

# The numeral system of Proto- Niger-Congo

A step-by-step reconstruction

Konstantin Pozdniakov

Niger-Congo Comparative Studies 2



Series information: ./langsci/seriesinfo/nc-info.tex not found!

# The numeral system of Proto- Niger-Congo

A step-by-step reconstruction

Konstantin Pozdniakov

Konstantin Pozdniakov. 2018. *The numeral system of Proto-Niger-Congo: A step-by-step reconstruction* (Niger-Congo Comparative Studies 2). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/191>

© 2018, Konstantin Pozdniakov

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

ISBN: no digital ISBN

no print ISBNs!

no DOI

Source code available from [www.github.com/langsci/191](http://www.github.com/langsci/191)

Collaborative reading: [paperhive.org/documents/remote?type=langsci&id=191](http://paperhive.org/documents/remote?type=langsci&id=191)

Cover and concept of design: Ulrike Harbort

Fonts: Linux Libertine, Arimo, DejaVu Sans Mono

Typesetting software: Xe<sub>La</sub>TeX

Language Science Press

Unter den Linden 6

10099 Berlin, Germany

[langsci-press.org](http://langsci-press.org)

Storage and cataloguing done by FU Berlin

no logo

Ирине Поздняковой



# Contents

<b>Preface</b>	<b>iii</b>
1    Introduction . . . . .	iii
1.1    Niger-Congo: the state of research and the prospects for reconstruction . . . . .	iii
<b>Acknowledgments</b>	<b>xiii</b>
<b>Abbreviations</b>	<b>xv</b>
<b>1    Reconstruction of Numerals in Niger-Congo</b>	<b>1</b>
1.1    ‘One’ . . . . .	1
1.2    ‘Two’ . . . . .	4
1.2.1    ‘Two’ . . . . .	4
1.2.2    ‘Two’ = ‘one’ PL? . . . . .	5
1.3    ‘Three’ . . . . .	7
1.4    ‘Four’ . . . . .	17
1.5    ‘Five’ . . . . .	20
1.6    ‘Six’ . . . . .	30
1.7    ‘Seven’ . . . . .	31
1.8    ‘Eight’ (‘Four’ and ‘eight’) . . . . .	32
1.9    ‘Nine’ . . . . .	39
1.10    ‘Ten’ . . . . .	40
1.11    Large numbers (‘twenty’, ‘hundred’ and ‘thousand’) . . . . .	44
1.12    Proto-Niger-Congo . . . . .	45
<b>Index</b>	<b>47</b>
Name index . . . . .	47
Language index . . . . .	47





# Preface

## 1 Introduction

### 1.1 Niger-Congo: the state of research and the prospects for reconstruction

It is quite predictable that the title of this book may be met with skepticism by specialists in the comparative-historical studies of African languages. The first question that may arise is whether a Niger-Congo (NC) reconstruction is achievable at all, considered that the reconstruction of proto-languages underlying particular families and their branches has not been completed (or even properly started, as is the case for some groups and branches of NC). Before we turn to the structure of the book, let us try to answer this fundamental question. To do so, it seems reasonable to very briefly outline the present state of affairs in NC comparative studies.

First, it should be noted that presently there is no general scientific discipline such as “NC comparative studies”. Instead, there are individual researchers who work on particular families, groups, sub-groups or branches of NC. Among these, comparative-historical Bantu studies has flourished the most. However, the Bantu languages comprise only a branch of the Southern Bantoid languages that (together with Northern Bantoid) go back to Proto-Bantoid. Hence Bantu is merely one of 16-17 Bantoid branches, as can be gleaned from the chart below (Table 1).<sup>1</sup>

The progress of comparative-historical studies of the Bantoid languages has been less impressive than that of Bantu studies. Proto-Bantoid, as well as a number of other proto-languages, goes back to the Proto-Eastern-Benue-Congo. In turn, the latter (along with Proto-Western-Benue-Congo and possibly some other

---

<sup>1</sup>This book does not investigate the genealogical classification of Niger-Congo as a whole, nor of the individual families of this macro-family. The schemes presented here take into account the most well-known classifications (sometimes with small deviations due to the specific purposes of our study). The scheme of Bantoid languages given here is based mainly on the classification in <https://mpi-lingweb.shh.mpg.de/numeral/Niger-Congo-Benue-Congo.htm>. It generally reproduces the John Watters’ classification (1989: 401) with some deviations, which are not considered here.

Table 1: Bantoid languages

Northern Bantoid:	Dakoid	Mambiloid	Fam	Tiba (Fà)
Southern Bantoid:	<b>Bantu</b>	Beboid	Yemne-Kimbi	Ekoid
	Jarawan	Mamfe	Mbam	Mbe
	Ndemli	Tikar	Tivoid	Wide Grassfields

languages that do not belong to these two major groups of Benue-Congo) goes back to Proto-Benue-Congo (BC). Hence, the Bantoid branch is merely one of 14-15 branches of Benue-Congo, as demonstrated by the chart below (Table 2).

Table 2: Benue-Congo languages  
Inventory of Benue-Congo groups is given mainly by **Williamson1989b** 266-269. The main difference in Table 0.2 is that Jukunoid is separated from Platoid, which allows us to better compare the forms of numerals of these groups, as well as the fact that Lufu has been added to isolated languages. The division of the BC into the Western and Eastern branches does not always reflect the genealogical characteristics of languages.

*Western BC	*Eastern BC	Isolated BC
Nupoid	Kainji	Oko
Defoid	Platoid	Akpes
Edoid	Cross	Ikaan
Igboid	Jukunoid	Lufu
Idomoid	<b>Bantoid</b>	

The traditional reconstruction of Proto-BC based on regular correspondences between the proto-languages underlying the separate branches listed in table 0.2 has developed rapidly in recent years. However (and I hope that my colleagues will take no offence at this statement), despite numerous brilliant studies dealing with the subject, this is still a relatively ‘young’ science.

Finally, in addition to Proto-BC there are probably more than ten proto-languages underlying other language families that together comprise the Niger-Congo macro-family (see Table 3).

Most of the works presently available in NC comparative studies do not reach

Table 3: Niger-Congo languages

The grouping of 12 families of NC into 5 geographical zones is convenient for technical purposes of generalization of data. So, it means nothing else. As for a genealogical tree of NC languages, as of today there are insufficient grounds for creating one, in my opinion.

beyond this point. Exceptions are rare, and examples of the comparative-historical approach to the NC reconstruction are few. Moreover, the most significant works of this kind (e.g. those of **Westermann**1927 **Greenberg**1966 **Sebeok**1971 etc.) are not that recent and usually date to the middle of the 20th century. Comparative studies of the African macro-families had a jump start but nearly had come to little by the end of the 20<sup>th</sup> century (important works such as **Bendor-Samuel**1989 including **Williamson**1988; **Williamson**1989a are few in this period).

So, what happened?

By the 1990s, our knowledge in the field of African languages had begun to grow exponentially. Hundreds of new language descriptions had been published, and the few dozen experts working in NC comparative linguistics were simply unable to digest this avalanche of new information.

The main problem in the 1960s was that we knew too little. From the 1980s on, we have faced the opposite problem: we know “too much”. Not only do scholars not have enough time to absorb new results, sometimes they do not even have enough time to acquaint themselves with those results. During the last four decades, amidst this dialogue between linguistic knowledge and language data, African linguists have remained in listening mode. But I am convinced that the time has come for linguists to say something new again. Unlike even ten years ago, today we are well equipped to do so.

Firstly, we have really exceptional databases. The best one is the RefLex database elaborated by Guillaume Segerer (Segerer & Flavier). It contains more than one million words from African languages (2017), and each entry contains a link to a PDF file of the corresponding source page. It provides a huge range of information and is maximally user-friendly to comparative linguists: it can be solicited for establishing regular phonetic correspondences, for reconstruction and for ranking reflexes as well as for various kinds of statistical data analysis. This new database is being constantly updated.

A big database is something much more than just a huge amount of data. When a database reaches certain degree of plenitude with respect to the main families and branches of the NC macro-family, it opens up prospects for both working with the distribution of words that do exist and with the distribution of **gaps** in postulated cognates. The distribution of filled cells and lacunes is a powerful

tool allowing 1) identification of important innovations, 2) targeted searches for unusual phonetic reflexes, 3) detection of diachronic semantic changes and 4) refinement of genealogical classification.

In my opinion, the opportunity to rely on both the apparent cognates as well as on the missing reflexes of reconstructed prototypes in particular languages dramatically changes the approach to the reconstruction itself.

The following case may serve as an illustration to this statement. Suppose we need to assess one of Greenberg's proposals, e.g. a Niger-Congo root meaning 'hill'. Among the reflexes quoted by Greenberg for this root are: "(2) Busa *kpi* 'mountain', Kweni *kpi* ; (4) Gã *kpa* ; Gwa *ogba* 'mountain'; (5) Nungu *agba*, Ninzam (Ninzo) *igbu*. Kordofanian: (2) Tagoi (*c*)*ibe*." (Greenberg1966). The phonetic correspondences underlying the comparison of these forms will not be discussed here (we will just assume that they are valid), for the main problem is elsewhere. A reader with no access to a representative lexical database on the NC languages is always uncertain about a number of key issues, including:

1. whether the root in question is widely attested in the families and groups for which the author postulates the reflexes?
2. whether the root is present in other NC families and groups and how widely it is attested in them?
3. are there any other roots possibly interpretable as NC terms for 'hill'?

The RefLex database establishes that:

1. there are plenty of forms phonetically similar to those of Greenberg (cf. e.g. Boko (in the same sub-group as Busa) *kpii* 'mountain', Gwari (Nupoid, BC) *ōpé* 'hill, mountain', etc), but the postulated root is at best only marginally attested in the families where Greenberg finds it.
2. The root is absent in other branches and families (even if the proposed phonetic correspondences are approached most liberally), although, if wished, its "reflexes" can be found in any of the NC families, cf. e.g. Ibani (Ijo) *kpókpó* 'hill', etc.
3. Most importantly, several other roots with the meaning 'hill, mountain' are distinguishable in the NC languages. All of them (unlike the one proposed by Greenberg) are valid candidates for the reconstruction of the NC prototype. One of these roots is presented in the chart below (0.4) (one could mention some other roots nearby):

Table 4: *\*tʷnd* ‘hill, mountain’ in Niger-Congo

		Dogon tórɔ́		Kordofan
Atlantic *tʷnd	Mande *tinti, *ton	Gur	Ubangi	Adamawa
Mel tul- ?	Kru tōdō	Kwa tu?	Ijo tʷndɔ́	Benue-Congo tʷndà

The exact correspondence between Proto-Bantu (*\*tʷnda*, zones HJKPMNRS > ( ?) *\*dʷndɔ́*, zones EGHJKLMNRS), Ijo (Ibani *tʷndɔ́*) and Atlantic languages (Atlantic Bak: Manjak *ntʷnda*, Atlantic North: Basari *e-tʷnd*, Bapen *ɛ-tʷnd*, Laala *tundə*, Fula *tulde*, Wolof *tund*) is reason enough to postulate the root *\*tʷnd* ‘hill, mountain’ at the Proto-NC level, especially since these languages have apparently been out of direct contact<sup>2</sup>. In addition, the absence of this root in Gur-Ubangi-Adamawa may prove to be a shared innovation in these languages.

Using the databases, the focus of our research could be redirected toward the basic meaning of the lexemes (rather than on the occasional phonetic similarities between the forms). This approach may help in answering the following question: if a Proto-NC term for ‘mountain, hill’ existed, how did it sound? The answer would probably be as follows: this word could sound like *\*tʷnd*, *\*kong/ keng* or *\*kudu* (‘hill, rock, stone’), but not like *dima* (PB *\*dìmə*, zone EGJ), *mut* (Proto-Jukunoid *\*muT*) or *pi* (PB *pìdì*, zone KLMN).

Upon arriving at these unconventional “results”, one could bring them to the attention of specialists in particular NC languages and branches for further evaluation. Without such professional evaluation there can be no hope for success. Moreover, in recent years it has become evident that this evaluation needs to be collaborative (i.e. made by dozens of specialists working together) for the simple reason that today no specialist can be proficient in the languages of more than one or a maximum of two NC families. Hence, it is important that these specialists are asked questions they can answer, so ideally the approach outlined above should be applied to every family within Niger-Congo. For example, according to the etymological database of the Atlantic languages (Pozdniakov-Segerer 3700 cognates, 2017) only *\*tʷnd* and *\*thəng* are potentially interpretable as the terms for ‘hill, mountain’ in Proto-Atlantic.

