Konstantin Pozdniakov,
INALCO-IUF-LLACAN

From Atlantic to Niger-Congo: three, two, one …

(a draft copie for participants of the Niger-Congo Congres, Paris, 18-21 september 2012)

Numerals deserve a special attention in the analysis of distant language relationship. On the one hand, in highly structured paradigms, as the numerals are by definition, regular phonetic correspondences work at a limited extent, as numerals undergo the process of mass analogical leveling which implies phonetic adjustment of the forms. On the other hand, this very fact is a powerful source for etymological hypotheses in case the regular phonetic correspondences between the studied languages are not yet discovered – hypotheses on the kinship of two forms from different languages can be confirmed (sometimes quite reliably) by their structural parallels, if not phonetics.

It is not by chance that isoglosses with a numeral are often chosen by researchers as significant for the genealogical classification – cf. the well-known division of the Mande languages into two groups depending on the form of numeral 'ten', "Mande-tan" and "Mande-fu", or even more famous division of Indo-European languages into *centum* and *satem* languages based on the isogloss for 'hundred'.

The hypotheses on the place of Atlantic roots for numerals 1, 2, 3 in Niger-Congo proposed below are - inevitably - preliminary. Moreover, their elaboration doesn't even meet the requirements for a publication. My task is, basing on the example of numerals, to attract attention of the participants of the Congress to some theoretical and methodological aspects of reconstructing Niger-Congo which I consider the most important.

Thanks to numerous researchers, I have obtained probably the richest database on numerals in Niger-Congo. It contains about 2,000 sources on about 1,000 Niger-Congo languages. First of all, this database combines several extensive databases, that is:

http://www.zompist.com/numbers.shtml - a comprehensive database (compiled by Eugene S. L. Chan), which combines data on numerals in 5,000 languages of the world (the data on Niger-Congo is in most cases reliable enough) - hosted by the Max Planck Institute http://lingweb.eva.mpg.de; http://sumale.vjf.cnrs.fr/Lexiques/reflex (Guillaume Segerer database for his famous RefLex project).

As for Atlantic and Mel languages, which I am working on in particular, my database is most probably **full** and unifies data from all sources published to date (Atlantic numerals were collected through the sources primarily by Guillaume Segerer with my contribution whenever possible). Data on numerals in Mande languages collected by Valentin Vydrin and his research group is almost exhaustive. In general, this database in undoubtedly

representative for the task of a tentative reconstruction and doesn't contain any serious gaps.

I would note two circumstances which determined the logic adopted in this paper while analysing the numerals.

The first circumstance is historical. The hypothesis of kinship between Niger-Congo languages didn't appear as a result of discovery of numerous related forms, for example, in Mande and Adamawa. It appeared as a result of comparison between the Bantu languages, for which the classical comparative method was possible to be applied and which were reliably reconstructed, with other African languages. Niger-Congo does not exist without Bantu. We need to say clearly that if we establish a genetic relationship between a form in Bantu and in Atlantic languages, or between Bantu and Mande, we have all grounds to trace this form back to Niger-Congo. If we establish such a relationship between Mel and Kru or between Mande and Dogon, we don't have enough reason to claim it Niger-Congo. In other words, all Niger-Congo languages are equal, but Bantu languages are "more equal" than the others.

The second remark concerns my personal experience. I've been working on Atlantic and Mel languages for years, which determines my interest for Niger-Congo.

Therefore, in this paper Atlantic (and Mel) languages, as well as Bantu languages are two focuses of reconstruction, while the languages of other branches are brought in less systematically and in less detail. Having this in view, I decided to bring to this discussion etymological materials and hypothesis which are not deeply elaborated, looking forward for precisions and critical remarks by the participants of this Congress which unites key specialists on major Niger-Congo branches.

Hereinafter, I examine numerals 1, 2, 3 in a reversed order, as reflected in the title. The reason is as follows. From the first three numerals in Niger-Congo numerals for 'three' are considered the most stable and the most "easy" for reconstruction (I will try to show, though, that this "easiness" is to a large extent illusory). Numeral 'two' is more difficult to reconstruct and the most difficult is 'one' which has the biggest root variability. Therefore, the problems of reconstruction of the numerals should be considered in this very order.

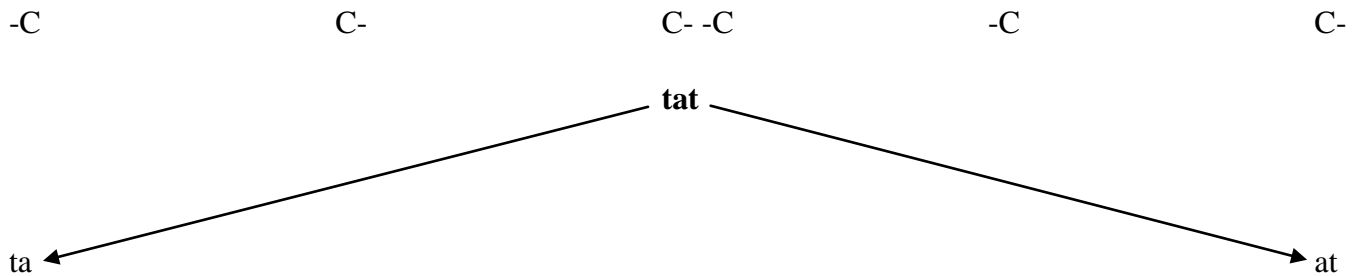**THREE**

The reflexes for numeral 'three' in Niger-Congo demonstrate a surprising stability and at the same time a surprising phonetic diversity. Here are some of the Bantu forms which can be traced back to the form *tato / caco* in Proto-Bantu:
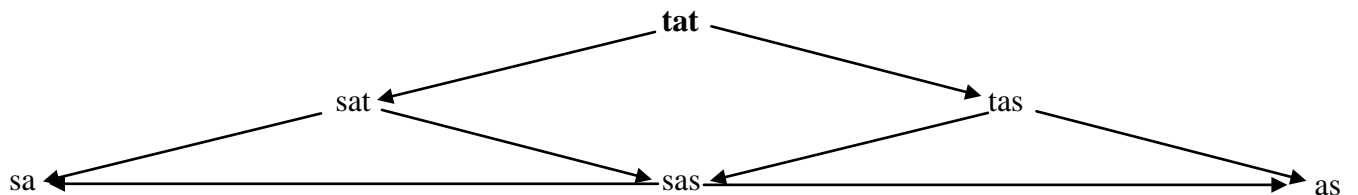
A	Nyo'o	tá	*proto-bantu		tatU / catU
A	Lundu	aru	D	Lega	sáro
A	Bonkeng	alu	E	Pokomo	hahu
A	Fang	lal	E	Embu	thatu
A	Ewondo	lá	E	Kahe	radu
A	Kpa	ráá	F	Sukuma	datu
A	Lombi	laso	G	Pemba	tatu
A	Bubi	cha	G	Tikuu	chachu
B	Yansi	taar	J	Konzo	satu
B	Mbede	tadi	J	Luganda	ssatu
B	Sira	reru	J	Nyankole	shatu
B	Kande	lato	K	Nyengo	ato
B	Galwa	ntʃaro	K	Mbwela	hatu
C	Bua	salu	L	Kete	sàcw
C	So	saso	S	Lozi	talu
C	Sakata	sâa	S	Venda	raru
C	Koyo	tsáro	S	Swazi	tsâtfu

Before comparing the Bantu data with the data from other NC languages, let's try to find out which processes in Bantu give way to such a diversity of phonetic variants (see schemes 1-3).
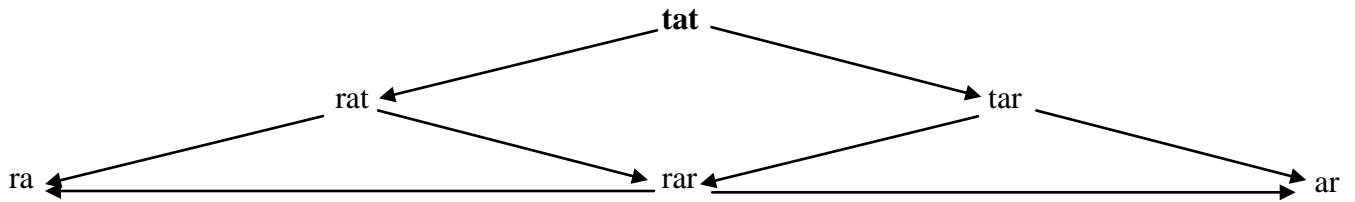
Scheme 1

-C   C-   C- -C   -C   C-

ta   at

Scheme 2

Схема 3


```
                              tat
                rat ←                 → tar
ra ←                         → rar ←                    → ar
```

Remarks to schemes 1-3:
The root contains two consonants. Putting aside the problem of the vowel in the second syllable, we shall call the two consonants C- and –C respectively. Each of them can drop, which gives way to the Bantu forms *ta* and *at* (scheme 1). Each of them can be transformed, for example, with a spirantisation *t > s*, or *t > r, *t > l*, can become voiced *t > d* and only after that may the drop of the second consonant have happened (schemes 2-3).

As a result, we have numerous sound forms, while the variation can be reduced to a limited number of processes:
- Voicing
- Palatalization
- Lenition – partial (spirantization) or full (transformation into Ø).

