

# **An Evaluation of Niger-Congo Classification**

**Kenneth S. Olson**

**SIL International**

**2004**

## **Contents**

Contents

Abstract

1. Introduction

2. Niger-Congo classification: major subgroupings

3. Kwa and Benue-Congo

4. Bantu

4.1. Inclusion of Bantu within Niger-Congo

4.2. Narrow versus Wide Bantu

4.3. The unity of Narrow Bantu

4.4. Bantu origins

5. Adamawa-Ubangi

5.1. Adamawa-Ubangi external relationships

5.2. Adamawa-Ubangi internal relationships

6. Discussion

6.1. Data and documentation

6.2. Methodology

6.2.1. Method of resemblances

6.2.2. Lexicostatistics

6.2.3. Shared innovations

6.2.4. The comparative method

6.2.5. Other methods

7. Conclusion

Acknowledgements

References

## Abstract

This paper examines the presently received classification of the Niger-Congo language family, with the aim of evaluating the classification and suggesting areas of further research. We find that the exact placement of Ijoid, Kru, and Dogon within the Niger-Congo tree remains to be determined and that the unity of each of the following nodes needs to be better established: Atlantic, (new) Kwa, (new) Benue-Congo, Wide Bantu, Narrow Bantu, and Adamawa-Ubangi. The macro classification is based on the method of resemblances, with refining based on lexicostatistics paired with the examination of shared innovations. The comparative method has mostly been applied at a micro level to specific subgroups. A comprehensive reconstruction of Niger-Congo, including the establishment of sound laws, remains the major future task.

## 1. Introduction

The Niger-Congo language family is the largest language family in Africa, comprising about 1,400 languages found in Sub-Saharan Africa. Its northern border stretches along a relatively straight line between Senegal and Kenya. In the south, in present-day Namibia, Botswana, and South Africa, Niger-Congo languages are intermixed with Khoisan and some Indo-European languages. The other major families to be found in Africa are Nilo-Saharan, Afro-Asiatic, and Khoisan (see figure 1).

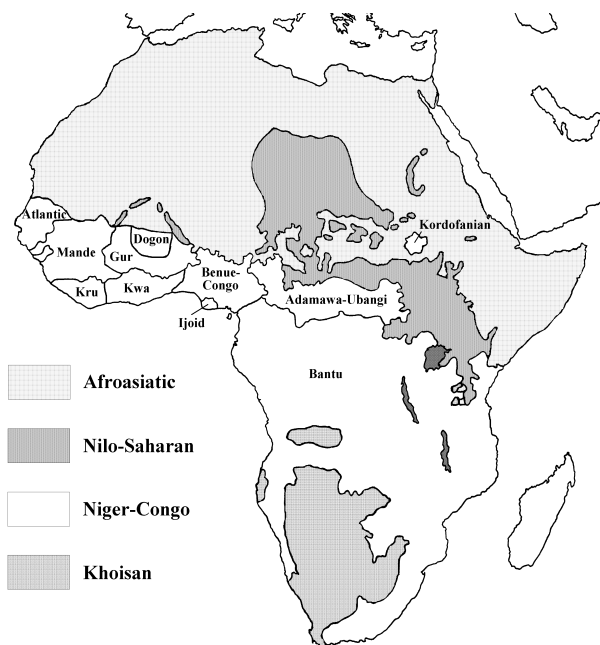


Figure 1. Classification of African languages. Data from Greenberg (1970) and Williamson and Blench (2000).

This paper aims to give an overview of the genetic classification of the Niger-Congo language family. In addition, I focus on certain salient issues in the classification, offering a critique of how these issues have been addressed in the past and indicating further research that is necessary in order to clarify outstanding problems.

Attempts at classifying the languages of Africa date from the early part of the nineteenth century. Both Cole (1971) and Williamson (1989a) offer good reviews of the development of this field from that period until the middle part of the twentieth century.

Since 1950, four particular works have been very influential in updating and summarizing the field of Niger-Congo classification. First, in the 1950s, Greenberg published a series of articles on the topic employing a controversial method called the METHOD OF RESEMBLANCES (which he later called *multilateral comparison*), which culminated in his 1963 work, *The Languages of Africa*. (I had access to the third edition: Greenberg 1970.) Greenberg's work defined the Niger-Congo family and refocused the direction of the field towards a classification based on genetic criteria. At the same time, it raised questions about what constitutes an appropriate methodology for establishing genetic relationship. Regardless of these issues, the Greenberg classification has taken on the role of a useful referential classification in much the same way that Guthrie's (1948) alpha-numerical classification has become the standard of reference for Bantu, even though Guthrie's subgroupings are no longer widely accepted as genetically accurate.<sup>1</sup>

The second major work on Niger-Congo language classification was *Current Trends in Linguistics 7: Linguistics in Sub-Saharan Africa* (Sebeok 1971). This volume had a broad scope, encompassing language classification, linguistic history, language policy, and a host of other topics related to language and linguistics in Sub-Saharan Africa. With respect to Niger-Congo language classification, it only made minor changes to Greenberg's work. This is probably due to the fact that only a short amount of time had elapsed since Greenberg's work and it had the broader goal of summarizing the state of the art rather than pushing the theoretical envelope.

The third major work on Niger-Congo language classification was an article by Bennett and Sterk (1977) entitled "South Central Niger-Congo: A reclassification." Their study does not claim to be an overview of Niger-Congo classification per se, but the conclusions that they draw have had widespread ramifications for the field. In it, they study the Niger-Congo family using a combination of lexicostatistics and evidence from commonly shared innovations.

The fourth major work on Niger-Congo language classification is *The Niger-Congo Languages* (Bendor-Samuel and Hartell 1989). This is the most recent definitive statement on Niger-Congo classification. It incorporates much of the classificatory research on the family from the preceding two decades and standardizes the nomenclature for the family (Williamson 1989a:18–20). One significant change worth noting is the broadening of the Niger-Congo umbrella to include the Kordofanian languages. In Bendor-Samuel and Hartell, "Niger-Congo" is effectively the same as what Greenberg alternatively calls "Niger-Kordofanian" or "Congo-Kordofanian."

---

<sup>1</sup>Despite the uncertainty about Guthrie's subgroupings, his reconstruction of Proto-Bantu is still widely accepted.

Implicit in the notion of language classification lies the question “On what basis is the classification made?” The received view among linguists is that the relatedness of languages is defined in terms of historical evolution from a common parent language, i.e., genetic relationship. They also tend to agree that the comparative method of historical reconstruction is the most “scientific” method for establishing genetic relationship (cf. Greenberg 1995, Newman 1995). However, in practice other methods have been employed, most likely due to their ease of use. Heine (1980a) identifies lexicostatistics, the identification of shared innovations (usually lexical), and Greenberg’s method of resemblances as alternative means for identifying a genetic classification. These methods have been used extensively on African languages, so in a sense, a discussion about the proper classification of Niger-Congo must out of necessity be intertwined with a discussion of the advantages and disadvantages of different methods of classification. I will look more closely at the question of methodology in §6.2.

## 2. Niger-Congo classification: major subgroupings

Presently, the most widely accepted general classification of Niger-Congo is in Williamson (1989a:21) as shown in (1).

- (1) Niger-Congo classification
  - A. Kordofanian
  - B. Mande [2]
  - C. Atlantic-Congo
    1. Ijoid (?) [4h]
    2. Atlantic (?) [1]
      - a. North
      - b. Bijago
      - c. South
    3. Volta-Congo
      - a. Kru (?) [4a]
      - b. (New) Kwa [4b]
      - c. (New) Benue-Congo [4c, 4d, 4e, 4f, 4g, 5, 6A3]
      - d. Dogon (?) [3b]
      - e. North Volta-Congo
        - i. Gur [3a, 3c, 3d, 3e, 3f, 3g]
        - ii. Adamawa-Ubangi [6 (except 6A3) ]

For reference, Greenberg’s corresponding subdivisions are given in square brackets in (1). Question marks indicate nodes whose placement is speculative. For the sake of consistency, I use Williamson’s nomenclature for the names of the subgroups throughout this paper, but the reader should be aware that the literature has not been consistent in the use of node labels. For example, Williamson’s “Niger-Congo,” “Kwa,” and “Benue-Congo” are significantly redefined from Greenberg’s use of these terms. On the other hand, Williamson’s “Atlantic” is essentially the same as Greenberg’s “West Atlantic.”

