2 The classification of African languages

This chapter discusses how the languages of Africa have been classified and introduces a few of the controversies that have troubled and indeed still trouble the field. Scholars agree in general as to the major families and even most sub-families, but there are decided differences as to how some languages and even as to how some groups should be assigned. Although no resolution to these many questions can be found here, the reader will at least become familiar with some of the issues. After some background, the several approaches to language classification are presented, followed by an introduction to the approach known as “mass comparison” championed by Joseph Greenberg. The next part deals with some of the problems that have arisen, dealing primarily with methodological issues. A few case studies illustrating these problems follow, one at the macro or highest level of classification, that of phyla (“family” is also used). The next two studies look at lower level groupings to illustrate first how geographical factors have influenced genetic classification and then how a typological feature has also held sway.

The sources of controversy are multiple and include (historically) at least prejudice on the part of Europeans, who have often allowed non-linguistic criteria to intrude on their classificatory tasks. Probably the greatest hampering factor, however, has been the lack of reliable data. At times the few available forms come from field workers not thoroughly trained in the collection of linguistic data. Egregious problems have arisen when workers have not been sensitive to tone; pitch differences function importantly in many African languages to mark lexical differences as well as grammatical ones (see Section 3.7). Other formal

18. The cultures of language classification vary. Africanists, for the most part, contrast sharply with their counterparts in Oceania and the New World. The former present a “tidy” picture with every language classified and few isolates, which process in linguistic jargon is known as “lumping”. In the latter two areas, the tradition is more of “splitting” and building relationships from the “bottom up” (Blench 1999a:8).
complexities have similarly hampered analyses: consonant mutation (see Section 3.6) and vowel harmony (see Section 3.5). But the problems are not always due to linguistic complexity, as the following anecdote will show.

The linguistic identity of a group in northern Ivory Coast has been controversial, more for reasons having to do with its analysts than with linguistic data. The Dieli have been considered by some an archaic group of Senufo, and by others an archaic group of Mande.

It is certainly ironic that, although the debate hinges in principle on whether the Dieli language is related to Sienare [a Senufo language] or to Manding [Mande], none of the scholars concerned is a trained linguist, nor has any linguistic evidence been cited in the literature substantiating any of the peremptory identifications which have been proffered with an air of authority. (Launay 1995: 155)

Launay continues that a trained linguist who has had a look at word lists, including earlier ones, finds that the evidence is inconclusive.

There is the further complication of multilingualism and the concomitant multiple ethnicities and identities of many Africans. Which is a multilingual individual’s mother tongue? What about a “first language” when it has all but been abandoned for a more prestigious variety (see Section 7.3.5)? As has been pointed out with regard to ethnicity,

...ethnic labels in the African context are at best meaningless, at worst the legacy of a colonial construction designed to control and oppress (Amselle 1985: 11–48). Thus, ethnicity might be more accurately viewed as one of a number of negotiable aspects of one’s identity. Identifying himself as Mandingo in the context of our interview (i.e. after having been accused of being Fula) was sufficient for my informant to distinguish himself as Mande from a quite separate ethnic category, [from] the Fula, as well as from other Sierra Leonean ethnicities such as Limba or Temne. It was only during the course of the interview that the leatherworker more precisely defined the origins of his family as Soninke (Frank 1995: 144).

In the townships of South Africa’s cities, ethnicity is uncertain if not unimportant; some would deny any ethnicity because it smacks of the classifications imposed by the apartheid government. Proclaiming oneself a “South African” can be an affirmation of support for a multiethnic (and multiracial) society. Language similarly poses a problem because of the mixing in the townships; even linguists have trouble identifying the provenance and identity of the extant
varieties, e.g., Calteaux 1994. Another problem arises when a language serves as a lingua franca and serves as a second language to many speakers, where pidginization and restructuring can occur, e.g., Zima 2001.

The next section presents the current classification of the four language phyla of Africa along with some details on each phyla’s constituent languages. In the sections following I detail some of the methods used in classifying African languages and some of the current controversies in classification.

2.1 The four phyla

The languages of Africa are generally divided into four major phyla (but see Section 2.4): Niger-Congo, Nilo-Saharan, Afroasiatic, and Khoisan, with Niger-Congo containing more than two-thirds of the 2,000 or so languages spoken on the continent (Bendor-Samuel 1989: vi). It must be remembered that from a classificatory standpoint, all phyla are equal, despite differences in size. See Map 2 for an indication of their rough geographical location.

For each language phylum in Table 1 is given a rough estimate of the number of languages belonging to each group along with some representative languages and the number of mother-tongue speakers after each named language in parentheses (in millions, except where noted as thousands by “K”). Note that Afroasiatic languages are not found exclusively in Africa; the figures here represent only those Afroasiatic languages spoken in Africa.

At least one caveat should be attached to Table 1: multilingualism is a common pattern if not the norm in many parts of Africa and may distort the figures (see Fasold 1984 for many other caveats about interpreting such estimates). For example, it is not unlikely that a typical citizen of Nigeria will speak at least three different languages in addition to Nigerian Pidgin English. What this means is that many people speak more than one language and will use different languages for different purposes (e.g., Myers-Scotton 1993). Thus they will adapt to the situation in which the (language) survey question is being asked. Speakers of Cangin languages in Senegal, for example, speak their languages only at home and were long considered to be speakers of Wolof, a much more widespread and politically important language (Pichl 1966).

19. Phylum overviews can be found in Heine & Nurse 2000; readers of German might also want to look at Heine et al. 1981.
This one-speaker-to-many-languages mapping is not unusual; nor is the pattern of one language having different meanings. In (3) is shown how using a single language can mean different things depending on the speaker and the situation. Hausa is an Afroasiatic language spoken as a first language primarily in northern Nigeria and southern Niger, but is used as a second language by many other speakers. All of the areas mentioned in (3) are found in West Africa.

(3) Variable significance attached to language choice (Fardon & Furniss 1994: 22–23)

A mother tongue Hausa using Hausa in Kano might subscribe to ideas of Hausa culture, Hausa centrality in the north, and Hausa nationalism that are widely shared in the community;
The classification of African languages

A speaker using Hausa in Adamawa, Tabara, or Borno might be using it as an expedient to allow communication between Suwa and Kanuri, or Kanuri and Fulani [Fula];

The use of Hausa might also signify personal allegiance to a particular group within Maiduguri or the wider north;

Using Hausa in Cameroon has totally different implications where the northern area has been variably Fulbe-ized; it is a way to finesse committing to Fulbe values.