Below is an ordered table of derived forms in Bantu (without arrows):

-C	C-	C- -C	-C	C-
		tat		
ta				at
	sat		tas	
sa		sas		as
	cat		tac	
ca		cac		ac
	rat		tar	
ra		rar		ar
	lat		tal	
la		lal		al
	hat		tah	
ha		hah		ah
	dat		tad	
da		dad		ad
	zat		taz	
za		zaz		az

However, the resource for changes in Bantu is not limited to the above. The derivational schemes mentioned above are constructed not only on the base of *tat*, but also from newly derived forms. For example, **tat > sat*, and as follows (scheme 4):

Scheme 4

-C	C-	**C- -C**	-C	C-

tat
↓
sat

zat   sad

za   zad   ad
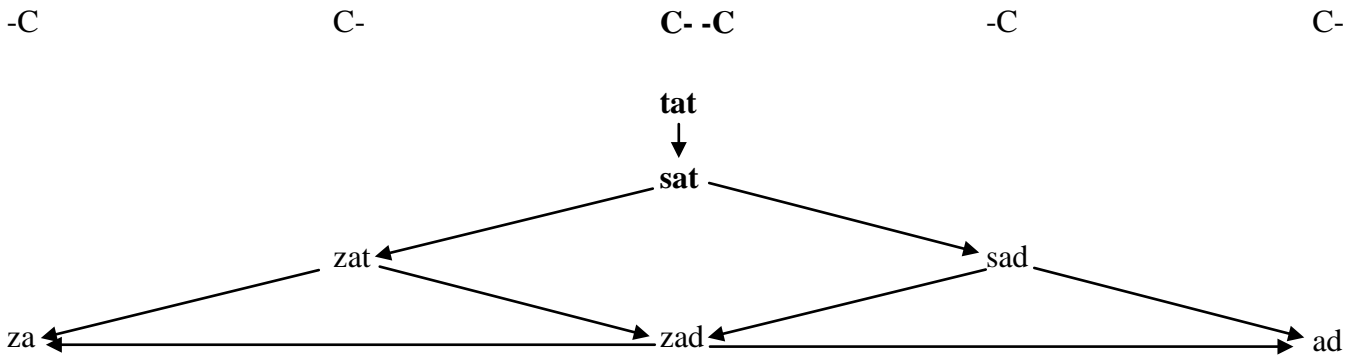
This is where the following forms, many of which are attested in Bantu, originate from (forms without square brackets):

	sat	cat	rat	lat	dat	zat
tas	**sas**	[cas]	[ras]	**las**	[das]	[zas]
tac	**sac**	**cac**	[rac]	[lac]	[dac]	**zac**
tar	**sar**	**car**	**rar**	[lar]	**dar**	[zar]
tal	**sal**	[cal]	[ral]	**lal**	[dal]	[zal]
tah	[sah]	[cah]	**rah**	[lah]	[dah]	[zah]
tad	**sad**	[cad]	**rad**	[lad]	**dad**	[zad]
taz	[saz]	[caz]	[raz]	[laz]	[daz]	**zaz**

Note:
We often do not know how one or another derived form appeared. E.g., form *las* in the first line of the table could have originated from **tas* (as a result of the change of the first consonant – the variation in the line) or from **lat* (the change of the second consonant – column). Many of the forms which are predicted theoretically are not attested in Bantu (they are given in square brackets).

The most amazing observation here is not the high degree of variation (which itself needs to be thought of), but the fact that we find absolutely the same variations in different branches of NC. As a result, in different branches of NC, that is, in languages with distant genetic relations, we find numerous identical forms while in every branch taken separately we find an "antimagnetic" landscape of forms which in closely related languages tend to be maximally differentiated. Examples from seven branches of NC are given below divided into two structurally identical tables:

	Bantu		**Adamawa**		**Atl.-Mel**	
TAT	Rundi	**tatu**	Yendang	**tat**	Fula	**tat-**
TAR	Yansi	**taar**	Bangunji	**taar**	Buy	**taar**
TAL	Lozi	**-talu**	Dadiya	**tal**	Gola	**tā̄ l**
TAD	Mbede	**-tadi**			Sereer	**tad-ak**
TAS			Kulaal	**tòòs**	Bapen	**ɓʌ-tas**
TAZ			Mom Jango	**tàáz**	Tanda	**-taaz**
TA	Nyo'o	**tá**	Tunya	**ta**		
SAT	Bushong	**-satu**	Kumba	**sa:t**		
SAR	Nzadi	**i-sár**				
SAS	So	**-saso**			Temne	**p̀é-sā s**
SA	Sakata	**i|sâa**	Mangbai	**bi-ssá-**		
AT	Nyengo	**-ato**			Nalu	**-at**
AR	Lundu	**-aru**			Kasanga	**-ar**
LAL	Fang	**lal**			Nyun	**ha-lal**
RAR	Venda	**-raru**			Sua	**-rar**
RA	Kpa	**-ráá**			Sherbro	**ra**
CAR	Orungu	**tʃaro**	Kam	**tshar**		
CA	Bubi	**-cha**	Galke	**cha-?a-**		
HAT	Nkoya	**-hatu**			Manjak	**go-hant**
DER					Baga Mb	**der**

	Bantoid		**BC**		**Dogon**		**Gur**	
TAT	Bankala	**tát**	Birom	**be-tat**	kolum so	**tã̀ãti**	Ditammari	**-tã̄ãtĩ**
TAR	Mambila	**tar**	Jiru	**i-tar**	bangeri-me	**ke-taro**	Senari	**tãre**
TAL	Kom	**tál**	Olulumo	**è-tál**	toro tegu	**taali**	Nateni	**tā lī̧, tā di**
TAD	Ngwe	**tád**	Upper-Cross	***-ttáD**	tommo so	**tadu**	Nateni	**tā di̧, tā lī**
TAS			ukaan	**tás**				
TAZ								
TA	Abon	**-ta**	Ibibio	**ì-tá**			Dagbani	**-ta**
SAT			Morwa	**sat**				
SAR	Mbe	**bé-sár**	Kugbo	**ì-sàr**			Lorhon	**sã:r**
SAS							Viemo	**saasi**
SA	Ekoi	**é-sá**	Oloma	**e-sa**			Kulango	**sã**
AT			Kohumono	**a-àtá**			Hanga	**ata**
AR								
LAL								
RAR			Abua	**ì-rààr**				
RA	Nkem	**í-rá**	Ukue	**è-rhá**				
CAR			Ufia	**kù-tshàr**				
CA			Bandawa	**ni-ca**				

We see, for example, that roots TAL and TAR are observed in all the seven branches.

To get a comprehensive idea of the presence of the forms is each of the 15 branches we are attracting attention to the following chart where the presence of the forms (at least in one language) is marked by a cross (the data is ranged in descending order in the summarising column as well as in the summary line):

	bantu	BC	atl	adam.	bantoid	gur	mel	kwa	ubangi	dogon	Kordof.	kru	ijo	Togo	mande	
TA	x	x		x	x	x		x	x		x	x		x		10
TAR	x	x	x	x	x	x			x	x			x			9
TAT	x	x	x	x	x	x	x			x						8
TAL	x	x		x	x	x	x		x	x						8
TAD	x	x	x		x	x				x	x					7
SA	x	x		x	x	x		x							x	7
AT	x	x	x			x		x			x					6
RA	x	x			x		x		x							5
SAR	x	x			x	x										4
SAS	x		x		x	x										4
LA	x	x							x			x				4
TAS		x	x	x												3
SAT	x	x		x												3
AR	x		x					x								3
HAT	x		x								x					3
RAR	x	x	x													3
CAT	x	x			x											3
CAR	x	x		x												3
TAZ			x	x												2
HA			x					x								2
LAL	x		x													2
DAT	x	x														2
CA	x			x												2
SAL	x															1
AL	x															1
AS						x										1
HAH	x															1
THAT	x															1
TSAR	x															1
RAH						x										1
DAR			x													1
TAH		x														1
TAC		x														1

DAD	x															1
DAZ						x										1
RAT					x											1
RAD	x															1
LAT	x															1
LAS	x															1
SAD		x														1
SAC	x															1
CAC	x															1
ZA								x								1
ZAC			x													1
	31	19	14	10	10	10	6	6	5	4	4	2	1	1	1	124

The following chart represents the number of groups (within the 15 groups of Niger-Congo) presenting the respective combinations of the first (the line) and the second (the column) consonants (the data is given in descending order):

	Ø	t	r	l	d	s	c	h	z	
t	10	8	9	8	7	3	1	1	2	49
s	7	3	4	1	1	4	1			21
c,ts	3	3	5				1			12
Ø		6	3	1		1				11
r	5	1	3		1			1		11
l	4	1		2		1				8
h	2	3							1	6
d		2	1		1				1	5
z	1						1			2
	32	27	25	12	10	9	4	3	3	125

As we can see, the most frequent consonants in the initial position are *t-* and *s-*, while the second consonant is one the the three: *-Ø, -t*, or *-r*.

If we reconstruct *tat-* on the NC level, following the majority of linguists, we will have to deal with quite a mysterious picture. In the majority of younger proto-languages we will also have to reconstruct *tat-*, because, as it has already been shown, it descends into more or less the same variation of forms. It means that during thousands of years from Proto-NC to the formation of proto-languages of separate branches the form remained phonetically unchanged and then suddenly the root *tat* independently started to explode giving reflexes in all possible phonetic variations.

I think a hypothesis that already in NC the root contained close but not identical consonants is much more typologically justified. The second consonant in that case was *-t,* while the first one was represented by a specific phoneme which we have no traces today, for example, *th- ?, *ṭ- ?,* ṭs- ?,*c- ?* As we tried to show in [Pozdniakov-

Segerer, 2007], the phonotactics of many languages (not exclusively in Africa) demonstrates a tendency: in CVC structures languages tend to avoid consonants constituting a minimal pair, for example, *fVp, bVp, sVz, lVr, rVl, sVsh, etc*. In the diachrony the existence of such combinations often leads to numerous irregular changes, in the course of which the consonants either become identical, for example, *lVr > lVl,* or, on the contrary, acquire a higher level of contrast, escaping the zone of "dangerous proximity", for example, *sVsh > sVh, *bVp > bVf.* In other words, too similar sounds being adjacent to one another are a constant zone of tension which provokes all possible irregular changes.

It is very likely that such a situation characterises the NC root for 'three'. In this case the considerable phonetic variability of the root on all the stages of its development from Proto-NC to contemporary languages can be typologically – phonotactically – explained.

Les us turn now to other forms for 'three' which we find in Atlantic and Mel languages.