Williamson bases much of her classification on Bennett and Sterk (1977). Bennett and Sterk's study consists of two parts. First, they set up the gross subgroupings of Niger-Congo using lexicostatistical percentages. They analyze an 87-item wordlist in fifty languages and examine data from previous studies. Second, they look more closely at the question of the relatedness of Kwa and Benue-Congo. For this portion of the study, they analyze lexicostatistical percentages for a 145-item wordlist in 150 languages. In addition, they try to support each subgrouping by identifying shared lexical and phonological innovations. I will look at the second part of their study in detail in §3. Here, I will briefly examine Williamson's classification, starting with the largest units and working to the right on the chart.

Greenberg includes Mande and Atlantic-Congo within Niger-Congo, but excludes Kordofanian. On the other hand, Williamson follows Bennett and Sterk in proposing a three-way split between the three branches. Her reasons are twofold: (1) Bennett and Sterk find that Mande and Kordofanian have only a few lexical similarities with Atlantic-Congo, and (2) Schadeberg (1981, cf. Williamson 1989a) shows that in one case Atlantic-Congo is closer to Kordofanian than to Mande in that the Kordofanian noun class system shows systematic resemblance to Atlantic-Congo, whereas Mande has no noun class system. Williamson concludes that Mande and Kordofanian both split from the rest of Niger-Congo at an early date, but that the split was not necessarily simultaneous.

Next, Williamson makes a subsequent split under Atlantic-Congo between Ijoid, Atlantic, and Volta-Congo.<sup>2</sup> The split between Atlantic and Volta-Congo is based on Bennett and Sterk's lexicostatistical percentages. In fact, their percentages suggest that the three sub-branches of Atlantic are so divergent that they are probably each coordinate branches with Volta-Congo. Williamson leaves this an open question. The placement of Ijoid at this level is more tentative. Greenberg places Ijoid within Kwa, but according to Bennett and Sterk, its lexicostatistical score is less than 18 percent with all languages except (new) Benue-Congo, and Bennett and Sterk attribute this similarity to borrowing. Williamson places it at this particular level for impressionistic reasons—"it seems to be outside Volta-Congo" (1989a:18).

Besides Ijoid, Greenberg also places Kru within Kwa. However, Williamson removes it from Kwa based on Bennett and Sterk's lexicostatistical percentages. I will discuss its placement in more detail in §5. What is interesting about the positions of Mande, Atlantic, Ijoid, and Kru in the classification above is that their new positions confirm the impressions of Greenberg himself:

The affiliation of Kru and Ijo to the Kwa group is to be considered tentative. Kwa and Benue-Congo are particularly close to each other and in fact legitimate doubts arise concerning the validity of the division between them. On the other hand West Atlantic seems more remotely related to the other group and Mande the most distant of all. (Greenberg 1970:39)

---

<sup>2</sup>The term Volta-Congo was coined by John Stewart; cf. Stewart (1976).

This quote also addresses the relationship of Kwa and Benue-Congo, which will be the subject of §3. The relationship between Kru, Gur, and Adamawa-Ubangi will be discussed in §5.1. It should be noted that Williamson's placement of Dogon is purely speculative.

The important point to highlight in this section is that the major subgroupings of Niger-Congo as posited by Williamson rely heavily on the lexicostatistical work of Bennett and Sterk. Shared innovations only come into play in defining nodes lower down in the tree, as we shall see for example in §3 for (new) Kwa and (new) Benue-Congo. In addition, no mention is made of the use of the comparative method at the higher levels.

### 3. Kwa and Benue-Congo

The relationship between Kwa and Benue-Congo has generated much discussion in the Niger-Congo classification literature. First, recent scholarship has cast doubt on the original division between the two branches as set up by Westermann and retained by Greenberg. Second, there is some evidence that the two groups should be considered a single branch under Volta-Congo, rather than two. I will examine these two points in turn.

Greenberg includes within old Kwa several languages which are today no longer considered to be a part of new Kwa: Kru [4a], Yoruba [4c], Nupe [4d], Bini [4e], Idoma [4f], Igbo [4g], and Ijo [4h]. In Williamson's (1989a) classification, Yoruba, Nupe, Bini, Idoma, and Igbo have all been moved to (new) Benue-Congo, while Kru and Ijo are now considered neither (new) Kwa nor (new) Benue-Congo.

Williamson (1989a:11) offers evidence for rejecting the old division between the two groups. First, she claims that there are no single lexical items that occur in all of the branches of old Benue-Congo and not in old Kwa. Greenberg (1970:32) suggests that the form *\*ana* 'child' is an old Benue-Congo innovation, but Williamson claims that the form should be *\*gwana*, and that variants of this form occur in Igbo and Yoruba. Greenberg states, "Many other such innovations could be cited," but unfortunately, he does not provide them. Second, there are no noun class innovations that occur in all the branches of old Benue-Congo which are unique to old Benue-Congo. Finally, Elugbe and Williamson (1977) argue that a certain putative typological distinction, the predominance of CV roots in Kwa, is invalid.

The accepted realignment of Kwa and Benue-Congo is based primarily on Bennett and Sterk's (1977) lexicostatistical study. In rejecting Greenberg's Kwa/Benue-Congo division, they note that the distinction between the two was originally on typological grounds. That is, the (old) Kwa languages were said to have no or limited concord systems, whereas the (old) Benue-Congo languages were said to have functioning concord systems. However, Bennett and Sterk claim that in reality the languages form a typological continuum from one extreme to the other rather than a dichotomy. In addition, their lexicostatistical and shared innovation evidence reject the distinction as well.

Instead, Bennett and Sterk offer a regrouping of the languages into what they call "Western South Central Niger-Congo" (i.e., new Kwa) and "Eastern South Central

Niger-Congo” (i.e., new Benue-Congo). They claim that the lexicostatistical percentages support this division, although they give neither their data nor their analytical charts for this part of their study. In addition, they give four sample lexical isoglosses that demonstrate a lexical border between the two groups:

(2)	(new) Kwa	(new) Benue-Congo	
	*tã	*tát(o)	‘three’
	*jɛ ~ jɔ	*bɛɛĩ	‘breast’
	*je ~ jɛ	*kónì	‘firewood’
	*kɔŋgo	*toŋ	‘neck’

Stewart (1989:218–219) examines the Kwa terms and concludes that the word for ‘firewood’ comes the closest to representing a Kwa innovation, though it is not convincing to him. He points out that the Tano subgroup of Kwa is clearly defined based on phonological innovations. However, “...no phonological innovation can yet be assigned with any confidence to Kwa...” It is clear from his prose that he doubts the integrity of the (new) Kwa group.

Williamson (1989b:249ff.) examines the Benue-Congo terms and concludes that only the word for ‘neck’ appears to be a Benue-Congo innovation. She proposes seven additional words that may be possible innovations, but she points out that “not a single one of them is attested in every division of Benue-Congo...” Her conclusion is that “Proto-Benue-Congo existed as a single language (if at all...) for a very short period of time.” Thus the integrity of (new) Benue-Congo is also doubtful.

It is interesting to note that Bennett and Sterk do not claim that these two sets of words represent common innovations for the two language groups, but it appears that Stewart and Williamson interpret them as making that claim. Indeed, Bennett and Sterk (1977:255) find no common innovations for (new) Benue-Congo, and they do not provide evidence for any (new) Kwa innovations.

In summary, (new) Kwa and (new) Benue-Congo are listed in the Niger-Congo chart in §2, but their status as units is by no means established.

The second question regarding Kwa and Benue-Congo is whether the two groups should be considered a single branch under Volta-Congo, rather than two. Even though Greenberg lists Kwa and Benue-Congo as separate subgroups of Niger-Congo, he himself notes that the two are closely related to the extent that perhaps they should be considered a single unit rather than two separate branches.

Bennett and Sterk (1977) posit a group comprising (new) Kwa, (new) Benue-Congo, and Ijoid, which they call South Central Niger-Congo (SCNC). They state, “SCNC is, as will be shown, a well-defined group.” They mean by “well-defined group” that there is clear evidence for lexical innovations found only within the group. Strangely, the promised data in support of the SCNC node are never given in the paper, as they focus instead on justifying the subgroupings under SCNC. In addition, they cast doubt on the inclusion of Ijoid within SCNC, since there are few obvious cognates.

