **j* in noun or verb stems, e.g. **jòjì* 3 'belly'

Group 2: **j* preceded by **i*

- Initial **j* in class 5 nouns, e.g. **jàdà* 'rubbish-heap', **jícò* 'eye'
- Initial **ji-C* and **jij*, e.g. **jíjɪb* 'know', **jíjì* 6 'water'

Group 3: **j* preceded by **n*

- Nouns supporting PB **ny*, e.g. **jókà* 'snake'
- Nouns supporting PB **nj*, e.g. **jògù* 'elephant'
- Nouns with mixed classes, e.g. **jíkì* 9/10 'bee' & 14 'honey'

So, first to be considered is **j* in the most neutral environment, i.e. at the beginning or middle of roots without a major conditioning factor. Then an examination of the consequences of the two major conditioning factors: a preceding *i* or a preceding nasal. Most roots only occur with **j* in one of these environments, but it will be useful to see what can be learned from those roots with allomorphic variants.

We will go through these environments in order, but in a summary fashion. Our goal is not to be exhaustive but rather to examine a few samples of each category as case studies and consider the issues the category presents.

2 *j not preceded by *i or *n

2.1 Unconditioned initial *j in noun roots

Our first category is one of the easiest: **j* reconstructed at the beginning of nominal roots in classes where a CV- prefix does not generally provide an environment conditioning a change.[15]

For example, **játò* 14 'canoe' (BLR 3252) is an old and widespread root, attested in all of Guthrie's zones, most frequently with the 14/6 (and 14/4) gender. This stem was treated in detail by Bulkens (2009) who lists the previous reconstructions: **ǵato* (Homburger), **ato* (Meinhof et al.), **átò* (Greenberg), **yátò* (Guthrie), **(j)átò* (Meeussen). In Bulkens' collection of 160 reflexes of this stem, only four languages attest a consonant-initial nominal stem and she shows how they developed, mostly due to reanalysis.[16] Otherwise, the stem always begins with a vowel, e.g. Lundu A11 *ádɔ̀*, Holoholo D28 *àtó*, Tsonga S53 *àtsò*.

So, the obvious reconstruction at the PB node 1 (and even earlier) is a return to Meinhof's vowel-initial root **átò* without Guthrie's **y* or BLR's **j* or even Meeussen's **(j)*. Bulkens (2009: 58) concludes that the data disproves the hypothesis according to which nominal stems in PB invariably had an initial consonant.

For **jákà* 3/4 'year' (BLR 3169, all zones, C.S. 1904), Guthrie gives 33 descendant forms, mostly in the 3/4 gender. Again, the great majority have the class prefix (often with glide formation **mʊ-* > *mw-*) followed by a vowel-initial stem, e.g. Tiene B81 *muáka* (Ellington 1977: 175), Lengola D12 *mwáka* (Stappers 1971: 275), Unguja Swahili G42d *mwaka*. The exceptions are a couple of cases in zone S where the plural class 4 prefix has crept into the singular.

Perhaps most demonstrative is **jéné* 1/2 'self, same' (BLR 3296, all zones; C.S. 1970). Not only are there no reflexes in Guthrie with an initial stop, but also the widely occurring variant **méné* 1/2adj 'self' (BLR 2171 zones ABCK+) suggests that **mʊ̀-éné* became **méné* and was reanalysed as an independent stem at a very early stage, perhaps even by PB. This early development is much harder to imagine with a putative PB **mʊ̀-jéné*. A similar history of incorporation and reanalysis must be the story with the doublet **jòngó* 14 'brain' (BLR 3571, zones BCE+) and **bòngó* 14 'brain' (BLR 274, zones ABG+), in this case with the noun prefix of class 14.

[15] That is to say classes 1, 2, 3, 4, 7, 14, but not classes 5 or 8 (because of prefix with close front vowel) or 6 (because of class 5 influence), 9 or 10 (because of non-syllabic nasal prefix), or 11 (because of class 10 plural influence).

[16] Bulkens' exceptions are Kota B25 *yàzí* 7/14 (probably not this root), Masaba JE31 *háárò* 5/6, Bukusu JE31c *járò* 5/6, and Pende L11 *wátó* 5/6. Most are due to reclassification of the noun with reanalysis of the former class prefix as part of the new stem.

At the PB stage, in these three roots for 'canoe', 'year' and 'self', there is simply no good evidence in descendant languages that would persuade us to reconstruct an initial stop, spirant, or even glide. There are not too many of these unconditioned **j* nouns, but enough to matter, including several other basic ones, e.g. **ánà* 'child' (BLR 3203), **ápà* 'armpit' (BLR 3237), **ògà* 'mushroom' (BLR 3257), **ʊ́mà* 'thing; bead' (BLR 3619). Bourquin (1923) listed over a dozen vowel-initial noun roots from earlier scholars and then added a dozen more. Creissels (1999: 305) lists 11 of these nouns where "the languages of subgroup S.30 (and in particular Tswana) demand to accept the possibility of variants of these reconstructions with no initial consonant."

2.2 Unconditioned **j* in verb stems

We will next look at the important group of verb roots reconstructed with an initial **j*. These 84 verbs account for almost half of the main entries in BLR3 beginning with **j*, and many are widespread through the Bantu area.

2.2.1 Typical reflexes

Following are some of the better attested roots, each with more than twenty languages cited in Guthrie's (1967–71) comparative series. To simplify the analysis, for each outcome of **j*, I have sorted them into what I have called "weak" outcomes (with no consonant, or with a glide) or "strong" outcomes (with stop, fricative or affricate, especially *j*, *z*). In parentheses, I have put the number of entries in Guthrie with that outcome. Because the strong reflexes are rather rare, occurring only in certain languages, I have explicitly cited those exceptional languages by their Guthrie number (and used Guthrie's orthography).

(1) **jác-(am)* 'open mouth; yawn' [BLR 3145/6, C.S. 1889(a) **-yác-(am)*][17]
Weak: ø (23), y (4), w (1)
Strong: j (P21, P22)

(2) **jàd* 'spread' [BLR 3147, C.S. 1890 **-yàd-*]
Weak: ø (13), y (10), w (2)
Strong: z (B22b, B82, M63, R24)

(3) **jánɪk* 'spread out (to dry in sun)' [BLR 3206, C.S. 1924 **-yánɪk-*]
Weak: ø (20), y (6), ny (2) – Bantoid: Tiv ø
Strong: y/j (B11a), z (M63, R24), j (P21)

[17] BLR (following Guthrie) only lists zones CEF+ for this verb, but its presence in zones AB is seen in Proto-Manenguba A15 **sám* 'sneeze' (Hedinger 1987: 247) and Bulu A74a *semele* 'sneeze'.

(4) **ját* 'split; separate' and derivatives [BLR 3242, C.S. 1945-6 **-yát-*]
Weak: ø (17), y (6) – Bantoid: Tiv ø
Strong: z (B82)

(5) **jégam* 'lean against' [BLR 3291, C.S. 1967ab **-yégam-*]
Weak: ø (13), y (5) – Bantoid: Ekoid y
Strong: z (B82), j (P21)

(6) **jó(o)g* 'bathe; wash; swim' [BLR 3525, C.S. 2107 **-yó(o)g-*]
Weak: ø (17), y (5) – Bantoid: Tiv ø, Ekoid y
Strong: j (A74), y/j (B11a), j (P21)

(7) **jɔ́m* 'be dry' [BLR 3616, C.S. 2161 **-yɔ́m-*]
Weak: ø (18), y (4) – Bantoid: Tiv ø, Ekoid y
Strong: y/j (B11a), j (B22b), z (B82, M63), j (P21)

This data is derived just from Guthrie's collection and some subclades are more heavily represented than others, but it is a broad survey of Bantu languages and enough to establish a *prima facie* case that the "weak" outcomes are the general rule and "strong" outcomes are the exceptions. According to Guthrie's data, about 90% of the many modern languages exhibit weak reflexes of **j* in these roots, especially *ø* but also a fair amount of *y*, which are supported by Tiv and Ekoid cognates. In other words, among about 70 languages tested in the samples above, there are only a few that ever show a consonant /j/ or /z/ (that is, something stronger than a glide in these roots). From the phylogenetic viewpoint, it is not only the quantity that matters, but also the distribution. These exceptional languages do not form a block supporting a strong reflex preserved in an early branch; rather, they are isolated or in small subclades deep in the phylogenetic tree in Grollemund et al. (2015: Fig. S1). Likewise, an argument that these few strong forms preserve some archaic heterogeneity would need to be based on some original phonological or morphological distinctions (e.g. their concentration in a certain tense), but that is also not the case. Rather, these occasional dispersed drops of *j* or *z* in a Bantu ocean of *ø* and *y* are a typical pattern for independent innovations in a large dataset.

In addition to Guthrie, we now have the data from the Grollemund Dataset, listing 75 common lexemes in each of 400+ Bantu languages. The only verb relevant for us is PB **jímb* 'sing'. Analysing all its forms in all zones, one finds that about 140 languages have weak reflexes and 16 have strong reflexes. The strong reflexes mainly come from the few pockets already seen in Guthrie – B11 (3 examples) and N10-P20 (4 examples) – as well as A80 (4 examples) which was sparsely recorded by Guthrie. Although this is only one lexeme and also not a complete

picture (**jímb* is missing a cognate in 200+ languages), the Grollemund Dataset confirms the distributional pattern of Guthrie's data and implies innovations in a handful of recent groups.

So, for the proto-phoneme at the beginning of these verbs it is easiest to posit an original *ø* from which *y* (or *w*) occasionally arose to resolve a hiatus or various prefixes were reanalysed and incorporated into the onset.[18] Accordingly, our primary interest here in considering Guthrie's exceptional languages does not concern reconstruction, but rather an examination of case studies to see some of the phonetic or phonological paths of development which are possible from PB stem-initial vowels and/or **y*.

2.2.2 Exceptional languages

North-Western Bantu (zones A, B10-30) and Central-Western Bantu (C, parts of D) The North-Western Bantu languages usually show weak onsets in Guthrie, e.g. **jót* 'warm oneself': Duala A24 *ɔl*, Yambasa A62 *ɔt-ɔbɔ*; **jígu* 'hear': Lundu A11 *ọk*, Bakoko A43b *ọx*, Bulu A74a *wok'*. Only two of his many languages in these important branches regularly show several strong reflexes, i.e. Mpongwe B11a in the Myene group and Ngom B22b in the Kele group.

For each Mpongwe example, Guthrie gives two forms, one with *y* and one with *j*, e.g. *yẹmb* & *jẹmb* 'sing', *yom* & *jom* 'become dry'. In his treatment of the PB reflexes in Nkomi B11e (a related variety of Myene), Rekanga (1994: 157–159) explains the doublets: the usual reflex of **j* is *ø* but the reflex *dy* (realised [dʒ]) occurs after the nasal prefix in class 9 (see also Grégoire & Rekanga 1994). The infinitive (class 10b) creates this same effect and so is also reconstructed as having once a nasal prefix. In short, the basic verb stem is that seen in the imperative and other forms with *y*, as one would expect. But the effect of a nasal prefix to create an affricate [dʒ] is a topic I will return to in considering class 9 nouns. For Ngom B22b, the reflexes are uniformly *j* (e.g. *jaə* 'spread', *jẹmb* 'sing', *jọm* 'become dry'), but Shake B251 *yemp* 'sing' and other forms in closely related languages from the Grollemund Dataset suggest that only a small group was affected by this development. For **jígu* (North-Western **júg*) 'hear', there are over 20 forms from North-Western languages in the Grollemund Dataset, with clearly strong reflexes only in the A80 group. For **jímb* 'sing', there are 7 weak and 3 strong reflexes.[19]

[18] For example, the irregular Lumbu B44 *ɣum* and Punu B43 *kum* (< **jɔ́m* 'be dry') reflect the **kʊ* prefix of the cl. 15 infinitive.