In Jaad-Biafada we find the root *jow / caw*. This is undoubtedly an innovation in the group which is represented by a remarkable isogloss which is an argument in favour of interpreting this group as a part of the northern branch of Atlantic family:

Biafada	-njo / -j. / -j.	jo / -j. / -j.
Biafada	bíí-co	co
Biafada	bií-yo	yo
Jaad	ma-cao	ma-cao
Jaad	ma-caw, má-dṣọu	ma-caw, ma-jọu
Jaad	má-cɔu	ma-cou

It is possible that we deal with an ancient borrowing of Proto-Jaad-Biafada from Mande (from *saba* 'three').

Borrowing from Mande is much more evident in the Balanta language where "three" is expressed by forms –*habi, yabi, jaabi, aabí*.

In theory, it is possible that forms attested in Cangin languages (*ka-hay / * ʔe-jɛʔ*), also originated from the Mande form (probably, already weakened to *habi / hawi*).

In this case, in numerous Northern Atlantic languages we find either reflexes of the Proto-NC form *THAT, or borrowings (taking into account the forms, very ancient) – from Mande languages.

On the contrary, in Central Atlantic languages we attest several innovations:

1) In Joola languages (including Bayot and Karon) numerous reflexes of a form which can be reconstructed as *feegi* are attested:

9

	« three »
Bayot	ʔi-fi:giʔ, figꞌi, fíːɣí, fɩjɩ, fɩʑɩ, i-fəəʒi
Joola banjal	gu-fiigir, siꞌ-fʰəɟi, si-foːɟi , si-fɤɤji
Joola bliss	sɩ-həːjɩ, haajut
Joola ejamat	si-həːjɩ, si/ku-hɤɤji
Joola fogny	si-feːgi, si-feegiir, sɩ-feːjɩ, sɩ-həːjɩ, ku-həəji, feejɩɩr
Joola gusilay	si-feegir, sifːəɟi, gu-fəəjir
Joola karon	hääji,sɩ-həːc, si-həəc, si-həəciil
Joola kasa	si/ku-həəji, si/ku-hɤɤjiʔ, hääji(s), -höji
Joola kerak	si-həəji, si-hɤɤji
Joola kwaatay	ki-əgi, si-həəji, háaji, ki-hɤɤjiʔ, höji
Joola mlomp	si-feːgi, si-hɤːɟil, si-həəjil

In other Atlantic languages this root is not attested in the meaning 'three' but is most probably related to the root *faaj / paaj* meaning 'six', which is attested in the other branch of the Bak group, that is, in Manjak-Mankany-Papel, as well as in the Balant language (the Central group) and in one language of the Northern branch, Biafada, which is in contact with the Bak languages.

	« six »
Balant kentohe	mfaacny
Balant-fca	faatʃ, fáaj, faac
Balant-ganja	faaj
Manjak	pã̄ gɟ paaj,pã̄dʂi,pááci, pay
Manjak-Bassarel	paaj
Manjak-Churo	paaẓ
Manjak-Pecixe	paay
Manjak-Tame	páági, pã̄ dʂɟ pááci
Mankany	paaj,paadj, pã̄ dʂɟpááci, paaḏ, pay, padʒɨ
Pepel	paaj, paaɟ, pã̄ dʂ páác
Biafada	mpáági, mpaaji, paji

2) The numeral "three" in the Manjak-Mankany-Papel group is expressed by a specific root which doesn't have parallels outside this group –*wants*:

Manjak	wants, waanṯ, ku-áánt, go hant, waa-ẓantʂ, wa-yintʂ, wóó-yant
Mankany	wa-yants, kó-ā yᶒntʂ wa-jenṯ, wajanṯ, ŋ̀ɨwà-dʒɛnt
Pepel	waa-jintʂ,ŋa-jens, n·gã̄-dʂint, -waajinṯ, ngáácint

In this case, the isogloss for 'three' is a good example of an innovation in the proto-language of the subgroup Manjak-Mankany-Papel.

At the same time, the situation is extremely complicated. There are many factors (which we will not list here) which complicate any possible hypothesis of the origin of the given forms. The most plausible reconstruction of their development can be deduced to the following scenario:

A. Obviously, the proto-language of the Central group (after the separation of the Nalu subgroup from the proto-Central) created an innovation for 'three': the Niger-Congo root *that* was replaced by the root *feegi / faagi/ faaj* the origin of which is unclear. In this period numeral 'six', along with the inherited model 6 = 5 + 1 (in the Proto-Diola dialectal zone), becomes associated with the model 6 = Class Marker for paired body parts (for instance, 'hand') + 3.

B. In Balanta, the Bak root for 'three' is replaced by a borrowing from Mande languages: in the meaning '6' this root in preserved. In proto-Manjak-Mankany-Pepel the Bak root for 3 is replaced by an innovation – the *wants* root, in the meaning '6' this root is also preserved. At the same time, in Balant as well as in Manjak-Mankany-Papel (that is apparently already in proto-Bak) the numeral 6 acquires a specific role which determined the emergence of rare models: 7 = 6 + 1, 8 = 6 + 2. In Biafada, the numeral 6 in borrowed from Manjak or from Balant. With all its complication, this reconstruction appears to be the most plausible.

3) In the Central branch, in the Bijogo language, we have one more innovation - *ɲ-ɲɔɔkɔ.*

Concluding the overview of forms for '3' in Atlantic and Mel languages, we shall cite the form in Southern Mel languages - Kisi and Krim – which competes with reflexes of the Niger-Congo form *that* : Kisi *yàá, ŋàá, ma* ? ~ Krim *yì-g(h)a.*

Neither of these forms has systematic parallels outside Atlantic languages.

TWO

The major challenge of the reconstruction of 2 in NC is related to the peculiarities of the segmentation and resegmentation of the lexical root. In general, the problem of the change of morphemic boundaries is the most important challenge for the reconstruction of the lexicon of NC, a language with nominal classes.

For Proto-Bantu, the reconstruction of the segmental root *bidi* or *badi* is fairly reliable. Les us examine several reflexes of this root in some Bantu languages considering numerals 3 and 4:

proto-bantu	proto-bantu	two	three	four
		bàɗɪ,bɪ̀ɗɪ, bɪ́ɗɪ	**tátʊ̀, cácò**	**nà,nàí,nnè,nèĵì**
bantu-E	Logoli	vi-viri	vi-vaga	vi-ne
bantu-E	Gweno	vi	raru	na
bantu-E	Kahe	shi-vi	shi-radu	shi-na
bantu-E	Kamba	i-li	i-tatu	i-nya
bantu-K	Luchazi	k-ari	k-ato	-wana
bantu-A	Bankon	bá!á	bí-á	bí-nán`

In closely related languages we attest different strategies of transformation of the proto-language form for 2:

1) a class prefix is added to a root – class *vi-* for all three numerals in Logoli, class *shi-* - in Kahe;
2) the proto-language root is reduced and its initial consonant is reinterpreted as a class marker - in Bankon, class *bi-* in all three numerals;
3) the initial consonant of the root is omitted and the root vowel is reinterpreted as a class prefix – in Kamba, class *i-* for all three numerals;
4) the initial root consonant undergoes an analogical change in favor of the consonant of the new class prefix – in Luchazi *k-ari* 'two' < **bari, k-ato* 'three' < **tato*,
5) the second syllable disappears – probably, influenced by the coincidence of the first syllable and the class prefix: in Gweno - **vi-viri > vi-vi > -vi* ?

All these different techniques of root reinterpretation often take place in the course of analogical change in phonetics or morphology of numerals 2-4.

In Atlantic and Mel languages, having a much bigger distance than between Bantu languages, we attest even more diverse root reinterpretation techniques with even more catastrophic consequences for a comparativist, because they open up an almost unlimited field for etymological comparisons while what we would like to get is quite the contrary. Let us consider several examples:

		two	three	four
atl-sua	Baga fore	si-di. si-li	si-tet	si-ne, -nəŋ
atl-center?-nalu	Nalu	bɛ-lɛ	pw-aat	bii-naɲ
atl-limba	Limba est	bi-le	bi-tat	bi-naŋ
atl-north-peul	Fulfulde	ɗiɗi	tati	nay
mel-north	Baga koba	pa-rã	pa-sas	pa-ŋere

In Nalu and Limba, we attest as a matter of fact a Proto-Bantu root – with the only difference: the first syllable is synchronically a class marker here. At the same time in Limba we can notice a systematic tendency of unification of class markers of numerals 2-4, that is, the same strategy that is manifested in numerous Bantu languages. This is

where probably the Nalu form comes from, *bile > bi-le*, by analogy with *bi-tat* and *bi-naŋ*. In Baga Fore, it is quite probable that *bili > si-li* is analogous to *si-tet* and *si-ne,* while in the Mel group, in Baga Koba in particular, *barã > pa-rã* is analagous to *pa-sas* and *pa-ŋere*.