Williamson (1989a) decides to abandon the SCNC node based on Schadeberg’s (1986) re-evaluation of Bennett and Sterk’s data using Nearest Neighbor, Furthest



Neighbor, and Branch Average methods. However, it appears that more work needs to be done to clarify this conclusion.

#### 4. Bantu

The term “Bantu” can be traced back to Bleek (1858, cf. Williamson 1989a:4).<sup>3</sup> Linguists have devoted much study to the Bantu language family, which covers most of the Niger-Congo region to the south and east of Cameroon. However, certain fundamental questions remain. For example, what exactly comprises Bantu, and having determined that, is it in fact a genetic unity? In this section, I will look first at how Bantu relates to the other Niger-Congo languages. Then, I will turn my attention to the questions of the domain and integrity of Bantu.

Below is the Bendor-Samuel and Hartell (1989) classification for the Bantoid group, including Bantu. Data are taken from Watters (1989:412), Hedinger (1989:425), and Watters and Leroy (1989).

- (3) Bantoid [5.D.]
  - A. Northern
    - 1. Mambiloid
    - 2. Fam
    - 3. Tiba
    - 4. Dakoid [6.A.3]
  - B. Southern (Wide Bantu)
    - 1. Tivoid
    - 2. Jarawan
    - 3. Mbe
    - 4. Ekoid
    - 5. Mamfe
    - 6. Beboid
    - 7. Wide Grassfields
    - 8. Tikar
    - 9. Ndemli
    - 10. Mbam
    - 11. (Narrow) Bantu
      - a. Northwest
      - b. Other
        - i. Central
        - ii. East

---

<sup>3</sup>Bleek included a much wider group of languages under Bantu than is usually included today. In fact, his Bantu group resembles today’s Niger-Congo group.

#### 4.1. Inclusion of Bantu within Niger-Congo

In the mid-nineteenth century, several researchers noted the genetic relationship between Bantu and West African languages (cf. Cole 1971, Williamson 1989a:4–6, Watters 1989:403). However, during the first half of the twentieth century, most scholars began to treat Bantu as a separate language family on typological grounds and were often influenced by paleontology and biology. This tradition became so established that when Westermann (1927) posited a “West Sudanic” group comprising the languages west of Lake Chad, he excluded Bantu from this group, even though he himself identified resemblances between Proto-West Sudanic and Proto-Bantu, both in basic vocabulary and in noun class structure (cf. Greenberg 1970:31 ff.). Westermann (1949) later mentions explicitly the genetic relationship of the two groups.

Greenberg was thus not the first researcher to identify the genetic relationship of Bantu to West Sudanic. Rather, his major contribution was identifying how Bantu is situated within West Sudanic. He places it within the Benue-Congo subgroup, and then renames the entire group Niger-Congo.<sup>4</sup> I will use the term “West Sudanic” to refer to the Niger-Congo languages excluding Bantu.

Greenberg uses evidence from Westermann to support his claim that Bantu should be included in West Sudanic. First, he points out that there are many resemblances between Proto-West Sudanic and Proto-Bantu in terms of fundamental vocabulary. These data show regular sound correspondences. Second, he notes that the noun class affixes of Proto-West Sudanic resemble those of Proto-Bantu both in form and meaning. In fact, he notes that the percentage of nouns in Bantu that show resemblance to Proto-West Sudanic is significantly greater than the percentage of nouns in English that can be related to Proto-Indo-European.

At the time of Greenberg’s work, there was resistance to recognizing a genetic relationship between Bantu and the West African languages. For example, Guthrie (1962) argued for maintaining a distinction between the two groups. He claimed that the two groups do not display the same regularity of correspondence as that which is found within Bantu. Rather, he attributed the similarities to borrowing. He hypothesized a “Pre-Bantu” people who lived between the Ubangi and Chari rivers. According to him, some of these people moved south and developed Proto-Bantu, while others moved west and lost their language, but some words were loaned into the languages of that region.

Subsequent research contradicts Guthrie’s claim by demonstrating that there are indeed regular sound correspondences between certain West African languages and Bantu. For example, Stewart (1965, cited in Schachter 1971), presents evidence of regular sound correspondences between Central Akan and reconstructed “Common Bantu.” The accepted view today is that Greenberg is right in placing Bantu within the broader framework of Niger-Congo.

---

<sup>4</sup>It appears that Greenberg (1970) was unaware of the previous work establishing a relationship between Bantu and West African languages. He writes (p. 37): “...all previous writers...accept the Bantu-Sudanese dichotomy as fundamental in African linguistics.”

Greenberg (1970:33–37) gives five reasons for rejecting a borrowing hypothesis, such as the one put forward by Guthrie. First, Greenberg shows that there is a high degree of agreement in the tonal systems of Efik and Proto-Bantu, to the extent that a borrowing hypothesis would be suspect. Second, certain borrowings would have dubious explanations. For example,

Bantu has a verb *vi-ala* ‘to give birth’. As a derivative from *vi* ‘child’ + *ala*, a verbal formative, it is quite understandable. But *\*vi* ‘child’ does not exist as a word either in Bantu or the Semi-Bantu languages, whereas it is the ordinary word for child practically everywhere else among the West Sudanic languages, and a Proto-West Sudanic form *\*bi* is generally assumed. The verb formation, on the other hand, is peculiar to Bantu. For the West Sudanic languages to have borrowed this word, would have required an analysis of the form *vi-ala* into its constituent elements and the abstraction of the form *\*vi-* in the meaning ‘child’. (Greenberg 1970:35)

Third, the supposedly borrowed words tend to be fundamental vocabulary, those terms that are putatively least subject to borrowing. Fourth, “some common Bantu words are found widely in West Sudanic, others are not found at all.” This situation is most easily explained if we consider the former to be Proto-Niger-Congo terms, while the latter are shared innovations unique to Bantu. If Bantu is not a part of Niger-Congo, it would be difficult to explain how the former terms were borrowed throughout West Sudanic. Greenberg’s fifth reason is that the supposed transitional languages are in fact Bantu. This issue will be dealt with in the next section.

#### 4.2. Narrow versus Wide Bantu

A second way in which Greenberg differs from Guthrie is in the question of what exactly constitutes Bantu. He points out that certain supposedly transitional languages in the northwest Bantu border area resemble Bantu more closely than they do the other Benue-Congo languages; he cites Bamum, Bali, Banen, and Jarawa as examples. He claims, “these languages show lexical innovations characteristic of Bantu languages as against the remaining Benue-Congo languages.” Unfortunately, he does not provide examples. Greenberg considers these languages to fit in the northwestern subgroup of Bantu that includes Duala and Yaunde.

Williamson (1971, cf. Watters 1989) picks up on this wider use of the term “Bantu.” She distinguishes between “Wide Bantu,” which is slightly more inclusive than Greenberg’s use of Bantu, and “Narrow Bantu,” which is essentially Guthrie’s Bantu. Wide Bantu is the same as Watter’s Southern Bantoid in the chart in (3).

The question that concerns us here is not which grouping should receive the label “Bantu,” but whether either Wide or Narrow Bantu as defined do indeed form a genetic unity. As mentioned above, Greenberg claims that his Bantu shows shared innovations, but he does not provide examples, so establishing Wide Bantu as a genetic unity requires further research. The question of the unity of Narrow Bantu will be discussed in the next section.

### 4.3. The unity of Narrow Bantu

Even though there is a long history of considering Narrow Bantu a group, its genetic unity has recently been called into question. Stewart (1976) notes that Guthrie does not give evidence in the form of common shared innovations to support the unity of Narrow Bantu.

Heine (1973, cf. Watters 1989) was the first to question the integrity of Narrow Bantu. He notes that Bube, a Narrow Bantu language (Guthrie's zone A.30), is more distant from the rest of Narrow Bantu than are Tivoid and Ekoid. Then, presentations at the Conference on Bantu Expansion in 1977 added evidence that additional languages needed to be excluded from Narrow Bantu. Heine (1980b:336) specifically mentions zones A.30, A.40, A.60, A.80, A.90, D.20, and D.30 as likely candidates for extraction from Narrow Bantu.