<sup>2</sup>We shall repeat that nearby there are some other candidates for ‘mountain’ in NC, which we do not treat here.

Initially I thought of numerals as of an ideal group of terms to test this approach. On the one hand, the core group of numerals must have existed in Niger-Congo. On the other hand, they represent a relatively compact lexical-semantic group with minimum potential for semantic shifts. My initial question seemed simple: what is the most probable Proto-Niger-Congo root for ‘two’? The term for ‘two’ (being the only numeral on the Swadesh list) is generally recognized as one of the most persistent numerals. Why not try reconstructing it on the basis of the NC evidence? It appeared, however, that such a reconstruction is beset with difficulties, so what was originally intended as an article turned into this very book. The structure of the book is described in the section below. As I hope to demonstrate, this structure is conditioned by specific issues encountered in the course of the reconstruction of NC numerals.

## **Sources and the monograph structure**

### **Sources**

Numeral terms included in the majority of lexical sources hold a privileged position. The information pertaining to the Niger-Congo numerals is more than extensive, it is nearly exhaustive. In addition to the above-mentioned RefLex database by Segerer-Flavier which contains over 17,000 entries marked as “numeral” (state April 2017)) a number of other databases with expansive coverage of the Niger-Congo languages are available. One of them is the “Numeral Systems of the World’s Languages” database created by Eugene S. L. Chan and edited by Bernard Comrie (Chan) The data regarding the number systems of about 4,300 languages (with hundreds of the Niger-Congo languages among them) is incorporated into it. Two or even three sources (often unique) are accessible for some of the languages via this neatly organized and user-friendly database. Another universal database that provides numerical data is “Numerals 1 to 10 in over 5000 languages” by Rosenfelder. It was consulted to a somewhat lesser extent because it only includes evidence pertaining to the first ten numerals, for which a simplified transcription is used. Finally, a number of unpublished databases that incorporate the evidence of specific Niger-Congo families and groups were consulted, e.g. the etymological databases of Atlantic (Pozdniakov-Segerer) and Mande (Valentin Vydrin).

As a result, a total of 2,200 sources for Niger-Congo languages were used in this study. This raises the issue of references, since it is impossible to provide a complete list of sources for every NC language. The language index at the end

of this book lists the nearly 1,000 languages cited. For these 1,000 languages, the main sources I used are indicated in Appendix 5. The index of sources in Appendix 5 is structured according to the NC main families in alphabetical order.

For each language, I provide not only the source(s) that can be found in the bibliography, but also the name of every contributor in Chan's database [Chan]. The list of contributors is many pages long, but their names should be known, even if their data are unpublished. This is the least I can do to express my sincere gratitude to each of them.

## Monograph structure

Noun class affixes are present in numerical terms in the majority of the Niger-Congo languages. At the same time, many forms that are considered primary at the synchronic level have frozen noun class affixes that are no longer productive. In such cases it is extremely difficult to distinguish the etymological root within a numerical term. Without it, however, both the comparison and reconstruction of roots is impossible. This is why the first chapter of this book is devoted to the study of various uses of noun class markers in numeral terms.

The second chapter deals with the alignment by analogy in numeral systems. As in other languages, numerals represent a lexical-semantic group that is especially subject to alignment by analogy due to its closed structure, where words are associated in a paradigm. A textbook example is the term for 'nine', with Indo-European \*n- irregularly reflected in Proto-Balto-Slavic as *d-* (Russian *dev'at* '9' instead of the expected \**nev'at*) by analogy with the term for 'ten' (Russian *des'at* '10'). This yielded a minimum pair *dev'at* ~ *des'at* that forms a "class of the upper numerals" within the first ten. Adjacent numerals may be aligned with each other in the NC languages by a similar formal marker. Thus, no satisfactory etymology can be suggested for the forms attested in Mumuye (Adamawa; *ziti* '2' ~ *ta:ti* '3' ~ *dě:ti* '4') without the analysis of alignment by analogy. The issues pertaining to both detection and analysis of such alignments are addressed in Chapter 2.

Chapter 3 offers a step-by-step reconstruction of number systems of the proto-languages underlying each of the twelve major NC families, on the basis of the step-by-step-reconstruction of numerals within each family. The term "reconstruction" related to numerals throughout this book calls for a definition. As mentioned above, the use of this term has been questioned, mainly because systems of regular phonetic correspondences between the languages within NC families remain unknown. This is why Kay Williamson opted for the term *pseudo-reconstructions* (marked with # instead of \*): "Reconstructions proposed by their

authors as based on regular sound correspondences are preceded by an asterisk. Pseudo-reconstructions based on a quick inspection of a cognate set without working out sound correspondences are preceded by a #” (Williamson1989b). In his numerous online publications Roger Blench uses # as well, but his terminology is different: he prefers the more neutral term of *quasi-reconstructions*. Modern comparative studies of the NC languages is a relatively young science, so the opposition between “real” and “pseudo-/quasi-” reconstructions seems irrelevant to me at this stage. The more so that nearly all of our reconstructions (maybe with the exception of Bantu and some other branches) should be marked with #, including the large proportion of reconstructions allegedly based on the evidence of historical phonetics. On the other hand, I think that many colleagues would agree with the following statement: although we do not know the regular phonetic correspondences between the languages that belong to different NC families, there is hardly any doubt that the NC root for ‘three’ sounded something like *tat*.

Throughout this book the term “step-by-step reconstruction of number systems” (e.g in the Atlantic family) is used in reference to the method that includes the following steps:

1. While comparing the forms of numerical terms attested in the languages under study, their most likely prototypes were established within both of the Atlantic groups, i.e. Northern (Proto-Tenda, Proto-Jaad-Biafada, Proto-Fula-Sereer, Proto-Wolof, Proto-Cangin, Proto-Nalu-Baga Fore-Baga Mboteni) and Bak (Proto-Joola-Bayot, Proto-Manjak-Mankanya-Pepel, Proto-Balant, Proto-Bijogo).
2. On the basis of these prototypes, the most likely forms of Proto-Northern Atlantic and Proto-Bak Atlantic numerals were suggested.
3. On the basis of these more ancient forms, the most plausible reconstruction of Proto-Atlantic numerals was offered.

Chapter 4 deals with the reconstruction of the Proto-Niger-Congo numeral system on the basis of the step-by-step-reconstructions offered in Chapter 3 for each of the twelve major families and a handful of isolates. The reconstruction described in Chapter 4 inspired the analysis of the distribution of reflexes of the NC proto-forms within each of the twelve families (as well as within the isolates) in order to establish:

- 1) the most archaic NC families / groups / branches (i.e. those that preserve the inventory of Proto-NC forms most fully);



2) NC families / groups / branches that are the most distant from Proto-Niger-Congo in what pertains to the reflection of numerals.

The results of this analysis are presented in Chapter 5.

To illustrate the logic of the complex structure of the monograph, let us consider one example.

In Chapter 3, along with other NC families, the numerals of the Atlantic languages are analyzed (section 3.12). Atlantic languages are divided into two main groups – North Atlantic (section 3.12.1) and Bak Atlantic (section 3.12.2).

In Sections 3.12.1.1.–3.12.1.7, systems of numerals are considered consecutively in the seven main subgroups of the North Atlantic languages. In particular, in §??, numerals in the Jaad-Biafada subgroup are considered and it is established that in these languages, for the numeral '10', the form *\*-po* is attested. In the final section of 3.12.1, namely in §?? the forms of numerals in the seven northern subgroups are compared, and in particular it is concluded that for Proto-Northern Atlantic, the most probable reconstruction for the numeral '10' is the reconstruction of *\*pok*.

In Sections 3.12.2.1–3.12.2.4, the numeral systems in each of the four subgroups of the second Atlantic group, namely Bak, are discussed consecutively. The final section concerning the Bak group (3.12.2.5) concludes that the only candidate for reconstructing '10' in the Proto-Bak (in addition to the possible model  $10 = 5 * 2$ ) is the root *\*-taaj*.

In the final paragraph of section 3.12, namely in 3.12.3, the systems of the North Atlantic languages and the Bak Atlantic languages are compared. This paragraph concludes that the comparative evidence points to the total absence of common roots present in both groups. The only exception to this is the root *\*tɔk / \*tVɔk* 'five'. Accordingly, it is concluded that it is impossible to reconstruct the Proto-Atlantic root for the numeral '10' without the Niger-Congo context.

In Chapter 4, reconstructions for each family are compared. Accordingly, Chapter 4 has a different structure. If in Chapter 3 each of the sections is devoted to a particular family of languages (in particular, §?? is devoted to the Atlantic languages), then in Chapter 4 each section is devoted to the prospects for the reconstruction of each Niger-Congo numeral. So, in §?? all intermediate reconstructions for the numeral '10' are considered. It turns out, in particular, that the form *\*-taaj* reconstructed for '10' in the Proto-Bak does not find parallels in other Niger-Congo branches. In contrast, the root *\*pok* '10', reconstructed for the North Atlantic languages, can be related to the roots reconstructed for the vast majority of Niger-Congo families (it seems to be missing only in Ijo, Dogon and

## *Preface*

Kordofanian). Based on the NC comparison, the root for '10' is reconstructed as *\*pu* / *\*fu*.

Chapter 5 traces the history of the numerals of Niger-Congo, reconstructed in Chapter 4, in each individual family of languages. Accordingly, each section, as in Chapter 3, is devoted to one of the NC families. So, §?? is devoted to the Atlantic languages. In particular, it is concluded that in the North Atlantic languages the term for '10' has been preserved in three sub-groups (Wolof *\*fukk*, Proto-Tenda *\*pəxw*, Proto-Jaad-Biafada *\*po*). In the other subgroups it is replaced with isolated innovations. The forms of the Bak languages are also innovated.

So, the basic logic of the chosen structure of the book is as follows: we will consistently move from reconstructions in individual families (Chapter 3) to the reconstruction of each Niger-Congo numeral (Chapter 4) and to the interpretation of each individual family in the Niger-Congo context (Chapter 5). We will take into account the provisions formulated in the preliminary chapters concerning noun classes in numerals (Chapter 1) and changes by analogy in systems of numerals (Chapter 2).

# Acknowledgments

Today the greatest benefit to being a researcher is the opportunity to directly contact leading specialists in the comparative studies of African languages. Even the best database does not ensure the proper interpretation of the results achieved by other scholars. In the course of my work on this monograph I have benefited from the help of many colleagues, whose comments and suggestions I greatly appreciate. My particular thanks go to Guillaume Segerer (Atlantic languages and ReFlex database), Valentin Vydrin (Mande languages), Raymond Boyd (Adamawa languages), Larry Hyman (Bantu languages and Benue-Congo in general), Mark Van de Velde (Bantu languages), Marie-Paule Ferry (Tenda languages), Pascal Boyeldieu (Bua languages and Laal), Marion Cheucle (Bantu A.80), Denis Creissels (Balant), Sylvie Voisin-Nouguier (Buy), Ekaterina Golovko (Baga Fore), Odette Ambouroue (Orungu) and many others. It is a great pleasure for me to thank you all!

My special gratitude is addressed to colleagues who read the first version of the manuscript of the book and made a number of valuable critical remarks. These are members of the editorial board of this series of Niger-Congo Comparative Studies: Valentin Vydrin, Larry Hyman, John Watters and Guillaume Segerer. I tried as much as possible to take their remarks into account. Naturally, all responsibility for inevitable mistakes and shortcomings lies with me.

I should like to express especially my gratitude for Sebastian Nordhoff for the layout of this book. Many thanks for my proofreaders – their comments were very useful for me.



# Abbreviations

## Language groups and proto-languages:

BC – Benue-Congo  
GD – Ga-Dangme  
GTM – Ghana-Togo Mountain  
Juk. – Jukunoid  
NC – Niger-Congo languages  
PB – Proto-Bantu  
PLC – Proto-Lower Cross  
PP – Proto-Platoid  
PTB – Proto-Potou-Tano-Bantu  
PUC – Proto-Upper Cross  
SE – South-Eastern Mande  
SWM – South-Western Mande

## Others:

CL – noun class  
CL.SG. – noun class of singular  
CL.PL. – noun class of plural  
CM – noun class marker  
dial. – dialect  
PL. – plural  
redupl. – reduplicated  
SG. – singular.