Depending on the situation, then, a speaker might or might not want to say he or she speaks a given language, i.e., admit to belonging to a particular ethnic group, for there is often a presumed isomorphism between language and ethnic identity. This reluctance is particularly common in cases of language shift where speakers do not want to admit they speak the language being shifted from (Brenzinger et al. 1991). With these facts in mind, we must be circumspect in interpreting the figures in both Table 1 and Table 2.

According to Bender 1996a, the most widely spoken African languages in 1993 are as in Table 2. Note that these numbers are in millions of both first- and second-language speakers; compare the figures here in Table 2 with the first-language speakers in Table 1, where only first-language speakers are considered. Note also that the most widely spoken language, Swahili, has barely five million mother-tongue speakers. This attests to its widespread importance as a lingua franca over much of eastern Africa.

---

**Table 1. The major language families (phyla) of Africa**

<table>
<thead>
<tr>
<th>Language Family</th>
<th>No. languages</th>
<th>Member languages (No. speakers in millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Niger-Congo*</td>
<td>1650</td>
<td>Bambara (3), Fula (13), Igbo (17), Mooré (11), Swahili (5), Yoruba (20), Zulu (9.1)</td>
</tr>
<tr>
<td>Afroasiatic (in Africa)</td>
<td>200–300</td>
<td>Arabic (180, all varieties), Amharic (20), Hausa (22), Oromo Somali (5–8), Songhai (2), Tachelhit Berber (3)</td>
</tr>
<tr>
<td>Nilo-Saharan</td>
<td>80</td>
<td>Dinka (all groups, 1.4), Kanuri (4), Luo (3.4), Maasai (883K), Nuer (840K), (Phylum total 30)</td>
</tr>
<tr>
<td>Khoisan</td>
<td>40–70</td>
<td>Nama (140K), Sandawe (70K), Kung (8–30K), !Xôô (3–4K)</td>
</tr>
</tbody>
</table>

Thus, as this display suggests, not only are Niger-Congo languages the most numerous, they are also the most widely spoken.

Figure 1 presents the classification of Niger-Congo, about which there has been some controversy (see Sections 2.4 and 2.5.2). Note that, as elsewhere, some of the nodes represent individual languages (e.g., Ijo, Pre/Beere), others represent relatively small groups of languages (e.g., Dogon), and still others represent quite large ones (e.g., Bantu). As in the other displays, the nodes are unequal in terms of language numbers but represent an identifiable and cohesive subset of the family as a whole, usually equated with a shared innovation. The asterisks (“*”) in the display indicate reconstructed languages; double lines indicate an “ancient dialect continuum”; a dotted line indicates a tentative placement. Although their representation departs somewhat from conventional formalizations, I have chosen to present the Williamson & Blench 2000 figure in its entirety.

Niger-Congo is the largest language phylum in the world covering a greater area than any other phylum (Williamson & Blench 2000). Not surprisingly, then, many of the most widely spoken languages in Africa belong to Niger-Congo; it is the primus inter pares of African language phyla. Although there has been some question as to the genetic identity of Niger-Congo (Dixon 1997) and indeed to the viability of establishing genetic groups in Africa as a whole (Tucker & Bryan 1956, Dalby 1970, Dalby 1977), most scholars unquestionably accept Niger-Congo as a genetic unity (Williamson 1989b, Williamson & Blench 2000). The proto-language has been estimated as having been spoken some 15,000 years ago (Ehret 2000: 293).20

---

20. Others differ: “The origin of Niger-Kordofanian (= Niger-Kongo) is around 10 kya [thousands of years ago], …Nilo-Saharan … a later origin according to the genetic tree (4 kya). Khoisan (!Kung) … on the order of 50 kya … Afro-Asiatic based on the distance of Berbers from Indo-Europeans is on the order of 15 kya. … The dates indicated here for the origin of linguistic
Despite the prodigious numbers of their speakers, the well-known Bantu languages are small beer in Figure 1. They are a sub-group of Bantoid which contains Bantu and Bantu-like languages. Nonetheless, Bantu languages form the largest group, are the most widely spoken, and are the most extensively studied. As the figure seeks to represent, Bantu is also a very late branching group; it is relatively “young” from this perspective, 4,000 years old in some estimates (Blench 2001a:9). To put some numbers on the size of Bantu in terms of numbers of speakers (with the usual caveats), see Table 3.

families are suggestions that should be taken with a ton of salt” (Cavalli-Sforza et al. 1994:104).
By comparison, Kordofanian, a much earlier branching from the Proto-Niger Congo stock, contains some 20 languages with only 165,000 speakers (see Map 3). The Kordofanian languages are spoken in the Nuba Mountains in the Republic of the Sudan.

In terms of classification, the unity of Bantu and indeed the Benue-Congo family was recognized early on (e.g., Doke & Cole 1961). To linguists working within other parts of Niger-Congo or in other parts of Africa in general, the uniformity to Bantu is impressive, in terms of both the lexicon (cognates) and shared grammatical structure. Word order, morphology and phonology in a typical language are all minor variations on a central Bantu theme, although there are some significant variations as one moves northwest out into the "Bantoid" area of languages (see Map 4) towards the posited Bantu homeland.

This next set of maps intends to make clear the distinctions between Bantu, (Northern and Southern) Bantoid, and Benue-Congo. The maps come from different contributors, as indicated, all in Bendor-Samuel 1989. The map entitled "Non-Bantu Benue-Congo" (Plate 11 in Williamson 1989a) shows mostly the non-Bantu/Bantoid portion of Benue-Congo. "Bantoid languages" (Plate 19 in Watters 1989) shows how the Bantoid part of Benue-Congo stretches from there over into East Africa and down to southern Africa. "Northern Bantoid languages" (Plate 20 in Hedinger 1989) is a relatively small group of languages, but Southern Bantoid includes all of Bantu (often known as

---

21. Niger-Congo has sometimes been called "Niger-Kordofanian" to underscore the inclusion of the Kordofanian languages within Niger-Congo (Greenberg 1963).

22. These and other languages were once called "Semi-Bantu"; Cole 1971 (p. 12) notes that Johnston 1919–22 “adopted the term 'Semi-Bantu' for the languages of West Africa, from Dakar to Cameroon, which manifest Bantoid noun class systems, but he conceived of these as mixed languages resulting from the impact of Bantu class languages on Sudanic nonclass languages, and not as indicating genetic relationship”. Thus, we see yet an example of classification by typological criteria and of Bantu bias.
"Narrow Bantu") as well as the Non-narrow Bantu languages (over 100) featured in the map entitled "Southern Bantoid languages" (Plate 21 in Watters & Leroy 1989). Narrow Bantu consists of some 500 languages.