[19] Weak: Kpe A22 *embà*, Yasa A33a *èhímbà*, Ewondo A72a *yia*, Bulu A74a *yia*, Fang A75 (Bitam and Minvoul) *əyiɛ*, Fang A75 (Medouneu) *əyee*. Strong: Eton A71 *jà*, Mkaa A15C *jém*, Elung A15C *jé*.

In the Central-Western branch, weak reflexes are the rule in the Grollemund Dataset: Babole C101 *emba*, Mboshi C25 *iyemba*, Bangi C32 *yémbá*, Soko C52 *hamba*, Mongo C61 *émba*, Bushong C83 *yéem*.

In sum, the great majority of the North-Western and Central-Western forms are weak, which supports the testimony of the other early branches for reconstructing a weak stem **ɪmb* or **yɪmb*. But the mixed evidence in North-Western sub-groups reminds us that there must have been a range of impacts from strengthening (and weakening),[20] nasal infinitive prefixes or subsequent front vowels, and analogy to verbal nouns, since some languages use phrases like 'make a song'. These processes are more clearly seen in other branches.

West-Coastal Bantu (zones B40-80, most of H) Confirming Guthrie, the extensive wordlist of twenty nearby languages (including Teke B70 and Kongo H16) compiled in Koni Muluwa & Bostoen (2015) typically shows initial *y*, *w*, or occasionally *ø* for these verbs.[21] The exception in Guthrie is Boma B82 which yields *z* or *j* at the beginning of these words: *zatɔ* 'split', *zi̱le* 'get dark' (< **jíd* 'get dark, black' BLR 6142), *zɔma* & *zu̱mi̱* 'become hard, dry', etc. But even Tiene B81, another language with Boma in the Kwa-Kasai North subgroup,[22] consistently has *y*, e.g. *yááta* 'split', *yíla* 'get dark', *yóma* 'become dry' (Ellington 1977: 175–176). So, Boma apparently has a language-specific development.

South-Western Bantu (zone R, parts of HKL) Weak reflexes of **j* are the rule. In Guthrie's data, the only exceptional language in this area is Ngandjera R24 which his inventory describes as "broadly similar" to Ndonga R22 and Kwanyama R21 but with a few distinctive changes including **j* > *z*. Guthrie's relevant data for Ngandjera was *zar* 'spread', *zanik* 'spread to dry', *zer* 'shine', *zon* 'spoil', etc. It is not clear what Guthrie's source was for Ngandjera and this variety of Wambo is not well attested, so for our purposes I will take the Wambo language R20 as

[20] In Eton A71, we see the possibility of lenition of fricatives: "the voiced alveolar fricative /z/ is realised by the voiced glottal fricative [ɦ] or simply not realised" (Van de Velde 2006: 28), although that does not affect the verb *jà* 'sing' which begins with an affricate.

[21] Nzadi B865 has variation in its reflexes of **j*: *o-yâŋ* 'spread to dry', *o-yûm* 'to dry', but *o-zwô* 'bathe (intr)', *o-zâŋ* 'to refuse' < **jáng* (zones CJRS), and nouns in *dz*. "There does not seem to be any regularity to this distribution, nor do the reflexes seem to line up consistently with any nearby languages" (Crane et al. 2011: 257). Since Bulu A74a also has an irregular onset in *jɔk* 'swim', one avenue to explore is whether some verbs were affected by the reflexive prefix *i-* ('to wash oneself'), which mutated *y* to *z/j*. For the nouns, *dz* in class 5 is merely a reflex of the prefixes (regularly Nzadi **di/dɪ-* > *dz-*).

[22] New groupings of West-Coastal Bantu can be found in Pacchiarotti et al. (2019).

a whole.[23] Unfortunately, I have not been able to find examples of *j > z in these verbs. Rather, Baucom (1975: 172) reconstructs Proto-Wambo *yoga 'swim, bathe' with y/w/ø reflexes of the initial *y in various daughter varieties, e.g. Ngandjera yoga ~ oga. Likewise, PB *jámu 'suck' yields Proto-Wambo *yama, with Ngandjera ama. Similarly, Ndonga and Kwanyama only have y as the reflex for initial *j in verbs.[24] If, indeed, z reflexes appear in some variety of Ngandjera, they must be a late local innovation.

Eastern Bantu (the broad area of Guthrie zones EFGJMNPS and part of D). Sorting through *all* of Guthrie's hundreds of entries from *all* of these languages, the only strong reflexes for these *j verbs are found in entries from two subgroups: Ruvuma and Botatwe.

Ruvuma group For these verbs in the closely related languages Yao and Mwera, Guthrie prints a double reflex: ɟ and zero.[25]

(8) Yao P21: *(ɟ)asam* 'open the mouth', *(ɟ)anik* 'spread to dry', *(ɟ)elajel* 'float', *(ɟ)egam* 'lean against', *(ɟ)oog* 'bathe, wash', etc.

(9) Mwera P22: *(ɟ)aam* 'open the mouth'

Ngunga (2000: 78–81) explains that in contemporary Yao there are two types of verbal roots: those with a "stable" [j], which is realised in all verb forms, and those with an "unstable" [j], which appears only in some verb forms. He concludes that the infinitive provides the underlying form and that the "unstable" [j] is an insertion in suffix-marked tenses. Ngunga's analysis is synchronic but it coincides with the obvious diachronic analysis: these *j verb roots historically had a vowel in root-initial position with a later hiatus-filler inserted after some tense markers,[26] whereas those verbs with stable [j] should have other origins.

[23] Maho (2007: 129): "The entire R20 grouping represents a single language, usually called Wambo or Oshiwambo. Kwanyama R21 plus all varieties coded R211 through R217 correspond to Baucom's (1975) northern dialect group, while the rest correspond to his southern group."

[24] Some examples from Ndonga (Fivaz 1986: 15, 99): *yala* 'spread' (*jàd), *yela* 'become bright clear' (*jéd), *yola* 'laugh' (*jòd), *yogá* 'swim' (*jóg). From Kwanyama (Turvey et al. 1977): *yala* 'spread (mat)' (*jàd), *yela* 'be, become bright' (*jéd), *yola* 'laugh, joke' (*jòd), *yota* 'warm hands at fire' (*jót 'warm oneself').

[25] Odden (2003: 529): "Yao and Mwera are very closely related, and might be treated as dialects." According to Guthrie's (1967–71: Vol. 2, 59) inventory for Yao: "*C_1: [...] *c, *j > s; *nc, *nj > s; *y > ɟ (in radicals)" and "*C_2: [...] as *C_1 [...] but *y > ɟ in stems". For Mwera: "Broadly similar to P.21, but *c, *j > ø."

[26] Odden (2003: 531): "Avoidance of hiatus is most strict in Yao (and Mwera), which have no V-V sequences within the word. Vowel fusion and glide formation are the rule within the word."

In short, Yao and Mwera do not provide relics of an early *j but rather support reconstructing a vowel in root-initial position for these verbs.[27] The larger lesson is that a palatal stop or even eventually an affricate can develop as one of the options for a hiatus-filler.

Botatwe group Guthrie has examples from two Botatwe languages:

(10) Ila M63: e.g. *zal* 'spread', *zambukil* 'spread', *zanik* 'spread to dry'

(11) Subiya K42: e.g. *zimb* 'sing', *lu-zimbo* 'song' 11/10

Bostoen (2009: 115) gives sample forms from most languages in this group:

(12)

			*jɔ́m 'to be dry'	*jímb 'to sing'[28]
a.	Western Botatwe			
	Shanjo	K36	*dʒûma*	*îmba*
	Fwe	K402	*ʒûma*	*ʒîmba*
	Totela	K411	*yuma*	*zimba*
	Subiya	K42	*zuma*	*zimba*
b.	Eastern Botatwe			
	Lenje	M61	*kú-yuma*	*kw-ímba*
	Ila	M63	*zuma*	*imba*
	Plat. Tonga	M64	*íkú-yuma*	*íkw-ímba*
	Soli	M62	*yuma*	*imba*

As Bostoen (2009: 115) notes, "[t]here is quite some variation in the realization of *j […]. For most lexical items, certain languages attest a fricative, while others have a zero reflex. The precise languages attesting zero (or glide) may differ, however, from one lexical item to the other." In short, whatever the source of the variation, the Botatwe data does not clearly lead to any internal reconstruction, even in subgroups.

[27] Almost all of the 39 stable-*j* verbs identified by Ngunga lack a clear origin, but many are verbs of noise or movement perhaps connected to ideophones. There are, however, two verbs with 'stable' [j] that are derived from PB roots in *j and require another explanation: *juman* 'quarrel' and *jiim* 'to not give' (which seems to have *j*-less variants and may be influenced by the common Bantu variant *nyím).

[28] Crane (2011: 78) gives *òkúyìmbà* '(to) sing' for the Zambian variety of Totela, while Bostoen (2009) mainly reports on the Namibian variety of the language.

2.2.3 Summary of initial *j in verb stems

Overall, the frequency and stability of "weak" forms is quite impressive, and a weak onset of these verbs is to be preferred for PB node 1. It is entirely possible that there are some PB verb roots which begin with *y and some with *ø, and considering the ease with which a glide can be inserted or deleted, further study will be needed to determine the best PB reconstruction for each root along with any allophones (including *w* before back vowels). Meanwhile, for these verbs I have adopted a convention of writing a parenthetical initial glide, thus: *(y)ác-(am)* 'open mouth' (BLR 3145/6), *(y)àd 'spread' (BLR 3147), *(y)ánɪk 'spread out (to dry in sun)' (BLR 3206), *(y)át 'split' (BLR 3242), *(y)égam 'lean against' (BLR 3291), *(y)ó(o)g 'bathe' (BLR 3525), *(y)ɛ́m 'be dry' (BLR 3616).

In addition to the specific subgroups with apparent strengthening (*y > z, j),[29] there are occasional exceptions scattered across other languages. Considering the several hundred forms cited by Guthrie for these verbs with initial *j, it is not surprising to encounter occasional variants or doubtful cases and I will not discuss them all here. Let it suffice to note a few examples of other languages with idiosyncratic forms for *jímb 'sing' in the Grollemund Dataset: Kaningi Nord B602 *o-lima*, Soko C52 *hamba*, Bira D32 *nyimbo*, Bembe H11 *kù-giùmbílà* (cf. Vili H12L *kw-imbilə*), Ha JD66 *uku-lilimba*. These are useful reminders that one can always expect exceptions in a large dataset, especially in a category when there are phonological opportunities like hiatus resolution and incorporation of various prefixes (especially nasal and infinitive prefixes) at morpheme boundaries.