Bennett and Sterk's (1977) study casts additional doubt on the unity of Narrow Bantu. Based on isoglossic evidence, they divide Narrow Bantu into two groups, Equatorial Bantu, consisting of zones A, B, C, and part of D, and Zambesi Bantu, consisting of the rest of Narrow Bantu. According to their analysis, Equatorial Bantu is more closely related to Jarawan, Ekoid, and Mbam-Nkam (a Grassfields group), whereas Zambesi Bantu is more closely related to Tivoid. Their evidence for this division are the isoglosses in (4).

(4)	Jarawan, Ekoid, Mbam-Nkam, Equatorial	Tiv, Zambesi	
	*wɔk	/ *uygwa	'hear'
	*-ɔŋ	/ *nyúélé	'hair'
	*-bay	/ (no agreement)	'red'

Watters (1989:409) notes that *\*uygwa* is found outside of Bantoid, so while perhaps it is an isogloss, it cannot be an innovation. In addition, *\*-bay* is found in only a subset of the languages of its group. Thus the evidence for this particular division is rather weak, and if so, then their argument for separating Equatorial and Zambesi Bantu is diminished significantly. Bennett and Sterk admit, "our data on this area have not yet been fully analyzed" (1977:261).

Additional studies cast further doubt on the integrity of Narrow Bantu. For example, using lexicostatistics, Bastin, Coupez, and de Halleux (1983) include Tivoid within Narrow Bantu, placing it most closely to Equatorial Bantu. However, they do not provide evidence from shared innovations for this claim.

Although certain questions remain about each of the aforementioned studies, it is clear that the traditional notion of what comprises Narrow Bantu has been called into question. Watters (1989) concludes that there are still many questions remaining with respect to Bantoid internal classification, including the integrity of both Wide and Narrow Bantu.

#### 4.4. Bantu origins

One outgrowth of the study of Bantu classification has to do with conclusions regarding the origin of the Bantu-speaking people (cf. Olivier 1979, Phillipson 1977, Heine 1980a). The Bantu languages have been somewhat of a puzzle, since they cover a wide area geographically, and yet they are closely related linguistically. Greenberg's placement of Bantu within Niger-Congo provides a linguistic clue to solving this puzzle. He suggests (1970:38) that the original location of the Bantu-speaking people was southeastern Nigeria or western Cameroon in the area where the Benue-Congo languages (excluding Bantu) are located.<sup>5</sup> From this source, he posits a rapid migration of Bantu-speaking people to the east and south, quickly spreading out to fill the areas where Bantu is presently spoken. This rapid expansion provides an explanation for the geographically distant yet linguistically close nature of the Bantu language area.

Three studies in the early 1970s support this claim. Henrici (1973) and Heine (1973) both performed lexicostatistical studies of Bantu, while Heine, Hoff, and Vossen (1977) used the method of resemblances. The details differ, but all three studies conclude that the Bantu languages in the northwestern part of the Bantu area have the most distant relationships linguistically, whereas those in the eastern half are closest linguistically. The most straightforward interpretation of this is that it indicates "a slow penetration of the equatorial forest region, followed by an extremely rapid expansion of Bantu people into the savanna regions to the east and south" (Olivier 1979:10).

Researchers have attempted to go into more detail about the exact nature of the Bantu expansion, positing one or more migrations to account for linguistic and archaeological data. However, due to space restrictions, I will not elaborate on this topic here. The following references offer more information about the topic: Hinnebusch (1989:454), Olivier (1979), Heine (1979, 1980a), Phillipson (1977), Bennett (1983b), and Williamson (1989b).

### 5. Adamawa-Ubangi

The Adamawa-Ubangi subgroup of Niger-Congo is located in an area centralized around the country of Central African Republic. It reaches to most of the surrounding countries, including Cameroon, Nigeria, Chad, Sudan, and the Democratic Republic of Congo. Greenberg (1970) named it Adamawa-Eastern, Samarin (1971) suggested the name Adamawa-Ubangian, and Boyd (1989) finally settles on the present name.

In this section, I will look first at the external relationships of Adamawa-Ubangi, discussing its place within Niger-Congo and its relationship to closely related languages. Then, I will turn to its internal relationships, focusing on the Ubangi group.

For reference, the classification that Boyd (1989) gives for Adamawa-Ubangi is given in (5).

---

<sup>5</sup>Johnston (1886) had earlier suggested that Bantu was originally spoken in this region on the grounds that "most common Bantu word-roots referring to the environment indicated a forest environment rather than a savanna environment" (Olivier 1979:8).

- (5) Adamawa-Ubangi classification
- Adamawa [6A]
    - I. Leko [6A2], Duru [6A4], Mumuye/Yendang [6A5], Nimbari [6A12]
    - II. Mbum [6A6], Bua [6A13], Kim [6A14], Day
    - III. (?) Waja [6A1], Longuda [6A10], Yungur (?) [6A7], Jen [6A9]
  - Ubangi [6B]
    - I. Gbaya [6B1]
    - II.
      - A. Banda [6B2]
      - B. Ngbandi [6B3]
      - C.
        - 1. Sere [6B6]
        - 2.
          - a. Ngbaka [6B5]
          - b. Mba [6B7, 6B8]
    - III. Zande [6B4]

He states that two groups, Daka [6A3] and Fali [6A11], which Greenberg classified as Adamawa, should be excluded from the group. In addition, he does not state how Greenberg's Kam [6A8] group fits into his classificatory scheme.

### 5.1. Adamawa-Ubangi external relationships

There are two issues of external relationship that need to be considered with respect to the Adamawa-Ubangi group. First is the question of whether Adamawa-Ubangi should be included in Niger-Congo. Second is the question of the relationship of Adamawa-Ubangi to its closest neighbors.

Westermann (1927) did not include Adamawa-Ubangi within his West Sudanic family. One of Greenberg's (1970) major claims, beside that of including Bantu in Niger-Congo, is that Adamawa-Ubangi should be included in Niger-Congo as well. He offers two types of evidence to support this claim.

First, Greenberg demonstrates that there are strong similarities both in form and meaning between the noun class systems found in parts of Adamawa-Ubangi and those found in the rest of Niger-Congo (e.g., the Bantu noun prefixes). In certain subgroups of Adamawa, the similarities are striking. For example, Longuda has a well-developed suffixal system marking both singular and plural. The plural *ba* of the personal class is found in forms such as *nji-re* 'man', *nji-b* 'men'. Longuda has a class marking parts of the body that come in pairs, e.g., *ju-a* 'breast', *ju-la* 'breasts'. It also has the *ma* class that marks mass nouns and does not have a distinction between singular and plural, e.g., *tu-ma* 'blood', *bɔ-ma* 'salt'.

Within the Ubangi group, Mba and Mondunga also have suffixes that show resemblance in form and meaning to Niger-Congo, but the situation is not as clear for the rest of the group. Greenberg claims that there are vestiges of the Niger-Congo noun class system to be found in the prefixes of the rest of the Ubangian languages, and he uses Banda to demonstrate this:

The situation in Banda is typical of most of the group. We have vowel prefixes in *o-tu* 'ear', *o-wu* 'nose', *a-ma* 'mouth', and similar words. That these are prefixes is, of course, suggested by comparative data: *to*, for

example, is the morpheme meaning ‘ear’ throughout most of the Niger-Congo family, combined with some classificational affix. That the *a-* in *a-ma* is a prefix is further shown within Banda itself by the occurrence of *ma* in place of *a-ma* in certain compounds. (Greenberg 1970:12–13)

However, it is not clear that these vowels are indeed prefixes. Olson and Schrag (2000) argue that the initial vowel in these forms is due to a minimal word restriction in Banda that requires a noun to have at least two syllables. Thus a noun such as *tu* is ill-formed in the language, and must be expanded to two syllables, in this case by the addition of a vowel in initial position. In compound nouns, the minimal word restriction is already satisfied, and thus forms such as *ma* are free to occur without augmentation. In addition, the form of the initial epenthetic vowel depends entirely on the form of the following vowel. As a result, the initial vowel does not bear direct resemblance to any specific prefix in the rest of Niger-Congo. Since resemblance in form is one of Greenberg’s requirements for relationship, a genetic affiliation cannot be established between Banda and the rest of Niger-Congo by this evidence.