On the other hand, an opposite direction of the proto-language evolution can also be suggested in the Niger-Congo context. The root *BADI / BIDI* is Proto-Bantu, but not necessarily Proto-Niger-Congo. If we assume that the proto-Niger-Congo root was *DI*, we will have to suppose that already in Proto-Bantu the ancient root was reinterpreted and a class prefix - of class *2? or *8? - was integrated into the root. In this case Atlantic roots considered above as well as Bantu forms can be compared to the Fulfulde form *ɗiɗi* (possibly, with a reduplication triggered by a pseudo-reduplicated root *tati*), as well as with numerous forms in other branches of NC-languages, including the following:

adamawa	Day (Buna dialect)	ɗīí
adamawa	Niellim, chad	nɗīdí
adamawa	Tunya (tunia)	àrī
adamawa-bua	Day	di-í
adamawa-bua	Day	ɗīí
adamawa-bua	Koke	le-di
adamawa-bua	Nielim	ndiri
adamawa-bua	Tunya	ali
atl-c-bak	Bayot	i-rigə
atl-c-nalu	Baga fore	di,li
atl-c-nalu	Baga fore	-di, -li
atl-limba	Limba	le
atl-limba	Limba est	le
atl-n-peul	Fulfulde	ɗiɗi, sogoro
atl-n-peul	Sereer	d`ik
atl-s-kisi	Kisi nord	dìíŋ, *le
atl-s-tem	Temne	pĕ-rəŋ
bantu-F	Nyilamba	-ele
bantu-G	Kagulu	ili
bantu-H	Suku	-óódi
bantu-H	Yombe	-o:le
bantu-R	Kwanyama	-ali
bantu-R	Ndonga	-áli
bantu-NW-A	Yambasa (Nugunu)	-àndé
bantu-NW-B	Bali	-ole
bantu-NW-B	Duma	ɔlɛ
bantu-NW-B	Mbama	-ele
bantu-NW-B	N teke	-ele
bantu-NW-B	Ndasanord	ɔlɛ

bantu-NW-B	Njebi	-oli
bantu-NW-B	Tiene	-elé
bantu-NW-B	Tsaangi	ɔlɛ
bantu-NW-B	Wandji	ɔlɛ
bantu-NW-B	Wumbvu	ɔlɛ
dogon	Dogon	lèy
dogon	Donno so	lɛy, le
dogon	Jamsay	ley, leiy
dogon	Toro tegu	ley,lei
dogon	Toro-so	ley
gur	Bimoba	–lè
gur	Chakali	álìɛ
gur	Ditammari	-dyá, ɗɛ́ɛ, diání
gur	Gurma	-lié
gur	Konkomba	–lèe
gur	Nateni	-ɗɛ́ɛ, ɗɛń
gur	Ntcham	-lí, ǹ-léé
gur	Yom (pila)	-li
gur-central	Oti-volta+kurumfe	*yi
gur-central	Oti-volta+kurumfe	*ɗe
gur-central-oti-volta	Basari (ntcham)	-lí
gur-central-oti-volta	Hanga	ayi
gur-central-oti-volta	Kusal	ayi
gur-southern	Sisala	lia
gur-southern	Siti	are
gur-southern	Tampulma	ale
kwa	Akebu, togo	jí
kwa	Aladian	aire

In any case, Atlantic forms for "two" listed above can be traced back rather to *DI[1] than to *BIDI.

Finally, let us cite some forms of NC-languages (beyond Bantu) which are directly comparable with those in Proto-Bantu – in this case it is not important if they are traced back to a class marker with a labial consonant and the root *DI* or need to be separated from the reflexes of *DI* and compared directly to the root in Bantu:

---

[1] Here we confine ourselves to indicating a hypothetical form of Niger-Congo having in mind that on the proto-Atlantic level the most plausible proto-forms are *DIK / DAK* with a velar final consonant. The basis for this reconstruction will be considered below.

adamawa-daka	Dirrim	bara
adamawa-daka	Gandole	barà
adamawa-daka	Taram	bara
adamawa-mbum-mundang	Mangbai	bati
adamawa-yungur	Yungur (bena)	fIte
bantoid-ekoid	Abanyom	bí-bâl
bantoid-ekoid	Nkem	í-bâl
bantoid-ekoid	Nkumm	í-bâl
bantoid-ekoid	Nnam	éb-bál
bantoid-grass?	Viya	bae
bantoid-jarawam	Bankala	b`àrí
bantoid-jarawam	Jaku (labir)	b`ár^
bantoid-jarawam	Mboa	ba:ï
bantoid-jarawam	Mbula	bari
bantoid-jarawam	Nagumi	ba:li
bantoid-jarawan	Kulung	b`a‖àli‖
bantoid-bamileke	Bafut	bá!á
bantoid-bamileke	Mankon (ngemba)	-bàé‖
bantoid-bamileke	Mbem (kaka)	bar
bantoid-bamileke	Ngie	ubie
bantoid-bamileke	Nkwen	bí!é
bantoid-nkambe	Limbum	bá!á
BC	Tchitchege	byele
BC-eastern-cross-delta	Eleme	òbèrè
BC-eastern-cross-delta	Kana	bàè
BC-eastern-cross-delta	Kugbo	ìwàl
BC-eastern-cross-delta	Ogbia	ìwàl
BC-eastern-kainji-western	Dakarkari	ìllè
BC-eastern-plateau	Bashar (yangkam)	bar
gur	Akaselem	mbìlé
gur	Bago-kusuntu	bààɭɛ
gur	Konni, ghana	àbɛ́lí / àɭ̃i
gur	Kulango	bíla
gur	Nuni northern	bìɭə
gur	Nuni southern	b`əɓ̃
gur-central	Buli	baye
gur-kulango	Kulango	bila
gur-southern	Lyele	bie
gur-southern	Nuni	balya
mande-proto	Mande-sud	FILA
mande-bobo	Sya	pila

mande-n	Jalonke	fíla
mande-n	Kono	fela, féla
mande-n	Susu	Firing, fíriŋ
mande-n	Vai	féla
mande-n	Yalunka	firin
mande-n-boz	Bozo-sorogama	pẹ de
mande-n-boz	Bozo-sorogama	pẹ ne
mande-n-boz	Bozo-sorogama	pẹ nde
mande-n-boz	Bozo-tiemaciewe	pẹ de
mande-n-boz	Bozo-tigemaxo	pẹ ɗe
mande-n-boz	Soninke	fillo, filo
mande-n-manding	Bambara	fíla
mande-n-manding	Dyula	flà
mande-n-manding	Malinke	fùla
mande-n-manding	Mandinka	fùla, fíla
mande-n-manding	Maninka (Kankan)	fíla
mande-n-manding	Xasonke	fila
mande-sud	Bisa	piiya
mande-sud	Busa	fla-
mande-sud	Dan	plè
mande-sud	Dan(yakouba)	péérɛ
mande-sud	Mano	pile, *péérɛ
mande-sud	Mwa	ple
mande-sud	Nwa (wan)	pilong
mande-sud	Tura	pìile-
mande-SWM	Bandi	fééle, *féla
mande-SWM	Kpelle	fele
mande-SWM	Kpelle de Guinée	pféérɛ, *féérɛ
mande-SWM	Kpelle du Liberia	felɛ
mande-SWM	Loko	fele, félee
mande-SWM	Loma	felego
mande-SWM	Loma	fééle
mande-SWM	Mende	felé
mande-SWM	Mende	féle
mande-SWM	Proto-SW-Mande	*fere

The following "sad" circumstance needs to be noted. If we didn't know anything about the classification of African languages and only analysed forms, we would have to compare NC forms listed above, including those in Bantu, with numerous similar forms in Chadic languages (cf. in Kofyar: *mél* 'one' – *vel* 'two' !), for example:

Ankwe (Goemai)	vìl
Chip (Miship)	vil
Daffo	fùl
Galambu	mbàal|
guus(sigidi)	mbə̀ɬì
Kirfi (Giiwo)	mbàlú
Pero	pelèyò
Ron (Bokkos)	'apì`l
Sura (Mwaghavul)	ful
zaar	mbə́ɬə̄ŋ

And finally, some Nilo-Saharan languages also demonstrate forms phonetically similar to the root *DI* as well as to *BADI / BIDI*:

nilo-saharian-east-sudanic	Erenga	war
nilo-saharian-east-sudanic	Kelo	wa:ti
nilo-saharian-east-sudanic	Merarit	wárre
nilo-saharian-east-sudanic	Tama	wari
nilo-saharian-komuz	Gojjam	mba:nd
nilo-saharian-komuz	Gumuz (Bega) Sai	mband
nilo-saharian-komuz	Kokit	mba:nda
nilo-saharian-kunama	Kunama	'bárè
nilo-saharian-maban	Kibet	mbaar
nilo-saharian-maban	Maba	mbàar
nilo-saharian-maban	Masalit	mbárá
nilo-saharian-central	Kaliko	írì
nilo-saharian-central	Logbara	ïrrì
nilo-saharian-central	Logo	írì
nilo-saharian-central	Avukaya	ärrì
nilo-saharian-central	Kenga	di:ió
nilo-saharian-central	Madi	rì
nilo-saharian-central	Moru	ärrì
nilo-saharian-central	Vale	dio
nilo-saharian-east-sudanic	Nera	arriga
nilo-saharian-nilotic	Akoli	aryo
nilo-saharian-nilotic	Alur	áryo|
nilo-saharian-nilotic	Anuak (Anywa)	arrio
nilo-saharian-nilotic	Burun	areo
nilo-saharian-nilotic	Lango	àryô
nilo-saharian-nilotic	Luo	ariyo
nilo-saharian-nilotic	Maasai	aare
nilo-saharian-nilotic	Shilluk	áryùu

nilo-saharian-nilotic	Teso	aarei
nilo-saharian-saharan	Kanuri	ndi

In conclusion, let us go back to the presumed Atlantic reflexes of the NC root *DI* and discuss the problem of its reconstruction for the proto-Atlantic level.

atl-centre	Bayot	i-rigə
atl-centre	Bijogo	n-dank, ruŋ?
atl-centre	Manjak-Mankany	ke-taw?
atl-centre	Nalu-BF-BMb	si-di,si-li, bi-le, sə-loŋ
atl-limba	Limba	bi-le
atl-nord	Cangin	ka-nak
atl-nord	Buy-nyun	naŋ, nak
atl-nord	Fula-sereer	ɗik-ɗik, ɗaq
atl-nord	Wolof	ñaar
atl-mel	Kisi-bullom-sherbro	dìíŋ, ŋìíŋ, t(s)iŋ, tring
atl-mel?	Sua	m-cen
atl-mel	Baga-temne-landuma	pa-rəŋ, mɛ-rəŋ
atl-mel?	Gola	tśé ʃ tȉel, cel

The root is widely represented not only in Atlantic languages (in Central, Northern branches and in Limba), but also in all Mel languages. In all North Atlantic and Mel languages its structure is CVC. In Central languages, together with *di / li* root, a root of CVC structure with a final velar consonant is also attested. In Atlantic languages and Mel, these forms could have developed from *di(n)k / da(n)k*. Taking into account the complete absense of forms with a final velar outside Atlantic (cf., however, a possible interpretation of Adamawa data at the end of the paper), we can presume for Proto-Atlantic a phonotactic transformation of the NC root caused by the formation of a root with CVC structure: *NC *di* > Atlantic *dik*.