Nonetheless, Bantu represents only a fraction of the genetic and typological diversity found in Africa, despite its scholarly importance (Hinnebusch 1989; see Lyovin 1997). Genetic classification has proven relatively straightforward with regard to Bantu, but outside this remarkably homogeneous core there has been some controversy (Watters 1989).
Map 4. Bantu, Bantoid, and Benue-Congo
Of the other language phyla, **Afroasiatic** has the next greatest number of speakers in Africa. The phylum contains over 300 languages (371 in the *Ethnologue* count (Grimes 1996) spoken by nearly 250 million people. The date of its proto-language, i.e., the date at which the proto-language began to diversify, “has been proposed by Diakonoff 1988 (p. 25) … to a period prior to 8,000 B.C.,” over 10,000 years ago (Hayward 2000: 75). Outside and within Africa can be found one of its most important members, Arabic; Hebrew and Aramaic also belong to the phylum, both spoken exclusively outside Africa. The vast majority of Afroasiatic languages, however, are found in Africa and the phylum as a whole originated there. The display in (4) shows how the languages of Afroasiatic are classified, probably the first African phylum to be recognized as comprising a genetic unity (Blench 1997: 95). The numbers in parentheses after each family name represent the number of languages in that family.

(4) The families of Afroasiatic (Hayward 2000)

- **Berber** (30): Languages spoken by some 10 million people in northern Africa (Morocco, Algeria, Tunisia, Niger, Mali, and Mauritania): Tachelhit, Tamazight, Kabyle, Tamezret, Tamashiq, Zenaga, etc.
- **Chadic** (192): Languages of Nigeria, Chad, and Cameroon including Hausa, Ngizim, Miya, spoken by some 28 million people.
- **Cushitic** (47–50): Languages of Ethiopia, Eritrea, Somalia, Sudan, Kenya, and Tanzania including: Dahalo, Iraqw, Oromo, Afar, Somali; 30 million speakers.
- **Egyptian** (1): Coptic (disappeared in the fourteenth century)
- **Semitic** (50): Middle Eastern as well as Ethiopian languages including (in Africa): Arabic (regional varieties), Gi’iz, Tigré, Tigrinya, Amharic, etc.; 140 million.
- **Omotic** (20+): Many languages spoken in Ethiopia. Three million speakers.

Among Africanists there is no great controversy as to the classification of Afroasiatic languages, although there has been some discussion of placing Omotic within Cushitic rather than as a separate family. Near Eastern scholars, however, are less happy with Afroasiatic, resenting the diminished attention accorded their languages in the accepted classification and the extrapolated African origin (Blench 1999b: 11). Some scholars consider them “minor
languages” and exclude them from the Semitic family tree (Blench & Spriggs 1999a: 12).

The next largest family in terms of both numbers of languages and numbers of speakers is Nilo-Saharan, probably the second oldest family in terms of the time depth of its differentiation. Bender 1996a has classified the languages of Nilo-Saharan into three major types along a continuum: Outlier-Satellite-Core. In Figure 2 the only languages to form a genetic group are the so-called “Core” languages to the right. It is the Outliers,23 the three nodes to the left, that have caused many headaches (Sections 2.4 and 2.5.1). Most of the names represent families, although For, Berta, Kunama, Gumuz are individual languages.

![Figure 2. Nilo-Saharan (Bender 1996b, Bender 1996a)](image)

The smallest and likely the oldest group is Khoisan, whose languages may not form a single group. The inclusion of the languages traditionally assigned to Khoisan is not uncontroversial, nor does everyone accept genetic relations between the parts. These problems are not surprising given the issues around the available data (see Traill 1995 for an overview). Not only do workers have the problem of incomplete data but there is also the fact that the family’s fragmentation is likely of a greater time depth than any of the other families, making it harder to relate the languages (see Section 2.5.3 for details). The group contains some thirty or more languages, most of which are threatened with extinction. All have clicks, although those in Hadza and Sandawe are less robust, and all but the same two languages are spoken in southern Africa (Namibia, Botswana, Angola, and South Africa). A recent consensus seems to

23. Such languages have been called isolates or ’floaters’ (flottants), problematic languages or groups inconsistently grouped on the basis of different methods (Bastin & Piron 1999: 150).
The classification of African languages

Northern (7): Languages spoken in Angola and Namibia.
Southern (6): Four in South Africa including !Xegwi; and two in Botswana, !Xóõ and #Hua.
Central (20): Includes Haiom, Kwadi, Nama, !Ora, Xiri, and others spoken in Botswana.
Hadza (Hatsa)
Sandawe


be that Hadza and possibly Sandawe should be classified as isolates (Sands 1997 as in Blench 1999b: 4, Sands 1998).24

With the idea of these four phyla in mind, we will now consider some of the controversies and issues that have arisen in the classification of African languages.25 Section 2.4 treats the problematic Nilo-Saharan family and how it can be related to the major African phylum, Niger-Congo. Section 2.5.1 treats two of its purported members, the languages Kuliak and Songhay, which may not be part of Nilo-Saharan at all. The following parts of Section 2.5 treat other unruly children.

Despite all of these complexities and resultant analytical shortcomings, there is considerable agreement, perhaps unwarranted, on how African languages should be classified. The section outlines the major approaches to classifying languages, all of which have been used in Africa. In the following section Greenberg’s method of mass comparison is discussed and contrasted with more traditional ones.

2.2 Approaches to classifying African languages

In some cases the criteria used for classifying African languages have been and continue to be typological and even areal or geographic. In other parts of the world, languages are said to be genetically related only if they share linguistic

24. The genes of the speakers are not Khoisan: “the Hadza and Sandawe … show little if any evidence of Khoisan genes.” (Cavalli-Sforza et al. 1994: 103). Note that the term “genetic” is used by linguists as well as by geneticists. Linguists use it to describe a relationship between languages that is characterized by a putative common parent: languages genetically related in linguistics were once unified but have since diverged (see Section 2.2).

25. There is no one review of African language classification, but see the phylum chapters in Heine & Nurse 2000. Cole 1971 contains a concise historical review. Williamson 1989b and other papers in the edited volume, Bendor-Samuel 1989, are also important.
properties and items which cannot have been borrowed, thereby demonstrating that they have a common "parent". Using the comparative method, the agreed-upon approach to language classification, the analyst seeks cognates, pairs of form-meaning correspondences between possibly related languages, in order to establish whether or not the two languages are related and in order to reconstruct their parent or proto-language (see further discussion below p. 31). Reconstructed Proto-Atlantic, then, would be the hypothetical language from which all attested Atlantic languages are descended (e.g., Pozdniakov 1993).

In the classification of African languages, additional, non-genetic procedures have been employed to supplement and sometimes to replace the comparative method, which procedures have not gone entirely unchallenged. It is not that non-genetic classifications are of no use, it is simply that it is inconsistent to use non-genetic criteria for making claims as to genetic relatedness. In fact, the validity and value of genetic classifications for African languages have been debated, e.g., Dalby 1970, that is, it has been questioned whether it is possible or worthwhile to perform such operations.

