Africa's Linguistic Diversity
Bonny Sands*
Department of English, Northern Arizona University

Abstract
African language classification in the latter half of the 20th century has been dominated by Joseph Greenberg's work classifying African languages into four linguistic genetic groupings: Afroasiatic, Niger-Kordofanian, Nilo-Saharan, and Khoisan. Current research indicates that there are a minimum of 20 unrelated African language families or isolates. Africa's linguistic diversity will remain poorly documented unless significant efforts are made to document some of the continent's most endangered, underdescribed languages and language families.

1. Introduction
The African continent is commonly characterized as having four language families or phyla. This characterization stems largely from the work of Joseph Greenberg, summarized in a monograph in which four language groupings (Niger-Kordofanian, Afroasiatic, Khoisan, Nilo-Saharan) are presented as ‘a complete genetic classification of the languages of Africa’ (Greenberg 1963: 1). This picture of African languages is inadequate for several reasons. First of all, it ignores the continent’s 27 or so sign languages, which belong to a minimum of eight unrelated lineages. Second, it ignores languages which belong to language families spoken primarily outside of Africa such as Indo-European and Austronesian. This means that Afrikaans and Malagasy are not considered to be ‘African’ languages, even though the majority of speakers of those languages live in Africa¹. Also, a number of language isolates have been identified which were not identified by Greenberg. Finally, and perhaps most importantly, a critical examination of Greenberg’s groupings (of spoken languages unique to the continent) shows that the number of unrelated lineages is at least 8, or twice the diversity commonly supposed.

African languages represent a rich set of resources for scholars interested in learning about language universals, processes of language change, and African prehistory. For instance, typological surveys based on outdated classifications may significantly undersample African languages. Linguistic classifications are often correlated with patterns found by archaeologists and geneticists, but if the classifications used are outdated ones, then the models of African prehistory that are generated are on shaky ground.
Classifications are important because they influence subsequent research in linguistics and other fields. For instance, a language classified as an isolate (e.g., Basque, Burushaski, Zuñi) is more likely to be included in a typological survey, or to be the focus of a research program, than is any individual language classified as part of a large language family. A number of African language isolates have recently been identified, and this number may increase as documentation improves. Since earlier classifications were often based on only a handful of lexical items, the mistakes made in classifications can be extreme. What was thought to be a dialect of one language may turn out to belong to another language family entirely. In other cases (e.g., Dogon, Songhay, Kunama), what were thought to be single languages turn out to be clusters of closely related languages. Perhaps shockingly, fieldwork even in the present day has the likelihood of uncovering previously unknown languages.

The field of African historical linguistics has undergone a shift in focus from language classification to the study of processes of language change, particularly those due to language contact. By better understanding how similarities between languages have arisen through such mechanisms as borrowing, areal diffusion and chance, researchers are increasingly questioning the older classifications found in reference works such as the online database Ethnologue (Gordon 2005). Researchers have also questioned Greenberg’s methodology (e.g., Fodor 1966; Campbell and Poser 2008). Greenberg and the generation of scholars immediately following him were largely interested in long-range comparisons and classifications, but currently, there has been more focus on reconstructing low-level groupings. This paper presents results from recent historical-comparative studies and places them against the backdrop of the commonly accepted picture of African language families. As will be seen, the linguistic genetic classification of African languages is far from settled. There are many cases where the classification of languages and language families is extremely uncertain and where additional research may yield very interesting results.

2. No classification without documentation

By one count, the African continent is home to some 2058, or 30% of the world’s languages (Gordon 2005). A survey of the specialist literature comes up with a smaller figure of 1441 different languages (Maho 2004). With such a large number of languages, a pragmatic, referential classification of African languages plays a useful role. Greenberg’s classification of African languages into four primary groups has played such a role. That is, it serves as an organizing tool for scholarly meetings, edited volumes, and introductory-level language surveys. For example, there are regularly held conferences for Nilo-Saharan, Afroasian, and Khoisan languages, but none for small language families such as Koman, Khoë, Aroid, Ron, or Kuliak. Most linguists, and indeed, many Africanists, are unlikely to have
heard of many of these smaller language families, and there are relatively small numbers of linguists who work on the various African language families of relatively shallow time depth.