This reconstruction (as well as any other made at the current level of development of NC studies) is inevitably dubious, in particular with respect to the possible inclusion of the forms with initial nasal consonants into the set – in Nyun-Buy and Cangin, as well as in Wolof. Moreover, the form for 2 in Nyun virtually coincides with the form for 4. Les us consider these problems in more detail.

TWO + TWO = FOUR?

In Proto-Bantu the reconstructed roots for 4 are *nà, *nàı́, *nnè, *nèjì*. This is one of the most stable roots in NC which has numerous parallels in almost every branch of NC. The problem is that a similar root is widely spread for numerals meaning 'two', including Atlantic languages, where it is attested in Nyun languages. In Proto-Nyun, the form for 2 is *ha-nak, i-nak*, the form for 4 is ***re**-nak > * ha-**re**-nak, i-**re**-nak > * ha-ren**ek**, i-ren**ek**. Therefore, the form for 'four' can be considered a plural form derived from 'two'.

However, it seems to be an illusion which is formed in the course of submorphic adjustments of "two" and "four", and more often, of "two", "three" and "four". Different mechanisms of this adjustment as a result of analogical change can be demonstrated on the Adamawa material. Citing the paradigms of numerals in Dii and Dugun languages, Lars Lode (14 November, 1994, http://lingweb.eva.mpg.de) formulates a questionable hypothesis that in these languages, 4 < 2 x 2?, basing on the following data:

	two'	four'
Dii	idú	ndaddʉ́ (2 x 2) ?
Dugun	irú	ndaró (2 x 2) ?

However, the comparison of these forms with corresponding forms in Adamawa demonstrates that most probably we deal with an analogical alignment. Let us cite the numerals from 2 to 4 in some Adamawa languages:

	two'	three'	four'
*adamawa	***do/du**	***tat/tar**	***nar/nat**
Dii	i-dú	tã̃ãnɔ́	nda-ddʉ́
Dugun	i-rú	tã̃ãnó	nda-ró
Kutin (Peere)	i-ro-	tu-u-re-	na-a-ro-
Sambaleko	ʔīī-rā̄ ~ ʔīī-rə̄	tō ө̄rə̄	nā̄ ā̄rā̄
Peere	i-ro	tã̃ã-ro	na-ro
Samba Leko	ii-rà	too-rà	naa-rà
Wom	i-ra	ta-ra	na-ra
Dirrim	b-ara	t-ara	n-ara
Taram	b-ara	t-ara	n-ara
Jenjo (Dza)	bw-əng / bwa-yung	bwa-tə	bwa-nyə
Duupa	ittó	tã̃ãtó	nattó
Gimme	idti-gè	taa-gè	náà-gè
Mumuye	ziti	ta:ti	ɗɛ̀:tì

In all languages listed above, with the exception of the first two, the numerals from 2 to 4 have a common feature while numerals 1 and 5 do not have it. In Peere, Samba and Wom the unification of the three forms is more transparent, in Dirrim and Taram it is attested at an even wider scale – synchronically, the real bearer of the numeral meaning is the initial consonant: *b-* '2'*, t-* '3', *n-* '4'. In Jenjo, on the contrary, the common initial consonant in the three forms is related to the class prefix, which characterises these particular numerals. In the last three languages, though, it is the final elements of the forms that become unified: in Gimme we deal with a suffix (which is attested in the form of "five" - *nɔngè*), while in Mumuye and Duupa, on the contrary, this time root elements are unified. All these diverse adjustments can create an illusion that formal similarities attested here have a derivational character.

Leaving aside a detailed analysis, let us make a general conclusion: the abovementioned Nyun form for '2' with an initial *n-*, as well as many other similar forms of NC languages, are apparently not original and appeared as a result of analogical change in parallel with the numeral '4' or are reflexes of forms merged with a nasalized noun class prefix. A bright example of this kind are forms 2-4 in Wolof: '2' – *ñaar*, '3' – *ñett*, '4' – *ñeent* (< *ñenent*). Initial *ñ-* in this language reflects an original plural noun class prefix denoting humans (cp. *nit ki* 'the person' – *nit ñi* 'these persons') which replaced an original root consonant at least in numerals '2' and '3'. This example is another evidence of the fact that the analysis of the forms of numerals '2-4' in some cases plays a crucial role for the reconstruction of plural human noun class markers in NC languages – very often numerals 2-4 display reflexes of the noun class 2.

Taking into consideration the evidence described above, the forms in Nyun and Wolof should be considered as reflexes of proto-Atlantic *dik / dak*.

Apart from the root already examined, three (?) following roots are attested in Atlantic languages:

centre-bak	joola fogny	sɪ-gaba		
centre-bak	joola banjal	sɪ-gaba	si-rubə	
centre-bak	joola ejamat	sɪ-gaba	si-lubə	
centre-bak	joola karon		si-subə	
centre-bak	joola bliss		si-lubə	
centre-bak	joola kasa		si-ɬubə	
centre-bak	joola gusilay		si-rubə	
centre-bak	joola kwaatay		ki-subə	
centre-bak	joola mlomp		si-subəl	
centre-bak	joola kerak		si-subə	
centre-bak	manjak		kə/gi-təb	
centre-bak	mankagne		nə-təb	
centre	balant kentohe		k-sib-m	
centre	balant-fca		sibi	
centre	bijogo		n-som,sòòbɛ́, súŋgb	
nord	tenda			kí / xí
nord	biafada			bi-he, ŋ-ke
nord	jaad			maa-ɛ

The roots with an intervocalic labial consonant are widely attested only in languages of the Central Atlantic group, while the root *ki / hi* can be considered as an innovation of the proto-language of the branch Tenda-Jaad-Biafada of the Northern group.

The origin of the latter root is unclear. Note in Basari (Tenda) a very probable relation of "two" with the exclusive 1pl pronoun: 1pl S: *-kɛ́...ɛ́* and/or the relation with *kɛɛ*

'another' in Diola languages. Outside Atlantic family this form for the numeral 2 is not attested, apart from the form *hi* in Kurumfe which is marginal even for Gur languages.

The roots which are phonetically close to Gaba, are attested only in some Adamawa and Benue-Congo:

adamawa	Mundang, Chad	gwa, gwa*
adamawa-mbum-mundang	Dama	ga?a
adamawa-mbum-mundang	Mono	ga
adamawa-mbum-mundang	Pam	ga?a
BC-eastern-kainji-western	Basa-Komo	-jebi
BC-eastern-kainji-western?	Bassa	jewi
BC-idomoid	Nupe	guba
BC-igbo	Ogbah	gwebo

The second root is attested only in some NC languages but is widely attested in Chadic (so widely that it is apparently reconstructed for Proto-Chadic), as well as in Nilo-Saharan. The material we have in hand is given below:

adamawa	Burak, Nigeria	rab
adamawa-waja	Tula	rop
adamawa-yungur	Roba	rap
bantoid-jarawam	Jarawa	rap
BC-eastern-kainji-western	Gurmana	e-ribu
kordofanian-heiban	Ebang (Heiban)	ram
kordofanian-heiban	Laru	rom
kordofanian-heiban	Logol	rab
afro-asiatique-chadic	Boghom (Burma, Bux)	rap
afro-asiatique-chadic	Buli	rowa
afro-asiatique-chadic	Dira	rop
afro-asiatique-chadic	Dwot (Dass)	rop
afro-asiatique-chadic	Guruntum	rab
afro-asiatique-chadic	Jara	rop
afro-asiatique-chadic	Kanakuru (Dera)	rap
afro-asiatique-chadic	Kariya	ra?a
afro-asiatique-chadic	Luri	rop
afro-asiatique-chadic	Pidhimdi	rap
afro-asiatique-chadic	Polchi (Pelci)	rop
afro-asiatique-chadic	Posi(polci)	rop
afro-asiatique-chadic	Tera	rap
afro-asiatique-chadic	Tsagu (Ciwogai)	ra?a
afro-asiatique-chadic	Warji	ra?a

afro-asiatique-chadic	Zeem	rapi
afro-asiatique-chadic	zodi(dott)	rop
afro-asiatique-chadic	zodi(dott)	rop
nilo-saharian-central	Kresh (Gbaya)	romo
nilo-saharian-east-sudanic	Bale	rama
nilo-saharian-east-sudanic	Me'en	rama
nilo-saharian-east-sudanic	Murle	ram
nilo-saharian-east-sudanic	Mursi	raman
nilo-saharian-east-sudanic	Suri	raman
nilo-saharian-nilotic	Agar	row
nilo-saharian-nilotic	Dinka	rou
nilo-saharian-nilotic	Rek	rou

Naturally, there is no reason to presume any direct influence of Chadic or Nilo-Saharan on Bak languages. The most plausible supposition in this situation will be that in Atlantic languages we don't deal with two roots but with one and only in the following forms: *si-ga-ba* and *si-(ru)-ba-I*, that is, here is another case of restructuring of the root and with a formal pre-prefixation which is characteristic for Bak languages (even the word 'Bak' is an example of restructuring of the root and its pre-prefixation). In such a case we have numerous new reflexes of the major "candidate" for being reconstructed on the proto-NC level for the numeral 2, *ba-di / *bi-di*, in which the original root is practically replaced by the original noun class prefix which attracts new prefixes. This scenario can possibly explain the "isolated" root in the Bom language of the Mel group: *bâ* '2'.

ONE

In Proto-Bantu, a basic form *mu-oti-(ga)* is reconstructed, where *mu-* is a class marker, -*ga* is probably a suffix and *oti* (*odi*?) is the root (cf., in particular, Vanhoudt).