Genetic classifications are only one, albeit one very useful, way of grouping languages, but much can be learned by adopting areal and typological perspectives, especially coupled with a good knowledge of socio-historical facts. Nonetheless, as opposed to the others genetic classification is the only method of the three which has the desirable properties of being necessarily exclusive (only one classification is possible), exhaustive (every language has a place, even if as only an isolate at some level), and non-arbitrary (based on common historical development). The coincidence of phonological and syntactico-semantic content guarantees this (Bender 1996a: 11).

In a typological as opposed to a genetic classification, analysts will use a prominent, often unusual, feature to group languages. For example, languages using clicks (see Section 2.5.3) are assumed to form a linguistic group; languages with noun class systems (see Section 4.2) have also been grouped together. On the basis of their having noun classes similar to Niger-Congo, the Kordofanian languages were classified with Niger-Congo, as an early-branching family (see Figure 1). This classification, however, was not accepted until formal similarities (shared sound-meaning correspondences) could also be demonstrated, to show that "the resemblance of the Kordofanian noun class system to that of Niger-Congo is highly systematic and not merely typological" (Schadeberg 1981: 123–24, as referenced in Williamson 1989b: 13). The same sort of detailed comparative work has recently been performed on some of the languages in Khoisan, somewhat to the detriment of its unity (see Section 3.1).
In another non-genetic approach, areal or geographic classification, the place where languages are spoken, is crucial. Languages are grouped together on the basis of being spoken next to each other and typically share features due to this proximity. The Atlantic Group of Niger-Congo was given its first name “West Atlantic” on this basis: its languages are spoken along the Atlantic seaboard of West Africa. At higher levels of classification geographic names are often used to designate a group, e.g., the Northern, Southern, and Central branches of Khoisan; the branches of Mande are known as Northern-Western, Northern, Southwestern, Southern-Eastern, etc. There is nothing wrong with this practice per se. What is misleading is when these geographical groupings are put forward or interpreted as genetic groupings.

A final technique is one rarely used to determine genetic relationships but is useful at the lowest level of classification, i.e., determining whether two varieties are dialects of the same language or two separate languages. The test is one of mutual intelligibility, evaluating whether speakers of Language A can understand the speakers of Language B and vice versa. It is a more finely grained measure than the others and is used primarily to measure the closeness of two varieties already known to be close, closer than varieties evaluated by the comparative method. In one of its many avatars, the approach plays a speech sample from Language B played to speakers of Language A. The Language A speakers then tested on how well they understand the speech sample. And then the reverse is done: speakers of Language B try to understand a speech sample from Language A. Although this method has not been put to widespread use by linguists in Africa (but see Ladefoged et al. 1972, Schuring 1985, Slabbert & Myers-Scotton 1997), it has been widely used by social psychologists and sociolinguists, e.g., Labov 1981. It features prominently, of course, in the toolbox of missionaries, particularly those intent on Bible translation, e.g., Casad 1974, Grimes 1990 (see Simons 1979 for a review). With the information gleaned from such an assessment, missionary analysts can make decisions as to how many different Bible translations are needed for a given area and even which dialect should be chosen as the standard.

What is appealing about the approach is that it offers a measure not of how well speakers hypothetically can understand each other on the basis of a structural analysis, i.e., shared lexicon and grammatical features, the regular

---

26. The “West” part of the name was subsequently dropped (Doneux 1975). Even earlier the group had been named “North-West Atlantic” (Koelle 1834a).
An Introduction to African Languages

data of linguists. Rather it is a very real measure of whether people can and want to understand the speakers of another variety. It allows the inclusion of social and psychological factors into the evaluation, both of which have been claimed to account for a sizable chunk of mutual intelligibility (15% in one estimate (Simons 1979: 186)). Because of this last feature, an analysis will often turn up surprising asymmetries that are not evident in a purely structural analysis, particularly when there are unequal social roles, e.g., between the Pygmy Efe foragers and the Nilotic Lese farmers (Grinker 1994). For example, speakers of Language A understand speakers of Language B at a rate of 80% but speakers of Language B understand speakers of Language A at a rate of only 60%.

Such measures allow insights into social processes and language change, insights absent from a purely structural analysis, ones that allow a more accurate extrapolation from the purely synchronic to the diachronic. These insights may also allow for extrapolations in the other direction: predictions can be made about language shift and even language death. For example, in the townships of South Africa and elsewhere, urban varieties have arisen that have been claimed by both insiders and outsiders to be unintelligible to outsiders, despite being just an age-graded dialect or slang.

A purely structural analysis shows one such variety, Isicamtho [isamtho], to be little different from Zulu, something like a youthful slang. Why, then, are outsiders unable to understand it, or rather, why do outsiders say they cannot understand it? The reason has to do with the associations of Isicamtho with criminality, violence, sexism, etc., as is the case with many urban vernaculars in Africa. Such varieties as Isicamtho are spoken by young unemployed males who in South Africa typically belong to gangs and have spent time in jail. Undoubtedly the variety is accorded a great deal of (covert) prestige in some circles, but the general populace does not accept the values associated with it and so makes little effort to understand it. In summary, then, an evaluation of intelligibility, in uncovering the social-psychological dimension, tells us much more than a purely linguistic analysis. Although such measures have not often been used in genetic classification, they do provide information on the relatedness of languages at the micro level. From the data they furnish one can make social-historical inferences and thus obtain information on past and future relations between language varieties, useful information for classification.

Before concluding this section, it must be admitted that non-linguistic factors have historically complicated the process of classification in Africa. In the nineteenth century and even beyond, racial factors were important: “races” were equated with language families and sometimes languages were placed on
an evolutionary scale. Some languages (and thus their speakers) were seen as less developed and more primitive than others. Non-European languages were not surprisingly considered to be more primitive, and thus lower on an evolutionary scale, than those displaying the grammatical features of the languages of the European analysts (see Greenberg 1959). This reasoning resulted in “the racist ‘Hamitic’ concept, which wrongly brought physical type and cultural traits such as pastoralism into linguistic classification and needlessly complicated the placing of such languages as Maasai” (Bender 1996b: 5). Languages such as Fula, Maasai, Somali, and Nama were grouped together in a group called “Hamitic” and were said to have varying amounts of “Negro admixture” and to belong to a superior Caucasoid racial type (Meinhof 1912). Picture 3 shows a Fula couple, Picture 4 a Maasai mother with her son and grandson, and Picture 5 a Nama elder.