# 1 Reconstruction of Numerals in Niger-Congo

## 1.1 ‘One’

The five stems present in Table 4.1 are the most likely candidates for the reconstruction of ‘one’ in NC (Table 1.1):

1. Commentary. The chart is used to demonstrate the distribution of roots across language families. It groups twelve families into five major branches, including Western NC (Atlantic, Mel), Northwestern NC (Dogon, Gur, Mande), Northern NC (Ubangi, Adamawa), Southern NC (Kru, Kwa, Ijo, BC), and Eastern NC (Kordofanian).

It should be stressed that this grouping has no implication for the genealogical classification of the NC languages and merely serves as convenient means of display for the isoglosses that will hopefully help to adjust the existing classification.

The chart demonstrates a variety of possible reconstructions. At the same time, some positive knowledge can be gleaned from it. First of all, it should be stressed that a step-by-step analysis of the forms for ‘one’ attested in the families and branches of NC strongly suggests that no other candidates, except for those displayed in the chart above, can be reconstructed. It should also be noted that the reconstruction of a tri- or even disyllabic root on the basis of the available evidence seems highly improbable, since all potentially reconstructible roots are monosyllabic. Moreover, the inventory of these roots is limited and merits special discussion. Such a discussion is essential, since many of the quasi-reconstructions presented above are not immediately apparent. The problems pertaining to the reconstruction of these roots were to some extent treated in the previous chapter. What follows is a brief survey of the basic facts.

The root *\*di*. This well-known root has received much scholarly attention as the major candidate for the reconstruction of ‘one’. It is manifestly absent only in Kru, Mande and Dogon. In addition to the families listed above, this root is

# 1 Reconstruction of Numerals in Niger-Congo

Table 1.1: Niger-Congo stems for ‘1’

<b>*di</b>		Dogon		Kordofan lel/ led?
Atlantic di(n)/li(n)/ne(n)	Mande	Gur de/ le/ re	Ubangi le/ ne	Adamawa di
Mel -lɛ?	Kru	Kwa di-	Ijo die ?	Benue-Congo (o-)di(n), ni/nye
<b>*in</b>		Dogon		Kordofan
Atlantic -in?	Mande	Gur	Ubangi	Adamawa in?
Mel -in	Kru	Kwa	Ijo	Benue-Congo hin/kin/cin/-in
<b>*do</b>		Dogon do		Kordofan
Atlantic don/ lɔŋ	Mande do	Gur	Ubangi	Adamawa do?
Mel	Kru do	Kwa	Ijo	Benue-Congo
<b>*ti</b>		Dogon ti(i)		Kordofan ɬe(k)/ lu(k)
Atlantic	Mande	Gur	Ubangi	Adamawa
Mel	Kru	Kwa	Ijo	Benue-Congo (o-)ti
<b>*gbo, *kpo</b>		Dogon		Kordofan
Atlantic mbɔ	Mande	Gur (k)po	Ubangi k(p)o(k)	Adamawa *(g)bunu, (mon)
Mel bul, mɔ	Kru (g)bolo	Kwa	Ijo gbéri?/n-kèni?	Benue-Congo gbon, m-o?



also attested in the Laal language isolate (*bídíl* (*bì-díl*?) '1'). At the same time, it is absent in the Sua, Gola and Limba isolates. It bears reminding that the reconstruction of this root in Benue-Congo and Bantu is only possible under the assumption that PB *mòdì* < \**m-ò-dì* '1' (with *m-* being a Proto-Bantu CL1, and *-ò-* being an archaic noun class marker (possibly < \**ko-*/ \**ʔo-*, i.e. NC class CL1 incorporated into the stem).

The root \**in*. Although this root is not attested outside Western NC, BC and possibly Adamawa, it is worth mentioning, especially in view of its possible etymological relationship with \**di* (see above).

The same is applicable to \**do* (best attested in Northern NC, Atlantic and Kru).

The reconstruction of \**ti* '1' is the least certain among the roots discussed above. The form *ha-nthe* '1' attested in the Limba language isolate is noteworthy.

The last root is a tentative representation of the forms with the initial labio-velar (or labial in the case of Western NC) that are not necessarily etymologically related. The root *guùŋ* '1' attested in the Gola isolate may belong here as well.

In addition to the five roots treated above, apparent innovations may be attested in particular families (or even in groups within them). Among these are Kordofanian *ʔn* (cf. Sua *sən*), Gur *túrí/ tumɔ*, Mande West *kelen*, and Atlantic Bak *-anor*, *əkon*.

## 1.2 ‘Two’

### 1.2.1 ‘Two’

A systematic comparison of the terms for ‘two’ attested in the NC families yields somewhat unexpected results. The only candidate for the reconstruction of the NC term is the root that can be tentatively recorded as *\*di*. At the same time, nearly every family has its own root (or, more often, roots) for ‘two’ that finds no parallel outside the branch/family in question. The distribution of *\*di*, as well as an overview of isolated roots, is presented in the chart below (Table 1.2):

Table 1.2: Niger-Congo stems for ‘2’

<b>*di/ni</b>		Dogon lé(y)/ ló(y)/ né(y)/ nó(y)		Kordofan
Atlantic di(k), nak	Mande	Gur nyi/ ne(n)	Ubangi	Adamawa du/ru, te/ re/ si
Mel dɪŋ/tsɪŋ/tɪŋ/rəŋ	Kru	Kwa	Ijo	Benue-Congo ba-di / ba-ji
<b>isolated roots</b>		Dogon		Kordofan kok/kek/cik, (can/ʔan, rak, rəm)
Atlantic ʔubəʔ, -təb/-təw, -pugut/pugus	Mande pila/ fila	Gur nyu/ ju, hin/ han	Ubangi si/ ʃi, (wa/gbwɔ, to/ so)	Adamawa ra(k)/ ra(p), gba/ gwa
Mel	Kru so(n)	Kwa ɲɔ	Ijo mamV	Benue-Congo pa ? ba(i) ?

1. Commentary. The isolated forms are as follows: Laal ʔĩsĩ (ʔĩ-sĩ?) (this root is comparable to that attested in Ubangi), Sua *cen*, Gola *tĩ-yèe/ tĩ-el/ cel* (the Gola and Sua terms may be related), Limba *ka-le/ kaa-ye* (this root may go back to NC *\*di*).

The unprecedented variety of forms exhibited by the term for ‘two’ is especially surprising because this notion has been viewed as one of the most persistent in language history (it is the only numeral on the Swadesh list). As we will see below, this term is the least stable in the Niger-Congo languages. However, the NC root *\*di* is well-attested across the families.

## 1.2.2 ‘Two’ = ‘one’ PL?

As can be gleaned from the evidence presented above, the only root for ‘two’ reconstructible in NC (*\*-di*) is suspiciously similar to the most likely reconstruction for ‘one’ (*\*-di*). This similarity was first observed by Raymond Boyd, one of the most renowned experts in the reconstruction of Adamawa. Before we turn to the discussion of the most promising (in terms of the NC reconstruction) forms, an overview of Raymond Boyd’s hypothesis regarding Adamawa and some of the BC languages is in order. Here is what Boyd writes about the reconstruction of ‘one’: “A rather complicated hypothesis would, in fact, cover most of the Cross River/Platoid 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 provided 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.” (Boyd1989). Boyd’s proposal is to reconstruct the Proto-Adamawa terms for ‘one’ and ‘two’ as *\*n-di* and *\*bà-dí* (with class 2 prefix) respectively (Boyd1989). According to him, “It was suggested above that the Cross River/Platoid 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” (Boyd1989).

It should be stressed that Boyd’s hypothesis explains the Proto-Bantu forms that underwent the following transformation over the course of time: *\*m* (CL1)-*o* (<*\*CL1*)-*di* > *\*mò-òdì* / *mòi* ‘1’ / *ba*(CL2)-*di* > *badi* ‘2’ (the dialectal Proto-Bantu form *jòdè* (zones BH) (< *\*jò*(CL5 ?)-*dí*?)). It bears reminding that our evidence favors the reconstruction of (*o*)-*di*(*n*) ‘1’ / *ba-di* / *ba-ji* ‘2’ at the BC level.

One of the major problems with this reconstruction is that synchronically the roots for ‘one’ and ‘two’ are the same in only a minority of the modern NC languages. This rare phenomenon is attested in the Ngabaka branch of Ubangi (Table 1.3):

As stated above, examples of this kind are exceptionally rare. A possible explanation for the overwhelming absence of the identical roots for ‘one’ and ‘two’ is that one of the classes is subject to the nasalization process (entailing further phonetic changes within the root), while the other is not. It bears reminding that, according to Boyd, a number of expanded forms such as *\*n-di* (with further de-

## 1 Reconstruction of Numerals in Niger-Congo

Table 1.3: The same stem in ‘1’ and ‘2’ (\**di*)

	‘one’	‘two’
Bayanga	bo-dé	bi-dé
Bomasa	bo-dé	bi-dé
Baka	kpó-de	bí-de
Gundi	po-dé	bi-dé
Ngombe	kpóo-de-	bí-de-

velopment to \*-*ni* / -*in* ‘one’) is reconstructible along with \*-*di*.

In view of this, the Oti-Volta numbers, thoroughly discussed in the previous chapter, are especially interesting. The pertinent Oti-Volta forms are as follows (Table 1.4):

Table 1.4: Potential reflexes of \**di* ‘1’ = \**di* ‘2’ in Gur

	i. Buli-Koma	ii. Eastern	iii. Gurma	iv. Western	v. Yom-Nawdm	*Proto-Oti-Volta
1	yén, ní	dènnì, yěnde/yènn(do), yòn, *de	den, ni	yen/ yin, dam	hén, nyǎŋ	den/ yen, ni, de?
2	yè, li	déé(ni), yēdē	le/ dé	yi(?)	li/ ré?/ *rya?	li/yi

The terms for ‘one’ and ‘two’ are similar within each of the branches, the differences between them being due to the presence of the nasal component in the term for ‘one’.

## 1.3 ‘Three’

Table 1.5: Niger-Congo stems for ‘3’

		Dogon taan		Kordofan tat/ t̩t/ ʔak
Atlantic North: ʔat̩	Mande	Gur tat/ ta(n)	Ubangi taar	Adamawa taat
Mel sas/ ra	Kru taa(n)	Kwa ta	Ijo tato	Benue-Congo tat
		Dogon		Kordofan (ʔitin/ ʔicin, hwaɣ)
Atlantic Bak: feegir, yant/ jent̩, habi/yabi	Mande sakpa/sagba/sawa, ʔààk̩ʔ/yààkáʔ	Gur	Ubangi	Adamawa kunun̩/ gbunun̩
Mel	Kru	Kwa	Ijo	Benue-Congo

As is well known, the term for ‘three’ is exceptionally persistent, with the same root attested in all of the major NC branches (except for Mande). The same root is also present in the Western NC isolates, cf. Sua *b-rar*, Gola *taai/tāāl*, Limba *ka-tati*. At the same time, some languages exhibit what are apparently innovative forms (see the downmost segment of the chart). An isolated root is also attested in Laal (*māā* ‘3’).

Although the relationship between the reflexes of the main root (*\*tath*) is unquestionable, their phonetics pose a problem. The issue is that each family exhibits a great variety of reflexes, while some of them cannot be explained as going back to either the initial *\*t-* or the final *\*-t* of the main root. In other words, reliable correspondences (with *\*t* preserved) are traceable in the majority of families, but not in the case of ‘three’. This forces us to assume that *\*t* may be irregularly reflected as *s*, *r*, *h* in particular families.

The table below (Table 1.6) provides an overview of the pertinent Bantu reflexes of *\*tātò* (ABEFGHJKLMNPRS) / *\*cātò* / *\*cācò* (CD) ‘three’ (these reconstructions follow BLR3):

The Bantu forms should be discussed in order to determine which processes in Bantu (and in Niger-Congo in general) give rise to such a diversity of phonetic variants.

## 1 Reconstruction of Numerals in Niger-Congo

Table 1.6: Reflexes of \*tátò ‘3’ in Bantu

zone	Language	Form	zone	Language	Form
A	Nyo’o	tá	*PB	*PB (dial.)	cátò
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	Mbere	tadi	J	Luganda	ssatu
B	Sira	reru	J	Nyankole	shatu
B	Kande	lato	K	Nyengo	ato
B	Galwa	ntfaro	K	Mbwela	hatu
C	Bua	salu	L	Kete	sàcw
C	So	saso	S	Lozi	talú
C	Sakata	sâa	S	Venda	raru
C	Koyo	tsáro	S	Swazi	tsâtfu

The root includes two consonants. Putting aside the problem of the vowel in the second syllable, we label the two consonants C- and -C respectively. Each of them may be dropped, yielding the Bantu forms **ta** and **at** (Scheme 4.1).