So, we can say that the four-way classification of African languages has been a success for the majority of Africa’s best-documented languages. These tend to be the ones with large numbers of educated speakers, and they are often major regional or national languages. The least-well documented languages tend to be those used by marginalized populations. Languages used by current or former residents of conflict zones and hunter–gatherers are among the most underdocumented languages of the continent. About 300 African languages are used by populations under 5000, and there are even more languages with larger population bases that are still under serious threat due to language shift. Linguists based at African universities often have little support for studying languages spoken by minority, refugee, or migrant populations when there is so much work to be done on national and regionally dominant languages. Languages used by very small populations, or by people in remote locations, are typically not very likely to have a native speaker with university training in linguistics available to document the language. They are also less likely to have been the focus of missionary linguistic efforts, which have been the sole source of linguistic documentation for a large number of African languages.

Africa’s least-well documented languages are often those which have the greatest ability to shed light on African linguistic prehistory because they are typically used by populations that are not part of recent population spreads. These languages tend to be less well-served by the four-way classification because they have typically been classified based on superficial documentation. When a language has been briefly documented by someone who does not speak the language or who has little training in linguistics, there are often errors in transcription and word segmentation. Often, borrowings and loan-translations from other languages are unidentified.

The documentation problem is especially serious for languages of southwestern Ethiopia, Southern Sudan and northeastern Nigeria, Chad, and Cameroon, and for particular language groups: Chadic, Omotic, Khoisan, Temein, Gumuz, Kadu, Dogon, and Kordofanian (Batibo 2005; Blench 2007a; Brenzinger 2007a,b; Connell 2007). To give a partial illustration of the documentation problem, I have listed the 19 Afroasiatic languages beginning with the letters ‘A’ or ‘B’ which have fewer than 5000 speakers (Sommer 1992; Gordon 2005). Of this small, but representative sample, more than half are virtually unknown. Only two can be called well-described, though even they have not been the subject of any phonetic study, discourse study, or linguistic anthropological study. Almost a third of the over 300 Afroasiatic languages have fewer than 5000 speakers.
The urgent need for African language documentation is not widely recognized. In part, this may be because two recent surveys of African languages (Heine and Nurse 2000; Childs 2003) lack chapters on endangered languages.

3. Survey of African language classifications by pragmatic groupings

In Sections 3.1 to 3.7, I survey the classification issues that have arisen in each of several sets of African languages. The discussion is organized along pragmatic lines, rather than by linguistic genetic groupings. For instance, linguistic isolates are discussed in a single section even though they are not related languages.

3.1. Sign languages

It is perhaps understandable that Greenberg neglected to include sign languages in his survey of African languages. During the 1960s, there was very little documentation of African sign languages, and even sign languages were still not widely regarded as ‘real’ languages. The number of independent sign language lineages appears to be large compared to the number of spoken language lineages in Africa, but the time depth reflected by the oral versus spoken lineages is likely quite different. Typological studies tend not to include sign languages, and the historical linguistic study of sign languages is in its infancy. Studies do, however, indicate that sign languages undergo the same types of processes of language change as do spoken languages (e.g., Rimor et al. 1984) and can be compared using lexicostatistics (e.g., Woodward 1993, 1996; McKee and Kennedy 1995). Comparisons of sign languages using the Comparative Method have been lacking, but this cannot be construed as evidence that sign languages are fundamentally different than spoken languages.

As languages used primarily in Africa, African sign languages should be included in any comprehensive survey of African languages. Since sign languages are still typically left out of African languages surveys, I will discuss them first.

There are 27 reported African sign languages (Gordon 2005; Kamei 2006), but the origins of only ten of these have been reported. Half of these are

<table>
<thead>
<tr>
<th>Two have grammars:</th>
<th>Burunge, Argobba.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Six have grammatical sketches:</td>
<td>Alagwa, Anfillo, Arbore, Bayso, Boni, Burji</td>
</tr>
<tr>
<td>Five have limited vocabularies:</td>
<td>Aasax, Baldemu, Bambassi, Birgit, Buwal</td>
</tr>
<tr>
<td>Six have no linguistic documentation:</td>
<td>Ajanci, Awjlah, Barain, Beele, Boon, Buso</td>
</tr>
</tbody>
</table>
Africa’s Linguistic Diversity 563

reported to be language isolates: Adamorobe Sign Language (Nyst 2007), Bamako (Malinese) Sign Language (Gordon 2005), Tanzanian Sign Language (Mreta and Muzale 2001), Zambian Sign Language (Serpell and Mbewe 1990) and Maganar Hannu, or Hausa Sign Language (Schmaling 2000). Seventeen African sign languages whose origins are not currently documented are listed in Table 1.