Greenberg 1949 shows that Fula (the language) should not be classified with Hamitic on the basis of the Fulfulbe’s (the people’s) looks but rather with the Atlantic Group, one of the earlier branchings of Niger-Congo (Figure 1). The Fulfulbe have non-Negroid looks but are centrally located, when linguistically classified from a genetic perspective, within the heart of the basically Negroid Atlantic group (Doneux 1978). Picture 6 and Picture 7 feature Kisi citizens of the town of Kpelloe, Liberia. Kisi also belongs to Atlantic.

Such notions as the primitiveness of African languages take a long time to displace and still hold sway in popular opinion. When non-linguists or beginning students ask me if I find African languages “primitive”, I can usually jolt them into some rethinking by saying that I know an African language with twenty-three genders! (Romance-intent undergraduates are particularly curious now that “gender” has replaced “sex” in politically correct and even popular usage.)

With this brief overview of general approaches to the classification of languages and one shameful admission (the Hamitic Hypothesis), we now turn to the mass-comparison approach initiated by Greenberg, the results of which transformed the classification of African languages. Note that the approaches discussed in the following section all represent genetic classification.

---

27. The implausibility of this grouping is underscored by the fact that it contained languages from all four genetic phyla (Bender 2000: 57).

28. Bender notes than an overview of the debate on Mischsprache (the idea that languages of the Sudan were mixed) and the concept of Hamitic-Nilotic can be found in Gregersen 1977; an expanded statement of Greenberg 1949 appears in Greenberg 1963.
Picture 3. Fula couple

Picture 4. Maasai mother, son, and grandson (courtesy of Doris Payne)

Picture 5. Nama elder (courtesy of A. Traill)
Picture 6. Kisi elder (courtesy of Bill Baron)

Picture 7. Kisi town chief and his family (courtesy of Bill Baron)
2.3 “Megalocomparison” or mass comparison?

The title of this section comes from an article that questioned Greenberg’s methodology when it was applied to Amerindian languages (Matisoff 1990; see Campbell & Mithun 1979), where his approach had nowhere near the success and acceptance it had with African languages. Much of the animosity from Amerindianists is likely attributable to cultural differences, especially the longstanding tradition of “splitting” (see note 18) in classifying New World languages.29

This section is limited to Greenberg, who, it should be noted, despite his new approach and considerable insights, relied on the work of others before him, especially Westermann (Williamson 1989b: 7) and Köhler (Bender 1996a:430). His proposal appeared in partial form in Greenberg 1949 and was expanded upon in subsequent publications, achieving something of a final version in Greenberg 1963.

Before Greenberg, the (genetic) classification of African languages was not based entirely on linguistic criteria, as it should have been. As mentioned above, typological and areal criteria were used, even racial ones. Although he can be faulted on some grounds, Meinhof was the first to apply the received comparative method to African languages, e.g., Meinhof 1899 [1910, 1932], a method which had so successfully been used to show the relatedness among Indo-European languages. Greenberg, however, adopted a new approach.

What Greenberg sought to do is what many others before and after have tried to do, namely, classify the linguistic groups in Africa according to genetic criteria, but he did it in a way some have considered too “quick and dirty”.31 For many reasons the traditional methods used by the Indo-European comparativists were of limited use in classifying African languages (Herbert 1997:ix),32 so Greenberg invented another approach. What some have found objectionable, roughly speaking, is that he did not follow the prescribed

29. The cause may also be a lack of agreement on methods or a lack of familiarity with statistical methods (Cavalli-Sforza et al. 1994:96).
31. See Winston 1966 for an early statement from an Africanist on this point; Goodman 1970 contains some pointed criticism.
32. There are time limitations to this approach, and the age of African language families may render it useless: Nichols 1992 has stated 8,000 years as a cut-off point for the viability of the comparative method; another source puts it at 10,000 years (Renfrew 1994:123).
The classification of African languages

methodology (the comparative method or Neogrammarian model) but rather used a new technique initially known as “mass comparison” (sometimes upgraded to “multilateral” comparison). Certainly Greenberg encountered opposition and even some heat when he made his original proposal, but he experienced nowhere near the vitriol as when he treated Amerindian (e.g., Campbell 1988, Chafe 1987); nor was his “lumping” (see note 18) in the Indo-Pacific (Greenberg 1971) widely accepted. Thus, the method does not have wide acceptance outside Africa. Nonetheless, it certainly put things to order in Africa.

In applying the technique of mass comparison, a great number of languages are compared; the analyst interprets a large number of shared cognates as representing a genetic relationship. It is not a comparison, as in the comparative method, typically limited to a few languages and the reconstruction of a proto-language, but rather involves forms from a great many languages. In later explicating his method, Greenberg sets up “a dichotomy between the traditional comparative method of ‘looking at many forms across only two or a few languages’ and his own method of ‘looking at few forms across many languages’” (Matisoff 1990: 118).

In addition to looking at many languages, Greenberg did not restrict himself to lexical items as had been done in the past, but also compared the form and function of grammatical morphemes. It was not enough that languages shared grammatical systems (typological comparison) but necessary also that the forms of grammatical morphemes be compared as well. He regarded the latter as being particularly strong evidence of relatedness. To most scholars such methodology seems unquestionably acceptable, at least in the preliminary stages when data and full analysis are unavailable. But many grumble that such groupings are incomplete and finally inadequate. Further analysis, however, has not dramatically changed his findings with regard to Africa; some solace can be taken in that his classifications have been generally accepted. Other approaches have not met with the same recognition and approbation, e.g., Mukarovsky 1976–77, Mann & Dalby 1987; Newman 1989 absolutely slams the latter.33

The one major revision is in the branch Greenberg called Kwa, one of the

33. Winston 1966 is a balanced early review of Greenberg’s classification; more critical is Oliver 1966 and Flight 1985, which come from a historian’s perspective; Greenberg 1972 is a response to the first. Bender 1996a contains highly critical comments on Greenberg’s work on Nilo-Saharan. Ruhlen 1987 defends the method of multilateral comparison. Newman 1995b provides a thorough overview and history of the controversy surrounding Greenberg’s methodology.
main branches of Niger-Congo (see Figure 1), the major phylum in Africa (see Section 2.1 for more details). What happened was that Eastern Kwa went to Benue-Congo, another major division of Niger-Congo, and Western Kwa became “(New) Kwa”. The reclassification originated in Bennett & Sterk 1977, which was subsequently criticized and re-analyzed in Schadeberg 1986, whose re-interpretation is widely accepted. Note that although this reclassification is a substantive one, it should not detract from the validity and considerable stability of Greenberg's comprehensive original work.

In the presentation of Schadeberg’s results given in Table 4, Kwa has been split into two, some of the languages are in “(New) Kwa”, while others are found in “(New) Benue-Congo”.