Different variants of reconstruction represented in the Tervuren database can be examined on the basis of the following dialectal changes in the proto-language:

Mu-oti-(ga) > modi > moi
Mu-oti-(ga) > moci / moca > moi
Mu-oti-(ga) > mUI, mUega, mUe, moi, mo

Reflexes of these forms can be found in all Bantu zones, including zones A, B, C. Here are some examples:

*moti

zone	language	form
bantu-NW-A	Tunen	-mòte
bantu-NW-A	kwakum	mɔ́tù
bantu-NW-A	Mande	móót
bantu-NW-B	Kande	moti
bantu-NW-B	Tiene	-móte
bantu-NW-C	Aka (Yaxa)	-mótí
bantu-NW-C	Bati	moto
bantu-NW-C	Bua	moti
bantu-NW-C	Lokele	-o-mwito
bantu-NW-C	Ngundi	e-moti

*moci

zone	language	form
bantu-NW-A	Ki	o:mwáàsi:í
bantu-NW-B	Duma	-mosi
bantu-NW-B	Duma	mosi
bantu-NW-B	Enenga	omori
bantu-NW-B	Galwa	mɔri
bantu-NW-B	Mbama	-mosh'
bantu-NW-B	Mpuono	mosə
bantu-NW-B	Myene	-mori
bantu-NW-B	Ngubi	mɔsi
bantu-NW-B	Orungu	mɔri
bantu-NW-B	Pinji	mɔ́tsì
bantu-NW-B	Punu	imosi, yimoosi
bantu-NW-B	Sangu	moosi
bantu-NW-B	Sira	ɣi-moosi
bantu-NW-B	Wandji	mosi
bantu-NW-C	Bushoong	mmɔ́cy
bantu-NW-C	Bwela	e-mosi
bantu-NW-C	Ngombe	-mosi

*moyi

zone	language	form
bantu-NW-A	Shiwa	mʷɛy
bantu-NW-A	Lue	mw-a:
bantu-NW-A	Nyo'o	mò:
bantu-NW-A	Yambasa	-mùè
bantu-NW-B	Bali	-mo
bantu-NW-B	Boma	mɔy

bantu-NW-B	Buma	möy
bantu-NW-B	C teke	-mu
bantu-NW-B	Duma	mɔ
bantu-NW-B	E teke	-mu
bantu-NW-B	Kaningi	mɔ
bantu-NW-B	Kaningi nord	mo
bantu-NW-B	Lempini	mɔ
bantu-NW-B	Mbama	ɔmɔ
bantu-NW-B	Mbede	-mo
bantu-NW-B	N teke	-mu
bantu-NW-B	Ndumu	-mo, mɔ
bantu-NW-B	Ne teke	-mu
bantu-NW-B	Njebi	-mo, mɔ
bantu-NW-B	Nzadi	mmɔ́
bantu-NW-B	S. Teke	mo
bantu-NW-B	Tekenord	mɔ
bantu-NW-B	Tekenord	mɔ
bantu-NW-B	Tsaangi	mɔ
bantu-NW-B	W teke	-mu
bantu-NW-B	Wandji	mɔ
bantu-NW-B	Wumvu	mo
bantu-NW-C	Foma	-mo
bantu-NW-C	Kusu	o-mo
bantu-NW-C	Mbole	-moye
bantu-NW-C	Mongo	-mõ
bantu-NW-C	Ntomba	mô
bantu-NW-C	Ombo	-mô
bantu-NW-C	Pande	-mo
bantu-NW-C	Sengele	-no
bantu-NW-C	So	-omwi
bantu-NW-C	Songomeno	-omo

*moti-ka

zone	language	form
bantu-NW-A?	ngul	mɔtuga, mɔg
bantu-NW-C	Bushong	moko
bantu-NW-C	Kela	o-mako
bantu-NW-C	Lingala	moko
bantu-NW-C	Nkutu	moko

In Bantoid, the same root, with the same prefix *mu-* is attested equally widely. Reflexes from major Bantoid groups are cited below:

Isu	mɔ̀ʔ
Kuk	mɔ̀ʔ
Mmen	mɔ̀ʔ
Vengo	mṳ̀ ʔ
Weh	mɔ̀ʔ
Wushi	mʷɔ̀ɔ̀ʔ
Zhoa	moʔ
Mboa	motu
Mbula	moshet
Bankala	mòk
Jaku (labir)	mòghóʔ
Jarawa	*mo
Jarawa	mòk
Aghem	mòʔ
Aghem	mɔ̀ʔ
Babanki	mó, mòʔ
Babungo (vengo)	mù'
Bafut	móʔ!ó
Bambui	móʔo:
Bamun	i-moʔ
Bamunka	mò~'ò~
Bangangté (medumba)	nchìʔ, mô:k
Bum	mɔ̀k
Dschang (yemba)	-moʔ
Kënsense (bamessing)	móʔ
Kensweinsei	mɔ̌ʔ
Kom	moʔ, mɔ̀
Lamnso	mo
Mandankwe	mòʔó
Mankon (ngemba)	-mòʔé
Mbem (kaka)	mo|:r
Meta	m2moɜ'
Mungaka	isn'ìn
Ndop	mù
Ngiemboon	mòʔó
Ngwe	mòʔfií
Nkwen	mòʔó
Oku	-mo, mɔ̀
Pinyin	mó!ʔó
Dumbo (kemezung)	mío

Adere (dzodinka)	-mòko
Kofa (mfumte)	mwõsú
Limbum	mò?ò-sé
Mbe'	mo'
Ndaktup	mo'ó
Ntem	mò~

However, in Bantoid-Nkambe a reflex without the prefix *mu-* is attested:

Yamba	waté

Such a situation characterizes zones A и B (but not C?), which, as we know, manifest the most important deviations from the class system in Proto-Bantu and in which reflexes without an initial nasal consonant, that is, without a class prefix or with another class prefix, are attested. In these languages, along with reflexes with initial prefix *mu-*, reflexes without this prefix or with another prefix are widely represented.

Examples without a velar suffix:

bantu-NW-A	Bekwel	wat
bantu-NW-A	Kako	wɛ́tɛ̀
bantu-NW-A	Ndambomo	i-woto, yoto
bantu-NW-B	Kele	yi-woto, nwoto
bantu-NW-B	Ndasa nord	yotu
bantu-NW-B	Koya	ooto
bantu-NW-B	Sake	woto
bantu-NW-B	Sekyani	-wote
bantu-NW-B	Wumbvu	ɔɔtu
bantu-NW-A	Mpiemo	-woro, wɔ̄rɔ̄
bantu-NW-A	Ngumba	vure
bantu-NW-A	Njem (koonzime)	nggw-ar
bantu-NW-A	Njyem	gwɔ́r
bantu-NW-A	Shiwa	vərə, vəri
bantu-NW-B	Myene	w̃ori

Examples with the original velar suffix:

bantu-NW-A	Lundu	e-oko
bantu-NW-A	Yasa	e-woko
bantu-NW-A	Yasa	èvɔ́ɔ̄kɔ̄
bantu-NW-B	Kande	pɔkɔ, pɔ́kɔ̀
bantu-NW-B	Kota	-oko, yɛkɔ
bantu-NW-B	Pinji	poko, pɔkɔ, pɔ́kɔ̀

bantu-NW-B	Simba	pɔkɔ
bantu-NW-B	Tsogo	poko, pɔkɔ
bantu-NW-B	Vove(pove)	pɔkɔ
bantu-NW-C	Bobangi	oko
bantu-NW-C	Koyo	-hogo
bantu-NW-C	Mbosi	ɸɔɔ

It is important to note that in other Bantu zones a reflex of the *woti / wodi / woci* without a nasal prefix is almost not attested. Let us cite all examples of this kind we have in hand:

Reflexes without a velar suffix:

bantu-G	Ngulu	bosi
bantu-G	Zigula	bosi
bantu-K	Luchazi	k-osi
bantu-L	Mbwera	k-usi
bantu-N	Kunda	posi
bantu-N	Nyungwe	posi
bantu-N	Podzo	posi
bantu-N	Sena	posi
bantu-S	Ndau	posa

Reflexes with a possible velar suffix:

bantu-M	Iwa	-ongga
bantu-M	Malila	hoka
bantu-M	Nyiha	yooka
bantu-M	Tembo	-oka
bantu-R	Nyaneka	-ike

So, in Bantu languages as well as in many Bantoid languages reflexes of *woti* with traces of noun class marker *mu-* are attested. At the same time, in many languages of zones A and B this prefix is absent.

It is well known that the nasality of noun class markers (with the exception of class 6N marker – proto-NC *ma) is not reflected in other branches of NC where nasal consonants in the classes 1, 3, 4, 9?, 10 correspond to oral consonants, for instance, initial *m- in the class 1 corresponds to the initial *k-. It does not matter how we reconstruct class 1 marker in NC - *ku or *mu, - reflexes of *(ku)-woti / wodo / woci* will be an exact correspondence to the forms in Bantu.

Let us consider possible parallels in other NC branches.

We will start with Atlantic languages. Complicated problems of genealogical classification of Atlantic languages will be examined in my presentation on the Congress. I will only note here that, first, there are no serious arguments in favour of placing Mel languages closer to Atlantic languages than to other NC branches, and, secondly, many languages that are traditionally considered as Atlantic, such as Limba, Gola, Sua, may also appear to be separate branches of NC.

The most probable correspondences of the Bantu root meaning «one» are lexemes from Balant. According to different sources, the forms of different dialects of Balant are as follows:

Balant kentohe	oda, ɔda , -ɔɔdaʔ, -ɔɔdn, f-hood-n
Balant-fca	woda, wɔda, -ɔdaʔ, wɔdi-bɔ
Balant-ganja	woda

An undisputably related root is attested in Bijogo. Let us cite the roots we have in hand:

Bijogo	nɔɔd
Bijogo-kajoko	-ɔ́ɔd
Bijogo (Ankaras)	modige
Bijogo-kagbaaga	óód, ɛ̀ɛt, èḍìgɛ́
Bijogo-kajoko	nɛt, -ɛ̀ɛt, -èḍìgɛ́
Bijogo-kamona	nɛt

The structure of these forms is not obvious. The most possible interpretation is that we deal with two probably related roots:

*óód (n-ɔɔd, *-ɛ̀ɛt (n-ɛt) and a form with a velar suffix *m-odi-ge consecutively reinterpreted as dige / è-ḍìgɛ́

The form *m-odi-ge is surprisingly similar to the Proto-Bantu *mu-odi-(ga), *mu-oti-(ga).