Five African sign languages are known to have developed from languages with origins outside of Africa: Ghanaian, Nigerian, and Guinean sign languages are related to American sign language (ASL); Madagascar sign language is related to Norwegian sign language (NSL), and South African sign language is related to British sign language (BSL). ASL, NSL, and BSL are not known to be related to each other. Additionally, ASL is one of nine ‘foreign’ sign languages used in some schools in Africa (Schmaling 2001). Foreign sign languages have been sources of loans to African sign languages and thus must be considered as contributing to Africa’s linguistic diversity. In some cases, however, languages such as ASL constitute threats to the maintenance of local sign languages.

The comparative-historical linguistic study of sign languages is dominated by well-documented languages such as ASL, but African sign languages can surely contribute greatly to our understanding of how sign languages change over time depending on their sociolinguistic settings. Nyst (2007) has shown that Adamorobe sign language, a language that has been used in a small Ghanaian village for a relatively long time (~200 years), resembles ‘young’ sign languages recently developed from home sign languages in some respects. In other respects, it resembles sign languages used by large groups of deaf users. 

---

Table 1. African sign languages of uncertain origin (given currently available documentation) (Gordon 2005; Kamei 2006)

<table>
<thead>
<tr>
<th>Sign Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algerian sign language</td>
</tr>
<tr>
<td>Bura sign language (Blench and Warren 2006)</td>
</tr>
<tr>
<td>Chadian sign language</td>
</tr>
<tr>
<td>Congolesean sign language</td>
</tr>
<tr>
<td>Egypt sign language</td>
</tr>
<tr>
<td>Ethiopian sign language(s)</td>
</tr>
<tr>
<td>Gambian sign language</td>
</tr>
<tr>
<td>Kenyan sign language</td>
</tr>
<tr>
<td>Libyan sign language</td>
</tr>
<tr>
<td>Mbour sign language (Senegal)</td>
</tr>
<tr>
<td>Moroccan sign language</td>
</tr>
<tr>
<td>Mozambican sign language</td>
</tr>
<tr>
<td>Namibian sign language</td>
</tr>
<tr>
<td>Sierra Leone sign language</td>
</tr>
<tr>
<td>Tunisian sign language</td>
</tr>
<tr>
<td>Ugandan sign language</td>
</tr>
<tr>
<td>Zimbabwe sign language</td>
</tr>
</tbody>
</table>

---

© 2009 The Author
Journal Compilation © 2009 Blackwell Publishing Ltd
The linguistic diversity represented by Africa’s sign languages alone is greater than that which is widely assumed for the continent’s languages as a whole. Documentation of these languages appears to be urgently needed.

3.2. AFROASIATIC LANGUAGES

There are a large number of Afroasiatic languages in Africa, including such well-known languages as Arabic, Hausa, Somali, and Amharic. Two of the main families of Afroasiatic are represented by Berber languages, and by Ancient Egyptian (and its descendant, Coptic, which is also no longer spoken as a mother tongue). Table 2 shows the major, commonly-accepted subgroups of the Afroasiatic phylum, along with citations of recent surveys. Recent overviews of Afroasiatic classification can be found in Gragg (2004), Bender (2003a), Crass (2006), and Hayward (2000).

Smaller Afroasiatic language families that have recently been the subject of historical-comparative studies include: West-Rift Southern Cushitic (Kiessling and Mous 2003), Agaw (Central Cushitic) (Appleyard 2006), Biu-Mandara A (Gravina 2007), and Ron (Blench 2003). Research, particularly in Chadic, has moved away from a focus on classification to a focus on reconstructing particular linguistic subsystems, such as: noun modifiers (Wolff 2006), locative phrases (Pawlak 2001), subject clitics (Tosco 2007), and tense-aspect marking (Jungraithmayr 2006a). The role of language contact in accounting for linguistic patterning within and across language families has been shown in a number of recent works, e.g., Jungraithmayr (2006b), Azeb and Dimmendaal (2006), and Heine and Nurse (2008). Because Afroasiatic languages are so numerous, they can provide many examples showing how present-day areal patterns can reveal older patterns of linguistic contact.

Given that the subgrouping of well-documented families such as Indo-European are, at times, hotly contested, it is not surprising that there are many different classifications of the six groups listed in Table 2, particularly since most of these groups lack thorough reconstructions for most of their major subgroups. What may be surprising is the wide range of classifications given to some of the lesser-known Afroasiatic languages such as Ongota and Omotic.