Table 4. Branching trees (Schadeberg 1986 as in Williamson 1989a:17)

| 1. Fula   | 5. Gur   | 9. (New) Benue-Congo | 9.5 Igboid   |
| 3. Temne  | 7. (New) Kwa   | 9.2 Idomoid | 9.7 Cross River |
|           |           | 9.4 Edoid   | 9.9 Bantoid |

The way the display in Table 4 is to be interpreted, as represented in Figure 3, is to see each member as representing a branching off from the parent stock (“Proto-Niger-Congo”). What we see, then, is that “Fula” (Fulfulde) is the first language to branch off, followed by Diola (“Dyola”) and so on. Note how the first three branches are all separate Atlantic languages. All three belong to a proposed single group (Atlantic). It would be expected, then, that the three would branch off together. They do not because of the cognacy cut-off point of 18%. According to this criterion, no languages can be grouped together that share less than 18% of their common vocabulary, a disabling blow to the purported unity of Atlantic (see Section 2.5.2). It is not just within Atlantic that each language constitutes a separate branch in this treatment, but also outside; the Nigerian language Igbo also constitutes its own branch. The nodes in language classification, then, can represent one or many languages (see p.30 for further examples).

34 The “New’s” are a convenient way to keep the familiar names but indicate their difference from the older groups.
Both studies on the re-alignment of Kwa were based on lexicostatistical analysis, that is, based on percentages of shared cognates between individual languages (see Hinnebusch 1999 for an introduction, Embleton 1986 for a review of the method). Lexicostatistics measures how much two languages have in common in terms of shared basic vocabulary. The technique is an important one in determining genetic relationships, albeit one of several that can be used (Bastin & Piron 1999: 149).

Lexicostatistics was based on the Neogrammarian model, which sees languages as related when they could be shown to have come from a common ancestor (see 2.2). The ways in which the languages under consideration are different from that common ancestor are assumed to be historical changes that they have undergone since they were part of that common ancestor. If two or more of the daughter languages share common changes or innovations, then they are seen as more closely related than are other daughter languages. This technique of “shared innovations” (see Newman 2000a) is an important tool for determining branchings, such as those presented in Figure 3 (the branchings arrived at by a different method, as discussed above).

As a last point, lexicostatistics should be clearly differentiated from glottochronology (Gudschinsky 1964), which, nonetheless, is also based on lexicostatistical data. Using lexicostatistical results, the glottochronologist estimates how long ago proto-languages split apart. Table 5 gives some numbers to show how the technique works. Note that the values in both columns are median values. That being said, one can calculate how far back languages diverged on the basis of how much shared basic vocabulary they have. For example, if languages share only 30% of basic vocabulary, they separated some 4,000 years ago.

The critical (and, to many, unwarranted) assumption is that languages change at a standard rate.\(^{35}\) One factor that can explicitly confound this

---

method is the degree and intensity of language contact, especially when borrowings cannot be identified (Hinnebusch 1999), a phenomenon discussed explicitly in Section 7.2.

With this information as background, we can now turn to some specific and often controversial cases.

### 2.4 And then there were three: Merging Nilo-Saharan and Niger-Congo

There are a number of issues that swirl around the group known as Nilo-Saharan, both internal and external. There are questions with regard to its constituency and its members’ relationships, as well as to its relationship to other phyla, particularly, Niger-Congo. At least one leading scholar considers Nilo-Saharan to be the least known of the four phyla (Bender 1996a: 10), although others argue that Khoisan deserves the honor. Nilo-Saharan has certainly engendered much classificatory interest. In this section I present some of the positions on whether or not Nilo-Saharan constitutes an independent group. Questions as to the group’s constituency are treated in 2.5.1.

Although the four African phyla are well established and by definition are separate and unique, there have been proposals to merge two. Under one proposal Nilo-Saharan would be seen as the older, earlier branch of the macro-phylum also containing Niger-Congo. The evidence has been accumulating as more is understood about Nilo-Saharan and many see the merger as inevitable.

---

**Table 5. Median dating and retention rates (Ehret 2000: 287–8)**

<table>
<thead>
<tr>
<th>Years [BP]</th>
<th>Retention rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>86</td>
</tr>
<tr>
<td>1,000</td>
<td>74</td>
</tr>
<tr>
<td>2,000</td>
<td>55</td>
</tr>
<tr>
<td>3,000</td>
<td>40</td>
</tr>
<tr>
<td>4,000</td>
<td>30</td>
</tr>
<tr>
<td>5,000</td>
<td>22</td>
</tr>
<tr>
<td>6,000</td>
<td>16</td>
</tr>
<tr>
<td>7,000</td>
<td>12</td>
</tr>
<tr>
<td>8,000</td>
<td>9</td>
</tr>
<tr>
<td>9,000</td>
<td>7</td>
</tr>
<tr>
<td>10,000</td>
<td>5</td>
</tr>
</tbody>
</table>
The proposal that Nilo-Saharan and Niger-Congo constitute a single phylum was first articulated in Gregersen 1972, although hints of the proposal are detectable in the work of Westermann (Williamson 1989b: 7). Gregersen proposes that two phyla be unified into a larger phylum called “Kongo-Saharan”. This claim was furthered with some changes in Bender 1981, and Blench 1995 adduces further data and proposes a new name “Niger-Saharan” on the basis of some non-lexical evidence.

Niger-Saharan, as presented in Figure 4 illustrates the configuration of the combined two phyla. Roughly speaking, Nilo-Saharan languages find themselves as various branchings at the top of the tree, with the entire Niger-Congo set of languages as a lower node coordinate with Central Sudanic. Instead of Nilo-Saharan being one large branching at the root of Niger-Saharan, the relationships are much more finely nuanced in Blench’s treatment: Niger-Congo forms a unity with only a subset of Nilo-Saharan. The rest of the Nilo-Saharan languages that have proved problematic (see 2.5.1) retain their relatively autonomous positions atop the tree.

![Figure 4. Niger-Saharan (Blench 1995: 88)](image.png)

The evidence that Blench finds the most compelling is morphological — what he considers identifiable remnants of the Niger-Congo system in Nilo-Saharan, for example, the ma- prefix for liquids or mass nouns corresponding

---

36. This proposal comes not from vainglory, but rather because Gregersen’s term “gives a misleading impression to non-specialists of the location of the families” (on a suggestion from Kay Williamson (Blench 1995: 84, 97)).
to η- in Kordofanian. In the way of phonology he finds ATR vowel harmony (see 3.5) and labiovelars (see 3.3) in uneven distribution, being confined to only some sub-groups. Despite wide disparities in the quality of the data, he finds that the lexical material confirms the analysis suggested by the phonological data. Note that such features as labiovelars have long been recognized as areal in their distribution (Greenberg 1983) and thus liable to areal diffusion. The picture he presents is given in Figure 4. The list of languages to the left constitutes a single node.