It is interesting to note the forms of "one" in Limba for which it is equally possible to presume a reflex of *wunte consecutively reinterpreted or simplified in the following diversity of attested forms: wunte, wu-nte, nte, wo-wuŋte.

And, finally, this root is probably reflected in Tenda group of the Northern branch of Atlantic languages – in Basari, Pen and Tanda (but not in Konyagi or Bedik):

Bapen	mat / ɓat
Basari	-ɓât / -mât, a-w̃àt
Tanda	baat

The root we analyze is probably attested in Mel languages, in the Bullom-Sherbro subgroup, to be more precise:

N. Bullom (Mmani)	m-ul
S. Bullom (Sherbro)	b-ul
Sei	bul, ní-mbul
Sherbro	bul

The final –l in this case reflects the original *-d.

The distribution of the root among the Atlantic languages (it is represented in all three major groups – Mel, Central languages, Northern languages), taking into account the parallels in Bantu and Bantoid languages, allows us to suppose that we deal with reflexes of a NC root.

How is it represented in other branches of NC?

In Adamawa languages we find probable reflexes in the following languages:

Niellim	ɓúdū
Kotopo	wate
Besme	mō ndā
Waka	bindi
Yendang	bInti
Yendang	ɓīntī

I am not a specialist in other branches of NC and I cannot estimate the plausibility of many other possible reflexes of this root. I just cite them here in hope that comparativists will estimate them:

BC-eastern-kainji-eastern	Chamo	wondi
BC-eastern-kainji-eastern	Kuda	wandi
BC-eastern-plateau	Eggon	òrí
BC-idomoid	Eloyi	-ònzé
BC-idomoid	Eloyi	kònzé
BC-idomoid	Yala	ósè
BC-igbo	Igbo	otu
kordofanian?-kadugli	Kadugli	ngôtòk
kordofanian?-kadugli	Katcha	nggoto
kordofanian-heiban	Laru	gwette
kordofanian-heiban	Logol	gwátte
kordofanian-heiban	Rere (koalib)	kwútte
kordofanian-heiban	Warnang	ngútto

kwa	Avatime	ò-lé
kwa	Avatime, ghana	ólè
kwa	Avikam	etõ
kwa	Avikam	ɛ́t͡ɔ́
kwa	Gbe-saxwe-	àɖé / ɖók͡pó
kwa	Gbe-xwla-	òɖě(lók͡pɔ̀)
kwa	Xwela-gbe	odé
ubangi	Baka	kpode
ubangi	Baka	kpóde
ubangi	Banda-Linda	bale
ubangi	Banda-Tangbago	bàlē
ubangi	Langbasi (Langbashe)	bàɛ̄
ubangi	Mbanza (Mabandja)	bale
ubangi-banda	Banda-Banda	bale
ubangi-banda	Gbi	bàle-
ubangi-banda	Langbasi	bàle-
ubangi-banda	Mbanja	bale|
ubangi-banda	Mbanza	bàle-
ubangi-gbanzili	Baka	gpo'dé
ubangi-gbanzili	Bayanga	bodé
ubangi-gbanzili	Bomasa	bodé
ubangi-gbanzili	Ngombe	kpóode-
ubangi-gbaya	gbaya-Kara	kpɔ́kɔ́ɗí
ubangi-gbaya	Proto-GBAYA	kpóm, kpɔ́k
ubangi-zande	Barambo	átsi

It is necessary to note that during the etymological analysis with this potential root in Niger-Congo, we need to take into account that many phonetically similar forms are attested in the Chadic family of the Afro-Asiatic phylum. Thus, for example, Proto-Bantu *modi* with a presumed noun class prefix coincides with a form in Bole, a Chadic language, *mod`ì* . I will cite here only the most remarkable forms:

Bole	mod`ì
Jimbin	wútí
Karekare	wed`i
Kariya	wúti
Kirfi (Giiwo)	móod`ì
Miya	wutÌn
Ngamo	mòd`i
Pa'a	waci
Siri	wùtí
Warji	wútí

30

We will go back to the problems of the reconstruction of the NC root at the end of the paper. Here we will examine other roots for 1 in Atlantic and Mel languages.

An exact parallel to numeral "one" with no relation to the root *WODI* examined above is attested both in Atlantic and Mel – I mean the root *IN* which is represented in many branches of North Atlantic as well as in the Northern group of Mel:

atl-n-cangin	Lehar	yin-o
atl-n-cangin	Ndut	yine
atl-n-cangin	Palor	yino
atl-n-cangin	Safen	yinor
atl-n-jad	Jaad	pa-ine, pa-inɛ, pá-ini
atl-n-nyun	Kasanga	-eena
atl-n-wol	Wolof	CL-enn
atl-c-nalu	Baga fore	b-en
atl-mel	proto-Temne-Baga-Landuma	*p-in
mel-north	Baga binari	p-in
mel-north	Baga koba	p-in
mel-north	Baga maduri	-iin
mel-north	Baga sitemu	-in
mel-north	Landuma	-iin, -in
mel-north	Temne	(p)-in

Outside Atlantic and Mel, possible parallels are attested in Adamawa, Gur and Benue-Congo. The data is given below:

adamawa	Bangunji (Bangwinji)	win
adamawa	Burak	kwín
adamawa	Lakka	ben
adamawa	Dadia	win`
adamawa	Tula	win`
gur	Bimoba	yènn
gur	Buli, Ghana	-yéŋ / wà-ɲī
gur	Dagaara,Northern	bõ-yen / bõe (thing-one)
gur	Dagbani (Dagomba)	yín-ó, yín-í
gur	Farefare	yénnó
gur	Hanga	–yɪnnɪ / ʊ̀ŋkʷɔ
gur	Kamara	yínè
gur	Malba Birifor	bõ-ɣén (bomɣén)
gur	Mampruli	ýɪnní / ndàm (in counting)
gur	Moba	jènǹ
gur-oti-volta	Dagaare (Dagara)	bòn yéní

gur-oti-volta	Hanga	yinni
gur-oti-volta	Nabte	ayeni
gur-oti-volta	Talne	ayen
BC-cross	Bekwarra	kìn?
BC-cross	Bete (Bendi)	ì-kìn
BC-eastern-cross-upper	Nkukoli	gínî
BC-eastern-cross-upper	Umon	wÌnì
BC-eastern-kainji-eastern	Gure	pi-in
BC-eastern-kainji-eastern	Sheni	hini
BC-eastern-plateau	Izarek	zini
BC-eastern-plateau	Jaba (Ham)	zhini
BC-eastern-plateau	Jari	z'iní
BC-eastern-plateau	Koro	è-ñí
BC-eastern-plateau	Rukuba (Che)	shin
BC-eastern-plateau	Yeskwa	ki-ñi
BC-jukun	Yukuben (Boritsu)	gí:en
BC-yoruba	Yoruba	ení, énɛ, ééni, ínɛ

We will return to the interpretation of Adamawa roots and to the hypothesis proposed by Raymond Boyd.

Note also *hina* in Dogon (Tomo Kan). Finally, for the etymological estimation of the Northern Atlantic root it is necessary to take into account the reconstruction for Proto-Berber: *yn 'one'.

Apart from the two examined roots *WOTI and *IN, in Atlantic we can find numerous roots for "one" which are very unlikely to originate from NC, taking into account the narrow territory of their spreading. The etymology of these roots which should be considered as innovations of separate Atlantic branches should be searched for in Atlantic.

Let us cite these roots.

We find an isolated root *gun* on the extreme south of Atlantic and Mel areas, in Liberia, in Gola which is traditionally considered a Mel language without any sufficient grounds.

Another root is represented in all branches of the Central group of Atlantic languages, apart from Bijogo:

Manjak-bassarel	pu-lɔɔl-ɛ
Manjak-churo	lɔɔl-ɔŋ
Manjak-tame	pu-lal-aŋ
Mankany	pu-lol-o, pu-lɔl-aŋ
Pepel	pu-loŋ < *pu-lol-oŋ

Joola karon	yɔ-nɔl
Joola mlomp	yɔ-nɔɔl, yo-noɭ
Joola banjal	ya-nur, ya-nor
Joola kerak	ya-nɔr
Joola ejamat	ya-nor, fa-nɔd
Joola fogny	ya-nɔr
Joola gusilay	ya-nor
Joola bliss	ya-nɔʔ
Joola kasa	ya-nɔʔ, ya-nor
Bayot	en-don < *en-doʔ-on?
Nalu	dɛnd-ɛk

It is phonetically comparable to the root represented in all Nyun dialects belonging to the Northern branch.

Nyun (Djibelor)	gu-mə-nduk, u-mun-duk
Nyun (Nyamone)	man-duk-a, u-duk-a
Nyun (Sonkodou)	man-duk, u-duk-a
Nyun (Tobor)	u-gɔn-duk, mun-duk
nyun-17th century	ma-nduk
nyun-gujaxer (Kasa)	(g)u-duk-a, u-mon-duk
Nyun-Jas	u-mon-duk
Nyun-Jfg	u-man-duk-(a)
Nyun-Kas	gu-mon-duk
Nyun-Nek	gu-duka
Nyun-Nes	gu-duka

In this case, the final velar consonant can originate from a determinative which is clearly seen in Nalu.

Finally, we cannot exclude the connection of this root with what attested in Sereer, *leng o, o-leinj,* and Bedik *dīyè / -rīyè.*

The origin of the root *DOL/NOR* is unclear. On the one hand, the kinship of forms is not evident even for the forms from the Manjak group, *LOL*, and Diola, *NOR*. The reason is that in Manjak we find a form of the 3sg indefinite pronoun **nul**, which is clearly unrelated to the forms **lol** in Manjak but is probably related to the forms **nor / nol** in Diola.