<table>
<thead>
<tr>
<th>Table 2. Major Afroasiatic language groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cushitic (Tosco 2000; Gragg 2007)</td>
</tr>
<tr>
<td>Omotic (Bender 2003b, 2007)</td>
</tr>
<tr>
<td>Berber (Kossmann 1999, 2007)</td>
</tr>
<tr>
<td>Egyptian (Takács 1999)</td>
</tr>
<tr>
<td>Chadic (Newman 2003; Schuh 2003)</td>
</tr>
<tr>
<td>Semitic (Rubin 2008)</td>
</tr>
</tbody>
</table>

© 2009 The Author
Journal Compilation © 2009 Blackwell Publishing Ltd
Omotic is the most controversial grouping typically characterized as Afroasiatic because the grammatical formatives ‘to which Afroasiaticists have tended to attach the greatest importance are either absent or distinctly wobbly’ (Hayward 1995: 14). Greenberg (1963) and others considered Omotic to be a subgroup of Cushitic, but there have been repeated questions as to whether Omotic can be considered Afroasiatic at all (e.g., Theil 2006). One view is that only the most divergent set of Omotic languages, the Aroid languages (including languages Ari, Hamer, and Dime), are actually Nilo-Saharan (e.g., Zaborski 2004; Moges 2007). Most specialists regard Omotic as a separate sub-branch of Afroasiatic, but this might lead some to underestimate the internal diversity of the 25 or so Omotic languages, which appears to be greater than that found in Berber and Semitic.

Ongota is a fascinating Ethiopian language that is on the verge of extinction. In 2000, it was only spoken by eight people out of an ethnic group that numbered less than 100 (Savà and Tosco 2000). There were no linguistic data available on Ongota until the 1980s. The difficulty in classifying a language on the basis of a very small amount of lexical and grammatical data led Lionel Bender to call Ongota one of the ‘mystery languages of Ethiopia’ (Bender 1994). Ongota has been classified as a separate, major branch of Afroasiatic (Fleming 2006). But, there are other competing hypotheses. Ongota may be a Nilo-Saharan language, or perhaps an Omotic language or some kind of creole (cf. Blažek 2007). The most convincing hypothesis is that Ongota is an East Cushitic language with a Nilo-Saharan substratum (Savà and Tosco 2003). In other words, it appears that the Ongota used to speak a Nilo-Saharan language but shifted to speaking a Cushitic language, while retaining some characteristics of their earlier linguistic system.

Ma’a (also sometimes referred to as Mbugu) was classed as a Southern Cushitic language by Greenberg, but it is now best classified as a Bantu (Niger-Congo) language. Thanks to recent fieldwork (Mous 2003), Ma’a is now one of the best-documented examples of a mixed language, and as such, it is of great interest to historical linguists. Ma’a people actually alternate between two speech varieties, one of which has a marked use of Southern Cushitic vocabulary, but both of which have Bantu grammar. The sociolinguistic use of special registers or vocabularies has only been documented for a handful of African languages, but in most cases, they are important sources of information about (1) the role of individual choice in language change, and (2) sociolinguistic contexts of earlier linguistic contact.

3.3. Niger-Congo languages

Recent surveys of Niger-Congo language classification include: Olson (2006), Williamson and Olson (2007), Bendor-Samuel (2006), and Blench
Other important overviews are: Williamson and Blench (2000) and Bendor-Samuel and Hartell (1989). Recent contributions on Niger-Congo include Hyman (2007) and Stewart (2002). The generally accepted major groups of Niger-Congo languages are given in Table 3. Given that there are between 1000 and 1500 languages classified as Niger-Congo, it is not surprising that they represent a wealth of data for historical linguistic study. This group contains many closely related language families, and several distantly, or questionably related language groupings.

Kordofanian languages are of particular interest to historical linguists as they appear to represent the earliest branch of Niger-Congo. But, these languages are very sparsely documented, in part perhaps because few linguists have been willing to work in southern Sudan. Overviews of Kordofanian classification include Schadeberg (2006) and Quint (2006). Schadeberg (2006) considers Kordofanian to contain four subgroups: Heiban, Talodi, Rashad and Katla, whereas Quint (2006) and Dimmendaal (2008) dispute the inclusion of Katla in Kordofanian. Further documentation may hopefully help determine whether Katla languages belong to a subgroup of Niger-Congo languages other than Kordofanian. Quint (2006) considers the sparsely documented language, Lafofa, to be one of the four main branches of Kordofanian languages while Schadeberg considers it to be part of the Talodi branch. Probably the greatest change in Kordofanian classification has been the exclusion of the nine Kadugli languages from Kordofanian where it was placed by Greenberg. It is now generally considered either to be a Nilo-Saharan (Stevenson 1991; Blench 2006a), or an isolate language family (Hall and Hall 2004; Dafalla 2006). Another hypothesis is that Kadugli languages are Niger-Saharan, i.e., that they belong to a family comprised of both Niger-Congo and Nilo-Saharan languages (Ehret 2000).
Mande languages are quite distinct from other Niger-Congo families (Childs 2004), and because of this, their inclusion in the phylum is sometimes questioned (Mukarovsky 1966). Recent comparative-historical work on Mande includes: Kastenholz (2002), and Vydrine (2004).