The next section deals with the splitting side of language classification, deciding whether a language belongs to a group or whether a set of languages constitute a group on their own.

2.5 Some local problems in language classification

This section illustrates some more specific problems that have arisen in the classification of African languages by the presentation of some fairly well described examples. Assuredly there are many more problems that could have been treated; what follows is a sample of the problems and issues and the way linguists have gone about resolving them.

2.5.1 Nilo-Saharan (again)

To give the reader a flavor of how various and variable classifications may be, I give the multiple classifications of two of the Nilo-Saharan “Outliers” in Bender 1996b (see p. 30 above), Kuliak and Songhay. After reading this section the reader will probably not want to hear about the other outliers, nor about the other problematic member, the Kado group in the Kordofan Hills of south-central Sudan (Blench 1995:88).37

The first of the two Outliers to be treated here is Kuliak, a set of languages spoken in the mountainous regions in the Karamoja area of northeastern Uganda (see Map 5). So, one of the Kuliak languages, is spoken on three extinct volcanoes in the Karamoja area of northeastern Uganda.

The list in (5) shows ways in which Kuliak has been classified. Not all of these classifications are, strictly speaking, mutually exclusive; they are nonetheless substantively different and certainly illustrate the differences that can arise, even with the best of intentions.

(5) Classifications of the Kuliak family

Fringe Cushitic in the Afroasiatic family (Tucker 1967, as referenced in Carlin 1993);
A distinct branch of Eastern Sudanic in the Nilo-Saharan family (Ferguson 1963, Greenberg 1963, as referenced in Carlin 1993);
Independent, a non-aligned language family (Tucker & Bryan 1966; Laughlin 1975; Heine 1976a, as referenced in Carlin 1993);
An “Outlier” of Nilo-Saharan (Bender 1996a);
An early branching from Proto-Nilo-Saharan (Blench 1999a).
Even more confusing is the classification of Kuliak’s fellow Outlier, Songhay. According to Williamson 1989b:9, the uncertainty of Songhay’s status was one of the reasons for the original lumping of Nilo-Saharan and Niger-Congo in Gregersen 1972 (see Section 2.4). If Songhay belongs to both families, then the two families should be united. (Few others advocating the merging of Nilo-Saharan and Niger-Congo use this argument.) Note that the classificatory positions of Songhay range from isolate to membership in every language phylum, save Khoisan (Songhay fortunately has no clicks!). Furthermore, the families to which it has been assigned in Niger-Congo, Mande and Gur, are quite separate from each other (see Figure 1). The final point to be made about the display in (6) shows that even linguists can change their minds.

(6) The many positions of Songhay

- An isolate, Westermann & Bryan 1952; Greenberg 1955, as in Williamson 1989b:9
- Nilo-Saharan, Greenberg 1963 as referenced in Dwyer 1989:9; the first branching in Nilo-Saharan, Welmers 1973 as referenced in Bender 1996b:59; an “Outlier” in Nilo-Saharan Bender 1996a:64
- Chadic, as referenced in Nicolaï 1990
- Mande, Delafosse & Caquot 1924 (1952), as referenced in Williamson 1989b:9
- Gur, Westermann 1927 as referenced in Williamson 1989b:9
- A mixed language based on Touareg and Mande; originated as a pidginized variety of Touareg superimposed on a Mande structural framework, Nicolaï 1990

A recent worker on the Songhay languages concludes that we might as well just throw our hands up in the air and accept the fact that Songhay resists easy classification. As he puts it, “The wider genetic affiliation of the Songhay language family has not been resolved” (Heath 1999:2; cf. Heath 1998). This is the position also suggested in Dimmendaal 1995 in the absence of adequate historical sources.

The next section treats the Atlantic family, a set of languages that perhaps should not be grouped together at all (Childs 2001c).

2.5.2 Location, location, location: Atlantic

The main point of this section is that areal factors often play and have played an important role in the classification of African languages. The second point is
that scholarly inertia reinforces mistakes, which are thereby perpetuated indefinitely, effectively forestalling any re-examination of the facts. The Atlantic Group likely proceeds from the lumping side of classificationists (see note 18). The reader might also note the relatively minor point in this section that arbitrary decisions in lexicostatistics are sometimes possible.

In Williamson & Blench 2000 (see Figure 1), Atlantic represents a collateral branching with Mande after the first offshoot, Kordofanian, and thus is one of the older branchings. The Atlantic languages were one of three families to be bullied about by the Mande warriors of the fifteenth century; the other two, Gur and Kru are equally interesting from the point of view of language classification and areal factors, a position that will be presented in greater detail in Section 7.2.3. It is likely true that contact with Mande, especially the Manding core, has structurally affected Atlantic languages (Childs 1995c, Childs 2001a); this contact may be obscuring an earlier, closer genetic unity.

The classificatory problems with Atlantic have hardly been resolved but have rather been abandoned in favor of a non-genetic classification. The most recent commentary on the group as a whole admits that the classification is not genetic: “The two features that make Atlantic a meaningful entity are typology and geographical distribution” (Wilson 1989: 81).\footnote{38} To many the group seems something of a “wastebasket group”,\footnote{39} the category to which all languages are assigned that do not fit into the more cohesive Mande family (but see note 39) found in the same area (see Map 6). To some extent the same exclusionary principle is used to group the Khoisan languages together, along with the typological criterion of using clicks (see Section 2.5.3). The internal diversity of Atlantic has been characterized as follows:

West Atlantic … is a very diverse group, containing at least three major subdivisions. It is possible that some language groups traditionally assigned to West Atlantic are in fact coordinate branches of Niger-Congo. There is no apparent common innovation linking West Atlantic, and evaluation of its status must await further detailed investigation. At present, all that can be said is that the lexicostatistical distance between branches of West Atlantic is nearly as great as that between West Atlantic and the remainder of Niger-Congo. (Bennett & Sterk 1977:248)

\footnote{38} The same claim has been made for Mande, e.g., by Köhler (Kastenholz 1998 p.c.), although the percentages within Mande are somewhat higher (17% or more Bimson 1978: 5 but 20–24% in Welmers 1958 as referenced in Kastenholz 1991/92: 110).