A major difference between the vowel-initial nouns and verbs is the frequent presence of glides before the verb stems. Besides the possibility of original glides, one likely reason is the greater range of morphological variation in verbs. For nouns, even with glide formation in the prefix, there are usually only one or two forms, e.g. *bɔ̀-átò 14 'canoe', *mɔ̀-ánà / *bà-ána* 1/2 'child(ren)'. But verbs have a large variety of prefixes of various shapes (ø, CV, V, N) that can lead to allophones in the root-onset. For example, the 'unstable *y*' in some Ganda verbs is so-called because the palatal element appears only at the beginning of the word (in the imperative), after non-high vocalic prefixes (*e*, *a*, *o*) and after *n* (as *ɲj*), e.g. for the stem *(y)egeka* 'support': *oyégeka* (2SG PRS), *yegeka* (IMP), *njégeka* (1SG PRS), but *twégeka* (1PL PRS), *okwégeka* (INF) (Meeussen & Tucker 1955: 175–176,

[29] In fact, Guthrie understood the basic development of these exceptional subgroups (1967: 62–63): "The question of *y is difficult, since in many languages its reflex is zero, although in Boma B.82, Subiya K.42 and Ila M.63 *ya > za, while in Yao P.21 *ya > ja. […] It is just conceivable that *y* was the sound in the source-pattern, and if it were, *y > j > z* is a not impossible development, on the one side, and *y > zero* on the other."

also Hyman & Katamba 1999: 369–376). In short, the glide does not appear after the high vowels because the prefixes themselves undergo glide formation, just as seen in PB nouns like **bɔ̀-átò* > **bwato*. One can assume that in some languages, in order to preserve morpheme stability, the glide variant of the verb was generalised throughout (and sometimes even strengthened). This development is seen in Ganda in other verbs, where only 'stable *y*' (*y* or *nj*) is found, especially before high-vowel stems, e.g. *yíta* 'call', *yíga* 'learn', *yîimba* 'sing'. Considering the possibility of cycles of addition and loss of glides and the conditioning factor of preceding prefixes, further study will be needed about the possibility of PB glides in these roots.

2.3 Unconditioned medial **j*

The suspicious paucity of early stems with unconditioned **j* in C_2 position reinforces our doubts about the existence of PB **j* as a standard consonant. There are no solid verbs in this category, but there are three well-attested nouns:

(13) **kájá* 5/6 (11/10) 'leaf; tobacco leaf' (BLR 1736, C.S. 1019 **káyá*, ABCD+)

(14) **jòjì* 3 'belly, abdomen' (BLR 3589, C.S. 2142, ABC) as well as **jòjò* 3 'life; spirit; heart' (BLR 3590, C.S. 2143-44, EF+), **jòjà* 3 'life' (BLR 3588, ps 550, HS)[30]

(15) **jòjá* 'fur, feather, bodyhair' 14, 3, 11/10 (ACEF+) combining **jòjá* 'fur' 14 (BLR 3587, C.S. 2141, FJL+), **jòjá* 5/6, 11/10 'feather' (BLR 3586, C.S. 2140, EGJ), **jòcá* 3 'feather; bodyhair' (BLR 7034, CJ).[31]

Almost all Guthrie's citations for these roots show *y* or ø in the C_2 position, with a few zone A languages only having one syllable. A few other BLR3 noun reconstructions are marked 'main' but without North-Western cognates, e.g. **bɛ̀jɔ́* 3/4 'baobab' (zones CGM+), *kɔ́jɔ̀* 3/4, 7/8 'fig-tree' (DE+), **jàjò* 11/10 'sole of foot' (DE+), **kɑ́ɑ́jà* (5, 9a) 'home village' (DEFGH+), all almost completely with *y* or ø reflexes.[32] It should be noted that in these roots either the vowels are the same

[30] When Guthrie did not have enough examples for a valid C.S., he created a "partial series", abbreviated as "ps". See Guthrie (1967: 42): "Frequently it has not proved possible to complete a valid C.S. but sufficient items have been discovered to make a partial series. Unless there are reasons to the contrary, such series are included in the main catalogue with a separate serial numbering, distinguished by the use of the abbreviation ps."

[31] An anonymous reviewer kindly added Duala A24 *ɲ-ɔ̀ɔ́* 'hair', Elip A62C *g^{y}-ɔ̀yá / b^{y}-* 'feather, hair'.

[32] Two of these words (baobab and fig) are flora, possibly added as certain species were encountered during the Bantu Expansion.

or both are low vowels, i.e. the conditions are not favourable for the simple formation of a glide from the first vowel, which is the standard treatment for resolving hiatus in vowel-initial roots. Accordingly, the easiest PB reconstruction here is *ø* for C_2 with the frequent development of epenthetic elements in various languages or branches but rarely with the strong **j* effects seen at morpheme boundaries. Early roots with this structure are rare in PB and, if no glide is reconstructed, one would want to understand their difference from long vowel roots. Other candidates having medial **ij* combinations will be treated later, e.g. **jíjà* 'fire', **jíjà* 1a 'mother', **jíjì* 6 'water'.

3 **j* conditioned by preceding **i*

In a significant number of cases, stems reconstructed with **j* are conditioned by a preceding **i*, either as part of the root or in a prefix. There are several ways for this to happen, especially:

- prefix *i-* before class 5 nouns, e.g. **jàdà* 'rubbish-heap', **jícò* 'eye'
- **jij* in stems, e.g. **jíjɪb* 'know', **jij* 'come'

Here, a distributional pattern appears that is very different from our previous categories. This environment is the major source for the strong reflexes of **j* and the tradition of reconstructing some palatal stop or affricate rather than *y* or *ø*. But these strong forms result from localised rules mostly in Eastern Bantu. Basically, what I have called weak reflexes (*y*, *w*, *ø*) are regular in the North-Western zones ABC, but strong forms (*j*, *z*) are occasional in the north-east and south-west Savannah zones (EFGJKR) and regular in South Bantu (N20-40, P30, S).

3.1 Initial **j* in class 5 roots

3.1.1 Typical reflexes

A fair number of class 5 nouns are traditionally reconstructed with **j* by both BLR and Guthrie (sometimes with doublets in **y*), for example:

(16) a. *jàdà* 5/6, (7/8) ‘rubbish-heap’ [BLR 1557, zones ABCDE+, C.S. 918]
b. *jàná* 5 ‘yesterday; tomorrow’ [BLR 1566, ABE+, ps 256]
c. *jánì* 5/6 + ‘leaf; grass’ [BLR 1567, ABCDE+, C.S. 926]
d. *jɔ́bà* 5 ‘sun’ [BLR 1614, ABCDE+, C.S. 955, 2147, ps 508]
e. *jɔ́ì* 3/4, 5/6 ‘voice; word’ [BLR 1612, ABCDE+, C.S. 954, ps 260]
f. *jɔ́dʊ̀* 3/4, 5/6 ‘nose, nostril’ [BLR 1620, ABCDE+, C.S. 960, 2151]

For PB, the class 5 prefix is reconstructed as the high front vowel **i*, with a pre-prefix (or augment) **dɪ-*, together forming the full template **dɪ-i-*ROOT.[33] It will be seen that these roots are best reconstructed with initial vowels to which the prefixes have attached themselves. Perhaps the strongest evidence for this comes from the fact that class 6 plurals almost never show any strong reflex.

A classic example of this category is **jɔ́bà* ‘sun’, which is attested in all zones and highlights the important evidence from the North-Western branches (sometimes with meanings ‘sky’ or ‘day’):

- from Guthrie: Lundu A11 *ɗ-ọɓa*, Duala A24 *l-oɓa* cl. 13(?), Mvumbo A81 *dụɔ*, Makaa A83 *düawɔ*, Ngom B22b *ð-ọɓa* cl. 11, Tsogo B31 *ọɓa* cl. 11.
- other A10-60: Kundu A122 *lóbà* (Atta 1993: 89), Batanga A32C *ɗóɓà* (den Besten 2016: 35), Abo A42 *lɔu*, Dibum A43a *lɔ́p*, Nen A44 *nìɔ̀f*, Kpa A53 *ɗíóó*, Baca A621 *ɲɔ̆·p* (Mous & Breedveld 1986: 227, 232).

The reflexes of **d* and **di* vary language by language, but all of these forms can be seen as descendants of a vowel-initial root with pre-prefix, **dɪ-(i)-ʊba*, with an initial *d / ɗ / l / n* from the conditioning and contraction of **dɪ-*/**di-* before the initial vowel of the root. The occasional forms in *j/dz/dj* apparently result from palatalisation before the initial vowel, e.g. **dɪ-V* > **d^{y}-V* > *jV*, hence Benga A34 *jọɓa*, Basaa A43 *jɔb*, Bulu A74a *jọp*.[34] In Ewondo A72a, this stem has two forms *yób ~ dzób* ‘sky’, which are apparently the results of the prefix or augment alone:

[33] There is possible influence from allomorphs in other classes which lack the *i-* environment (especially the class 6 plurals) or which have N-conditioning. So, in selecting class 5 nouns for analysis, I have excluded any which have class 9 or 10 by-forms, to ensure that there is no influence of those **nj*, **ny*, **nz* forms on the class 5 forms. Accordingly, an analysis of this type would need to be more detailed, especially since the distribution of strong forms varies by lemma.

[34] There are probably a number of phonological and morphological factors in each language. For example, there are different conditioning factors in Bulu A74a: in C_1 unconditioned **d* > *y*, but **di* (or **dí-i*) > *d* (e.g. *dim* ‘extinguish’ < **dim*; *dis/mis* ‘eye(s)’ < **jícò*; *di/mi* ‘fireplace(s)’ < **jíkò*), and **dɪ-VC* > *j-VC* (*jal/mal* ‘village(s)’ < **jàdá*) (Yanes & Moise 1987: 10–14).

**i-ɔ́bà*, **dɪ-ɔ́bà* (Essono 2000: 197). A key point is that nowhere do we find forms that reflect an augment plus a consonantal onset like **i-jɔ́bà* or **dɪ-jɔ́bà*. The substantial North-Western evidence for the PB reconstruction of a vowel onset for this class 5 root is matched with straightforward data elsewhere ('sun': Nande JD42 *erị-uβa*, Luba-Kasai L31a *di-ūba*, Mwera P22 *li-uβa*, Herero R30 *e-yuβa*).

But it is also important to understand the different changes in certain Eastern branches that led previous scholars to generalise the strong onsets. To understand the general path of development, it is useful to look at a few special nouns reconstructed by BLR3 with initial **ji*, which are also likely to be vowel-initial:

(17) a. **jícò* 5/6 'eye'
b. **jíkò* 5/6 'fireplace; country'
c. **jínò* 5/6 'tooth'
d. **jínà* 5/6 'name'

These class 5 nouns show an unusually wide variety of onsets across the Bantu area. However, if we assume that these were also roots with an initial vowel *i* (as supported by Bantoid forms of 'eye': Ekoid *e-yɨd*/*a-mɨd*, Tiv *i-ʃə*/*a-ʃə*), then the variety is quite understandable. The contact of the class 5 prefixes **i-* and **dɪ-* with the initial vowel inevitably led to certain mergers that blurred the morpheme boundaries. We see three types (examples from Guthrie C.S. 2030 **yícò* 'eye', using his orthography):

- Contraction: **dí-(i)-íco* 'eye' > **dííco* or **dííco*. From the full PB augment and prefix, we can expect a contraction of the sequential front vowels. The impact of the vowels on the initial **d* depends on the quality of the contracted vowel, the consonant rules of specific languages, and analogy:
  - languages with a form of *d* conditioned by the vowel *i*, or a prevocalic reflex (typically *dʒ*), rather than the unconditioned reflex (typically *l*, *y* or *ø*). Often, we can assume an intermediate **dii*, due to a contraction of the augment and prefix and the root beginning with *i*. For example, Duala A24 *dʼisɔ*, Ngom B22b *dʼiʃ*/*miʃ*, Bali-Teke B75 *dziu*, Bongili C15 *diʃọ*/*miʃọ*, Boloki C36e *dʒiọ*/*miọ*, Bushong C83 *diiʃ*/*miiʃ*, Manyanga H16b *diisu*/*meeso*, Luba-Katanga L33 *jiiso*/*meeso*.
  - languages which show the unconditioned reflex of **d*, most likely because the onset was generalised from other class 5 modifiers. For example, Sukuma F21 *liiso*/*miiso*, Luvale K14 *liso*/*meso*, Yao P21 *liiso*/*meeso*, Southern Sotho S33 *lẹihlɔ*/*mahlɔ*.