Ijoid languages have not been the focus of much recent investigation. Greenberg (1963) had classed Ijoid as Kwa, but it is now generally considered a distinct subgroup of Niger-Congo. Recent surveys include: Lee and Williamson (1990), and Harry (1995).

Atlantic languages are a fairly diverse group of languages that has a very low rate of lexical cognacy. That is, the percentage of basic words that are shared by the languages is much lower than that found in Mande, and low enough to question whether Atlantic languages can even be considered to be a family since chance resemblances between unrelated languages may produce pseudo-cognates in a similar proportion of items. Estimates of cognacy range from 3% (Pozdniakov 2007) to 8% (Childs 2003). It is possible that Atlantic languages could actually be three distinct early branches of Niger-Congo (see Childs 2003, pp. 40–41 for a discussion).

Dogon languages are of special interest to historical linguists because the subgrouping of Dogon in Niger-Congo is in question (Bendor-Samuel 2006). Formerly considered a single language, Dogon is now considered to be a cluster of languages. One of these varieties, Bangeri-me, may not be a Dogon language at all, but may be a language isolate (Blench 2005).

The languages currently accepted as Kwa were only one of the subgroups of languages considered by Greenberg (1963) to be Kwa. Comparative-historical research in Kwa today is focused on Kwa subgroups rather than on any attempts to reconstruct Proto-Kwa. Some work has been done on segmental reconstruction of various Kwa subgroups (e.g., Snider 1990; Capo 1991, 1992; Stewart 1993); little work has been done on other linguistic subsystems. Kwa languages are of particular interest to linguists interested in how the Suriname creole of South America developed (Huttar et al. 2007).

Benue-Congo is the largest subgroup of Niger-Congo. A sub-branch of a sub-branch of the Benue-Congo languages is the well-known, large family known as Bantu. Despite their seemingly minor position within Niger-Congo, Bantu languages are at the forefront of much historical linguistic research. They have been used to test classificatory techniques borrowed from the biological sciences (e.g., Rexová, et al. 2006; Atkinson, et al. 2008). The expansion of Bantu languages has been the focus of numerous studies comparing archaeological and linguistic evidence (e.g., Eggert 2005; Bostoen 2007). Recent overviews of Bantu classification include Nurse and Philippson (2003) and Maho (2003). Some recent, significant contributions to comparative Bantu include: Nurse (2000, 2007), Mouguiama-Daouda (2005), Mouguiama-Daouda and Van der Veen (2005), Marten, et al. (2007), Masele (2001), and Schadeberg (2002).
Kru languages are among the most sparsely documented groups of Niger–Congo languages. Little has been done on either the internal or the external relations of the family, even though both are considered tentative (Marchese 1989).

A few comparative studies have been done recently on Volta–Congo: Gbéto (2002), Stewart (1995), and Casali (1995).

Gur languages have been the focus of a small number of studies of language contact, e.g., Kleinewillinghöfer (2000), and Dombrowsky–Hahn (2004). Recent comparative work on Gur includes Sambiéni (2005), and Miehe, et al. (2007).

The unity of Adamawa–Ubangi languages has recently been called into question (Olson 2006; Dimmendaal 2008). It is even questionable whether Ubangian languages should be considered Niger–Congo at all (Dimmendaal 2008). As with other Niger–Congo groups, recent comparative work has either been on small subgroups (e.g. Moñino 1995; Elders 2006), or on language contact patterns (e.g. Zima 1991; Cloarec–Heiss 2002). Adamawa–Ubangi languages lie in the ‘Macro–Sudan belt’, a region in which many linguistic patterns are shared across language families (cf. Güldemann 2008), so comparative studies of these languages are likely to reveal complex patterns of language interaction and change. Ubangian languages are also of particular interest because of their role in the development of Sango, one of the few non-European language-based creoles in the world.

3.4. NILO–SAHARAN LANGUAGES

Recent surveys of Nilo–Saharan (Dimmendaal 2006; Cyffer 2007) present competing classifications by Greenberg (1963), Bender (1996) and Ehret (2001). These classifications differ in their subgrouping, but not in their overall constituent members which are basically the same apart from the inclusion of Kadugli languages (which Greenberg classed as Kordofanian). There seems to be no consensus among scholars between the Nilo–Saharan classifications. Since Bender’s classification is the one that has been argued most explicitly (cf. Bender 2000; Blench 2006b), I present it in Table 4.

