Konstantin Pozdniakov

# From Proto-Atlantic to Proto-Niger-Congo: the root structure

It would be useful to get some clear ideas about lexical root structure in Proto-Niger-Congo before trying to advance concrete lexical reconstructions. Meanwhile, we know very little about the root structures in different branches of Niger-Congo. A deplorable state of our knowledge in this area could be illustrated by the last publication (Blench 2016) on the subject where we read, in particular, the following:

« A common assumption about the shape of Proto-Niger-Congo roots is that stems were disyllabic. This is supported by a vision of Bantu-like roots for nouns with the canonical shape (C)V-CVCV. <...> This paper puts forward a radical alternative, that many early Niger-Congo roots were trisyllabic, CVCVCV. »[1]

This quote raises a lot of questions. Is this assumption really common? With regard to the Bantu languages, what is "the canonical shape" for Verbs and how could we explain the striking differencies of Noun and Verb structures in Bantu? Does the simple scale (monosyllabic, disyllabic, trisyllabic roots) suffice and is it suitable for a systematic comparison of root structures?

The main focus of my presentation is a comparison of the root structures in the Atlantic languages. I also discuss some structure problems that arise when comparing Atlantic roots with those in other branches of the Niger-Congo (NC) and, in particular, the vocabulary of the proto-Bantu (PB).

I suggest the following model for comarison of various root structures:
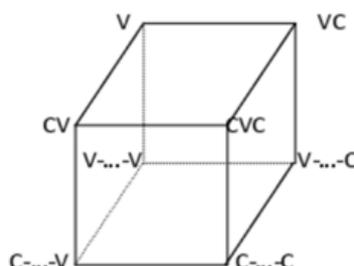
The front / back faces of the cube: structures 'Initial C-… / Initial V-…'
The top / bottom faces: structures 'Monosyllabic / Polysyllabic'
The left / right faces: 'Final -V / Final -C'.

Among these 8 structures, those 4 which are on the front face of the cube are more relevant than 4 others, because in general the roots with initial vowels are much less typical for NC languages.

Here I will demonstrate some problems treated in my topic by using the PB statistical data. Firstly, it is easy to control them because they were calculated on the basis of BLR reconstructions[2]. Secondly, for specialists in Atlantic these well-known PB data are particularly interesting, if we accept the following postulate: as proto-Atlantic and PB represent two polar zones of NC without any contact between them, any reliable Atlantic-Bantu cognate can be attributed to proto-NC.

Basing on the BLR reconstructions, we get the following frequencies of the structures: CV 6%' CVC 23% ; CVCV, CV ... CV 51% ; CVCVC, CV ... CVC 20%. However, these figures vary considerably if we choose the most reliable (the most stable) Bantu roots represented in

---

[1] A draft circulated for comment with the intriguing title « Supposing we have been completely wrong about the shape of early Niger-Congo roots ? » REF. The very title of the text contains a statement of the existence of a common position on this issue, which, in my opinion, is doubtful.

[2] http://www.africamuseum.be/collections/browsecollections/humansciences/blr

seven zones or more ("Bantu 7+"). We find 948 such reconstructions in BLR and the structure frequencies are very different there: the CVCVC frequency is lower by half (only 11%). In general, more stable is a root, more is it "shorter" (Table 1).

The next interesting point is that PB structures are represented unequally in different Bantu zones. If we select PB roots represented in different zones, we could conclude, for example, that the CVCVC structure has 25% in the zone J and only 6% in the zone A, where 953 PB roots are represented (Table 1).

The main point is that the structure frequencies in Nouns and in Verbs are strikingly different (Table 1).

Tabl. 1

	N	N	N	N	%	%	%	%	%	%
	Bantu ALL	bantu 7+	zone J	zone A	Bantu ALL	bantu 7+	zone J	zone A	Nouns ALL	Verbs ALL
CV	570	101	218	87	6	11	5	9	8	2
CVC	2217	288	880	270	23	30	22	28	0	49
CVCV	5045	450	1887	538	51	47	47	56	91	6
CVCVC	1988	109	995	55	20	11	25	6	0	43
	9820	948	3980	953	100	100	100	100	100	100

It is clear while regarding the data in 2 last columns of the Table 1, that before proposing any reconstruction of the root structure in NC, we need a coherent interpretation from bantuists concerning possible scenarios in which these differences could arise: 92% of Bantu Verb roots have a final consonant, while it is totally absent in Noun roots. I am going to discuss this indispensable problem.

A comparison of the root structures is extremely important also for the Atlantic reconstruction. Here I give just some illustrations to this point.

While comparing frequencies in mono- and polysyllabic roots on Nouns and Verbs of an « average » Atlantic language, we get the following distribution (Table 2):

Table 2. Monosyllabic structures opposed to polysyllabic structures (the opposition "top – bottom").

Average	Monosyllabic	Polysyllabic	SUM
Nouns	23% (–)	**77% (+)**	100%
Verbs	**47% (+)**	53% (–)	100%

In the majority of the examined Atlantic languages there are many more monosyllabic Verbs than Nouns. That means that Nouns are longer than Verbs. Why? Only one explication is possible: the presence of fused morphemes in Nouns (markers of noun classes).

Table 3. Structures with final consonants vs. structures with final vowels

%	Basari	Jaad	Gola	%	Basari	Jaad	Gola
Verbs -	33	13	**97**	Nouns -	34	**82**	**81**
Verbs -	**67 (+)**	**87**	3	Nouns -	**66 (+)**	18	19
	100%	100%	100%		100%	100%	100%

This table presents very interesting data. The situation in Basari reflects a typical situation for Atlantic languages: words with final consonants represent the majority of the words in the dictionary (approximately two-thirds of the entries) as the most frequent structure of the lexical stem is CVC. This prototypical situation is valid both for Verbs and Nouns.