- Inverted augment: **dɪ-i* > **i-dɪ* or **ɪ-di* etc. > *ili* etc. Certain Eastern Bantu languages have reversed the order and/or function of the class 5 augment and prefix, probably to create paradigmatic regularity with V-CV structures in other classes.[35] Conditioning of the **d* can lead to *r* or *dʒ* instead of *l*. For example, Nande JD42 *eriĵso*/*ameeso*, Nyoro JE11 *eriiso* /*amaiso*, Luyana K31 *ilito*/*amiyo*.
- Loss or disuse of the augment: **(dí)-i-íco* 'eye' > *iso* etc. This development is not uncommon in zones C and D and is characteristic of zone R, e.g. Umbundu R11 *iso*/*oβaso*, Ndonga R22 *exo*/*omexo*. It also forms the basis for some South Bantu changes seen below.

Once again, the categories above are explicable by reconstructing the class 5 forms of a vowel-initial PB root **íco* 'eye'. Likewise, throughout the Bantu languages we see several options in their class 6 plurals based on a vowel-initial PB root:

- (*a*)*me(e)so*, from a coalescence of **(a)ma-íco*;
- (*a*)*ma(a)so*, from a contraction of **ma-(i)co* (favouring the first vowel) or a reanalysis of the stem as **co*, perhaps based on a singular form **dico*. This is the standard form in zones AB;
- (*a*)*mi(i)so*, from a contraction of **m(a)-íco* (favouring the second vowel);
- (*a*)*ma-iso*, uncontracted, likely an analogical restoration (rare outside JE10).

3.1.2 Eastern cases of class 5 strengthening

In addition to the straightforward development of class 5 vowel-initial roots in most of Bantu above, there are two sub-branches where fricative or other strong onsets developed: South Bantu and North-East Coast Bantu.

3.1.2.1 South Bantu strengthening: class 5 forms with j, z, ž, etc.

In South Bantu languages (zones NPS), we see several types of paradigms in these common nouns:

[35] For combinations of cl. 5 prefixes in Eastern Bantu, see Kamba Muzenga (1988).

- Vowel-initial stems in singular and plural, e.g. Tswana S31 *lì-íná/mà-íná* 'name', *lì-ínɔ́/mènɔ́* 'tooth', *lì-ítɬɔ́/mà-tɬɔ́* 'eye' (Creissels 1999: 325) – notice that there are three different types of plural formation – or Zulu S42 *i(li)so/amehlo*.
- Strong onsets in singular and plural, e.g. Zulu S42 *izinyo/amazinyo* 'tooth'.
- Mixed onsets in the same paradigm, e.g. Shona group S10 *zísó/mèsó* + *màzísó* + *màsó* 'eye'; *zínó/mènó* + *màzínó* 'tooth, teeth' (Hannan 1974).

Several languages in the region have some mix of types, so analogical processes must be at work. The class 6 plurals (aided by contraction) often preserve vowel-initial stems and we can surmise that the occasional strong onsets in the plural are by analogy to the singular.

What is the source of the several South Bantu strong onsets? An obvious option would be a development from the class 5 augment and prefix **dí-i*, as seen above, and that may be a factor in some languages. But that does not seem to work for languages like Shona where the strong *z* reflex here is not derivable from any version of the prefixes.[36] Rather Shona *z* matches the onsets in class 5 forms from PB **g*. In general, PB **g* was lenited to Proto-South Bantu **y* and eventually lost in most languages. After the class 5 prefix **i-* there arose a special set of changes for all the stops, e.g. Shona *dákó/màtákó* 'buttock' < **tákɔ̀*. For **g*, we see **i-g* > **i-y* > Chewa N31b *(d)z*, Shona *z*, Venda S21 *d̠*, Zulu *z*, Tswa S51 *t*, for example, **gɔ̀dɔ̀* 5 'sky, top' > Zulu *ízulu* 5 'sky, heaven'. This phonological change is also seen inside roots, e.g. **tɔ̀ìgà* 'giraffe' > **tɔ̀ìyà* > Shona *twìzà*. These are the same reflexes seen for the **j* nouns in class 5. It is for this reason that Meinhof et al. (1932) began many of these class 5 stems with **ɣ* (the graphic predecessor of **g*) rather than **y* (now **j*), and Guthrie had a doublet series in **g* for some of these words: C.S. 831 **gína* and C.S. 2068 **yínà* 'name'; C.S. 828 **gíkɔ̀* and C.S. 2056 **yìkɔ̀*.

In short, the strong reflexes of **j* in South Bantu nouns appear to reflect stems which had initial *y* at some stage, perhaps because they were the inherited forms in some stems or, more generally, because the glide was inserted to resolve the hiatus between a prefix and a vowel-initial root. In fact, the augment **i-* may have sometimes become that glide and then was reanalysed as part of the root and assigned the root anew, i.e. **i-ɔ́bà* 'sun' > **yɔ́bà* > **i-yɔ́bà*.[37]

[36] In Shona, **dí-C, *dì-V* > *dz* (**dím-a* 'extinguish' > *dzíma, *dì-ama* 'sink' > *dzàma, *dì-ik-a* 'bury' > *dzika*) and **dɪ-V* > *dy* (**dí-a* 'eat' > *dyá*).

[37] Similar is the development of a glide and then glide strengthening in Ganda JE15, where the class 5 prefix generally causes gemination, e.g. **jíbà* 'pigeon' > *ejjibá* 5 / *amayíba* 6.

3.1.2.2 North-East Coast Bantu strengthening: class 5 forms with j, z, ž, etc.

There are a number of languages in the Sabaki group (E70, G40) and nearby that frequently show strong forms in class 5 (and by analogy in class 6), for example:

(18) a. Dawida E74a *iziso/meso* 'eye(s)' *ijani/mani* 'leaf/leaves'
b. Unguja Swahili G42d *jicho/macho* 'eye(s)' *jani/majani* 'leaf/leaves'
c. Ngazija G44a *dzitso/matso* 'eye(s)' *wani/mani* 'leaf/leaves'

In many North-East Coast Bantu languages, the only class 5 prefix is a single vowel *i-* and often it is deleted, leaving a *ø*-prefix for polysyllabic consonantal stems, e.g. **pácà* 'twin' 5/6 > Swahili *pacha/ mapacha*. But for monosyllabic stems, a variety of prefixes are found in the Sabaki languages, e.g. from **bú* come *ivu*, *ɟivu*, *vuu*, *livu*, *rivu*. A number of hypotheses (including retention of the prefix **dɪ-*, and analogic reformation) led Nurse & Hinnebusch (1993) to reconstruct a series of local changes to explain these monosyllabic stems, as well as our class 5 vowel stems: pre-North-East Coast Bantu **(ɟ)lɟ-* > Proto-North-East Coast Bantu **(ɟ)zɟ-* > Proto-Sabaki **ɟjɟ-*.[38]

3.1.3 Summary of class 5 effects

I have given some attention to the South Bantu and Sabaki groups, because the impact of certain coastal languages (e.g. Zulu and Swahili) on the early Bantuists was high and inclined them to propose some consonantal onset for these stems. But in other branches as well, there are examples of both strong and weak reflexes which suggests that they co-existed for many years, as the form of the class 5 prefixes varied, with possible analogy from class 6 forms in *ma-*. The Kikongo Language Cluster (part of the West-Coastal branch) provides examples of this variety of prefixes and onsets (*y ~ z)* for forms of **jɔ́dɔ̀* 'nose' (with variant **jídò*): Vili H12L *liyilu*, Yombe H16c *yilu*, Soonde H321 *múzulu*, Mbala H41 *muzulu*, Sikongo H16a *zúúnu*, Solongo H16aM *dizunu*, Woyo H16dK *yiilu*, etc. This is paralleled by a variety of class 5 forms in PB **g*: for example, **gɔ̀dɔ̀* 'sky, top': Vili *liyilu*, Yombe *yilu*, Lumbu B44 *diyuulu*, Yaka H31 *zúlu*, Laadi H16f *zúlù* (from the Grollemund Dataset, itself taken from de Schryver et al. (2015) for the

[38] The problems of **j* and class 5 forms in the Sabaki group are discussed in Nurse & Hinnebusch (1993: 108–112, 186–196). The process of strengthening in Comorian G44, discussed at pp. 133–145, parallels that found in South Bantu. See also Nurse (1979: 149–153) on Chaga E60 and the North-East Coast.

KLC).[39] The various explanations depend on the individual languages and lexemes. For our purposes, it suffices to say that the developments involved are all at intermediate to late stages of Bantu history.

In sum, all the class 5 nouns reconstructed with **j*are best reconstructed with initial vowels for PB node 1.[40] The general absence of consonantal reflexes in the class 6 plurals of these nouns is a significant problem for reconstructing a consonantal onset.[41] Rather, various phonological processes affected the singular class 5 prefixes **dɪ-* and *i-* before vowels with results that were sometimes reanalysed as strong onsets for the roots, especially in Eastern Bantu. Likewise, there is no need for **j* in Meeussen's (1967: 97) reconstruction of the augments **ju* (cl. 1), **jɪ* (cl. 9) or **ji* (cl. 10), which were based on Eastern innovations. For class 10, a coronal seems more likely, e.g. **di* (cf. C.S. 2225a).

This is also a convenient time to clarify one important point. Sometimes references are made to Bantu Spirantisation of PB **j*, based on *z* in some of the singular forms of these special words, see for instance, for Kalanga S16, Mathangwane (1999: 82–83, 88, 213). However, these are more easily explained by class 5 effects or reformation. If indeed these PB roots had had an initial **j* and if there had been an effect of the subsequent **i* on it, we should see it in both the singular and plural. But the fact is that we often see some change in the singular but not in the plural. Why would **j* not spirantise systematically before high vowels? Because it is actually zero or a glide.

3.2 Initial **ji-C* and **jij*

Long ago, Meeussen pointed out that his Bantu reconstructions had a surprisingly large number of verb roots beginning with **ji* (Meeussen & Tucker 1955: 177). Perhaps out of deference to tradition, Meeussen (1967: 86, 90) himself later hesitated about **ji-C* structures, reconstructing a parenthetical onset in forms like **(j)íjɪb* 'know', and an examination of the specific modern reflexes now shows that the first **j* is not needed.[42]

[39] Similar variation can be found under the entries for 'sky', 'fireplace', 'nose', 'eye' or 'tooth' in Koni Muluwa & Bostoen (2015: 72, 99, 127, 130, 181).

[40] The roots were likely vowel-initial at an earlier stage too. Cf. Eastern Grassfields **lɪ-ít`*/*\*mà-ít`* 'eye', **díŋ`* 'name' (Elias et al. 1984: 38).