Few linguists have specialist expertise in three or more of the major Nilo–Saharan groupings, so it is difficult for anyone to give an overall critique of the various proposals. However, the Nilo–Saharan classification of each of Bender’s outlier language groups have been strongly criticized by specialists: Songhay (Nicolaï 2003), Kuliak (Heine 1976), and Saharan (e.g., Petrácek 1989). Other specialists seem to take no personal stand on the classification of the outlier groups, but merely present Greenberg’s classification as something that is generally accepted (e.g., Cyffer 2000).

Deep skepticism exists for the Nilo–Saharan affinity of two of the satellite–core groups, For (Jakobi 1993), and Central Sudanic (Boyeldieu 2006). Even one of Bender’s core groups, Gumuz, has been questioned (Mikkola 1999), and recently, even by Bender himself (Bender 2005a). It
seems quite possible that the Nilo-Saharan affiliation of some of the other groups is unchallenged, not because the classifications have met with positive approval, but because a careful re-evaluation of the evidence has been neglected.

The core of Nilo-Saharan appears to be the East Sudanic group, which includes the following families: Nubian, Nera, Jebel, Nyimang, Temein, Tama, Daju, Surmic, and Nilotic. The internal relationships between these language families must be better understood before any external relationships can be evaluated. The least well accepted members of Nilo-Saharan are Songhay and Kuliak. Numerous competing classifications of Songhay and Kuliak have been proposed (cf. Childs 2003: 45–46). The growing consensus is that Nilo-Saharan represents a minimum of four distinct lineages. For the present, Nilo-Saharan is best considered as a referential grouping, and not a language phylum or family.

Recent comparative work on Nilo-Saharan has included reconstructions of relatively low-level groups: Boyeldieu (2000), Bender (2005b), and Heusing (2004); and studies highlighting the role of language contact: Dimmendaal (2007), Storch (2005, 2007), and Moges (2006). Some scholars recently brought up the old suggestion that Nilo-Saharan and Niger-Congo languages also form a family (Bender 2000; Blench 2007b). Like Boyd (1996), I question the utility of such a super-phylum classification in shedding light on processes of language change given the shaky ground on which Nilo-Saharan itself stands.

3.5. KHOISAN LANGUAGES

Greenberg’s Khoisan is not accepted as a genetic grouping by most Khoisanists working in historical linguistics now (cf. Güldemann and Vossen
2000; Sands and Güldemann 2008), and in the past (e.g., Westphal 1962; Traill 1986). Rather, it is considered to be a pragmatic grouping of four or five unrelated language families, as shown in Table 5. However, specialists in long-distance relationships (e.g., Ruhlen 1994; Ehret 2003; Starostin 2003) continue to propose reconstructions involving the three major Southern African Khoisan language groups: Khoe, Ju, and Tuu.

Reconstructions of Khoisan (or, Khoesan) language families are heavily dependent on the documentation of the individual dialects of these endangered languages. For example, the !Xóõ language was undocumented and unclassified, presumed because of its name to be a dialect of !Xóõ, a Tuu language, but recent documentation provides evidence to link it to the Ju languages (Honken 2008). Dialect variation is key in helping untangle complex patterns of language contact, such as have been recently detailed (e.g., Güldemann 2004, 2006; Honken 2006). Recent documentation has led to reconstructions that would not have been possible even a decade ago (e.g., Elderkin 2004; Sands 2008).

### 3.6. AFRICAN LANGUAGE ISOLATES (SPOKEN LANGUAGES)

The number of African language isolates has been underestimated in the past. Greenberg (1963) mentions no isolates as contributing to African linguistic diversity. Table 6 lists 10 languages that currently have no widely agreed upon classification. The languages listed in Table 6 have at least a small amount of published data.

There are likely to be other African language isolates for which no published data exist. In some cases, all we have are reports that people once spoke a distinct language, but we know next to nothing about that language. Some languages, like Bung (Connell 1998), may be extinct before they are ever described in enough detail to classify them. Other languages, such as Ongota (Savà and Tosco 2003) and Koegu (Hieda 1991) have now been convincingly classified as Cushitic and Surmic, respectively, on the basis of recent fieldwork. These languages, both spoken by small populations of hunter-gatherers in southwestern Ethiopia, were difficult to classify because they have been heavily influenced by neighboring languages.