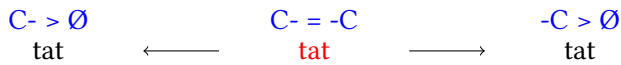


Figure 1.1: Scheme 4.1

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 can the second consonant be dropped. (Schemes 4.2-4.3).

As a result, we have numerous forms, while the variation can be reduced to a very limited number of processes:

- Voicing ( $*t > d$ )

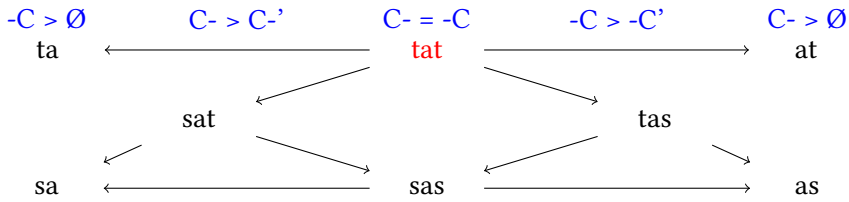


Figure 1.2: Scheme 4.2

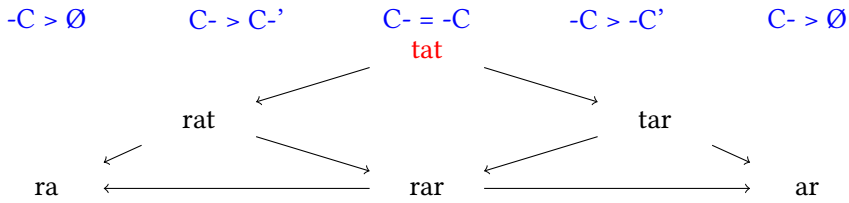


Figure 1.3: Scheme 4.3

- Lenition – partial (spirantization: \*t > s, \*t > r) or full (> Ø).

The following table (Table 1.7) provides a structured overview of the derived Bantu forms (with no arrows):

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

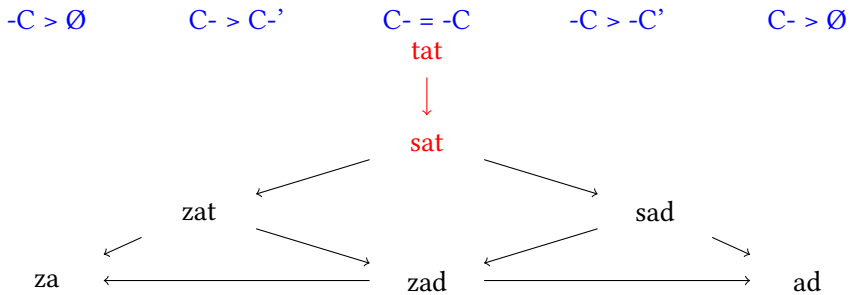


Figure 1.4: Scheme 4.4

This is where the following forms (Table 1.8), many of which are attested in

# 1 Reconstruction of Numerals in Niger-Congo

Table 1.7: Phonetic variations of *\*tat-*

-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

Bantu, originate (forms without square brackets):

Table 1.8: Reflexes of *\*tat-* attested in Bantu

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

1. We often do not know how one or another derived form appeared. For example, the form *las* in the first line of the table could have originated



from *\*tas* (as a result of the change in 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; these are shown in square brackets.

The most amazing observation here is not the high degree of variation (which itself needs to be considered), but the fact that we find precisely 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 and divided into two structurally identical tables (Table 1.9-4.10):

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

To get a comprehensive idea of the presence of the forms in each branch 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 arranged in descending order in the summarising column as well as in the summary line) (Table 1.11):

The following chart represents the number of groups (within the 14 branches of Niger-Congo) presenting the respective combinations of the first (the line) and the second (the column) consonants (the data is presented in descending order) (Table 1.12):

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

If we reconstruct *\*tat-* on the NC level, in line with the majority of linguists, we will have to contend 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 in separate branches, the form remained phonetically unchanged. Then, suddenly the root *\*tat* independently started to explode, giving rise to much phonetic variation in its reflexes.

I think that a hypothesis stating that the root already contained close but not identical consonants in NC is far more typologically justified. The first consonant in that case was *\*t-*, while the second one was represented by a specific phoneme for which no traces remain, for example, *\*-th*?, *\*-t*?, *\*-tʂ*?, *\*-c*? As we tried to show in (PozdniakovSegerer2007), the phonotactics of many languages (not exclusively in Africa) demonstrates the same tendency: in CVC structures languages tend to avoid consonants constituting a minimal pair, for example, *fVp*,

# 1 Reconstruction of Numerals in Niger-Congo

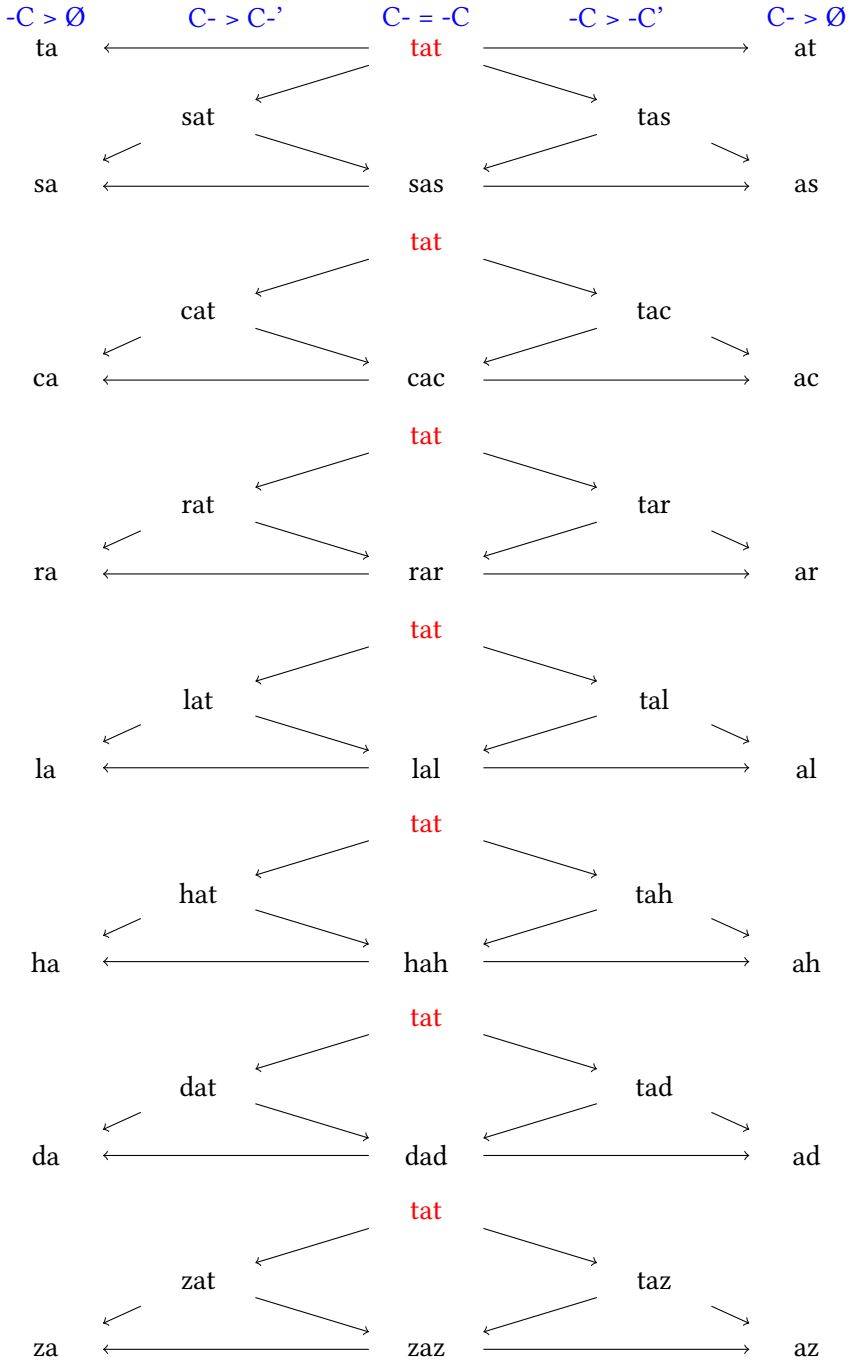


Figure 1.5: Scheme 4.5

Table 1.9: Reflexes of \*tat- in Niger-Congo (1)

Bantu		Adamawa		Atlantic-Mel		
TAT	Rundi	tatu	Yendang	tat	Fula	tat-
TAR	Yansi	taar	Bangunji	taar	Buy	taar
TAL	Lozi	-talu	Dadiya	tal	Gola	tā'l
TAD	Mbere	-tadi			Sereer	tad-ak
TAS			Kulaal	tòòs	Bapen	ḃḃ-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	pě-sās
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	t̪faro	Kam	tshar		
CA	Bubi	-cha	Galke	cha-?a-		
HAT	Nkoya	-hatu			Manjak	go-hant
DER					Baga	der
					Mboteni	

*bVp*, *sVz*, *lVr*, *rVl*, *sVf*, etc. In diachronic perspective, 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, 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 in all the stages of its development from Proto-NC to contemporary languages can be typologically – phonotactically – explained.

Table 1.10: Reflexes of *\*tat-* in Niger-Congo (2)

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

Table 1.11: Distribution of different reflexes of \*tat- in the Niger-Congo families

	Bantu	Benue- Congo	Atl	Adam.	Bantoid	Gur	Mel	Kwa	Ubangi	Dogon	Kordof.	Kru	Ijo	Mande
TA	x	x		x	x	x		x	x		x	x		9
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	123

Table 1.12: Number of different phonetic structures for ‘3’ in 14 NC branches

	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

## 1.4 ‘Four’

Table 1.13: Niger-Congo stems for ‘4’

		Dogon nay(n)		Kordofan
Atlantic Nord: nak	Mande náání / nãñi	Gur naan	Ubangi naar	Adamawa naX, pɛn/ nɪŋ, nda
Mel	Kru na	Kwa na	Ijo néin	Benue-Congo nai
		Dogon kɛɛso		Kordofan -ɽɔŋ /-ɽando/ -rɔm?,-gálɪm)
Atlantic Bak: baakər/wakər, tasala	Mande	Gur	Ubangi ( syɔ), lu)	Adamawa
Mel Nord: ’-ɲkɪlɛ/- nɪɛ, Sud: hiɔl	Kru	Kwa	Ijo	Benue-Congo

Just like the term for ‘three’, the term for ‘four’ is exceptionally persistent in NC. It is represented by the same root in all the families (except for Mel and Kordofanian), as well as in the Western NC isolates, cf. Sua *b-nan*, Gola *tii-nàŋ*, Limba *ka-naŋ*. At the same time, a number of innovations are attested in some of the families (see the downmost segment of the chart) and in the Laal isolate, cf. *bĩ-sãn* (*bĩ-sãn?*) ‘4’.

This root is not present in Nilo-Saharan (including Songhai), nor in Afroasiatic or Khoisan. In light of this, the root can be viewed as one of the best isoglosses indicating the genetic relationship of languages within NC. Used together with the isogloss for ‘three’, it becomes a powerful means of classification, i.e. if the term for ‘three’ has (or goes back to) *t-* as the initial consonant in a given language, whereas the term for ‘four’ starts with *n-*, this language must belong to the Niger-Congo family. Hundreds of the NC languages match this description, while, as far as I am aware, none of the languages from other families meets these requirements.

There will probably be no objection from the specialists in the field to the statement that the main root for ‘four’ begins with *\*na-*, e.g. this form is reconstructed for Proto-Potou-Akanic-Bantu by John Stewart. At the same time, many languages show that the root initially included two vowels, *\*i* being the

second of the two. The major issue, however, is establishing whether the root included another consonant (i.e. whether *\*nai* or *\*naCi* should be preferred) and if so, what it was. Stewart suggests *\*na~ɲi~* ‘4’ as the Proto-Potou-Tano-Congo form (Stewart1983), but his reconstruction is not applicable to NC.