[41] There are also nouns like **jánì* 'leaf, grass' (BLR 1567, C.S. 926, 1928) which is commonly cl. 5/6 but its initial vowel is clearly seen in other classes: Lundu A11 *ẹ-anị* 7/8, Bubi A31 *s-anyị* 19/13, Maande A46 *nụ-any/tụ-any* 11/13, Luba-Kasai L31a *lw-anyi* 11, Tswa S51 *by-anyi* 14.

[42] Among dozens of **jij* verb reflexes in the data from Guthrie (1967–71) and the Grollemund Dataset, we find an element before the *i* only in Teke Yaa B73c *yir* 'come', Yao P21 *(j)íis*, and Manyanga H16b, where they are resolving the hiatus of vowel-initial stems. Initial *y* is some-

In *ji-C* verbs, the South Bantu consonant changes are similar to what we saw with class 5 reflexes, for which we reconstruct the prefix as *i- not *ji-. For example, *jì-kad 'dwell, sit' > Manyika S13 *gara*, Makhuwa P31 *khala*, just as *i-kádà 'ember' 5/6 > Manyika *gara/makara*, Makhuwa *ni-khala*.[43] Thus, we would do better to return to the simpler version of Meeussen's (1967) reconstructions: **ikad* 'sit', **ɟgad* 'shut', **im* 'stand (up)', and *ɟji* 6 'liquid'.

So, if the descendant languages almost never show any consonantal remnant of the proposed first **j*, why was there a reconstruction of **ji* in these verbs instead of simpler **i*, and **jij* instead of simpler **ij*? If I understand the scholarly history, the prefix **ji-* (earlier **ɣi-*) was reconstructed to explain some verb forms which occasionally show *i* at the beginning of the stem or some consonant mutation. Meinhof et al. (1932: 179) state, "But *ɣi* can also be what remains of an old infinitive prefix, which has been retained in a few languages only. E.g. **-ɣikala* 'sit, remain', Shambaa *-ikata*, Herero *-kara*, Swahili *-kaa*." Meinhof's suggestion that **ɣi-* is what remains of an old infinitive prefix which later merged with the class 5 prefix has not been accepted. A better source morpheme of the appropriate shape and position is the reflexive pronoun **i-*, which Meinhof et al. (1932: 43) wrote as *yî*. The incorporation of reflexive particles into verb forms is well attested cross-linguistically and seen in Bantu languages in Tswana, Ganda and others.[44] The fact that many Bantu languages lose or change the reflexive particle allows this particular morpheme to be lost or reanalysed as part of the verb stem. Thus, the initial consonant in **ɣi-* seems to be due to two factors: Meinhof's early etymology of the infinitive prefix from a verb *ɣa, ɣe* or *ɣia* 'go' (ibid. 43), and the occasional forms in *ji/yi* in languages like Sango G61 and Kongo H16.[45] Accordingly, the reconstruction of the initial **j* in these roots seems to be a relic of Meinhof's early work and can be removed.

times also found in other **jiC* verb reflexes, e.g. Mpongwe B11a *yir*/*jir* 'pour' < **jit*, Makonde P23 *yigal* 'open' (but *id* 'come'). It is particularly common in the verb **jíb* 'steal' which has many zone AB reflexes with *yib* or *jib*.

[43] Botne (1991) gives a wide set of reflexes and an analysis for **jikad* 'dwell, sit'.

[44] For Bantu reflexives, see Marlo (2015); for a discussion of the lexicalisation of reflexives, especially with **kada*, see Botne (1991: 252).

[45] But certain sample languages dominated. Already in Meinhof's (1899: 153) *Grundriß*, two of the four reflexes given for **ɣi*-ama, *ɣi*-ma 'stand' have what looks like a consonantal reflex: Northern Sotho S32 *yema* (*ema*, *yama*) and Sango *jima*. Later Laman's data for Manyanga H16b had a major role in the sample languages, with **ji-C* reflexes like *yikal* 'dwell', *yimit* 'become pregnant'. Thus, Meinhof et al. (1932: 161) analysed the Kongo *-y-* as a preservation (even though they provided the evidence to show it is actually resolving the hiatus): "**ɣi* > *yi*, e.g. *yiza* 'come' < B. **ɣiɣa* […] In some instances, *ɣ* is completely lost, e.g. *iṅgi* 'many' < B. *ɣiṅgi*, *kw-iza* 15 infin. of *yiza* 'come'. Sometimes *k* appears for *ɣ* […] e.g. *kima* (dial.) 'stand fast' < B. *ɣima*."

There is a small but important group of PB nouns and verbs reconstructed with **jij*, with parallel reflexes:

- **jíjì* 6 'water' (BLR 3433, ACDE+), **jíjà* 'fire' (BLR 5884, ACDH+), **jíjà* 1a 'mother' (BLR 3513, ABCFG);
- **jìj* 'come' (BLR 3425, ABCE+), **jíj-ad* 'be full' (BLR 3429, ABCD+).[46]

Obviously, the initial **j* in all these stems can be omitted from the reconstructions. As usual, the noun reflexes are fairly stable: **jíjà* 1a 'mother' has five forms in zones ABC, all with *iy*. For **jíjà* 'fire', we see mostly *y* (many *eya*) but also some strong forms in zone A.

In the group of verbs reconstructed with **jij*, the shortness of most roots makes it sometimes hard to be certain of cognates or distinguish other effects. Two of the better documented verbs are **jíjad* 'be full' (**jíjud* 'become full') and **jíjɪb* 'know'.[47] In the reflexes of these lexemes, we typically see three types of initial sequences with examples of 'know' from the Grollemund Dataset:

- Weak (*y*, ø) – frequent everywhere: Yasa A33a *èyíbà*, Kuria JE43 *iβa*
- Strong (*z*, *j*, *dz*, etc.) – frequent everywhere, regular in South Bantu: Chewa N31b *dziwa*
- *i-* + Strong – frequent in Eastern Bantu: Mwani G403 *iʤiwa*

These outcomes are somewhat similar to the pattern that was discussed for class 5. Since the South Bantu languages share common reflexes of **jij* with what was reconstructed as class 5 **i-*strengthening of initial *y* (Shona *z*, Southern Sotho *tɬ*, Venda *ḓ*), it seems reasonable to tentatively consider that sort of **iy* structure for these words too. But in this case, **iy* would have to be already present at the PB level.

Let us begin with some examples of **jìj* 'come' from the North-Western branches: Kundu A122 *iya*, Mkaa A15C *yà*, Kpe A22 *jâ*, Kako A93 *njạ*, Tsogo B31 *e-y-a*, with an extended stem *yak/zak* seen in several B20 languages. For **jíjɪb* 'know': Wumbvu B24 *u-yiba*. In Central-Western languages, 'come' and 'know': Mboshi C25 *i-yaa* and *i-yeɸa*, Bunji C25A *i-jaa* and *i-jéβa*, Mongo C61 *yá* and *eb*, Libobi

[46] Cf. also the Eastern compound noun **jíjɔ̀kɔ̀dɔ̀* 'grandchild' 1/2 (BLR 3435, DEF+).

[47] Cf. C.S. 2047 **yíjad* 'become full'; 'know': C.S. 938, 968, 2001. I have not included very reduced forms of 'know' like Abo A42 *jĩ* or Basaa A43a *yi* because of the possible relationship to the stem *yem/jem* 'know' seen in A70.

C412 *bo-yéi* and *bo-yebi*. In general, for these branches, there seems to be a majority of weak reflexes but enough strong ones to need more study before making a generalisation. Likewise, West-Coastal and South-Western Bantu have a mixture of weak and strong forms, with much variation even inside subgroups.

In Eastern Bantu, almost all the reflexes are strong, but with such variation (*j*, *c*, *s*, *ʃ*, *ts*, *tʃ*, *z*, *dz*, *ʒ*, *ʤ*, etc.) that it is not easy to describe a common phonological development for the branch (although perhaps for subgroups like South Bantu). The same can be said for **bàij* 'work wood' (BLR 8930, C.S. 32, 86), a rare example of medial **ij* in a verb stem: there are no citations for zones ABC so the reconstruction must be attributed to an intermediate node of the Bantu Expansion, by which time some relevant phonological developments might have taken place.

In short, **jij* has become the traditional reconstruction for several stems regularly showing strong reflexes or *i* + strong in Eastern Bantu and frequently elsewhere. Since there are only a few of these roots (just as with medial **j* in general), this **iX* structure probably arose from the juncture of other elements in the language. At present, I might propose **iy* insofar as it is a common reflex and plausible source for some of the other forms. But one would need to explain the source of the glide, and how to distinguish the evolution of **V-iy-a*, **V-i-ia*, and **V-i-a*.

4 **j* conditioned by preceding **n*

Our final group is reconstructed **j* when pre-nasalised or in nasal combinations. Although BLR3 does not have **y* as a separate phoneme from **j*, it does distinguish **ny* from **nj*.[48] Altogether, there are several categories we could consider here (each followed by the number of main reconstructions in the BLR3 database):

- **N-j* (stem-initial **j* with the class 9 prefix) – 25 nouns
- **N-ny* (stem-initial **ny* with the class 9 prefix) – 3 nouns
- **ny* (other stem-initial, or final **ny*) – 7 verbs, 3 nouns
- **nj* (stem-medial or final **nj*) – 4 verbs, 8 nouns

[48] I maintain the graphic convention (used by Guthrie and BLR) of writing **ny* in these reconstructions, although **ɲ* may have been the case, as seems more likely in **nyàmà* 'animal' below. The emphasis in the discussion is rather on distinguishing reconstructions of **ny*/**ɲ* from those with a stop or fricative under the cover term **nj*.

Whether or not all of these reconstructed categories are correct, there must have been occasional cross-influence and reanalysis. Not surprisingly, BLR3 shows variation between *(*N*)*j* and **ny* in some stems, e.g. Main 7055 **nyóòtà* 5, 9 'thirst' ~ Variant 3580 **jótà* 9 'thirst' and Main 3273 **jéd* 'shine, be clear' ~ Variant 2324 **nyényè* 'star'.

We will mostly look at class 9/10 forms, which have nasal prefixes, but also some forms with nasals in other classes. The patterns are more consistent if we consider them by groups based on reflexes: (1) those with weak reflexes, pointing to **n-y*; (2) those with strong reflexes, pointing to **n-j/z*; (3) those with mixed classes.[49]

4.1 Nouns supporting PB **ny*

There are several nouns reconstructed with **nj* or **ny* that regularly have palatal nasal reflexes in both Bantu and Grassfields languages.[50]

***jókà* 9 'snake; intestinal worm'** (Guthrie both **yókà* and **jókà*) is attested in all zones. All citations from zones A and B (which are half of the Bantu family tree) have reflexes with *ɲ* (or occasional *n*) and the preservation of *ɲ* (or *n*) in zones H, L, R and S confirm that **n-yókà* ought to be reconstructed for PB. But in some other zones there frequently arose fricatives, affricates, and palatal stops, e.g. zones C (*ndz*, *nz*, *nj*, *ɲ*), DEF (*nz*, *nj*, *ɲ*, *nc*, *nʃ*, *ʃ*, *ch*), M (*nz*, *nj*). This range of mutations shows how **ny* could evolve into strong forms, and the individual variants were probably affected by the developing non-pre-nasalised phonemic inventory in those sub-branches.