The only clear nominal prefix in Banda is *a-*, which marks animate plural (e.g., *gbolo* ‘child’, *agbolo* ‘children’), and which Greenberg suggests is an additional resemblance with the rest of Niger-Congo. However, it is not clear that this prefix corresponds directly in meaning with any of the general Niger-Congo noun class affixes, and thus it is weak evidence for genetic affiliation.

Greenberg rightly states that the “absence of the affixes does not prove lack of connection.” For example, he points out that the nominal affixes have been entirely lost in Mande and parts of Kwa, but that these groups are still considered to be part of Niger-Congo. On the other hand, evidence for the affiliation of Ubangi with Niger-Congo from the noun class system is in reality weaker than Greenberg’s portrayal, and thus casts some doubt on the affiliation.

The second type of evidence that Greenberg uses to argue for the inclusion of Adamawa-Ubangi in Niger-Congo is lexical resemblances. He provides a 49-item wordlist<sup>6</sup> that shows putative cognates between Adamawa-Ubangi and the rest of Niger-Congo. Unfortunately, Greenberg only gives a sampling of his data,<sup>7</sup> so it is impossible to determine precisely the cognate scores that would result from his data. In the case of Banda, only eight out of the forty-nine words (16 percent) are listed as showing resemblance with other Niger-Congo languages, a percentage that does not reach Greenberg’s own criterion (20 percent) for removing chance or symbolism from consideration as the source of the resemblance (cf. Greenberg 1957).

On the other hand, in the case of Gbaya, another Ubangian language, 43 percent of the words are listed as showing resemblance (21 out of 49), giving firmer evidence of historical relationship. But it should be noted that the Adamawa-Ubangi languages have

---

<sup>6</sup>Although the wordlist contains items from all the branches of Niger-Congo, Greenberg labels it “Adamawa-Eastern Comparative Word List.” A more accurate name would be “Niger-Congo Comparative Word List.”

<sup>7</sup>From footnote 19: “In general I have restricted citations to three languages from each group. Both the number of languages cited and the number of etymologies on this list could be very greatly extended” (p. 40).

short roots for the most part, a factor that increases the possibility of chance as the source of resemblance.

On a related note, some scholars have pointed to a connection between Adamawa-Ubangi and the Nilo-Saharan family, which is immediately to the north and east geographically. Boyd (1978) identifies a substantial number of resemblances between the two groups, to the extent that he claims it is difficult to determine whether a given Adamawa-Ubangian language belongs to Niger-Congo or to Nilo-Saharan. In the particular case of Banda, Cloarec-Heiss (1995) points out that it shares many properties with the Central Sudanic branch of Nilo-Saharan. She describes phonological, lexical, morphosyntactic, and syntactic convergences between the two groups. Her particular hypothesis is that Proto-Banda was a pidgin, with Central Sudanic as the substrate and Ubangian as the superstrate (i.e., the lexifier). What is clear is that there are significant resemblances between the two language families that require an explanation in any account of their linguistic history.

In summary, then, Greenberg's evidence for the inclusion of Adamawa-Ubangi within Niger-Congo is substantial, but it is weaker than he claims, especially for the Ubangi branch. In addition, there are significant resemblances with Nilo-Saharan that must be accounted for.

The second major issue with respect to the external relationships of Adamawa-Ubangi is its relation to its nearest linguistic neighbors, Gur and Kru. Based on lexicostatistics, Bennett and Sterk (1977:249–250) tentatively posit a group called North Central Niger-Congo (NCNC) that includes Gur, Adamawa-Ubangi, and “probably” Kru. Lexical innovations offer weak support—\**sõ* ‘two’ is found in Kru and parts of Adamawa-Ubangi, while \**du* ‘head’ is found in Kru and Adamawa-Ubangi and has the form \**yu* in Gur. All three groups have suffixal noun class markers instead of the typical prefixes in most of Niger-Congo.

Williamson (1989a:15) casts doubt on the inclusion of Kru within NCNC. She points out that Schadeberg's (1986) reanalysis of the lexicostatistical data never shows Kru grouped with Adamawa-Ubangi and Gur. Second, she quotes Boyd (personal communication) who doubts the reconstruction of \**sõ* and \**du* for Adamawa-Ubangi. Third, she suggests that the common suffixing of noun class markers may not have been a shared innovation, since such suffixing must be posited elsewhere in Niger-Congo.

In fact, there is some doubt that Adamawa-Ubangi forms a linguistic unit at all, but rather that it should form a group with Gur. Bennett and Sterk (1977) point out that some Adamawa languages (e.g., Longuda and Tula) show higher cognate scores with Gur than with some other Adamawa-Ubangi languages. They suggest that Adamawa-Ubangi and Gur form a continuum or dialect chain. Bennett (1983a) reasserts this claim, based on evidence from lexicostatistics and shared innovations. He is able to find no phonological innovations and only a handful of weak lexical innovations to support the unity of Adamawa-Ubangi. On the other hand, he states that there are a significant number of lexical items shared by Adamawa-Ubangi and Gur. However, he admits that the evidence for an Adamawa-Ubangi-Gur group is not as solid as he would like.

Thus, the evidence points toward a possible node comprised of Gur and Adamawa-Ubangi, which Williamson (1989a) calls North Volta-Congo. But it is unclear that a branching into the Gur and Adamawa-Ubangi subgroups is justified.



## 5.2. Adamawa-Ubangi internal relationships

Much work remains to establish the internal relationships within Adamawa-Ubangi. On the Adamawa side, Boyd (1974) has produced a comparative study, but he makes no claims about the internal classification of the subgroup. On the Ubangian side, several classifications have been posited, but on weak grounds. Samarin (1971) suggests that Greenberg's groups B1, B3, B5, and B6 be grouped together, but he offers no evidence. Barreteau and Moñino (1978) offer the same classification as Samarin. They state that it is based on typological criteria, but they do not state what their criteria are. Boyeldieu and Cloarec-Heiss (1986) offer a classification based on the dialectometric method (cf. Guarisma and Möhlig 1986), but their study is incomplete in that it does not consider groups B6, B7, and B8.

For his classification listed above, Boyd (1989) does not give firm evidence. For example, in discussing Adamawa, he states that he is basing his subdivisions on available wordlists, but he does not state his methodology in analyzing them. With respect to Ubangi, he says, "My understanding of available lexical data...leads me to propose a tentative overall classification of Ubangi..." Once again, he fails to explain what analytical method leads him to his classification. As a result, we cannot place much weight on his classification.

The division of Adamawa-Ubangi into two subbranches, Adamawa and Ubangi, has been useful as a referential classification, but in fact this division is itself not genetically well established. Boyd (1989) offers the following evidence for the division: "This division can be justified both by typological features of phonology (e.g., difference of syllable structure) and by characteristic lexical items (e.g., a hypothetical *\*no-* 'eye' in Adamawa and *\*te* 'to fall' in Ubangi)" (p. 178).

This is weak evidence for establishing a genetic division. First, it is generally accepted that typological features are not suitable for arguing for a genetic relationship. Second, it is unclear whether Boyd is claiming that these "characteristic lexical items" are to be taken as shared innovations within each subgroup.

Bennett's (1983a) study of Adamawa-Ubangi makes use of evidence from lexicostatistics (102-item wordlist in fifty languages) and shared innovations. His conclusions are tentative, but certain observations are worth noting. First, the lexicostatistic evidence casts doubt on the inclusion of Gbaya within Ubangi. Indeed, Greenberg himself was unclear about this affiliation, stating (1955:12, footnote 17), "I assign Gbaya to the Eastern [i.e., Ubangi] Branch with some hesitancy since it displays evidence of affiliation with the Adamawa branch."

Bennett (1983a:29–33) discusses whether Ubangi, minus Gbaya, may be considered a unity. He gives evidence of a number of isoglosses that separate Ubangi from the rest of Adamawa-Ubangi. In the case of the words for 'breast', 'man', and 'leaf', the Ubangi languages preserve an initial *\*k*, while the other languages show innovations. In the case of the words 'fat', 'dog', and 'bone', the Ubangi languages appear to be the innovators. Bennett then considers several potential problems with this hypothesis, but in the end he concludes that Ubangi, minus Gbaya, is a unity.