At the same time, the reconstruction of the proto-form for ‘four’ is not an easy task. The problem is that a given form does not define the languages it is attested in as members of the same group. Nearly every group has an inventory of phonetically similar forms (just like in case of ‘three’). The Bantu languages may provide a good illustration for this phenomenon.

The most frequently attested Bantu forms include *na*, *nai*, *nayi*, *ne*, *nei* and *ni* (six in total). They are found in 276 of 355 Bantu sources that include a form for ‘four’ available in our database. Their zonal distribution is as follows (Table 1.14):

Table 1.14: Distribution of the main n- forms for ‘4’ in Bantu zones

zone	na	nai	nayi	ne	nei	ni	SUM	SOURCES
A	13	3	2	6	1	7	32	52
B	31	8	10	7	1	1	58	65
C	2	2		2	18	1	25	28
D	1	1		4			6	14
E	4			4		1	9	19
F				9		3	12	13
G	2			18		1	21	26
H	7						7	11
J	10			15		1	26	27
K	6			7		1	14	15
L	6	1	2				9	12
M	3	1		11		5	20	20
N	2	3	2	2			9	12
P	2	2					4	11
R				3			3	7
S	7			14			21	23
SUM	96	21	16	102	20	21	276	355

As can be gleaned from the table, the six forms discussed above are commonly attested in our sources stemming from zones as diverse as C, F, J, M, and S. For instance, pertinent forms are attested in 26 out of 27 sources available in our



database for the J zone (the last source, namely the Luganda language, has *nya* ‘four’ that probably goes back to the same root).

The problem, however, is that this (or a nearly identical) set of forms is attested within the other NC families as well, cf. e.g. the Kwa evidence (Table 1.15):

Table 1.15: Main n- forms for ‘4’ in Kwa

Agni (Anyin)	n-na
Abron	n-nai
Baule	nu-ne
Eotile (Beti)	a-ni

The Adamawa evidence is as follows (Table 1.16):

Table 1.16: Main n- forms for ‘4’ in Adamawa

Tupuri	na
Mundang	nai
Gula	nay
Waja	ni

My suggestion is that the variety of similar forms attested in the majority of the NC branches may be due to the complex inter-relationship between the terms for ‘four’ and ‘eight’ in NC. We will return to this hypothesis later, in the section dealing with ‘eight’.

## 1.5 ‘Five’

The term for ‘five’ is typically based on the lexical term for ‘hand’ in Mel and Atlantic. At the same time, the term for ‘ten’ is often derived from ‘five’ or, like ‘five’, directly from ‘hand’ in the plural. Multiple examples illustrating this phenomenon will be provided below. At this point I will limit myself to merely stating that the attestation of this pattern throughout the NC branches is inconsistent. Thus, it is virtually unattested in Bantu (as well as in BC on the whole). According to NursePhilipppson1975 the Usseri dialect of Rombo (Bantu E) is a unique exception in this respect, cf. *ku-oko* ‘hand’ (Proto-Bantu *\*bókò*) yielding *ku-oko* (‘5’) and *ku-oko ka-vili* (‘10’, ‘5\*2’). At the same time, the reflexes of the Proto-Bantu roots for ‘five’ (*tanu*) and ‘ten’ (*i-kumi*) are attested in this language along with the irregular forms discussed above. These two patterns are barely attested in Kwa, Gur, Kru, or Ijo. On the contrary, they are common not only in Atlantic and Mel but also in Ubangi (Gbaya in particular), in some of the Adamawa languages, in a number of Kordofanian branches and possibly in Mande. In view of this distribution, the existence of these patterns in NC seems unlikely. Apparently, the terms for ‘hand’ should be considered when trying to establish the NC etymology for ‘five’ and ‘ten’.

Our discussion will start with the unrelated roots for ‘hand’ and ‘five’ attested within the same branch. Then we will turn to the evidence of those groups where both terms go back to the root for ‘hand’. This approach will allow the accumulation of data that will enable us to suggest a likely diachronic explanation for the phenomenon.

We will start with the Bantu evidence. The Bantu languages (like the majority of the NC groups in general) are characterized by the presence of multiple roots for ‘hand’ and ‘arm’. The most persistent of these according to BLR3 are the following roots (Table 1.17):

I would like to stress that these roots are virtually unattested in Bantu with the meaning ‘five’ or ‘ten’. According to BLR3, the only primary root for ‘five’ commonly attested in Bantu is *\*táànò*. At the same time, the root *\*dòngò*, which probably goes back to *\*dòngò* ‘line, row’ (zones: ABCDEGHJKLMNRS) deserves our attention as well.

The initial consonant in *\*táànò* is the same as in *\*tátò* ‘three’, which is probably a coincidence. However, this fact can still be used for establishing the genetic relationship of the NC forms for ‘five’. The possibility that the languages (or language groups) are related to the reconstructed Bantu forms is stronger if the terms for ‘three’ and ‘five’ attested in them have the same initial consonant. The

Table 1.17: Distribution of the stems for ‘hand’, ‘arm’ in Bantu zones

PB	meaning	regions (5)	zones (16)
bókò	arm; hand; front paw	5: NW SW Ce NE SE	14: A B C D E G H J K L M N R S
gànjà	palm of hand; main	5: NW SW Ce NE SE	14: A B C D F G H J K L M N P S
pí	palm of the hand; slap	5: NW SW Ce NE SE	14: A B D E F G H J K L M N R S
kónò	forearm; arm; hand; leg; hoof	4: SW Ce NE SE	10: E F G J K L M N P S
nàmà	limb: arm; leg; thigh	4: NW SW Ce NE	8: A B C E H L M R
jádà	nail (> finger > ‘hand’)		> ‘hand’ A D E F G J L N P S

following Bantu evidence (Table 1.18) is illustrative of this admittedly unconventional approach (further BC evidence will be quoted later in this chapter).

Table 1.18: Identical initial consonants in ‘3’ and ‘5’ in Bantu

	Language	‘3’ - *tátò	‘5’ - *táànò
Bantu-J	Rwanda	tatu	tanu
Bantu-B	Punu	reru	ranu
Bantu-E	Gusii	sato	sano
Bantu-G	Swahili	tatu	tano
Bantu-R	Herero	odatu	odano
Bantu-A	Bubi	ca	cio
Bantu-A	Tunen	lal	lan

This rule is irreversible, i.e. the diversity of the initial consonants is not indicative of either form not being a Proto-Bantu reflex (Table 1.19):

The fact that the same consonants are reflected differently may have several explanations, e.g. that the noun class prefixes (especially the nasal marker of class 9) may have impacted the process. A number of other phonotactic factors may also be involved (some of which are treated in detail in the section dealing with ‘three’).

Table 1.19: Different initial consonants in ‘3’ and ‘5’ in Bantu

	Language	‘3’ - *tátò	‘5’ - *táànò
Bantu-F	Bungu	tatu	(zi)sano
Bantu-G	Pogoro	tatu	mhanu
Bantu-S	Sesotho	taro	hlano
Bantu-G	Komoro	traru	canu
Bantu-D	Holoholo	satu	tano
Bantu-J	Haya	-satu	i-tanu
Bantu-K	Mbwela	-hatu	-tanu
Bantu-E	Kahe	si-radu	si-tanu
Bantu-A	Kpa	-ra	-tan
Bantu-G	Tikuu	-cacu	-tano
Bantu-K	Mwenyi	-atu	mu-tanu
Bantu-A	Balong	be-lal	be-tan
Bantu-B	Kele	-lali	-tani
Bantu-L	Mbwera	k-atu	-tanu
Bantu-E	Digo	-hahu	cano
Bantu-E	Taita	i-dadu	i-sanu
Bantu-N	Manda	ji-datu	mu-hanu
Bantu-S	Ronga	-rjarju	tlhanu

The pairs of BC terms with the same initial consonant attested outside Bantu will be our primary concern in further discussion. Some of them are quoted in the table below (Table 1.20): As can be gleaned from the table, the root *\*tanV* / *\*taVn* is systematically attested in nearly every BC branch, hence its reconstruction at the Proto-BC level seems certain. Moreover, it is widely attested in many other NC branches as well. The following forms of ‘three’ and ‘five’ (with the same initial consonant) are comparable to *\*BC* root (Table 1.21):

The Table 1.21 shows peculiar forms attested in one of the Southern Mel languages (Bom) that are virtually identical to the BC reconstructions. Thus, we have every reason to reconstruct the term for ‘five’ as *\*tan* (unrelated to ‘hand’) at the NC level. The distribution of this root is illustrated in the following chart (Table 1.22):

The attestations of this root in Southern NC (namely in BC, Kwa and Ijo) are

<sup>1</sup>Elugbe1987

Table 1.20: Identical initial consonants in ‘3’ and ‘5’ in Benue-Congo

BC	Language	‘3’ - *taT	‘5’ - *tan
Bantoid	Tiv	-tar	-tan
Bantoid	Mambila	tar	tin
Bamileke	Bamun	i-tet	i-ten
Chamba	Chamba	tera	tuna
Daka	Dirrim	tara	tona
Daka	Gandole	tara	tuna
Bamileke	Kom	tal	tain
Beboid	Dumbo	te	ten
Grassfieldss	Mmen	ta	taij
Jarawan	Jarawa	tat	towun
Nkambe	Mbe’	tei	tan
Idomoid	Gade	i-ta	i-to
Jukun	Proto-Jukunoid	*tat (i-)	*ton (i-)
Ikaan	Ikaan	tas	ton
Lower-Cross	Anaang	i-ta	i-tien
Upper-Cross	Olulumo	e-tal	e-tan
Kainji	Amo	n-tat	n-taun
Platoid	Horom	tat	ton
Ekoid	Nkem	i-ra	i-ron
Jarawan	Mboa	sai	sian
Edoid	Proto-Edoid	*i-caGi <sup>1</sup>	*i-ciNeni
Edoid	Ukue	e-rha	i-rhini
Edoid	Okpamheri	esa	iseni
Idomoid	Eloyi	e-la	e-lo
Jukun	Wapan	cara	cwana
Jukun	Jukun Jibu	sara	sona
Upper-Cross	Korop	bu-nan	bu-neg
Upper-Cross	Kiong	o-nan	o-nen
Platoid	Irigwe	ciæ	co
Platoid	Morwa	sat	suon

## 1 Reconstruction of Numerals in Niger-Congo

Table 1.21: Identical initial consonants in ‘3’ and ‘5’ in Niger-Congo

Family	Language	‘3’	‘5’
Kwa	Ewe	eto	ato
Kwa	Fon-Gbe	a-to	a-to, *ta
Kwa	Fon	a-tɔn	a-tɔ́n
Kwa	Tuwuli	ɛ-lale	e-lo
Kwa	Kebu	ta	to
Kwa	Igo (Ahlon)	ita	uto
Adamawa-Bua	Gula	tar	tiŋ
Adamawa-Bua	Bolgo	teri	tiso
Adamawa-Bua	Koke	teri	tiso
Adamawa-Mbum	Mambai	bi-saa	bi-sape’e
Ijo	Defaka	tato	tuno
Mel	Bom	tat	tan

Table 1.22: \**tan* ‘5’ in Niger-Congo

		Dogon		Kordofan dinin/ dulin?
Atlantic tɔk, tən?	Mande **tan? (> ‘10’?)	Gur tɔ	Ubangi	Adamawa sa?
Mel kə-ʈamaʈ (<*kə- ʈa ‘hand’?), tan?	Kru	Kwa ton	Ijo túnó	Benue-Congo tan/ ton

more systematic. In Western NC the root is reliably attested as well, despite the fact that the Northern Mel form *kə-ʈamaʈ* allows a two-fold interpretation (i.e. as a derivative of either *ʈam-* or \**kə-ʈa* ‘hand’).

The Bom form is a direct reflex of *tan* ‘five’. It bears reminding that the final velar in the Northern-Atlantic forms is regular. In the Gur languages, the pertinent form is attested in particular branches only. As attested in Western Mande, the form implies a semantic innovation, i.e. \*‘5’ > ‘10’. The relationship of the Kordofanian forms is not immediately apparent.