In the Nouns of Jaad the opposite situation is encountered: approximately 80% of the Nouns present in the dictionary have final vowels but not the Verbs (only 13%). In this language this particular distribution was influenced by two factors. First of all, thanks to this statistical data I paid attention to the fact that in the derivation of Nouns from Verbs (of CVC structure) you have not only a class prefix but also an additional final vowel: *cid* 'to cook' > *ka-cid-e* 'kitchen'; *pees* 'to sweep' > *ka-mpees-a* 'broom'; *puuf* 'to blow' > *ka-mpuuf-ɛ* 'bellows'; *raf* 'to make old' > *ka-ntaf-ɛ* 'old age', *etc*. This is characteristic also for other languages of the group Tenda-Jaad. Compare the closely related forms in Konyagi: *i-pas* 'to sweep' > > *æ-fas-a* 'broom' ; bedik *u-wuf* 'to blow' > *gi-mbuf-e* 'bellows'; bedik *u-raf* 'to be old' > *ndaf-a* 'old man'. But in Jaad there was also another cause for the appearance of the final vowel in Nouns. In Jaad there are a lot of loanwords from Mande languages with an initial CVCV structure and the Nouns are much more often borrowed than are Verbs[3].

Statistically, in Verbs Gola is different from Basari-Jaad and in Nouns Basari is different from Jaad-Gola. Summing up the data, it is clear that the opposition of Basari to Jaad and Gola in Nouns is not of a genetic nature. There are no doubts that Basari and Jaad belong to the same group in the Northern branch of the Atlantic languages. Gola instead apparently does not belong with the Atlantic languages, representing an independent branch in the NC macro-family. This is confirmed by the distribution of frequencies in Verbs: Basari and Jaad are opposed to Gola where (the only one in the three languages) all the Verbs have a final vowel.

How could such a strange distribution of frequencies occur? It reflects two different independent diachronic changes which took place in Gola, on the one hand, and in Jaad, on the other hand.

In Gola proto-language roots systematically lost the second consonant. Some examples are reported as follows:

- ATL. **jeb* 'cure' (Mankanya *p-jeb,* Nyun *jeb,* …) ~ **Gola *jwɛɛ***;
- ATL. *ɲamb* 'elephant' (Joola **ɲaab*, Basari *ɲàmb*, …) ~ **Gola *ó-ɲ̣a̱ɡ̣***;
- ATL. **deng* 'thorn' (Wolof *deg*, …) ~ Sua *deng-en* ~ **Gola *é-d̰ɛ̰é***;
- ATL-North. *ɗɪɪg* 'cow without horns' (Palor *ɗɪɪg*, Sereer *ɗiik*, …) ~ G**ola *ó-dɪ̀ɪ̀***;
- Balant *tɔg* 'push' ~ **Gola *tɔɔ***;
- ATL. *ɓɔŋ* 'thigh' (Joola **bɔŋ*, Ndut *ɓaŋ*) ~ **Gola *o-gbḁ̀ḁ́***;
- Balant *tɔŋ* 'show' ~ Sherbro *ʈonki*, Bom *tɔngi* ~ Nalu *tɔŋ-el* ~ Limba *tɔŋ-ina* ~ **Gola *tɔɔ***;
- ATL. **nof* 'ear' (Bijogo *kɔ-nnɔ*, Cobiana *si-nuf*, Basari *a-nʌ̀f*, Palor *nʋf*, Wolof *nɔpp*, Fula **nof-ru*, Baga Mboteni *ɛ-nʌ́f*, Baga Fore *í-nɔ́p*, Nalu *nɛɛw,* …) ~ Limba *ku-luh-a* ~ **Gola *ké-núù***;
- ATL-CENTRE **sʋn* 'horn' (Joola **sʋn*, Nalu *seen*) ~ Limba *kɔ-se* ~ **Gola *é-sii***.

This list of examples can be significantly extended. The loss of the final consonant in Gola regularly gives the compensatory length of the vowel. The loss of the final vowels in Gola in the words with the CVC structure can be found differently as well. For example, Table 4 shows a comparison of the frequencies of the monosyllabic words in two different languages:

---

[3] I would like to thank Guillaume Segerer who drew my attention to this important characteristic of loans influencing frequency distribution.

Table 4

	Gola	Wolof
CV	**38%**	2% (–)
CVC	10% (–)	**41%**
VC	0%	1%
SUM	48%	44%

In both languages the percentage of monosyllabic words is mainly the same – a little bit less than one half of a dictionary. The two concrete structures have a complementary distribution: apparently the majority of words with CVC structures in Wolof should correspond to words with CV structure in Gola.

In *Table 5* the frequencies of structures in an average Atlantic language are presented.

Table 5

Average (%)	*Verbs*	*Nouns*
CVCVC,	14 (–)	**25 (+)**
CVCV, CVCVCV	26 (–)	**34 (+)**
CVC	**41 (+)**	19 (–)
CV	5	3
VCVC, VCVCVC	10	11
VCV, VCVCV	4	7
VC	1	1
V	0	0
	100%	100%

The transformation of *CVC in **CV**-CVC in Nouns (integration of noun classes in the roots) is very clear. Despite this, the sum of these two structures in the Verbs (41+14=55%) is much higher than the same sum in the Nouns (19+25=44%). This means that the present explanation is not enough. We can suppose the change *CVC > CVC-**V** where the last vowel is a noun class suffix, a determiner or a derivational morpheme.

Some other interesting aspects of comparison of the root structures in NC languages will be discussed as well.