***játí* 9 'buffalo'** is compiled by Guthrie (and followed by BLR3) in two series: **(n)yátì* (zones ABCEGMNPRS) and **játí* 9/10 (zones BCMN).[51] It is hard to believe that there were really two concurrent stems for a morphologically invariable and semantically stable item (and no single language preserves a doublet). Guthrie's data has *ny* in all 11 forms from the North-Western branches, and the majority elsewhere – leading us to reconstruct **nyátì* for PB node 1.[52] Once again it is interesting to note the half-dozen scattered forms in *n-j* or *n-ʤ* cited

[49]The most extensive study of this category is Bostoen (2005: 182-88) who focuses on **jɔ̀ngɔ́* 'cooking pot', but includes **jɔ̀ndɔ̀* 'hammer, anvil', **jénjé* 'cricket' and many other relevant lexemes. He assumes these class 9 nouns had a strong *NC in the C_1 position and shows how Meinhof's Rule plays a significant role in producing weak reflexes in Eastern Bantu.

[50]In this section, unless otherwise specified, Bantu language data comes from Guthrie (1967–71) and the Grollemund Dataset; Grassfields from the Grollemund Dataset.

[51]C.S. 927, 1947, ps 495; BLR 1569.

[52]Frequent nasal-initial weak forms in Grassfields would tend to push the reconstruction back further.

by Guthrie. Several of these are clearly late innovations (distinct even from close neighbours) but useful evidence that the development *ny > nj/ndʒ/nz is quite possible in independent languages.

***jùmá* 9 'back, rear'** is primarily listed by Guthrie under C.S. 2060 *yi̯mà.[53] For this stem, Guthrie's data is almost unanimously in favour of a weak onset, with numerous variations on the initial sequence displaying the range that is possible inside what I have called "weak": *nyi* in zones ADGHJKLR, *ni* DHKR, *ngi* HL, *nyu* BDFGLMP, *nu* DEFM, *nnyu* E. Occasional forms in other classes (e.g. Tikuu G41 *mma* 5, Mbundu H21 *r-ima* 5, Kwambi R23 *oku-nima* 15) show that the initial nasal in class 9 could be perceived as the class 9 prefix or as part of the stem. What is striking is the absence of strong forms (i.e. *n-j, n-ʒ, n-dʒ*) in Guthrie's evidence, even in the presence of the high front vowel, which has a spirantising effect in only a few cases, e.g. Sangu B42 *nzîmə̀* 'back, behind' in contrast to *ny* before the back vowels in the Sangu words for meat, god, snake, bird, and body (Idiata-Mayombo 1993: 102).

Guthrie (followed by BLR) considered the basic classes of ***jɔ̀nì* 'bird'** to be 7/8 or 12/13. However, the zone A and Bantoid evidence shows that the basic classes were 9/10, with the diminutive 'birdie' as an alternative formed in Bantu classes 7/8 or 12/13 (class 19 in Grassfields). The Grollemund Dataset lists over sixty forms of this word from zone A, Jarawan and Grassfields languages—all of them with *n*, *ɲ*, or *ny* (likewise Tiv and Ekoid). The later diminutives in other classes sometimes add prefixes to a stem with initial nasal, e.g. Shi JD53 *a-nyonyi* or Oku (Grassfields) *fə-nʊ́n*, or without, e.g. Luba-Katanga L33 *ky-onyi* (or *koni*), Tumbuka N21 *chi-yuni*.

One of the words most widespread in Bantu languages can be confidently reconstructed at PB node 1 as ***nyàmà* 9 'animal, meat'**, with palatal nasals also frequent in Bantoid cognates. But the internal structure of the form is less clear. It might seem simplest to reconstruct the PB root as *yàmà with a nasal class marker and assume reanalysis led to occasional forms with PREFIX-*nyàmà* in other classes (especially the animate class 1 *mu-*). But several factors argue for treating the palatal nasal as part of the PB root itself, as BLR reconstructs here exceptionally: **nyàmà*. First, it seems there are apparently no strong onsets of this word in Bantu languages. Also, unlike the word for 'snake', where some Grassfields and Beboid languages elide the initial nasal, the word for 'animal' always maintains an initial nasal in those languages. Possibly a pre-Bantoid proto-form

[53] BLR 3653 prefers **jùmá*, but the Grassfields, Tiv and Ekoid cognates argue for reconstructing the front vowel for both Proto-Bantoid and PB, which was then sometimes affected by the subsequent bilabial.

had an *i*-prefix in some of these lexemes, e.g. 'snake **i-noka* or **in-oka* or **in-noka* or **ni-oka*, but the persistence of the palatal nasal in 'animal' suggests it must have been part of the stem itself before Bantu.

4.2 Nouns supporting PB **nj/nz*

There is also a group of nouns with consistent strong reflexes like *nz*, *ndz*, *ndʒ*, and *nj* in descendant Bantu languages. Some examples are:

****jògù* 9 'elephant'** uniformly has strong reflexes: Mbonge A121 *njɛku*, Basaa A43a *ndʒɔk*, Mbula (Jarawan) *ǹzû*, West Kele B22a *nʒɔk*, Bangi C32 *nzɔku*, Kongo H16 *nza*, Ganda JE15 *enjovu*, Xhosa S41 *indlovu* (all from the Grollemund Dataset).

****jàdà* 9 'hunger; famine'** is recorded in all Bantu zones, consistently with strong reflexes: Akoose A15C *nzàà*, Bubi A31 *ecalá*, Mpongwe B11a *ndʒana*, Mongo C61 *njala*, Pende L11 *nzala*, Jita JE25 *injara*, Hehe G62 *inzala*, Zezuru S12 *nzara*; as well as Grassfields Fefe *nžiɛ̀* and Aghem *dzɨ̀ŋ*, and Tiv *ịjẹ̀n* (all from the Grollemund Dataset).

****jìdà* 9 'path'** is recorded in all Bantu zones, consistently with strong reflexes: Manenguba A15 *nzè*, Kulung (Jarawan) *njéré*, Eton A71 *zěn*, Ngom B22b *nzɛla*, Punu B43 *nzilə*, Rundi JD62 *inzira*, Lenje M61 *nshila*, Tsonga S53 *ndlela*; as well as Grassfields Fefe *má-ǹ-ʒì* and Aghem *dʒì* (all from the Grollemund Dataset).

Although our best examples of roots supporting PB **ny* occasionally develop strong forms, roots supporting PB **nz/nj* almost never weaken to *ny*. Accordingly, class 9 roots with mixed reflexes are best reconstructed with **ny*.

4.3 Nouns with mixed classes

So far, we have considered class 9 singular nouns that pair with class 10 plurals, and both classes are reconstructed by Meeussen (1967: 97) with prefix **n-*. But a good way to test the conditioning of **j* is to look at nouns which have allomorphs in different classes, i.e. in the phonological environments of different class prefixes.

Some of the best cases for testing nasal and non-nasal environments are nouns with singular cl. 11 prefix **dʊ-* and plural cl. 10 prefix **n-*. An example is ****jádà* 11/10 'fingernail'**, for which forms in classes 7/8 and 5/6 are also recorded, often with a semantic difference, e.g. 'finger' or 'hand' (BLR 1558, C.S. 919-20, 1893-4). In those languages which maintain some form of the cl. 11 prefix (either fully or integrated into another class), we sometimes see the original weak nasal-less stem, e.g. Mbole D11 *lwála*, but also the nasal incorporated, e.g. Wumbvu B24

liŋala, Bangi C32 *lɔ́nzáli*, and sometimes apparently even the cl. 7 prefix incorporated, e.g. Songola D24 *lù-chálà*. These nasal intrusions into cl. 11 show that analogy played a strong role in paradigm levelling, but the motivation might also be resolving an original hiatus from something like **dʊ-(y)ala*, hence Tetela C71 *lòkálá*.

North-Western class 10 (or 9) forms with palatal nasal reflexes, e.g. Mbuu A15A *nyàn*, Kulung (Jarawan) *nyáálí*, Njem A84 *nyâ*, as well as non-nasal forms in other classes, e.g. Abo A42 *tʃ-ăt*, argue for reconstructing a PB weak stem also for the nasal variants, e.g. **n-(y)ada*. But strengthening of **ny* > *nz* is seen in certain languages and groups like B50-80 + H16, where almost all the nasal forms are strong. Thus, class 11/10 pairs like Bali-Teke B75 *liyala*/*ndzala*, Nilamba F31 *lọala*/*nzala*, and Zezuru S12 *rwàrá*/*nzàrá* support PB **(y)àdá*, with some form of post-nasal strengthening (generalised in Nilamba *nzoka* 'snake', but not in Zezuru *nyóká*). This post-nasal strengthening or analogy must be localised because a mixture of its presence and its absence is seen among related languages: Kaningi Nord B602 *leŋara* and Atsitsege B701 *liŋala*, but "Teke d'Ibali (Congo)" B71aIb *lindzala* and Wuumu B78 *linzál*.

The lexemes ***jíkì* 9/10 'bee' and 14 'honey'** provide another set of allomorphs. Guthrie gives more than thirty forms for 'honey' from every zone, yet none of them has a stop or even a glide as an onset to the root: e.g. Bubi A31 *b-ọẹ*, Bulu A74a *w-ọẹ*, Mfinu B83 *bʉ̈ɪ̈ʉ*, Kuyu E51 *ọ-ọkẹ*, Manyanga H16b *bw-iki*, Luba-Katanga L33 *bu-uki*, Yao P21 *u-uci*, Xhosa S41 *uß-usi*.[54] In that sense, the data looks like that of the vowel-initial nominal roots discussed earlier, for example, **bɔ̀-átò* 14 'canoe'. For 'bee' (with the nasalising prefix of classes 9/10, and by extension 11), Guthrie provides evidence only for forms in *ny-* in zones A and B: Bubi A31 *lọ-nyọẹ*, Mpongwe B11a *nyọẹ*, Ngom B22b *ða-nyọi̯*, Lumbu B44 *nyosi*, Nzebi B52 *nyu̯x*(*i̯*), Bali-Teke B75 *nnũũ*. Similar forms in *ny-* are found throughout all regions of Bantu. So, the uniform testimony of the North-Western languages, with parallels in other zones, supports a PB weak onset for both words, e.g. **bʊ(ʊ)kɪ* 'honey' and **nɪ-ʊkɪ* 'bee'(or **bʊ-yʊkɪ* or **n-yʊkɪ*).[55] In that case, the strong forms of 'bee' in a number of Bantu languages (e.g. Bangi C32 *lọ-ndzọi̯*, Nande JD42 *en-zuki*, Ila M63 *in-zuki*) must once again be due to some post-nasal

[54] C.S. 962, 2003-4, 2113, 2156-7, 2159 (Guthrie 1967: 124–125, §74.31-4).

[55] The original character of the root's first vowel is unclear. It could be a front vowel which was affected by the back vowel of the cl. 14 and 11 prefixes, or it could be a back vowel which was affected by the glide *y* or V_2. The editors of BLR3 reconstruct the front vowel, but the evidence of most zones (including AB) argues for the back vowel at PB node 1. But cf. Jarawan *i* in 'honey': Mbula *nyì*, Jaku *bínyì*, Bankal *nyí* (Gerhardt 1982: 92).

strengthening of *ny* > *nz*, occasionally leading to mixed paradigms like Rundi JD62 *uru-yuki/in-zuki* 11/10.