The distribution of the alternative reconstructible root \**nu/ nun* is described in the chart below (Table 1.23):

A comparison to Kru implies the labialization of dentals in the vicinity of a

Table 1.23: \**nun* ‘5’ in Niger-Congo

		Dogon núnéé(n)/ nü:(yn) / nûm		Kordofan
Atlantic	Mande	Gur nu(n)	Ubangi	Adamawa nu(n)
Mel	Kru mm	Kwa nu(n)	Ijo	Benue-Congo

back vowel. As the Dogon and Gur evidence suggests, the root is possibly derived from the term for ‘hand’. In Dogon the forms of ‘five’ and ‘hand’ differ in all languages / sources. Interestingly, the term that means ‘five’ in one Dogon language may be used with the meaning ‘hand’ in another (and vice versa, see **HochstetlerEtAl2004** cf. the following evidence (Table 1.24):

Table 1.24: ‘Hand’ and ‘5’ in Dogon

Group	Language	‘hand’	‘5’
Central	Tommo So	numɔ	nʔnɔ
Central	Donno So	numɔ	nɔʔ
Northern	Dogulu Dom	numɔ	nnɔ
South-East	Jamsay	numɔ	nui
Central	Toro So	nonnɔn	numonron
Central	Kolum So	nuwɛn	numu

In light of this, the fact that, according to some sources, similar distribution of the same root is attested in a number of Gur languages is intriguing, cf. e.g. the following data (Table 1.25):

This raises the question, are we dealing with direct Dogon-Gur contact or with the reflexes of an additional NC root for ‘hand’? The following roots may be considered potential correspondences: Proto-Bantu \**nàmà* ‘limb: arm; leg; thigh’ (Regions 4: NW SW Ce NE ; Zones 6: ABEHMR) or \**nðe* ‘finger, toe’ (Regions 5: NW SW Ce NE SE ; Zones 9: ADJKLMPRS), (cf. Bantu, zones MN – Nyiha-Malila-Lambya (**NursePhilipppson1975**) *i-nyove*, cf. (**Koelle1963**) Aku (De-foid) *ɲɔwɔ* ‘hand’. The Bak (Atlantic) root *ñen* ‘hand’, ‘five’ discussed above may belong here as well. The Gola root *nɔ̀nɔ̀nɔ̀* should also be mentioned here. The meaning ‘hand’ is not attested for this root in Kwa and Adamawa.

## 1 Reconstruction of Numerals in Niger-Congo

Table 1.25: ‘Hand’ and potential reflexes of *nun* ‘5’ in Gur

Group	Source	Language	‘hand’	5
Bariba	Koelle1963	Baatonum	nóma	nǒwu
Bwamu	BloemartsdeRasby2010	Bly	núumánnu	
Grusi	Koelle1963	Tem		nónūa
Grusi	CLNK1999	Kabiye		naanūwa
Grusi	Koelle1963	Kiamba	noon / noozi	noonuua
Grusi	Koelle1963	Sisaala Tumulung		ínǫm
Oti-Volta	Koelle1963	Mosi	nurɔ	nu
Oti-Volta	Koelle1963	Gurma	unu / inui	mu ~ mmu

The following Atlantic roots attest to the semantic development of ‘five’ (and consequently ‘ten’) < ‘hand’ (Table 1.26).

Table 1.26: ‘Hand’ > ‘5’ in Atlantic

Group	Language	‘hand’	‘5’	‘10’
Atlantic-Bak	Balant	f-cef / k-	cef	f-cef meen
	Kentohe			(‘whole hands’)
Atlantic-Bak	Bijogo	kɔ-ɔkɔ /	nde-ɔkɔ	n-rua-kɔ
	Kagbaaga	ɲa-akɔ		
Atlantic-Bak	Bijogo	kɔ-kɔ /	ɲu-duβ-kɔ	ɲɔ-rúnɲa-kɔ
	Kamona	ɲa-kɔ		
Atlantic-Bak	Mankanya	ka-nyɛn	ka-nyɛɛn	e-nyɛn
Atlantic-Bak	Manjak	ka-ñen	ka-ñen	ka ñen
Atlantic-Bak	Pepel	ɲenɛ	ɲenɛ	dise-ɲenɛ
Atlantic-North	Nyun	si-lax	ci-lax	haa-lax
	Djibonker			
Atlantic-North	Nyun	ci-lax / xa-	ci-lax	xa-lax
	Gujaxer			
Atlantic-North	Biafada	gə-bəda /	gə-bəda	
		ma-bb-		
Atlantic-North	Jaad	ko-bəda	ko-bəda	

This data is especially interesting in view of the BC evidence discussed above.



As we have seen, the phenomenon of ‘five’ and ‘ten’ being based on the term for ‘hand’ is attested in both Atlantic groups (Bak and Northern). Moreover, this pattern is observable in a wide variety of roots with the meaning ‘hand’ attested in the languages under study (e.g. five roots with this meaning are attested in eight languages represented in the table above; the derivation pattern is the same in each case). In view of this, it is not surprising that the reconstructed NC root is not traceable in Atlantic.

The same pattern is also attested in the Northern Mel languages (that are in contact with Bak) for ‘five’ (but not for ‘ten’), cf. (Table 1.27).

Table 1.27: ‘Hand’ &gt; ‘5’ in Northern Mel

Group	Source	Language	‘hand’	‘5’
Temne-Baga-Landuma	<b>Wilson2007</b>	Baga Koba	kə-tsa/ε-	kə-tsa-mat
Temne-Baga-Landuma	<b>Ganong1998</b>	Baga Sitemu	kε-ca	kə-ca-mət
Temne-Baga-Landuma	<b>Wilson2007</b>	Landuma	kə-ca/cə-	kə-caa-mət
Temne-Baga-Landuma	<b>Wilson2007</b>	Temne	kə-ta/mə-	ta-math

However, we may be dealing with the secondary alignment of the terms for ‘hand’ and ‘five’. The pattern CV-stem-VC (with CV- and -VC being a noun class prefix and suffix respectively) is characteristic of this language group, e.g. the Temne form may go back to *ta-m-ath* with the lexical root *\*-mV-* as its base. This pattern could also explain the similarity between the Temne terms for ‘five’ and ‘ten’: in this language *tɔfɔt* ‘10’ probably goes back to *tɔ-f-ɔt* and hence to the NC root *\*fu* ‘10’.

Some of the Atlantic languages (e.g. various Joola and probably Proto-Joola as well) developed a separate root for ‘five’, while the term for ‘ten’ still remained a derivative of ‘hand’. As expected, this root corresponds to Southern NC *\*tan/ ton* ‘5’ discussed above (Proto-Atlantic: *\*tok* ‘five’: Kasanga-Kobiana *ju-roog*, Sereer *be-tak / be-tuk / be-tik* (cf. also Limba *bi-sɔhi* ; Sua *sungun*), cf. Table 1.28.

The etymological link between the terms for ‘five’ and ‘ten’ and their source (‘hand’) is not always explicit, e.g. different roots for ‘hand’ are attested in some of the sources for Mankanya-Manjak (Atlantic) and Temne (Mel), along with the derived form for ‘five’. Such innovations are quoted in bold in the table below (Table 1.29).

Some of the forms of the term for ‘five’ go back to the root *\*ko* in a number of the Ubangi languages (and possibly in some of the Mande languages as well, see Chapter 3 for details). Here we may be dealing with a NC root, cf. e.g. ‘hand’:

Table 1.28: 'Hand' > '10' in Joola (Atlantic: Bak)

Language	'hand'	'5'	'10'
Joola_Banjali	ga-ɲen / gu-ɲen	fu-tox	gu-ɲen
Joola_Fogny	ka-ɲen / u-ɲen	fu-tək	u-ɲen
Joola_Gusilay	ga-ɲen / u-ɲen	fu-tək	u-ɲen
Joola_Kasa	ka-ɲen	hu-tək	ku-ɲen
Joola_Kasa_Esuulaalu	ka-ɲen	hu-tək	ku-ɲen
Keeraak	ka-ɲen-ak / ʊ-ɲen-aw	hʊ-tək	kʊ-ɲen
Joola_Kwaatay	ɛ-ɲomu	hu-tək	si-ɲomu
Joola_Kwaatay	ɛ-mɔŋo	hu-tək	su-muŋo
Joola_Mlomp	ɛ-bɛ:s	ɲa:-suwarɲ	sɛ-bɛ:s

Table 1.29: 'hand' > '5'/'10' in some Atlantic and Mel languages

Branch	Language	'hand'	'5'
Atl.-Centre-Manjak	Mankanya	ka-nyen	ka-nyɛɛn
Atl.-Centre-Manjak	Manjak	ka-ɲien	ka-ɲien
Atl.-Centre-Manjak	Manjak	<b>kádʂāg</b>	kányan
Atl.-Centre-Manjak	Mankanya	<b>úlōl</b>	kányɛn
Atl.-Centre-Manjak	Manjak Bassarel	<b>pëndänd</b>	kañan
Atl.-Centre-Manjak	Manjak Tame	<b>wúepalōl,</b> <b>pl. n-gípalōl</b>	kényān
Temne-Baga-Landuma	Temne	kə-ta / mə-	ta-maɖ
Temne-Baga-Landuma	Temne	<b>a-loŋk (i), ma-</b>	ɖamaɖ
Temne-Baga-Landuma	Temne	<b>à.loŋk</b>	-tàmath

Proto-Gbaya *kɔ̃*, Proto-South Mande *kɔ̃*, Proto-Eastern Mande *gɔn* (?), Dida (Kru) *kɔ̃*, etc.

The following Kordofanian terms that attest to the development of ‘hand’ > ‘5’ are also noteworthy: Dagik (Kordofanian) *si-s-ɜlɔ* ‘5’ (lit: ‘one hand’): “The *si* in 5 comes from the word ‘hand’. So 5 is ‘one hand’”<sup>2</sup>, Acheron *zəguŋ zulluk* (lit: ‘one hand’): “The number ‘five’ is literally ‘one hand’: *zəguŋ* = ‘hand’, *z-ulluk* = ‘one’”<sup>3</sup>.

To summarize, the primary root for ‘five’ (\**tan*) probably existed in Proto-NC. Over time it was independently replaced with the derivatives of ‘hand’ in some branches and various languages. In turn, the original term for ‘hand’ was replaced with innovations (with the term for ‘five’ in particular) in a number of languages, cf. Atlantic *rib/ ʔi:p*, Mel *wan/wen*, Mande *dúuru/ sɔ́ru*, Kru *gbə / gbo*, Gur *mwan/ bwa*, Ubangi *du(w)/ lu(w)*, Kordofanian *ŋer-/ per-*. As a rule, these innovations (not quoted here exhaustively) are only attested in particular branches of the families under study.

<sup>2</sup>John Vanderelst, <https://mpi-lingweb.shh.mpg.de/numeral/Dagik.htm>

<sup>3</sup>Russell Norton, <https://mpi-lingweb.shh.mpg.de/numeral/Acheron.htm>

## 1.6 ‘Six’

The explicit pattern ‘6=5+1’ is present in the vast majority of the families. Primary terms for ‘six’ are attested in some of the NC families (or, more precisely, in their particular branches). However, they cannot be reconstructed at the NC level (see Chapter 3 for their detailed treatment). Selected forms of this kind include Atlantic *paag/paaj* (‘7=6+1’), Kwa *golo / kolo, kua, ciε* (‘7=6+1’), Adamawa *jup, gu*, Ubangi *zala/ zya*, Dogon *kuro/ kule*, Gur *do(b)*, Mande *t(s)um?* (the examples are quoted by family without further detail). The pattern ‘6=3 redupl.’ is rarely attested. It is found in BC (possibly as a Proto-BC innovation attested in Bantoid, Cross, Edoid, Kainji?, and Platoid) and Kordofanian only.

## 1.7 'Seven'

The main pattern is '7=5+2' (or '7=X+2' if the term for 'five' is replaced with an innovation). Primary roots are rare, being attested in BC (Defoid *\*byē* (cf. Edoid *ghie?*), Idomoid *renyi* (cf., however, Ikaan *h-ráněfi* ('6+1')), Adamawa (*bir/ bil, rɪŋ, nbutu*), Ubangi (*sílânā, lè-rəzi*), Dogon (*suli/ soli/ soye*), Gur (*pɛ(n)*) and Atlantic Bak (*jand/ jaan?/ cand* (Pepel)).

The rare patterns of '7=6+1' and '7=4+3' are limited to Atlantic Bak, Kwa, BC Platoid, and Kordofanian.

## 1.8 ‘Eight’ (‘Four’ and ‘eight’)

In the majority of the NC families the term for ‘eight’ is historically based on the term for ‘four’ (with the exception of Mel, Kru, Dogon, Mande and Western NC isolates).

The pattern ‘8=4+4’ is normally implemented via the reduplication of the root for ‘4’. In some cases an ‘entire’ reduplication (affecting the conjunction and the noun class marker) is employed (Table 1.30).