---

**Table 5. Major Khoisan language groups (after Güldemann and Stoneking 2008)**

<table>
<thead>
<tr>
<th>Language Group</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hadza (language isolate)</td>
<td></td>
</tr>
<tr>
<td>Sandawe (single language)</td>
<td>(possibly related to Khoe-Kwadi)</td>
</tr>
<tr>
<td>Khoe (Central Khoisan) + Kwadi (single language)</td>
<td></td>
</tr>
<tr>
<td>Ju (Northern Khoisan) + ‡Hoan (single language)</td>
<td></td>
</tr>
<tr>
<td>Tuu (Southern Khoisan)</td>
<td></td>
</tr>
</tbody>
</table>

© 2009 The Author


Journal Compilation © 2009 Blackwell Publishing Ltd
3.7 OTHER AFRICAN LANGUAGES

Africa has much to offer of interest for historical linguists that might be missed if the focus is primarily on language classification, such as: urban youth vernaculars, non-European lexifier creoles, and mixed languages.

The languages of certain groups tend to be left out of discussions of African languages, even when languages are unique to Africa or have been used in Africa for centuries. Malagasy, an Austronesian language spoken in Madagascar, is an African language even though the other languages in the phylum are not. Malagasy has been used in Madagascar for some 2000 years and has had at least two periods of contact with mainland Bantu languages (Berchem 1989) that have resulted in shared lexical items in at least some Malagasy dialects (Kent 1968; Dahl 1988). Afrikaans developed from Dutch, an Indo-European language, into a distinct language and is now learned as a first or second language by millions of Africans. Afrikaans has influenced the vocabulary of numerous southern African languages.

South Indian languages have been used in East and South Africa for generations by tens of thousands of people. These include Tamil, a Dravidian language spoken in South Africa, and Indo-Iranian (Indo-European) languages such as Gujarati, Kachchi, Urdu, and Bhojpuri-Hindi (e.g., Mesthrie 2002; Gordon 2005).

Other Indo-European languages used in Africa include former colonial languages such as French, English, Portuguese, and German. There are also Indo-European based creoles, e.g. Angolar Creole Portuguese (Sãotomense); Kabuverdianu, Nigerian Pidgin English. These languages are learned as first languages by significant numbers of Africans and should be considered as part of Africa’s linguistic diversity.

Table 6. Some possible African language isolates

<table>
<thead>
<tr>
<th>Name</th>
<th>Citation</th>
<th>Possibly related to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hadza</td>
<td>(Sands 1998)</td>
<td></td>
</tr>
<tr>
<td>Shabo</td>
<td>(Bender 2000)</td>
<td></td>
</tr>
<tr>
<td>Jalaa</td>
<td>(Blench 2006b)</td>
<td></td>
</tr>
<tr>
<td>Laal</td>
<td>(Blench 2006b)</td>
<td></td>
</tr>
<tr>
<td>Bangime</td>
<td>(Blench 2005)</td>
<td>Chadic</td>
</tr>
<tr>
<td>Mpre † (?)</td>
<td>(Cardinall 1931)</td>
<td></td>
</tr>
<tr>
<td>Dompo</td>
<td>(Blench 2007c)</td>
<td></td>
</tr>
<tr>
<td>Kujarge</td>
<td>(Blench 2006b)</td>
<td>Khoe-Kwadi</td>
</tr>
<tr>
<td>Sandawe</td>
<td>(Blench 2006b)</td>
<td></td>
</tr>
<tr>
<td>Meroitic †</td>
<td>(Rowan 2006; Rilly 2007)</td>
<td>Nilo-Saharan or Afroasiatic</td>
</tr>
</tbody>
</table>
4. Discussion

The picture of African linguistic diversity presented by Greenberg (1963) has had a major impact on African linguistics for more than half a century. Support for Greenberg’s classification of Afroasiatic (e.g., Newman 1995) and Niger-Congo families has been greater than that for his Nilo-Saharan and Khoisan families. Since many more linguists study Afroasiatic and Niger-Congo languages than Nilo-Saharan or Khoisan languages, positive critiques may have outweighed the negative ones. But, also, some of the long-range comparative work in the generation immediately following Greenberg seemingly supported his results. This, however, was because these works were characterized by some of the same questionable methodological practices. For instance, earlier comparisons tended to give lists of similar-looking lexical items, without demonstrating regular sound correspondences. Or, where sound correspondences seemed to be regular, semantic correspondences were iffy. Grammatical items were often compared singly, or in pairs, rather than as parts of entire paradigms. In order to make roots across languages seem more comparable, morpheme boundaries were often posited with very slight evidence.