Another example of apparent nasal strengthening would be **jɔ̀ndò* 'hammer, anvil' (C.S. 965, 2171, 706), with weak reflexes in various classes: Pongo A26 *ẹ-ọndọ* 7/8 'axe', Ngombe C41 *ẹ-yọndọ* 7/8 'hammer', and Ngom B22b *y-ọndọ* 19/13 'axe', but strong in Benga A34 *nɟọndọ* 9/6 'hammer'. Cf. Tiv *nọndọ/ị-nyọndọ*.

4.4 Summary of **ny* and **nj/nz*

There are two sets of nasal patterns for **j* with distinctly different onsets: palatal nasal (*ny*) and stronger combinations (*nj*, *nz*, *ndz*, etc.). In fact, in the dozens of languages in zones A and B which have reflexes of both 'snake' (apparently **yókà*) and 'elephant' (apparently **zògù* or **jògù*), none has the same onset for the two words. The same distinction in zones G and S shows that this is not an areal phenomenon and should be reconstructed for PB.[56]

If one wanted to reconstruct both these sets under one proto-phoneme, one would likely start at some pre-Bantu stage with the palatal nasal form and generate the strong nasal form as a conditioned allophone, since that is the directionality seen in the examples above: strong PB **nz/nj* forms (seen uniformly in **jògù* 'elephant', **jàdà* 'hunger', **jɪdà* 'path') rarely weaken in Bantu languages, whereas PB **ny* was often strengthened in various ways. This strengthening is seen both in class 9/10 lexemes like **yókà* 'snake' and lexemes of mixed classes like **(y)ʊ́kɪ* 9/10 'bee' and 14 'honey'. For the lexemes considered in this section, neither the influence of tone nor a subsequent vowel would give us a phonological rule to generate the strong reflexes. A possible rule could be based on C_2:[57] that voiced C_2 leads to a strong reflex of C_1 after nasal prefixes, e.g. **jàdà* 'hunger; famine', **jàdí* 'lightning', **jɪdà* 'path', **jògù* 'elephant', **jɔ̀gʊ́* 'groundnut'; and the lack of C_2 would also need to qualify, e.g. **jɔ̀* 'house' and **jâì* 'outside'. But apparent exceptions can be found, and the status and age of each lexeme would need to be studied. Any phonological rule would also need to account for variations in strong and weak reflexes of **nj* in C_2 as well. Even if a rule for allocating allophones could be found, it would have started in some pre-Bantu stage to account for parallels in other Bantoid groups, and it is not clear how long it operated or when the allophones eventually phonemicised.

[56] For nouns maintaining this distinction in Tswana, where the contrast is between weak *n* and strong *tl*, see Creissels (1999: 306–307).

[57] This is the approach of Meeussen (1973: 9-10). A phenomenon like Meinhof's Rule (nasal assimilation of N-C_1 before nasal or nasalised C_2 in nouns) in some Bantu languages supports the consideration of C_2 influence on C_1.

Accordingly, for BLR's nasalised or post-nasal *j, at the stage of PB node 1, it seems simplest to separately reconstruct initial *ny and *nj/nz. For 'animal, meat', one may also maintain a structure like BLR's *n-nyàmà.

5 Conclusions

In looking at environments where PB *j has been reconstructed, we have seen that it is a collection of distinct stories which require separate reconstructions, some clearer than others. Most often *j is really just a placeholder for various effects that occurred at morpheme boundaries and needs to be deconstructed, not reconstructed. To summarise, I have proposed replacing BLR3 *j and *nj with a PB inventory of this sort:

- initial ø (in most nouns and some verbs)
  e.g. *átò 'canoe', *(y)át 'split', *ícò 'eye';
- both ny and nz (in cl. 9 or 10 onsets and sometimes medially)
  e.g. *nyókà 'snake', *nzògù 'elephant', *nyànzá 9 'lake';
- y (in some onsets and medially)[58]
  e.g. *yíb 'steal', *káyá 'leaf', *íyad 'be full'??

This would mean removing *j from the reconstructed consonant chart in Meeussen (1967: 83), and in all his reconstructed forms. Likewise, there is no need for *j in the reconstruction of the pronominal prefixes (augments) of classes 1, 9 and 10 (*ju, *jɪ, and *ji respectively) nor in the demonstratives built on them (Meeussen 1967: 97, 107).

What are the implications for PB phonology and its evolution?

Vowel-initial roots The reconstruction of vowel-initial roots is an old idea, which was never really refuted. The Homburger-Coupez tradition put initial *g'/j in these roots and led to an expectation of CV-syllable structure in Bantu lexemes, but certain PB inflectional prefixes have always been reconstructed with initial vowels and thus inflected forms are often vowel-initial. It is clearly easier from the phylogenetic viewpoint to explain the exceptional strong (z/j) forms in a few languages than the weak (y, w, ø) forms in the great majority of languages

[58] BLR3 has already addressed other types of stems where Meeussen (1967: 82) considered it "difficult to distinguish VV from VjV, e.g. -béjad-/-bé(j)ad-/-béad- «plant, sow»" – in this case reconstructing BLR 165 *bíad.

across most major branches. The failure of **j* to undergo Bantu Spirantisation and the extreme rarity of **j* at C_2 in roots is not surprising if reconstructed **j* is understood as a construct based on later effects seen at morpheme boundaries in various languages or groupings. A number of these roots are found in Bantoid languages and further study may justify a reconstruction of some words with a consonant at some pre-Bantu stage, but our goal here was simply to clarify PB node 1.

Distinguishing *ny and *nz Whatever the pre-Bantu history, for PB one should make a distinction between **ny* and some other nasal sequence. While [nz], [ndʒ] and [nj] all frequently occur as "strong" reflexes of BLR's **nj*, the most common is perhaps [nz], so **nz* is a reasonable choice for the PB symbol, and it has the advantage of being detached from the conflations of the current symbol **j*. Of course, the specific phonetic features of any symbol will depend on further study of Bantoid data and directional tendencies in sound changes involving these sorts of fricatives. Since [s] has been seen as the likely phonetic value of **c*, it might be useful to remove the palatal series altogether and follow Greenberg in relabelling both **c* and **j* as **s* and **z*. The presence of **ny* and **nz* in the PB inventory might suggest that independent **y* and **z* were more frequent at some pre-Bantu stage, just as they were later in many Bantu branches.

Is *y part of the PB phonemic inventory? Many contemporary Bantu or Bantoid languages have semi-vowels, so it would not be surprising to include them in the PB system. Or perhaps the better question is at what stage(s) to reconstruct them.[59] The strongest cases for an early *y* that we have seen are in medial position in a few nouns, verbs in **iya*, and in the initial position of some verb stems. Also, if we are reconstructing **ny* (**ɲ*) for PB, it would not be a surprise to include a palatal glide. Its initial frequency might not have been high, but various processes have increased its frequency. The extent to which /y/ or /w/ should be reconstructed either as a phoneme or allophone (and at what stages) needs fresh study, free from the legacy of current unitary **j*. One might ask whether PB had rules for vowel contraction or hiatus resolution.[60]

[59] Nurse & Hinnebusch (1993: 61) in their overview of the phonological system of Proto-Sabaki: "the glides *w* and *y* are unchanged from earlier proto periods." Meinhof et al. (1932: 28) also reconstructed allophonic semi-vowels **ŷ* and **ŵ* (from **î* and *û*).

[60] Cf. Meeussen (1967: 82): "A closed vowel (ị, ụ; i, u; e, o) followed by a more open vowel (i, u, e, o, a) is sufficient to account for the occurrence of semi-vowels in the present-day languages. It is often difficult to distinguish VV from VjV (which will usually be written here as V(j)V."

Glide creation and strengthening Several times, we have seen variation between strong (*z*, *j*, *ʒ*, etc.) and weak (*ø*, *y*) reflexes in closely related groups of Bantu languages. This reinforces the cross-linguistic evidence discussed at the beginning that glides can often become fricatives and sometimes vice-versa. Environments that favour strengthening in the history of Bantu are preposed *i* and *n* from a variety of inflectional prefixes, e.g. **nyókà* 'snake' > Ngombe C41 *ndʒɔ*, Chewa N31b *njoka*. But languages can also make changes elsewhere, e.g. Eastern Bantu **kɔ́yɔ̀* 'fig-tree' > Yao *kuɟu*.[61] Faytak (2014) presents several examples of "high vowel fricativization" by which front high vowels change to coronal fricatives, i.e. [i] → [z] or [ʐ]. This process "that ends in complete fricativization of reconstructible **i* and **y*" (2014: 60) could be one of the routes of what appears to be strengthening of glides.[62] Glides, nasals, stops or fricatives could also arise at morpheme boundaries as incorporations of class or infinitive prefixes (**n* or **kɔ*) or other analogical processes.

Acknowledgements

I am grateful to the participants at the Bantu 6 Conference for their comments and questions, and particularly to Koen Bostoen, Gérard Philippson, and the editors and reviewers of this volume for their many helpful suggestions.

References

Atta, Samuel Ebongkome. 1993. *The phonology of Lukundu (Bakundu)*. Yaoundé: Université de Yaoundé I. (MA thesis).

Baker, Gary K. & Caroline R. Wiltshire. 2003. An OT treatment of palatal fortition in Argentinian Spanish. In Ana Teresa Pérez-Leroux & Yves Roberge (eds.), *Romance linguistics: Theory and acquisition. Selected papers from the 32nd Linguistic Symposium on Romance Languages (LSRL), Toronto, April 2002* (Current Issues in Linguistic Theory 244), 33–48. Amsterdam: John Benjamins.

[61] See Mortensen (2012) for the emergence of obstruents after high vowels as a recurring sound change, including examples in Grassfields Bantu.

[62] Following up on Connell (2007), who reviews fricative vowel phonemes in a Mambiloid lect, Faytak (2014: 64–78) discusses fricativisation in Grassfields and other Bantoid languages but mostly with regard to back vowels. Cf. also Hall (2014) for Westphalian German Spirantisation: "the change from an original prevocalic long vowel to the corresponding short vowel plus fricative (i.e. [ɣ])."

Bastin, Yvonne, André Coupez & Michael Mann. 1999. *Continuity and divergence in the Bantu languages: Perspectives from a lexicostatistic study* (Annals – Human Sciences 162). Tervuren: Royal Museum for Central Africa. https://www.africamuseum.be/en/research/discover/human_sciences/culture_society/lexicostatistic-study-bantu-languages.

Bastin, Yvonne, André Coupez, Evariste Mumba & Thilo C. Schadeberg (eds.). 2002. *Bantu lexical reconstructions 3 / Reconstructions lexicales bantoues 3*. Tervuren: Royal Museum for Central Africa. http://www.africamuseum.be/collections/browsecollections/humansciences/blr.

Baucom, Kenneth L. 1975. The phonology of proto-Wambo. *African Studies* 34(3). 165–184.

Bostoen, Koen. 2005. *Des mots et des pots en bantou : une approche linguistique de l'histoire de la céramique en Afrique* (Schriften zur Afrikanistik / Research in African Studies 9). Frankfurt am Main: Peter Lang.

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

Bostoen, Koen. 2019. Reconstructing Proto-Bantu. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 2nd edn. (Routledge Language Family Series), 308–334. Milton Park, Abingdon: Routledge.

Bostoen, Koen & Yvonne Bastin. 2016. Bantu lexical reconstruction. In *Oxford handbooks online*. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199935345.013.36.