The reduplication can also be ‘partial’ (as a rule the reduction of the first syllable is involved), cf. Table 1.31.

This pattern can also be used when the original root for ‘four’ is replaced by another one, cf. the Balant (Bak) evidence: *tahla* ‘4’ ~ *ta-ta(h)la* ‘8’. The same is observable in Yungur (and possibly in Burak (Adamawa)), cf. *net* ‘4’ ~ *nat-at* ‘8’ (Boyd1989).

Sometimes ‘eight’ is derived from ‘four’ not via the reduplication, but by means of a simple replacement of CL.SG with CL.PL (or by adding the Pl. marker), cf. Table 1.32.

In Dii (Adamawa-Duru) a step-by-step replacement of classes is used as a derivation mechanism, i.e. ‘2’ > ‘4’ > ‘8’: *i-dú* ‘2’ > *nda-ddú* ‘4’ > *ka-ʔa-nda-ddú* ‘8’.

A rare pattern is ‘8=4\*2’, with the direct involvement of the term for ‘two’, cf. Viemo (Gur) *jumĩ* ‘4’, *niinĩ* ‘2’, *jumĩ-jo niinĩ* ‘8’.

When considering the reconstruction of ‘four’, it should be noted that if the term for ‘four’ (on which a reduplicated term for ‘eight’ is based) has any vowel other than [a] (typically [e] or [i]), the reduplicated form either preserves the vowel present in ‘four’ or has [a] in the first syllable. This mechanism is confirmed at least in the case of Bantu (Table 1.33).

The latter fact leads to at least two conclusions: 1) the reduplication mechanism was used to derive ‘eight’ from ‘four’ at the Proto-Bantu level; 2) [a] that which is preserved in ‘eight’ should be reconstructed in the first syllable of ‘four’, where it was lost.

Moreover, there is a considerable body of Bantu examples of a Proto-Bantu root being preserved in the reduplicated term for ‘eight’, but lost in the term for ‘four’ (Table 1.34).

One of the factors that could explain the emergence of the second nasal in the term for ‘four’ is the alignment of ‘four’ and ‘eight’ by analogy, followed either by the replacement of the term for ‘eight’ with a composite term (‘5+3’ or ‘10-2’, see Table 1.35) or with an innovation (Table 1.36):

The evidence presented above strongly suggests that the pattern ‘8=4 redupl.’

Table 1.30: ‘8’ &lt; ‘4+4’ (entire reduplication)

Branch	Languages	‘4’	‘8’
Bantoid-Ekoid	Ekoi	ni	e-ni-ga-ni
Bantoid-Ekoid	Kwa	ni	a-ni-ka-ni
Bantoid-Ekoid	Ndoe	ne	be-ne be-ne
Bantoid-Ekoid	Nkem	ni	a-ni-gi-ni
Bantu-Central-E	Chaga	na	nana
Bantu-Central-E	Embu	nya	i-nyanya
Bantu-Central-E	Kamba	nya	nya-nya
Bantu-Central-E	Kikuyu	nya	i-nyanya
Bantu-Central-G	Sango	na	m-nana
BC-Edoid	Okpamheri	ni	e-ni-e-ni
BC-Edoid	Urhobo	ne	e-nene
Bantoid-Grass	viya	na	ge-nana
Bantoid-Jarawan	Mbula-Bwazza	i-ne	i-ne i-ne
Bantu-Central-D	Enya	na	ce-nana
Bantu-NW-B	kande	na	ge-nana
Bantu-NW-B	Lumbu	na	di-nana
Bantu-NW-B	Punu	na	i-nana
Bantu-NW-B	Sira	na	gi-nana
Bantu-Central-J	haya	na	omu-nana
Bantu-Central-J	Nyankole	na	om-nana
Bantu-Central-J	Nyoro	na	om-nana
Bantu-Central-J	Gwere	na	mu-nana
Bantu-Central-J	Nkore-Kiga	na	mu-nana
Bantu-Central-J	Soga	na	mu-nana
BC-Cross	Alege	ne	e-nene
BC-Cross	Bokyi	je	je-ri-je
BC-Cross	Kukele	na	i-na-mi-na
BC-Bantoid	Esimbi	mō-jī	mō-jī-ō-jī
BC-Jukunoid	Mbembe	nyε	é-nyεnyε
Bc-Ikaan	Ikaan	nā <sup>1</sup> / nā	nà:ná <sup>1</sup> / nà:ná
Adamawa-Fali	Fali	ná:n	nàn nán
Adamawa-Duru	Gəunəm	náárək	náárək àp náárək
Gur-Southern	Lamba	nasa	nasì-nasa
Gur-Southern	Lyele	na	nana
Laal	Laal	ḃīsān	ḃīsān.ḃīsān

Table 1.31: '8' < '4+4' (partial reduplication)

Branch	Language	'4'	'8'
Bantoid-Jarawan	Kulung	i-nin	i-ni-nin
Bantu-NW-B	Enenga	nai	e-na-nai
Bantu-NW-B	Myene	nayi	e-na-nayi
Bantu-NW-B	Orungu	nayi / i-nayi	e-na-nayi / na-nayi
BC-Eastern-Platoid	Boyawa	nas	na-nas
BC-Eastern-Platoid	Kwanka	nas	na-nas
BC-Eastern-Platoid	Idong	enar	na-nar
BC-Eastern-Platoid	Kadara	er-nar	ir-na-nar
Ijo	Nembe	i-nei	ni-nei
Atl-Centre	Balant	tahla-	ta-tahla-
Adamawa	Yungur	kurun	kun-kurun

Table 1.32: '8' = 4PL

Branch	Language	'4'	'8'
Kwa-Nyo	Lelemi	í-né	máá-né
Kordofanian Heiban	Warnang	ɣɛ̀lám̀làn	ɣelamlaaŋ-ɔ
BC Platoid	Ikulu	ín-nāā	nín-nāā
Adamawa Leko-Nimbari	Yendang	nâ:t	ḃṵ-lá-nâ:t
Adamawa Mbum-Day	Niellim	ɲɛ́ní	twā:-ɲɛ́ní
Adamawa Waja-Jen	Waja	nu	wu-nii
Ubangi Sere-Ngbaka-Mba	Gbanzili	ḃṵ-nā	sá-nā
Gur Grusi	Delo	a-naara	gya-naara
Gur Grusi	Tampulma	a-naasi	ɣmɛ-naasa



Table 1.33: ne/ ni '4' ~ nane/ nani '8' (Bantu)

Zone	Language	'4'	'8'
Proto	PB	ne	nane
NW-B	Vove (Pove)	nai	nanai
NW-B	Sira	ne	gi-nane
NW-B	Punu	ne	yi-nane
NW-B	Lumbu	ne	nane
NW-C	Kela	nei	i-nane
NW-C	Kusu	nem	e-nanem
NW-C	Ombo	nei	i-nanei
Central-E	Pokomo	ne	nane
Central-E	Zanaki	i-nye	i-nyanye
Central-F	Bende	i-ne	mu-nane
Central-F	Kimbu	ji-ne	mu-nane
Central-F	Mbugwe (Irangi)	ne	i-nane
Central-F	Nyamwezi	ne	m-nane
Central-F	Sukuma	ne	nane
Central-F	Sumbwa	i-ne	m-nane
Central-G	Bondei	ne	nane
Central-G	CAsu (dial.)	ne	nane
Central-G	Kami	ne	nane
Central-G	Komoro	ne	nane
Central-G	Kutu	ne	nane
Central-G	Ngulu	ka-ne	m-nane
Central-G	Pangwa	i-ne	nane
Central-G	Shambala	ne	m-nane
Central-G	Swahili	ne	nane
Central-G	Tikuu	ne	nane
Central-G? E?	Tubeta (Taveta)	i-ne	nane
Central-G	Zigula	ne	m-nane
Central-J	Hunde	i-ne	mu-nane
Central-J	Konzo	ne	omu-nane
Central-J	Luhya	ne	mu-nane
Central-J	Masaba	ci-ne	si-nane
Central-J	Nande	ne	omu-nane
Central-J	Vinza	ka-ne	mu-nane
Central-M	Mambwe	vi-ni	ci-nani
Central-M	Pimbwe	i-ne	nane
Central-M	Rungu	vi-ni	ci-nani

Table 1.34: '8' < '4' ~ '4' is lost (Bantu)

Zone	Language	'4'	'8'
Central-G	Mbugu	hahi	nane
Central-G	Bena	tayi	fi-mu-nana
Central-G	Hehe	tayi	i-mu-nana
Central-G	Ndamba	mceci	nani
Central-G	Pogoro	msei	nani
Central-H	Kikongo	kuya	e-nana
Central-H	Yaka	ya	nana
Central-H	Yombe	ya	di-nana
Central-N	Manda	cece	nani
Central-N	Matengo	sesi	nani
Central-N	Mpoto	sesi	nani
Central-P	Matuumbi	sese	nani
Central-P	Ngindo	cece	nani

Table 1.35: '8=4+4' > '8=5+3'

Group	Language	'4'	'8' ('5+3')
Atlantic	Baga Fore	si-neŋ / ci-neŋ	sak-tet
Atlantic	Baga Mboteni	i-neŋ	ib-ader
Atlantic	Wolof	jenet	jurom-ŋeta
Gur	Birifor (dial.)	anan	anu-ni-ata
Gur	Teen	nan	to sanr
Mande	Vai	nani	sog sakpa
Adamawa	Karang	niŋ	tòŋ ndók sé'de ('10-2')

Table 1.36: ‘8=4+4’ &gt; ‘8’ innovated

Family	Languages	‘4’	‘8’
Bantu-A	Bafo	benin	wam
Bantu-A	Bankon	bi-nan	mwam
Bantu-A	Fang	nɪn	mwom
Bantu-A	Ndambomo	li-nanji	li-mwabi
Bantu-B	Kota	napi	mwabi
Bc-Platoid	Mabo	nen	hur
Dogon	Tene Kan	nani	silá
Dogon	Tene Kan	nani	sira
Kwa	Abron	nain	ɲocie
Kwa	Akan (Akuapem Twi)	anan	awotcyɛ /tw/
Kwa	Baule (Baoulé)	nan	nmocue
Kwa	Foodo	naŋ	dukwe / dukoi
Kwa	Mbato	ne-ni	o-gbi
Mande	Mandinka	náani	segi
Mande	Looma	náanĩ	dosawa

was already in use at the Proto-NC level.

It should be noted that in those languages where this reduplication mechanism (or the pattern ‘8=4PL’) is observable most clearly, another pattern is often used along with ‘8=4+4’, namely ‘6=3+3’ (or ‘6=3PL’) (Table 1.37):

As expected, numerous languages that belong to different families exhibit a variety of patterns that are reused along with the one discussed above (including the general pattern ‘8=5+3’ as well as ‘8=10-2’ and even ‘8=6+2’). It seems, however, that such a wide distribution of this pattern (‘8=4 redupl.’) within the NC languages is genetic rather than typological.

Primary roots for ‘eight’ are also attested. However, their attestations are usually limited to one or two families or to particular branches within a family, cf. e.g. ‘8’ in Defoid (BC) \**jo/ ro* (cf. in Kainji *ro/ ru*), Kwa *kwe/ kye*, Kordofanian *bɔ, tɛŋi-*, Mande *seki/ segi*, Dogon *sele/ sagi* (< Mande ?), *gá(a)rà*, Atlantic Bak \**ʋas-*. These forms (as well as some additional ones) are interpreted as local innovations.

Table 1.37: '8' < '4', '6' < '3'

Branch	Language	'3'	'6'	'4'	'8'
Bantoid-Ekoid	Ekoi	e-sa	e-sa-g-asa	e-ni	e-ni-ga-ni
Bantoid-Ekoid	Kwa	e-sa	a-sa-ka-su	i-ni	a-ni-ka-ni
Bantoid-Ekoid	Ndoe	be-ra	be-ra-ba-ra	be-ne	be-ne be-ne
Bantoid-Ekoid	Nkem	i-ra	i-ra-ra	i-ni	a-ni-gi-ni
Bantu-E	Embu	i-tatu	i-ta-tatu	i-nya	i-nya-nya
Bantu-E	Kamba	i-tatu	ta-tatu	i-nya	nya-nya
Bantu-E	Kikuyu	i-tatu	i-ta-tatu	i-nya	i-nya-nya
Bantu-F	Nyamwezi	datu	ta-dato	ne	m-na-ne
Bantu-F	Sukuma	datu	ta-datu	ne	na-ne
Bantu-G	Gogo	datu	m-ta-datu	ni	mu-na-ne
Bantu-G?E?	Tubeta (Taveta)	tatu	ta-datu	i-ne	na-ne
Bantu-G	Zigula	ka-tatu	ta-datu	ne	m-na-ne
BC-Edo	Okpamheri	e-sa	e-sa-sa	e-ni	e-ni-e-ni
BC-Cross-River	Bokyi	bé-ciaat	ɲá-ciaat	bé-ɲii	ɲí-rii-ɲi
BC-Cross-River	Alege	é-ce	é-ce-e-ce	é-ne	ee-né-ne

## 1.9 ‘Nine’

The main pattern for ‘nine’ (‘9=5+4’) is self-explanatory. This is the only pattern that can be reconstructed for Proto-Niger-Congo.