Recent comparisons have tended to be based on more accurate and thorough linguistic descriptions. They have generally focused on closely related groups of languages or on particular linguistic subsystems. Often, the sociolinguistic and ecological conditions of the languages being compared are better known than they were in Greenberg’s time. There has also been an increased awareness of the role of language contact and convergence and chance in creating similar structures across languages. Grammatical morphemes that may once have been seen as strong proof of relatedness may now be seen as likely borrowings.

5. Conclusion

While Greenberg’s (1963) classification of African languages has been successful as a referential classification, it has not been entirely successful as a linguistic genetic classification. Khoisan is no longer accepted as a genetic grouping by specialists, and there are serious doubts about several of the major subgroups of Nilo-Saharan and Niger-Congo. A number of sign languages and spoken language isolates must be regarded as independent linguistic genetic lineages. The linguistic genetic diversity of Africa is at least five times greater than commonly presumed.

6. Acknowledgements

I would like to thank the people at Document Delivery at Northern Arizona University for their tireless assistance. Thanks also to Sheena Shah of Georgetown University, Joe Eska, and the two anonymous reviewers for their comments.
7. Short Biography

Dr. Bonny Sands is a linguist who combines fieldwork and laboratory phonetics research, with typological and historical-comparative linguistic studies. Her special focus is on Africa's click languages. She has authored or co-authored papers on African language classification that have appeared in: The New Encyclopedia of Africa, Language History and Linguistic Description, and Language, Identity, and Conceptualization among the Khoisan. Her book Eastern and Southern African Khoisan: Evaluating Claims of Distant Linguistic Relationships argues that the Khoisan languages do not meet the criteria for a language family. Sands has taught at the University of Michigan, and currently holds an adjunct position in the Department of English at Northern Arizona University in Flagstaff, Arizona. She has conducted fieldwork in Eastern and Southern Africa on a dozen different languages, on various projects funded by the National Science Foundation. She is currently funded by the US National Science Foundation to investigate the phonetics of !Xung and Hoan languages in Namibia and Botswana. She holds a BA in Linguistics from Yale University and a PhD in Linguistics from the University of California, Los Angeles.

Notes

* Correspondence: Dr. Bonny Sands, PO Box 6032, Flagstaff, Arizona 86011, USA. E-mail: sands@lowell.edu.

Re-use of this article is permitted in accordance with the Creative Commons Deed, Attribution 2·5, which does not permit commercial exploitation.

1 Africa is taken to mean the African continent and surrounding islands. This usage is commonly found in atlases and in introductory texts, e.g., Martin and O’Meara (1995).

Works Cited


—— 2003a. Afrasian overview. Selected comparative-historical Afrasian linguistic studies, in


Africa’s Linguistic Diversity 575

   and Linguistics 13.41–58.
Casali, R. F. 1995. On the reduction of vowel systems in Volta-Congo. African Languages and
   Cultures 8.2.109–21.
   ———. 2004. The Atlantic and Mande groups of Niger-Congo: A study in contrasts, a study in
Connell, B. 1998. Moribund languages of the Nigeria-Cameroon borderland. Endangered
   Brenzinger, 163–78. Berlin, Germany: Mouton de Gruyter.
Cyffer, N. 2000. Linguistic properties of the Saharan languages. Areal and genetic factors in
   München, Germany: LINCOM Europa.
Dafalla, R. Y. 2006. A phonological comparison in the Katcha-Kadugli language group of the
   Nuba mountains. Insights into Nilo-Saharan language, history and culture: Proceedings of the
   9th Nilo-Saharan linguistic colloquium, Institute of African and Asian Studies, University of
   Köln, Germany: Rüdiger Köppe.
   in the prehistory of arid Africa: proceedings of the SFB conference, Königswinter 1–3
   Germany: Rüdiger Köppe.
   Afrikanistische Beiträge zum 65. Geburtstag von Franz Rottland. ed. by R. Vossen, A.
   Köln, Germany: Rüdiger Köppe.
   ———. 2003. Towards reconstructing Proto-South Khoisan (PSAK). Mother Tongue (Newsletter
   of the association for the study of language in prehistory) 8.65–81.
   Forschungen 64). Wiesbaden, Germany: Harrassowitz.
Fodor, I. 1966. The problems in the classification of the African languages (Methodological
   and theoretical conclusions concerning the classification system of Joseph H. Greenberg).
   Budapest, Hungary: Center for Afro-Asian Research of the Hungarian Academy of
   Sciences.


Africa’s Linguistic Diversity 579