Botne, Robert. 1991. Variation and word formation in Proto-Bantu: The case of *-YIKAD-. *Afrika und Übersee: Sprachen, Kulturen* 74(2). 247–268.

Bourquin, Walther. 1923. *Neue Ur-Bantu-Wortstämme, nebst einem Beitrag zur Erforschung der Bantu-Wurzeln* (Beiheft zur Zeitschrift für Eingeborenen-Sprachen 5). Berlin: Dietrich Reimer (Ernst Vohsen).

Bulkens, Annelies. 2009. Quelques thèmes pour 'pirogue' dans les langues bantoues. *Africana Linguistica* 15. 27–58.

Bybee, Joan L. & Shelece Easterday. 2019. Consonant strengthening: A crosslinguistic survey and articulatory proposal. *Linguistic Typology* 23(2). 263–302.

Connell, Bruce. 2007. Mambila fricative vowels and Bantu spirantization. *Africana Linguistica* 13. 7–31.

Coupez, André. 1954. Les phonèmes bantous g et j non précédés de nasale. *Zaïre : revue congolaise / Zaïre: Congolees tijdschrift* 8. 157–161.

Coupez, André, Yvonne Bastin & Evariste Mumba (eds.). 1998. *Bantu lexical reconstructions 2 / Reconstructions lexicales bantoues 2*. Tervuren: Royal Museum for Central Africa.

Crane, Thera M. 2011. *Beyond time: Temporal and extra-temporal functions of tense and aspect marking in Totela, a Bantu language of Zambia*. Berkeley, CA: University of California. (Doctoral dissertation).

Crane, Thera M., Larry M. Hyman & Simon Nsielanga Tukumu. 2011. *A grammar of Nzadi [B.865]: A Bantu language of the Democratic Republic of Congo* (University of California Publications in Linguistics 147). Berkeley, CA: University of California Press.

Creissels, Denis. 1999. Remarks on the sound correspondences between Proto-Bantu and Tswana (S.31), with particular attention to problems involving *j (or *y), *į and sequences *NC. In Jean-Marie Hombert & Larry M. Hyman (eds.), *Bantu historical linguistics: Theoretical and empirical perspectives* (CSLI Lecture Notes 99), 297–334. Stanford, CA: Center for the Study of Language & Information.

de Schryver, Gilles-Maurice, 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 + supplementary material online.

den Besten, Margaret G. G. 2016. *A phonology of Batanga (Bantu A30); The Banɔ'ɔ dialect*. Leiden: Leiden University. (MA thesis).

Elias, Philip, Jacqueline Leroy & Jan Voorhoeve. 1984. Mbam-Nkam or Eastern Grassfields. *Afrika und Übersee: Sprachen, Kulturen* 67(1). 31–107.

Ellington, John Ernest. 1977. *Aspects of the Tiene language*. Madison, WI: University of Wisconsin–Madison. (Doctoral dissertation).

Essono, Jean-J. Marie. 2000. *L'ewondo : langue bantu du Cameroun : phonologie – morphologie – syntaxe*. Yaoundé: Presses de l'Université catholique d'Afrique centrale.

Faytak, Matthew. 2014. High vowel fricativization and chain shift. *UC Berkeley Phonology Lab Annual Report* 10. 52–100.

Fivaz, Derek. 1986. *A reference grammar of Oshindonga (Wambo)* (African Studies of the Academy 1). Windhoek: Department of African Languages.

Gerhardt, Ludwig. 1982. Jarawan Bantu: The mistaken identity of the Bantu who turned north. *Afrika und Übersee: Sprachen, Kulturen* 65(1). 75–87.

Greenberg, Joseph H. 1969. Review of I. Fodor's (1966) The problems in the classification of the African languages: Methodological and theoretical conclusions concerning the classification system of Joseph H. Greenberg. *Language* 45. 427–432.

Grégoire, Claire & Jean-Paul Rekanga. 1994. Nouvelles hypothèses diachroniques sur la cl 10b du myene-nkomi (B11e). *Africana Linguistica* 11. 71–77.

Grollemund, Rebecca. 2012. *Nouvelles approches en classification : application aux langues bantu du nord-ouest*. Lyon: Université Lumière – Lyon 2. (Doctoral 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 National Academy of Sciences of the United States of America* 112(43). 13296–13301.

Guthrie, Malcolm. 1953. *The Bantu languages of Western Equatorial Africa* (Handbook of African Languages). London: Oxford University Press (for the International African Institute).

Guthrie, Malcolm. 1965. Comparative Bantu: A preview. *Journal of African Languages* 4(1). 40–45.

Guthrie, Malcolm. 1967. *Comparative Bantu: An introduction to the comparative linguistics and prehistory of the Bantu languages. Part I, Volume 1: The comparative linguistics of the Bantu languages*. Farnborough: Gregg Press.

Guthrie, Malcolm. 1967–71. *Comparative Bantu: An introduction to the comparative linguistics and prehistory of the Bantu languages (4 volumes)*. Farnborough: Gregg International.

Hall, T. Alan. 2014. The analysis of Westphalian German spirantization. *Diachronica* 31(2). 223–266.

Hannan, Michael. 1974. *Standard Shona dictionary*. 2nd revised edition. Salisbury: Rhodesia Literature Bureau.

Hedinger, Robert. 1987. *The Manenguba languages (Bantu A.15, Mbo cluster) of Cameroon*. London: School of Oriental & African Studies, University of London. (Doctoral dissertation).

Homburger, Lilias. 1913. *Étude sur la phonétique historique du bantou* (Bibliothèque de l'Ecole des hautes études. IVe section, Sciences historiques et philologiques 209). Paris: H. Champion.

Hualde, José Ignacio. 2011. Sound change. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology* (Blackwell Companions to Linguistics), 2214–2235. Malden, MA: Wiley-Blackwell.

Hyman, Larry M. & Francis X. Katamba. 1999. The syllable in Luganda phonology and morphology. In Harry van der Hulst & Nancy A. Ritter (eds.), *The syllable: Views and facts* (Studies in Generative Grammar 45), 349–416. Berlin: Mouton de Gruyter.

Idiata-Mayombo, Daniel-Franck. 1993. Eléments de phonologie diachronique du isangu (B. 42). *Pholia* 8. 67–108.

Kamba Muzenga, Jean-Georges. 1988. Comportement du préfixe nominal de classe 5 en bantou. *Annales Æquatoria* 9. 89–131.

Koni Muluwa, Joseph & Koen Bostoen. 2015. *Lexique comparé des langues bantu du Kwilu (République démocratique du Congo): Français – anglais – 21 langues bantu (B, C, H, K, L)* (Grammatical Analyses of African Languages 48). Cologne: Rüdiger Köppe.

Maho, Jouni F. 2007. *A referential classification of the Bantu languages: Keeping Guthrie's system updated*. Preliminary (online) version # 6 (dated 28 January 2007). Accessed 31 January 2007, but no longer online.

Marlo, Michael R. 2015. Exceptional properties of the reflexive in Bantu languages. *Nordic Journal of African Studies* 24(1). 1–22.

Mathangwane, Joyce T. 1999. *Ikalanga phonetics and phonology: A synchronic and diachronic study* (Stanford Monographs in African Language). Stanford, CA: Center for the Study of Language & Information.

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

Meeussen, Achiel Emiel. 1969. *Bantu lexical reconstructions*. Unpublished ms., 28 pp. Tervuren: Royal Museum for Central Africa.

Meeussen, Achiel Emiel. 1973. Comparative Bantu: Test cases for method. *African Language Studies* 14. 6–18.

Meeussen, Achiel Emiel & Archibald N. Tucker. 1955. Les phonèmes du ganda et du bantou commun. *Africa: Journal of the International African Institute* 25(2). 170–180.

Meinhof, Carl. 1899. *Grundriß einer Lautlehre der Bantusprachen, nebst Anleitung zur Aufnahme von Bantusprachen. Anhang: Verzeichnis von Bantuwortstämmen* (Abhandlungen für die Kunde des Morgenlandes 11(2)). Leipzig: F.A. Brockhaus.

Meinhof, Carl. 1910. *Grundriß einer Lautlehre der Bantusprachen, nebst Anleitung zur Aufnahme von Bantusprachen. Anhang: Verzeichnis von Bantuwortstämmen*. 2nd revised edition. Berlin: Dietrich Reimer (Ernst Vohsen).

Meinhof, Carl, Alice Werner & Nicolaas J. Van Warmelo. 1932. *Introduction to the phonology of the Bantu languages*. Being the English version of *Grundriß einer Lautlehre der Bantusprachen*. Berlin: Dietrich Reimer (Ernst Vohsen).

Mortensen, David R. 2012. The emergence of obstruents after high vowels. *Diachronica* 29(4). 434–470.

Mous, Maarten & Anneke Breedveld. 1986. A dialectometrical study of some Bantu languages (A. 40 – A. 60) of Cameroon. In Gladys Guarisma & Wilhelm J. G. Möhlig (eds.), *La méthode dialectometrique appliquée aux langues africaines*, 177–241. Berlin: Dietrich Reimer.

Ngunga, Armindo S. A. 2000. *Phonology and morphology of the Ciyao verb* (Stanford Monographs in African Languages). Stanford, CA: Center for the Study of Language & Information.

Nurse, Derek. 1979. *Classification of the Chaga dialects: Language and history on Kilimanjaro, the Taita Hills and the Pare Mountains*. Hamburg: Helmut Buske.

Nurse, Derek & Thomas J. Hinnebusch. 1993. *Swahili and Sabaki: A linguistic history* (University of California Publications in Linguistics 121). Berkeley, CA: University of California Press.

Nurse, Derek & Gérard Philippson. 2003. Towards a historical classification of the Bantu languages. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages* (Language Family Series 4), 164–181. London: Routledge.

Odden, David. 2003. Rufiji-Ruvuma (N10, P10-P20). In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages* (Language Family Series 4), 529–545. London: Routledge.

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

Philippson, Gérard & Rebecca Grollemund. 2019. Classifying Bantu languages. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 2nd edn. (Routledge Language Family Series), 335–354. Milton Park, Abingdon: Routledge.

Rekanga, Jean-Paul. 1994. Les réflexes du protobantou en myene-nkomi, langue bantoue du Gabon (B11e). *Africana Linguistica* 11. 149–167.

Schadeberg, Thilo C. 2003. Historical linguistics. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages* (Language Family Series 4), 143–163. London: Routledge.

Sihler, Andrew L. 1995. *New comparative grammar of Greek and Latin*. New York, NY: Oxford University Press.

Stappers, Leo. 1971. Esquisse de la langue lengola. *Africana Linguistica* 5. 255–307.

Teil-Dautrey, Gisèle. 2004. *Lexiques proto-bantous : étude des cooccurrences segmentales et supra-segmentales*. Lyon: Université Lumière – Lyon 2. (Doctoral dissertation).

Turvey, B. H. C., W. Zimmermann & G. B. Taapopi. 1977. *Kwanyama – English dictionary*. Johannesburg: Witwatersrand University Press.

Van de Velde, Mark. 2006. *A description of Eton: Phonology, morphology, basic syntax and lexicon*. Leuven: Katholieke Universiteit Leuven. (Doctoral dissertation).

Yanes, Serge & Eyinga Essam Moise. 1987. *Dictionnaire boulou – français, français – boulou; avec grammaire*. Sangmelima: P. Monti.