The alternative pattern ‘9=10-1’ is much less common, whereas the pattern ‘9=6+3’ (attested in Atlantic Bak) is exceptionally rare. The Platoid pattern ‘9=12-3’ seems to be unique, cf. Birom, ‘15=12+3’, ‘9=minus 3’, ‘10=minus 2’. Primary roots are attested in those languages (branches) that have a full set of primary terms covering the sequence from ‘one’ to ‘ten’ (which is a rare case), e.g. Bantoid *bukV* (if indeed primary), Akpes *ɔ̌-kpɔ̌lɔ̌f(i)*, Defoid *\*sá(n)*, *dà* (cf. Edoid *cien/sin*), Igboid *totu/tolu*, Ubangi *kùsì*, *me-newá*, Laal *yàṅjāṅ*, Dogon *túwɔ́*, Mande *kònonɔ́to/kòṇnɔ́ndɔ́(n)* (historically perhaps ‘10-1’).

## 1.10 ‘Ten’

The root *\*pu/ fu* is the most likely candidate for the NC reconstruction. The distribution of its reflexes is shown in the chart below (Table 1.38).

Table 1.38: *\*pu/ fu* ‘10’ in Niger-Congo

		Dogon		Kordofan
Atlantic pok	Mande pu/ fu	Gur fu/ po	Ubangi ḃú / fu?	Adamawa boo/ fu?
Mel pu/ tɔ-f-ɔt ?	Kru pu	Kwa fo / wo	Ijo	Benue-Congo pu/ fu

The roots listed in this chart are obviously related. The root is lacking in Kordofanian, where a variety of terms for ten are attested, e.g. *tu(l)*, *rakpac*, *fəŋən*, *tiəɾum*, 5PL. This probably indicates that in Proto-Kordofanian the root for ‘ten’ was not present. The Dogon form *\*pérú/ pélú* has the same initial consonant, but our evidence is inconclusive as to whether it is related to the roots above. Finally, the Ijo form (w)ójí allows a twofold interpretation. If it is taken as (w)ó-jí based on *\*ji*, it is comparable to *ziìyà* ‘10’ attested in the Gola isolate. Alternatively, it can be analysed as a complex root *\*(w)o* ‘10’ plus *ji* (< *\*1*). If so, it may be related to the roots quoted above (or at least to one of its allomorphs (?) attested in Kwa).

The presence of forms with the voiced **b**- in Adamawa-Ubangi requires an explanation. The evidence suggesting a connection between the **b**- and **f**- forms attested in these languages is insufficient. In view of this, it can only be noted that a similar phenomenon is observable within the Mande family: the form *\*bù* is reconstructed in the Southern group of the South-Eastern Mande branch, whereas in Western Mande (as well as in the Eastern group of South-Eastern Mande) the reconstructed form is *\*pu/fu*.

It should be noted that the Adamawa root with the initial voiceless labial is only marginally attested (e.g. in Munga (*fuə*) and Pere (*fób*)).

Raymond Boyd tentatively suggests that *fob* is related to the main Adamawa root *\*kop*: «The Kutin group has *fóp* which may be related to *\*kóp*» (Boyd 1989). However, an alternative explanation exists. A brief study of the Adamawa number systems shows that numerical terms attested within this family (unlike those found in other NC families) often end in **-p** or **-b**. The Tula system, one of the first quoted by Boyd in his excellent article, may serve as an example (Table 1.39).

Table 1.39: Labial suffix in Tula numerals

1	-iñ	6	nukuñ
2	rɔp	7	nibiñ
3	táa	8	na’á-rəp
4	naa	9	túrúkup
5	nu	10	kwəp

The final -p in ‘eight’ is easily explainable (possibly due to ‘8=4\*2’). However, at least in the case of ‘two’ and ‘ten’, the final -p is attested in non-compound terms. In his discussion of the final -p in the Adamawa terms, Boyd suggested that we may be dealing with the suffix *\*(a)p* (or *\*(a)b*, with the devoicing characteristic of a reduced consonant inventory in the final position). <...> The same suffix also appears in group 1 in *\*naar-ap* ‘eight’, derived from *\*naar* ‘four’. <...> Compare this situation with ‘Bantoid’ Vute: *būrúrp* ‘two’, *nà:sùrp* ‘four’ (Boyd1989). Furthermore, he challenges Kay Williamson’s opinion on whether this morpheme was an original suffix or a suffix that developed out of a noun class prefix. The most important result of this discussion is that the suffix *\*-p/-b* found in numerical terms allows us to trace the Adamawa forms directly to NC *\*pu/po* without the intermediate *\*kop/kob*. As for the isolated Adamawa forms of *bo* ‘ten’, Boyd suggests a Chadic origin for them, although alternatively they may be related to the similar Ubangi root and reflect the NC root *\*pu / fu*.

The main Adamawa root *\*kop/kob* ‘10’ should be discussed in a wider NC context as well. In view of the secondary nature of the final -p/-b in Adamawa (see above), this root is comparable to the NC roots *ko* ‘ten ; hand’.

Direct BC parallels for this root (with the final labial) should be discussed first. We refer here to the hypothetical relationship of a number of forms discussed in Chapter 3, including Delta-Lower-Cross *-kəp/ du-op/ du-ob* (Dimmendaal1978 *\*lùgòp*) (cf. Bendi *kpu* ‘10’, nearby *fo/ hwo*), Yukuben-Kuteb (Jukunoid) *kuwub*, Kainji *\*kop / ʔup / kpa* (together with *\*pwa/ pa*), and Platoid *\*kop*. This evidence suggests that more attention should be paid to the reconstruction of the allomorph *\*kop* in both Proto-BC and Proto-Adamawa. This root should probably be compared to the Kru root *kɔgba* ‘10’, unless it is a non-compound root that goes back to *ko* (see below).

In view of Boyd and Williamson’s interpretation of the final labial as a suffix, the forms quoted above should probably be treated together with the root *ko* ‘10’, which is sporadically attested in multiple families. As noted above, it most proba-

bly goes back to the lexical root *\*ko* ‘hand’, that represents one of the alternative Proto-NC reconstructions of this term. Its distribution with this meaning is as follows:

First of all, it is reconstructed by Moniño for Proto-Gbaya as *kɔ́* ‘hand’. This root is also attested in Mande (at least in the Southern group of the South-Eastern Mande branch, cf. Vydrin’s evidence: Proto-South-Eastern Mande *\*kɔ́* ‘hand, arm’). In Kru, this root is attested not only in the Eastern group (Dida *kɔ́* ‘hand’), but in the Western group as well (Glio-Oubi *hō*, Krumen *hō*). Finally, it is (admittedly only marginally) attested in Bantoid (as an alternative to the wide-spread root *kómi* ‘10’): according to Larry Hyman (in Paulin1995) this root is distinguishable in Kom (*ə-kə*) and Narrow Bantu, e.g. in zones B (Mpur *kɔ*, Yansi *kwɔ*) and E (Mashami *oko*, Meru *uko*, NursePhilippson1975). The Limba root *koh-* ‘10’ probably belongs here as well.

It is difficult to say whether this evidence is sufficient for the Proto-NC reconstruction. However, when choosing between the two possibilities for the reconstruction of the term for ‘ten’ (i.e. from *\*pu/ fu* and *\*ko*) the first one should be preferred.

Among other roots relevant to our discussion, the following two roots (whose attestations are not limited to one family) are of interest: Gur *gba/kpa* ‘10’ (cf. the BC root *gwo/ jwo*) and Kwa *du* ‘10’ (possibly related to the Adamawa root *d(u)o*; cf. also Kordofanian *ru* and Gur *nu/ nyu?*). The latter root may be compared to Bantu *\*dòngò* ‘10’. It is attested in seven zones (i.e. EGJMPR according to BLR3, but a number of attestations from D.62 are available, hence it is found in all five regions). BLR tentatively suggests a Bantu etymology for this root (‘*spécilisation de “ligne” dòng?*’). However, it has parallels in other BC branches, namely in Cross River (Connell1991) and probably Idomoid (Table 1.40).

The use of numerous other roots for ‘ten’ is limited to one family, i.e. they are apparent innovations, such as in Bantoid *kum/kam* ‘10’ (Bantu *kómi/ kámá*). The latter form (that sometimes coincides with the term for ‘hundred’) has an internal Bantu etymology: its tentative relationship to the lexical root meaning ‘touch’ is assumed in BLR 3 (BLR3: ‘see also *kóm* ‘touch’ - zones DHJLM’). However, the nasalization of the final segment in the Bantoid proto-form cannot be excluded. If this process indeed took place, this form becomes comparable to *\*ku(b)* as well as others discussed above.

Other isolated forms for ‘ten’ include Atlantic (*n*)*taaj*, *taim*, *-suwan*, Mel *wi-tɔʔ*, Western Mande *tan* (< *\*ʔ5ʔ*), Gur *kɛ(n)*, Kwa *bula* (cf. Ubangi *bale*), Ubangi *busa*, *sui*, Kordofanian *tu(l)*, *di*, *rakpac*, *fəŋən*, *tiərum*, Adamawa *kutu(n)* (< *\*kutu(n)*, cf. Laal *tūū*, Kordofanian *ɬAA*, Sua *tɛŋi* etc.



Table 1.40: Parallels for Bantu \**dòngò* ‘10’ in Cross River and Idomoid

Branch	Language	Form
Cross River	Ebughu	lùgò
Cross River	Efai	dùgù
Cross River	Ekit	dùgò
Cross River	Enwang	lùgù
Cross River	Etebi	dùgù
Cross River	Ilue	lògù
Cross River	Okobo	lùgù
Cross River	Oro	lùwù
Cross River	Uda	lùgù
Idomoid	Eloyi	dộn· & ndộn· (Koelle1963)

### 1.11 Large numbers ('twenty', 'hundred' and 'thousand')

It is better to treat large numbers together for the following reasons:

First, these terms were probably lacking in Niger-Congo, so it comes as no surprise that they are often borrowed from European languages, Arabic, Hausa, Lingala or other "languages of influence".

Secondly, these roots are often identical, i.e. the root that means 'thousand' in one language may mean 'hundred' or even 'ten' in another. Some of the forms simply denote 'a large number'. The well-known migrating root *keme* that has the meaning 'hundred' in the majority of the Mande languages may be used with the meaning 'eighty' or even 'sixty' in other Mande languages.

However, each of the roots has its own characteristics.

In the majority of the NC languages, the term for 'twenty' goes back to lexical roots that mean 'person', 'leader', 'body', 'head', 'grain', 'sack' and 'large number'. Numerous examples of this kind are discussed in Chapter 3. The etymology of those terms for 'twenty' that seem to be primary at the synchronic level should be sought with this in mind.

It can be safely stated that the terms for 'hundred' and 'thousand' were absent in Proto-Niger-Congo. Thus, the pattern 'twenty' = 'person' remains the only reconstruction possibility for large numbers in Proto-Niger-Congo.

1.12 Proto-Niger-Congo

The reconstruction of the Proto-Niger-Congo number system may be summarized as follows (Table 1.41):

Table 1.41: Proto-Niger-Congo numeral system

1	ku-(n)-di (> ni/ -in), do, gbo/ kpo	7	5+2
2	ba-di	8	na(i)nai (< 4 redupl.)
3	tat/ tath	9	5+4
4	na(h)i	10	pu/fu,
5	tan, nu(n)	20	< ‘person’
6	5+1		

This table summarizes our discussion. However, it is tempting to apply our conclusions to the evidence pertaining to particular families in order to identify the most archaic families, groups and branches within NC. Such a review of data within a wider NC context could also help, enhancing the intermediate reconstructions suggested in Chapter 3.