On the other hand, while analysing the Diola forms, we should not exclude from the analysis their possible etymological connection to the Mel root *NO* 'person' (in Bullom, Sherbro, Kisi and Gola) which is also attested outside Atlantic:

| Oti-Volta | *nu |
| Upper Cross | *-nòD |

We would like to point out that the dialectal form of Bijogo **nod** 'one', which is phonetically identical to the form meaning 'person', for example, in Upper Cross, apparently reflects another root, that is, the root *ODI* analyzed above with an initial nasal noun class marker.

One more interesting root can be attested in another language of Nalu group, Baga Mboteni: **mbo.** If we admit that the original root descends from a determinative (in Atlantic, we attest numerous forms **wo** meaning 3sg. as well as 'this', 'that'), the Baga Mboteni root can be associated with the one attested in Fula, *go'o*. In this case, in Fula we deal with an operation of strengthening of the consonant alternation degree in nouns and dependent forms of the noun class 1. However, the Fula root can also originate from the root meaning 'cola nut'.

The "isolated" Kisi root **pum** is undoubtedly related with Bullom **pum** 'some' and Sherbro **poom / mpum** 'some'.

In conclusion, let us see other isolated roots with the same meaning:

biafada	nnəmma
jaad	pa-kan
buy	teena, gu-rena
sua	sɔn

TWO = ONE ?

Now we will go back to the two roots which, as it was noted above, have parallels outside Atlantic and Mel, that is, to the roots **(W)OTI / *(W)ODI* and *EN / *IN*.

The comparison of these forms inside the paradigms of Atlantic and Mel languages, as well as those of other branches, brought me to a supposition that in NC we should reconstruct a common root for numerals 1 and 2 which attracts different noun class markers: the numeral 1 attracts the marker of noun class 1 (sg. human), while the numeral 2 attracts that of noun class 2 (pl. humans). There are many serious arguments in favour of reconstructing prefix *ku-/ko-* for class 1 in NC. The forms in question, therefore, should be reconstructed as follows:

*NC *ku-di* 'one ~ *ba-di* 'two' (with a variant *bi-di*, which could have appeared as a result of a regressive assimilation of the vowel or of a replacement of class 2 by class 8).

It is tempting to trace the prototypical situation in Ubangi forms:

		'one'	'two'
ubangi-gbanzili	Bayanga	bo-dé	bi-dé
ubangi-gbanzili	Bomasa	bo-dé	bi-dé
ubangi-gbanzili	Baka	gpo'-dé	bi'-dé
ubangi-gbanzili	Gundi	po-dé	bi-dé
ubangi-gbanzili	Ngombe	kpóo-de-	bí.-de-

The forms of the Tamari language which belongs to the Oti-Volta branch of the Gur language family where noun class markers are suffixed, are also interesting in this respect:
'one'- *le-mu*, 'two' – *le-n*.

In this case, the Proto-Bantu for "one" results from restructuring the root and including the original class prefix into it: ***o-di > *mu-odi*, which conforms with the hypothesis of nasalisation of certain class markers in Proto-Bantu, including that of class 1.

Can this model be attested in modern NC languages? It is possible that in Edoid, a BC group, the form for "two" with another root can be interpreted as a derivate from "one" by means of historical replacement of the noun class marker:

		one'	two'
BC-edo	Okpamheri	o-vo	e-vo
BC-edo	Engenni	á-vù	í-vè
BC-edo	Degema	ò-vú	í-vè
BC-edo	Urhobo	o-vo	í-vè

Forms from Are, a Kordofanian language, are given below:

		one'	two'
Kordofanian-heiban	Shirumba (Shwai)	ittí-Bó	ítti-a

The most plausible reconstruction of noun class 2 marker in Proto-Atlantic is *be-. For the numerals, preferable reconstructions are: Atlantic and Mel *o-dV- 'one ~ be-di 'two' .

At the last stage of this research, I read a very interesting article by Raymond Boyd (Boyd, 1989) on numerals in Adamawa which, unfortunately, I hadn't known of before. It turned out that many years ago Raymond Boyd came to this conclusion basing on absolutely different data. In particular, this is what Boyd writes about the reconstruction of 1:

« A rather complicated hypothesis would, in fact, cover most of the Cross River / Plateau data : Let us assume a single root, ~*DI (sometimes ~*DU) and two affixes, (V)K(V) and (V)N(V), which can appear, separately or together, as either prefixes or suffixes, or both. <…> Some support for this hypothesis is provides by the frequently observed inversion of the coronal and velar features : in most cases, where we find a term with initial velar, we find a final coronal nasal ; and where we find an initial coronal, we find a

final velar nasal. This can be explained by assuming the prefixation of *KV-N- in the former case, and suffixation of *-N-K(V) in the latter.» (p.151-152).

The numeral 1 in *Adamawa Boyd proposed to reconstruct as *n-di, while the reconstruction of the numeral 2 is *bà-dí with class 2 prefix (p. 156). One more citation: « It was suggested above that the Cross River / Plateau root for 'one' was *DI. We may now hypothesize that the root for 'two' in the proto-language for these groups was the plural *BA.DI, and that, when Proto-Bantu developed its more complicated class system, this term, whose prefix may have been invariable, was reinterpreted as mono-morphemic » (p. 157).

Two points should be noted here. First, the Atlantic reconstruction, which was performed independently, coincided with the reconstruction for Cross River / Plateau and Adamawa languages reproduced here. At the same time, not only noun class 2 marker can be traced in the numeral 2, which Boyd also mentioned, but also noun class 1 marker in the numeral 1 which Boyd does not indicate clearly taking into account the ambiguity of the reconstruction of this class marker in NC.

Secondly, Boyd's hypothesis about the existence of suffixes lets him, in particular, trace back to the same root the form *KWIN*, typical for Adamawa: following Boyd, *KU + DI + N > KUJIN > KUYIN > KWIN*. If we accept his hypothesis, the second root for the numeral 1 which is widespread in Atlantic languages and in other NC branches, that is, **IN/EN*, can also be reinterpreted as a reflex of the root *DI with a nasal suffix.

CONCLUSION

Les us summarise some of the key conclusions, both "practical" and methodological:

- numerals 1-3 in Niger-Congo: '1' *-*ku-di, *ku-di-n, *ku-n-di, *ku-di-ək* (the last of them, most complicated variant is probably related to the structure of the definite form which includes a confix noun class marker: CV-(class prefix) + CVC (root) + VC (class suffix) ; '2' - **ba-di, *be-di* ; '3' - **thati*. The existence of amalgamous prefixes and suffixes with different functions certainly does not simplify the reconstruction;
- Numerals 1 and 2 in Proto-NC have the same root and are differentiated by a noun class prefix – sg. human for '1' ~ pl. human for '2';
- Data related to the numerals provides complimentary arguments in favour of the reconstruction of **ku-* as the noun class 1 marker in NC. In general, the reconstruction of numerals provides precious material for the reconstruction of noun classes in NC;
- With respect to the numerals we are ready to discuss not only the reconstructed roots, but also the reconstructed lexemes which means that collected material is enough to try to understand in particular which was the class agreement for the numerals in NC.
- If we find it reasonable to assume that, say, *–iin* 'one' in Temne (Mel) (from *di-n) and *felego* 'two' in Looma, (SW Mande) (from *be-di-gə-o) and *mɔ̀zì* 'one' in Lega (Bantu D) (from *mu-o-di) reflect a common NC root, we need to admit that the degree of etymological approximation here is so high that, with certain skills it becomes possible to match any two forms, which is deplorable. The only hope is

the combination of "*pro*" and "*contra*" factors which need to be taken into account in any tentative reconstruction, and the intuition of professionals.

- The data analyzed here can be of interest for a workshop on etymologisation of the nasalised noun classes which will be organized by Larry Hyman as part of the Bantu Congress in June 2013 in Paris. My position is that the data examined here is one of the indirect evidence in favour of the hypothesis on the nasalisation of noun class 1 prefix, an innovation of proto-Bantu. Here it is neither the time nor the place to recommence this eternal discussion. I will only note that, in the light of what has been presented above, the Proto-Bantu form for 1 crystallises into *mu-o-di, which means that it is highly probable that during the process of nasalisation of some of classes in Bantu the prefix *o-* (which is a reflex of **ko/ku*) ceased to be interpreted as a class morpheme and was included into the root: *mu-odi < *odi* < **o-di (note that *mu-otị is the form reconstructed by Bettie Vanhoudt for Proto-Bantu [Vanhoudt, p.16]) ;

- For the etymologisation of numerals in NC, it is important to take into account numerous analogy changes some of which were demonstrated above;

- Finally, we need to keep in mind the fact that a number of roots which could be considered as reflexes of numerals in NC, have good parallels in Afro-Asiatic and in particular in Chadic languages. They are too numerous to be associated with late borrowings.

References:

- Boyd, Raymond. Number Systems in the Adamawa Branch of Niger-Congo // African Languages and Cultures, Vol. 2, No. 2 (1989), pp. 149-173.
- Pozdniakov, Konstantin. Сравнительная грамматика атлантических языков. Москва, «Наука», 1993 (Comparative Grammar of Atlantic Languages)
- Pozdniakov, Konstantin; Guillaume Segerer. Similar Place Avoidance: A Statistical Universal // *Linguistic Typology*, 12, 2, 2007, p.307-348 (в соавторстве с G.Segerer)
- Vanhoudt, Bettie. L'expression de "un" dans la numération référentielle en protobantou // Africana linguistic, XI, 1994, p.p. 215-221
- http://www.zompist.com/numbers.shtml
- http://lingweb.eva.mpg.de;
- http://sumale.vjf.cnrs.fr/Lexiques/reflex
- unpublished sources of Guillaume Segerer, Valentin Vydrin, Konstantin Pozdniakov