\footnote{39} Gur is also “a rather heterogeneous and ill-defined group” (Bendor-Samuel 1971:143).
Bennett and Sterk found levels of cognacy so low that they posited each of the Atlantic languages they treated (Fula, Diola, and Temne (see Figure 3) as
The classification of African languages

49

separate branchings (see p. 40), and gave each a “0” in their scoring.40

Wilson 1989 presents the lexicostatistical data, which is pretty grim for anyone championing a cohesive Atlantic Group. Within several subgroups of Atlantic, e.g., the Baga languages, Sherbro, the Cangin group, there are more credible shared lexicons (above 30%), but for the group as a whole the percentage of shared basic lexicon is below 10% (see note 41).

I. Northern Branch
   A. Senegambian languages: Fulfulde, Serer; Wolof
   B. Cangin: Lehar, Safen, Non; Ndut, Palor
   D. Eastern Senegal-Guinea
      1. Tenda*: Tenda Mayo, Basari; Onian, Bedik; Konyagi
      2. Bifada, Badyara
      3. Buy, Kasanga; Bainouk
   E. Nalu: Nalu, Mbulungish, Pukur

II. Bijagó (n = 1)

III. Southern Branch (n=17)
   A. Mansoanka
   B. Mel languages
      1. Temne; Baga Binari, Baga Maduri, Baga Sitemu, Baga Tchitem, Baga Koba, Landuma
      2. Bulom languages: Kisi; Mmani, Sherbro, Krim, Bom
      3. Gola
   C. Limba

* Tenda may also be a non-genetic group based on geography, referring to languages spoken by minorities in eastern Senegal and Guinea-Conakry (Ferry 1991).

Figure 5. The classification of Atlantic

Some analysts have gone so far as to say that individual languages within the group have closer genetic relationships with languages outside the group than with languages within, e.g., Mukarovsky 1966, Wilson 1989. A more recent proposal deals with Bijagó, an agreed upon isolate within Atlantic. The Bijagó, for the most part, live on the islands of the Bijagó Archipelago off the coast of Guinea-Bissau and on the nearby mainland, where they have been relatively isolated for centuries (Henry 1994). The proposal is that Bijagó shows closer

40. This has been seen as an overreaction; arbitrarily replacing the 0’s with 17% vastly overstates the level of cognacy among these three languages by as much as threefold (Schadeberg 1986). Seeing this figure of 17%, some have inferred that Atlantic as a whole has a 17% level of cognacy! (Podzniakov 1999 p.c.).
links to languages outside Atlantic than within, namely, to a “pre-Bantu” substrate, “un stade anterieur au bantou commun” (Segerer 1997:6).

The lesson from this section, as indicated in (7), is that areal features may be used as a first step in language classification but cannot be relied upon for a truly genetic classification, even if the languages are related to no other group. There are other revisions to the classification in Figure 5 that will be discussed in a later section. Mansoanka, for example, is a Mande language or at least a mixed language combining both Atlantic and Mande components (Segerer 1999 p.c.). As will be seen in Section 7.2.3, the history of the interaction between Mande and Atlantic has led to all manner of blends.

(7) The lexicostatistics of Atlantic and some interpretations

- Shared basic lexicon is around 8%;
- Languages of the group do not rise to the 18% level of cognacy needed for languages to be considered related in Bennett & Sterk 1977;
- Papel is closer to Common Bantu than to other Atlantic languages (Wilson 1989);
- Bijagó may have a pre-Bantu substrate (Segerer 1997).

In conclusion, neither lexicostatistical data nor shared innovations (nor reconstructions) seem to favor treating Atlantic as a genetic group. The culture of lumping and inertia are the only factors favoring its continuance. From the next section we will draw a similar lesson, this time with regard to typological classification: typological features are not reliable for genetic classification.

2.5.3 Clicks and classification: Khoisan

The Khoisan language group, one of the four major language phyla of Africa, has provided analysts with problems of both data and interpretation (see Sands 1995 for a review). What is particularly fascinating about the phylum is that two purported members, Sandawe and Hadza, are located in Tanzania half a continent away from the major locus of the family. Map 7 shows just the “Southern” Khoisan languages in pre-colonial times (excluding Sandawe and Hadza). In the recent past it has been estimated there were more than a 100 languages — today less than 30 exist (Güldemann & Vossen 2000:99), and many of them are bound for extinction (see 1.1).

This distance, as well as their speakers’ physical similarities to their northern neighbors rather than to the distinctive southern African type (Bender 1996a:9),
The classification of African languages has engendered some doubts as to their genetic relatedness to greater Khoisan. Recent work has shown these suspicions to be valid. If indeed these two click languages cannot be shown to be related to the Khoisan languages of southern Africa, the question then arises as to which languages Sandawe and Hadza are related to, or are they simply linguistic isolates, as was suggested for Bijagó of Atlantic in the preceding section? The general lesson to be learned from this
section is that a typological feature, however distinctive, cannot genetically define a linguistic group.

In addition to the question of the relatedness of the two distant candidate members, a more general question is whether the click languages constitute a genetic group at all. They obviously share a highly marked linguistic feature, the use of the velaric speech mechanism to produce five basic sounds (clicks at five different places of articulation; see Table 6) which, along with their many accompaniments, have led to enormously large phonemic inventories. From Westphal 1971 on to the present, “Khoisan” has been used as a cover term “to refer to those African languages which use clicks as regular speech sounds and which are not obviously members of one of the other families” (Sands 1998: 75). We have already seen how the latter condition and areal factors led to the promotion of Atlantic as a genetic group. The same factors are at work here with the addition of the presence of clicks (and perhaps a racial type). This combination has led analysts to group the languages together and even to include two languages which do not satisfy the areal criterion. The implicit assumption is that clicks (typology) are more important than location (area).

It is hard to come by evidence from shared cognates or reconstructed proto-forms to show a true genetic relationship among the Khoisan languages (Traill 1986). In addition to the question of the group as a whole, there is some question as to the genetic relatedness of the two major sub-groups, the Khoi and the San, who may be only culturally differentiated (Wilson 1986). In a statement about the Khoisan languages of South Africa, Traill baldly states, “The South African Khoisan languages fall into two linguistically unrelated [my underlining] groups” (Traill 1997a).

Although it will be impossible to enter into the evidence in any great depth here, it seems likely that both Sandawe and Hadza do not form part of Khoisan. Work on noun class systems suggests that Hadza may not belong (Sands 1998). The implication here is that just as Atlantic may not form a coherent group, so may not Khoisan; in fact it is much more likely. Africanists must simply learn to live with the fact that these isolates are possible if not common outcomes of language change and diversification, determined to stymie neat classification.
and tidy formalizations, particularly as greater time depths are considered. These findings, incidentally, do not hold promise for those interested in Proto-World, the ultimate lump, the mother of all languages, a language from which all languages are descended.

As this chapter ends with a denunciation of clicks as inadequate data for genetic classification, the next will begin with their exaltation as one of the most phonetically interesting phenomena to be found in Africa.