Another major claim of Bennett's paper is that Greenberg's groups B2, B5, B6, and B8 form a subgroup under the Ubangi node which he labels *Ka*. Bennett claims that this is supported by lexicostatistical evidence (unfortunately, he gives neither his data nor his cognate percentages) and shared innovations, of which he provides one: *kã* 'breast'. The other branches of Ubangi are coordinate with *Ka*, except Gbaya (B1), which Bennett excludes from Ubangi as mentioned above. This differs significantly from Boyd's (1989) classification listed above, but it has the benefit of being supported by an explicitly stated methodology and analysis.

## 6. Discussion

The study of how Niger-Congo classification has progressed brings up two important issues: (1) the completeness of the documentation and availability of data for supporting claims which are made and (2) the role and validity of the different methodologies which are used in making the classifications.

### 6.1. Data and documentation

With respect to data and documentation, I will look at three works to exemplify the issues involved: Greenberg (1970), Bennett and Sterk (1977), and Bennett (1983a). All three works provide brief overviews of the type of data used, but fail to provide complete references as to the source of the data. Bennett and Sterk (1977) state: "It would be impossible to list all the language sources consulted." This leaves the reader with no means within the published literature of checking and verifying the claims made in the paper. It also means that the reader has no way of assessing the integrity of the data. Regarding Greenberg, Fodor (1969) states, "There are many controversial, ambiguous or, to be candid, incorrect data in the material of Greenberg...To avoid misunderstandings it would have been more fitting to indicate in each case the sources Greenberg relied on."

Let me illustrate the point. Bennett (1983a:27) makes passing reference to the fact that he had access to the dictionary of Banda by Tisserant (1931). I can presume from my knowledge of published material at the time that this was also the source of Banda data for Greenberg. Now, though Tisserant's dictionary is certainly a valuable resource, it has serious flaws, especially in its completeness regarding dialectal variation and its accuracy concerning vowel quality and tone. From my own library and field research on Banda, I have access to more recent data on the language family, both published and personally elicited. As a result, I have the means of assessing the accuracy of the data which Greenberg and Bennett cite for Banda. However, the case of Banda is the exception, since for most of the languages cited, the reader does not know the source of the data unless he is able to find out directly from the author.

In the cases where data are provided, they are often incomplete. Greenberg (1970:13–24) provides a 49-item wordlist for Niger-Congo. In it, he only lists forms that he determines are cognate across most of the groups, and he limits citations to three from each group. As a result, there is no way to check cognate percentages (e.g., for lexicostatistical purposes) from his data. In addition, there is no way to identify possible

lexical innovations for each group or subgroup. These limitations thus reduce the usefulness of the wordlist for research purposes.

Bennett and Sterk (1977) do not provide wordlists in their paper. Rather, they list only the names of the languages for which they have wordlists, the glosses for which they have words, and the percentages of shared cognates. From this, the reader can check the glosses for the presence of cultural vocabulary, but the reader is unable to verify the cognate percentages. These data are only provided for the part of their study that gives an overview of Niger-Congo. For the part of their study that focuses on Kwa and Benue-Congo, they provide no data. To their credit, Bennett and Sterk do give a good description of the procedure they use in analyzing their data.

Bennett (1983a) provides neither wordlists nor cognate percentages, but only lists the languages and glosses. Though the paper claims a certain classification based on lexicostatistics, the reader is left without any knowledge of the cognate percentages on which this classification is made.

## 6.2. Methodology

The second major issue in Niger-Congo classification is one of methodology. The accuracy of the genetic classification is of course dependent on the accuracy of the methodology upon which the classification is made. To date, the conclusions regarding Niger-Congo classification have been based predominantly on the method of resemblances, lexicostatistics, and evidence from shared innovations. Very little has been based on historical reconstruction using the comparative method. Some authors, such as Bennett and Sterk, offer reconstructed lexical items as evidence for shared innovations, but they unfortunately do not provide a detailed account of how they arrive at these reconstructed forms. In the following sections, I will briefly discuss issues related to the assorted methodologies that have been used in establishing the Niger-Congo classification.

### 6.2.1. Method of resemblances

First, Greenberg's (1970) classification of Niger-Congo is based on his method of resemblances (Greenberg 1957). The method is often referred to as "mass comparison," but this term refers to only one part of this classification method. The method has been both extremely influential in African linguistics and also the source of much controversy. Because of its importance in the field of African linguistics and thus its relevance to the topic of this paper, I will give a brief overview of the method.

Greenberg considers the method of resemblances to be a preliminary step that makes hypotheses about the genetic relationship of languages. Its goal is more to determine *if* languages are related rather than the *degree* to which they are related. Once the method of resemblances has established that languages are related, the comparative method may be used to perform an historical reconstruction of the proto-language and in the process extract sound laws that establish the genetic relatedness of the languages.

There are two basic principles underlying the method of resemblances. First, one identifies cross-linguistic resemblances that involve both *form* and *meaning*. These

resemblances may be between lexical items or between grammatical elements of the languages compared. Consider a trivial example that demonstrates the relevance of both form and meaning. Both English and Mono (D.R. Congo) have words pronounced [ʃu]. The Mono word means ‘to be bitter’, whereas the English word refers to an item worn on the foot. In this case, there is a resemblance in form, but no resemblance in meaning. Indeed, there is no known historical connection between the two words and the likelihood of a connection is slim.

However, a simple resemblance in form and meaning is not enough to establish a genetic relationship. Greenberg (1957) lists four possible sources of such a resemblance. First, the forms may indeed be related genetically. Second, it is possible that one language borrowed the form from another language. This is often the case in languages that are in close geographic proximity. Third, the resemblance may be due to sound symbolism, as in the case of onomatopoeia. For example, in Mono, the word for ‘cat’ is [ɲáō]. The pronunciation of this word bears a striking similarity to the English word *meow*, and is likely due to onomatopoeia. Fourth, the resemblance may be due to pure chance. For example, in the Australian language Mbabaram, the word for ‘dog’ is [dag] (Dixon 1997:16).

The first step, then, is to remove nonhistoric factors, i.e., symbolism and chance, as possible explanations of the resemblance. Greenberg suggests three diagnostics. First, if the percentage of resemblance between the two languages is greater than 20 percent, then these factors may be eliminated from consideration. Second, the longer a form, the less likely it is due to chance. Third, the presence of similar suppletive morphological alternants is strong evidence for an historical connection. For example, the odds are rather low that the resemblance between the English paradigm *good-, bet-, be-* (good, better, best) and the German one, *gut-, bes-, be-*, is due to pure chance.

The next step is to remove borrowing as a factor. Greenberg (1957) claims, “it is always possible to tell whether a mass of resemblances between two language is the result of borrowing” (p. 39). The main means to reduce the chance of borrowing is to eliminate cultural words from consideration and to rely on basic vocabulary and grammatical morphemes, items that are assumed to be the most resistant to borrowing. Greenberg’s claim may be a bit overstated as basic vocabulary and grammatical constructions are not always immune to borrowing. For example, in Mono, the negation marker *nene* is being replaced in the language by the Lingala negative marker *te*.

The second major principle underlying the method of resemblances is *mass comparison* (or “group comparison”). This is basically the notion of identifying resemblances across a broad scope of languages rather than isolated comparisons of pairs of languages. Greenberg claims that the larger the number of languages that exhibit a certain resemblance, the less likely that the resemblance is due to chance, symbolism, or borrowing.

The method of resemblances has been widely criticized in the literature. In some cases, such as in Dixon (1997), it is discounted as being simply typological in nature, but this is a mischaracterization. Other researchers have posited more substantial criticisms. First, Bennett and Sterk (1977) accept that Greenberg’s method is adequate for demonstrating relationship, but they state that it “is not best suited for investigation of degrees of relationship and subgrouping” (p. 242). Indeed, this appears to be a problem

for Greenberg. In the quote in §2 above, he notes that Mande and West Atlantic are likely more remotely related to the other branches of Niger-Congo than his classification would imply. Greenberg's classification does not include deeply-nested branches in the genetic tree; rather, it is mostly flat.

A second criticism of Greenberg's method is his avoidance of positing sound laws. Both Schachter (1971) and Fodor (1969) point this out. His critics argue:

Greenberg, although he has presented long lists of putative cognates among the languages for which he claims genetic relationship, has not specified precise sound correspondences, and thus has failed to produce the only proof of cognation that is acceptable in standard comparative-linguistic practice. (Schachter, p. 33)

Both of these criticisms ignore one essential caveat that Greenberg (1957) states explicitly. He insists that the conclusions drawn from his method are to be considered tentative; he sees them as hypotheses to be verified and expanded upon by implementation of the comparative method. In other words, the value of his method is not in the firm conclusions that it draws, but rather in the creation of a scaffolding from which other research may build. Greenberg states:

The establishment of valid hypotheses concerning genetic relationships among languages is a necessary preliminary to the systematic reconstruction of their historical development. The appropriate techniques cannot be applied to languages chosen at random but only if preliminary investigation has already indicated the likelihood of the success of such an enterprise. (1957:35)

Greenberg thus does not intend for his method to replace the comparative method, but rather to complement it.

### *6.2.2. Lexicostatistics*

Much of the classificatory work on Niger-Congo is based on a technique called LEXICOSTATISTICS (cf. Gudschinsky 1956, Crystal 1997:333), in which one calculates the percentage of cognates in the basic core vocabulary of two languages. This technique is used in glottochronology, a study that attempts to determine the rate of change of languages over time. As originally defined, the term lexicostatistics refers to the analytical technique and glottochronology refers to the general study, but today the term lexicostatistics is often used in the broader sense. Hinnebusch (1989) states that it is useful for identifying potential starting points for reconstructive comparative work.

Gudschinsky (1956) lists four basic assumptions of lexicostatistics. First, basic core vocabulary is less subject to change than cultural items. As I noted above, the method of resemblances makes this assumption as well. Second, the rate of retention of basic vocabulary in a language is constant through time. Third, the rate of loss of basic vocabulary is the same in all languages. Fourth, if we know the percentage of true cognates within the basic vocabulary between a pair of languages, we can calculate the

length of time that has elapsed since the two languages began to diverge from a single parent language.

Bennett and Sterk (1977) are concerned about the use of lexicostatistics to identify lower-level branching in a genetic tree. They state:

Lexicostatistics, while useful for preliminary gross subgrouping, is not—if used alone—adequate for indication of fine degrees of relationship. The nature of lexicostatistics is such that geographic and social proximity tends to increase cognacy scores significantly. (p. 245)

In addition, many scholars question the basic assumptions of lexicostatistics. For example, Dixon (1997) states, “The rate at which a language changes is not constant and is not predictable” (p. 9). He then gives examples of languages with accelerated change. He also states, “There is no universal principle that core vocabulary...is less likely to be borrowed than non-core items” (p. 10). He states that in Australia, the percentage of shared vocabulary between a pair of languages is the same regardless of the number of items compared. So while lexicostatistics may be useful as a rough estimate, the accuracy of the method is in serious doubt.

### *6.2.3. Shared innovations*

Evidence from shared innovations has been used, often in conjunction with lexicostatistics, in order to justify certain nodes in a classification. It is assumed that if a language produces an innovation at a certain point in time, then all the descendents of that language will have that form whereas all externally-related languages will not.

This of course runs the risk of skewing due to borrowings, chance, or symbolism, just as in the method of resemblances and lexicostatistics. Williamson (1989b:249) points out another problem: “There is a certain problem in the use of lexical innovations in that they most often come about through semantic shift. Since this process may occur repeatedly and independently, it is not fully reliable.”

In addition, in this study we have seen how there can be confusion between isoglosses and innovations. An isogloss indicates a linguistic boundary, but this may not necessarily indicate that the opposing words both represent innovations at the same genetic level.

### *6.2.4. The comparative method*

The comparative method has been traditionally considered the most accurate means by which to establish the genetic relationships between languages. The method involves comparing lexical items and grammatical forms between languages, setting up correspondences between the phonemes of the languages, and then making hypotheses about the structure of the proto-language and the sound laws that led to the development of the present-day languages in the family.

The use of the comparative method has been neglected in Niger-Congo classification. It has been used on a micro level to establish the genetic relationship of

certain subgroups within Niger-Congo, e.g., Bantu, but to date there has been no concerted effort to establish a proto-system for Niger-Congo as a whole, along with the sound laws which led to the modern languages in the family. Dixon (1997:32–35) does not mince words in criticizing the Africanists for this, but to their credit, the Africanists have stated all along the tentative nature of their conclusions and the need for the application of the comparative method to verify their findings.

Dixon (1997) points out that there are certain problems associated with the comparative method. First, it cannot be used blindly as a discovery procedure to mechanically produce a reconstruction. Dixon provides a couple of hypothetical examples to illustrate this point. Second, he quotes Bloomfield, who states:

The comparative method, then—our only method for the reconstruction of prehistoric language—would work accurately for absolutely uniform speech-communities and sudden, sharp cleavages. Since these pre-suppositions are never fully realized, the comparative method cannot claim to picture the historical process. (1933:318)

This having been said, Dixon (1997) calls into question the appropriateness of applying the comparative method to Niger-Congo at all. According to his punctuated equilibrium model of language change, the Niger-Congo languages have been in a state of equilibrium in which areal features have diffused over the geographical region through borrowing, effectively masking the type of language change typically associated with historical reconstruction. As a result, he doubts that an accurate reconstruction of Proto-Niger-Congo is possible.

#### *6.2.5. Other methods*

Two other types of classifications, typology and dialectometrical, have been occasionally employed in Niger-Congo classification, but I have not focused on them in this paper. According to Heine (1980a), “The typological method is nowadays rejected by most linguists since its basic assumption that structural comparisons necessarily lead to the discovery of genetic relationship has been proved wrong in a number of cases” (p. 298). The dialectometrical method has been the basis of one study on African languages (Guarisma and Möhlig 1986), but I will not discuss it here for lack of space.

## **7. Conclusion**

It is clear from this overview of Niger-Congo language classification that much more work needs to be done in this realm. The exact placement of Ijoid, Kru, and Dogon within the Niger-Congo genetic tree remains to be determined. Whether several linguistic groups—Atlantic, (new) Kwa, (new) Benue-Congo, Wide Bantu, Narrow Bantu, and Adamawa-Ubangi—are each a unity also remains to be established.

Until now, the Niger-Congo classification has been influenced predominantly first by work using the method of resemblances, and then by work focusing on lexicostatistics and shared innovations. While these methods may be useful for approximating gross

groupings, there are serious questions about the precision that can be obtained by their application.

While the comparative method has occasionally been applied to small language families within Niger-Congo, particularly Bantu, its use has so far been neglected in tying the language family together as a whole. A comprehensive reconstruction of Niger-Congo, including the establishment of sound laws, remains the major future task in Niger-Congo classification.

### Acknowledgements

I wish to thank Patrick Bennett, Bill Darden, John Goldsmith, Joseph Greenberg, Karen Landahl, Constance Kutsch Lojenga, Paul Newman, John Stewart, and Kay Williamson for helpful comments on this paper. All errors are my responsibility. *SDG*.

### References

- Barreteau, Daniel, and Yves Moñino. (1978). "Les langues oubangiennes." *Inventaire des études linguistiques sur les pays d'Afrique noire d'expression française et sur Madagascar*, D. Barreteau (ed.), 195–208. Paris: CILF.
- Bastin, Yvonne, André Coupez, and B. de Halleux. (1983). "Classification lexicostatistique des langues bantoues (214 relevés)." *Bulletin de l'Académie Royale des Sciences Outre-Mer* 27(2):173–199.
- Bendor-Samuel, John, and Rhonda L. Hartell (eds.) (1989). *The Niger-Congo languages*. Lanham, Maryland: University Press of America.
- Bennett, Patrick R. (1983a). "Adamawa-Eastern: Problems and prospects." In *Current Approaches to African Linguistics* 1, I. R. Dihoff (ed.), 23–48. Dordrecht: Foris.
- Bennett, Patrick R. (1983b). "Patterns in linguistic geography and the Bantu origins controversy." *History of Africa* 10:35-51.
- Bennett, Patrick R., and Jan P. Sterk. (1977). "South Central Niger-Congo: A reclassification." *Studies in African Linguistics* 8:241–273.
- Bleek, Wilhelm H. I. (1858). *Catalogue of the library of Sir George Grey*. London: Trübner.
- Bloomfield, Leonard. (1933). *Language*. New York: Holt, Rinehart and Winston.
- Boyd, Raymond. (1974). *Étude comparative dans le groupe adamawa*. (Bibliothèque de la SELAF 46.) Paris: SELAF.
- Boyd, Raymond. (1978). "À propos des ressemblances lexicales entre langues Niger-Congo et Nilo-Sahariennes." *Études comparatives*, 43-94. (Bibliothèque de la SELAF 65.) Paris: SELAF.
- Boyd, Raymond. (1989). "Adamawa-Ubangi." In Bendor-Samuel and Hartell, 178–215.
- Boyeldieu, P., and France Cloarec-Heiss. (1986). "Dialectométrie lexicale dans le domaine oubanguien." In *La méthode dialectométrique appliquée aux langues africaines*, G. Guarisma and W. Möhlig (eds.), 331–394. Berlin: Dietrich Reimer.



- Cloarec-Heiss, France. (1995). "Emprunts ou substrat? Analyse des convergences entre le groupe banda et les langues du Soudan central." In *Proceedings of the Fifth Nilo-Saharan Linguistics Colloquium, Nice, 24–29 August 1992*, Robert Nicolai and Franz Rottland (eds.), 321–355. Köln: Köppe.
- Cole, Desmond T. (1971). "The history of African linguistics to 1945." In Sebeok, 1–29.
- Crystal, David. (1997). *The Cambridge encyclopedia of languages* (2nd edition). New York: Cambridge University Press.
- Dixon, R. M. W. (1997). *The rise and fall of languages*. Cambridge: Cambridge University Press.
- Elugbe, Ben Ohiomamhe, and Kay Williamson. (1977). "Reconstructing nasals in Proto-Benue-Kwa." In *Linguistic studies offered to Joseph Greenberg 2*, A. Juillard (ed.), 339–363. Saratoga: Anma Libri.
- Fodor, István. (1969). *The problems in the classification of the African languages: Methodological and theoretical conclusions concerning the classification system of Joseph H. Greenberg*. (Studies on developing countries 5.) Budapest: Center for Afro-Asian Research of the Hungarian Academy of Sciences.
- Greenberg, Joseph H. (1955). *Studies in African linguistic classification*. New Haven: Compass Publishing Company.
- Greenberg, Joseph H. (1957). *Essays in linguistics*. (Viking Fund Publications in Anthropology 24.) New York: Wenner-Gren Foundation for Anthropological Research.
- Greenberg, Joseph H. (1970 [1963]). *The languages of Africa* (3rd edition). Bloomington, Indiana: Indiana University.
- Greenberg, Joseph H. (1995). "The concept of proof in genetic linguistics." *Mother Tongue (Journal of the Association for the Study of Language in Prehistory)* 1:207–216.
- Guarisma, G., and W. Möhlig (eds.) (1986). *La méthode dialectométrique appliquée aux langues africaines*. Berlin: Dietrich Reimer.
- Gudschinsky, Sarah. (1956). "The ABCs of lexicostatistics (glottochronology)." *Word* 12:175–210.
- Guthrie, Malcolm. (1948). *The classification of the Bantu languages*. London: Dawsons; Oxford University Press for the International African Institute.
- Guthrie, Malcolm. (1962). "Bantu origins: A tentative new hypothesis." *Journal of African Languages* 1:9–21.
- Hedinger, Robert. (1989). "Northern Bantoid." In Bendor-Samuel and Hartell, 421–429.
- Heine, Bernd. (1973). "Zur genetischen Gliederung der Bantu-Sprachen." *Afrika und Übersee* 56:164–185.
- Heine, Bernd. (1979). "Some linguistic observations on the early history of Africa." *Sprache und Geschichte in Afrika* 1:37–54.
- Heine, Bernd. (1980a). "Methods in comparative Bantu linguistics (the problem of Bantu linguistic classification)." In *L'expansion bantoue 2*, Luc Bouquiaux (ed.), 295–

308. Actes de Colloque International du CNRS, Viviers, 4–16 avril 1977. Paris: SELAF.
- Heine, Bernd. (1980b). “Some recent developments in the classification of Bantoid: The problem of defining Bantu.” In *L’expansion bantoue 2*, Luc Bouquiaux (ed.), 333–340. Actes de Colloque International du CNRS, Viviers, 4–16 avril 1977. Paris: SELAF.
- Heine, Bernd, Hans Hoff, and Rainer Vossen. (1977). “Neuere Ergebnisse zur Territorialgeschichte des Bantu.” In *Zur Sprachgeschichte und Ethnohistoire in Afrika*, W.J.G. Möhlig, F. Rottland, and B. Heine (eds.), 57–72. Berlin: Dietrich Reimer.
- Henrici, Alick. (1973). “Numerical classification of Bantu languages.” *African Language Studies* 14:81–104.
- Hinnebusch, Thomas J. (1989). “Bantu.” In Bendor-Samuel and Hartell, 450–473.
- Johnston, H. H. (1886). *The Kilimanjaro expedition*. London.
- Newman, Paul. (1995). *On being right: Greenberg’s African linguistic classification and the methodological principles which underlie it*. Bloomington: Institute for the Study of Nigerian Languages and Cultures, African Studies Program, Indiana University.
- Olivier, Roland. (1979). “Cameroun - the Bantu cradleland?” *Sprache und Geschichte in Afrika* 1:7–20.
- Olson, Kenneth S., and Brian E. Schrag. (2000). An overview of Mono phonology. In *Proceedings of the 2nd World Congress of African Linguistics, Leipzig 1997*, H. Ekkehard Wolff and Orin D. Gensler (eds.), 393–409. Köln: Rüdiger Köppe.
- Phillipson, D. W. (1977). “The spread of the Bantu language.” *Scientific American* 236:106–114.
- Samarin, William J. (1971). “Adamawa-Eastern.” In Sebeok, 213–244.
- Schachter, Paul. (1971). “The present state of African linguistics.” In Sebeok, 30–44.
- Schadeberg, Thilo C. (1981). “Das Kordofanische.” In *Die Sprachen Afrikas, Band 1, Niger-Kordofanisch*, B. Heine, T.C. Schadeberg, and E. Wolff (eds.), 117–128. Hamburg: Helmut Buske.
- Schadeberg, Thilo C. (1986). “The lexicostatistical base of Bennett and Sterk’s reclassification of Niger-Congo with particular reference to the cohesion of Bantu.” *Studies in African Linguistics* 17:69–83.
- Sebeok, T. A. (ed.). (1971). *Linguistics in sub-Saharan Africa*. (Current Trends in Linguistics 7). The Hague, Paris: Mouton.
- Stewart, John M. (1965). “Three Akan-Bantu sound correspondences.” Paper presented at the Fifth West African Languages Congress, Legon, Ghana.
- Stewart, John M. (1976). *Towards Volta-Congo reconstruction (inaugural lecture)*. Leiden: University Press.
- Stewart, John M. (1989). “Kwa.” In Bendor-Samuel and Hartell, 217–245.
- Tisserant, R. P. Charles. (1931). *Dictionnaire banda-français*. (Travaux et mémoires de l’Institut d’Ethnologie 14.) Paris: Institut d’Ethnologie.

- Watters, John R. (1989). "Bantoid overview." In Bendor-Samuel and Hartell, 401–420.
- Watters, John R., and Jacqueline Leroy. (1989). "Southern Bantoid." In Bendor-Samuel and Hartell, 431–449.
- Westermann, Diedrich. (1927). *Die westlichen Sudansprachen und ihre Beziehungen zum Bantu*. Berlin: de Gruyter.
- Westermann, Diedrich. (1949). *Sprachbeziehungen und Sprachverwandtschaft in Afrika*. (Sitzungsberichte der Deutschen Akademie der Wissenschaft zu Berlin.) Berlin: Akademie-Verlag.
- Williamson, Kay. (1971.) The Benue-Congo languages and Ijo. In Sebeok, 245–306.
- Williamson, Kay. (1989a). "Niger-Congo overview." In Bendor-Samuel and Hartell, 3–45.
- Williamson, Kay. (1989b). "Benue-Congo overview." In Bendor-Samuel and Hartell, 247–274.
- Williamson, Kay, and Roger Blench. (2000). "Niger-Congo." In *African languages: An introduction*, Bernd Heine and Derek Nurse (eds.), 11–42. Cambridge: Cambridge University Press.