A Survey Report for the Bantu Languages

by

Derek Nurse

Cartographer: Irene Tucker

SIL International

2001

Contents

On classifying Bantu languages
General
Referential classifications
Historical classifications
Appendix
The Bantu languages of Africa
The Bantu languages of Africa—northwestern area
References

On classifying Bantu languages

General
According to the most recent estimate (Grimes 2000) the world has 6,809 languages, of which 2,058, approximately 30%, are spoken in Africa (an additional 44 are described as “extinct”). Africa is home to the world’s largest language phylum, Niger-Congo, with 1,489 languages (the next largest being Austronesian, with a mere 1,262). If we accept the figure of 750 million as Africa’s population size today, then some 400 million Africans speak Niger-Congo languages, of whom about 240 million have a Bantu language as their first language (the figure includes Grassfields). That is, nearly a third of all Africans speak a Bantu language as their native language.

Readers should treat any claim about the number of languages as having general but not absolute validity. The main difficulty is, what is language and what is dialect? The conventional answer says that a language tends to be the standard variety, be written, have more speakers, have some form of official status, have prestige, and not be intelligible to speakers of other “languages”. By contrast, dialects are not the standard, not written, have fewer speakers, have no official status, have little or no prestige, and are mutually intelligible. In sub-Saharan Africa, as elsewhere, these distinctions are only partly true and in any case any distinction between language and dialect is part linguistic, part political, part prestige-related. There is a cline of linguistic difference between the similar and the dissimilar, and since no one knows exactly where to cut a cline, it is
hard to state with accuracy the total of Bantu “languages”. This uncertainty can be seen by considering recent estimates of “Bantu language” numbers: Guthrie (1967–1971) names some 440 Bantu “varieties”, Grimes (2000) has 501, Bastin et al. (1999) have 542, Maho (forthcoming) has some 660, and Mann, Dalby et al. (1987) have ca. 680.

Bantu-speaking communities live in Africa south of a line from Nigeria in the west, across the Central African Republic (CAR), the Democratic Republic of Congo (DRC: formerly Zaire), Uganda, and Kenya, to southern Somalia in the east. Most language communities between that line and the southern tip of Africa are Bantu. The exceptions are pockets: in the south, some small and fast dwindling Khoisan communities; in Tanzania one, maybe two, Khoisan outliers; in the northeast of the area, larger communities speaking Cushitic (part of Afro-Asiatic); and along and inside the northern border many communities speaking Nilo-Saharan languages and Adamawa-Ubangian (Niger-Congo but non-Bantu) languages. Communities speaking Bantu languages are indigenous to twenty-seven African countries: Angola, Botswana, Burundi, Cameroon, CAR, Comoros, Congo, DRC, Equatorial Guinea, Gabon, Kenya, Lesotho, Madagascar, Malawi, Mayotte, Mozambique, Namibia, Nigeria, Rwanda, Somalia, South Africa, Sudan, Swaziland, Tanzania, Uganda, Zambia, and Zimbabwe. Non-Bantu Niger-Congo languages are spoken north and mainly west of Bantu. Starting in the north of the DRC, they stretch west across the CAR, Cameroon, Nigeria, and right across all west Africa as far as Senegal.

While agreed in general that Bantu is one of several families that make up Niger-Congo, linguists are not agreed on what exactly defines Bantu within Niger-Congo. The main difficulty lies in eastern Nigeria. Current interpretations suggest that some five millennia ago (Vansina 1995) the ancestral Bantu community left its homeland astride the Nigeria-Cameroon borderland and diffused south and east across the rainforest, reaching roughly its full contemporary distribution by the early centuries of our era. The descendants of the communities which moved away, those who now live in central, eastern, and southern Africa, lost contact with their old Niger-Congo neighbors. Those who stayed behind, or didn’t move very far, the so-called northwestern Bantu languages, have interacted linguistically with these Niger-Congo neighbors for five thousand years, or longer. The result is that the northwestern languages have become less like their Bantu siblings and more like their Niger-Congo cousins, to the point where it is hard to draw an unambiguous line between them. This is yet to be resolved (see Williamson and Blench 2000), for an overview and a very recent statement).

Classifications of the internal relationships of language families are of various types: areal, typological, genetic/historical, and referential. This article deals only with the last two. The period starting with Guthrie’s Comparative Bantu (1967–1971) has seen well over thirty attempts to classify some or all of the Bantu languages, excluding recent work concerned with the position of Bantu within Niger-Congo. Most are either explicitly nonhistorical/referential or cover only part of the area. Only a few have tried to provide a picture of the historical development of the whole Bantu area and a majority of languages.

**Referential classifications**

Referential classifications aim primarily at providing a practical referential taxonomy, something obviously necessary for such a large family. While the last two decades or so have seen other
internal referential classifications of Bantu (e.g. Mann, Dalby et al. 1987, Grimes 2000), undoubtedly the most influential has been that of Guthrie. His thinking went through various adjustments from 1948 to 1967–1971 (of which an updated version can be found in Maho (forthcoming)). The final version divided the (Narrow) Bantu area into 15 “zones” of roughly equal size, labelled A, B, C, D, E, F, G, H, K, L, M, N, P, R, S, to which Belgian scholars later added a J, by combining bits of D and E. The zones in turn consist of up to nine groups, numbered 10, 20, 30, etc., and each group also has up to nine members, mostly very similar to each other. Thus zone A is the first of his zones (it happens to be in the northwest), A20 is a group of very similar languages within A, and A24 refers to the language Duala (small letters after the number refer to dialects, e.g. A11a, or A15b). His method was partly geographical, partly linguistic, in that he started with a language or a small group of languages and looked around for languages having “similar” features. These were fitted into groups, and the groups into zones. If the zones got too large or unwieldy, another zone was formed. The accompanying map shows these zones. The linguistic features have little genetic validity, which was as he intended—it was a practical, not a historical statement. And since they are based on nongenetic features, the zones themselves have little historical validity, although the smaller groups often have more validity, both typological and historical. Many scholars feel it is useful to keep Guthrie’s referential classification (or a modified version of it), while searching for a more accurate representation. To modify it once is to modify it many times, leading to confusion.

**Historical classifications**

Different scholars have different purposes and methods when they talk of historical or genetic classifications. A good historical classification of a language family ought to show two things. First, it will represent the genealogy of its members. That is, assuming that most or all of the members of a language family derive from a common ancestor, a historical classification will represent this, and the various splittings and branchings that occurred since that ancestor. The commonest way of representing is via what is called a family tree diagram. But such a tree diagram is a rather static model, so the second thing that a historical classification should try to reflect is the various changes that have occurred and the various contacts that the family members have undergone—it is hard to do this via a family tree.

Seeking to correct Guthrie’s historical perspective, or lack of it, a series of scholars have proposed alternatives over the last thirty years (for an overview, see Nurse 1994). Many have relied on the use of lexicostatistics (e.g. Heine 1973, Bastin et al. 1999). A language may split into two (or more) dialects, later two languages. As they move through time, they share progressively less vocabulary. The more vocabulary they share, the more recent the split; the less they share, the more distant the split. Lexicostatistics is based on counting these shared words and is thus a measure of lexical similarity and retention. Other scholars have relied on the use of shared lexical innovations (new words) and on shared loanwords (borrowed words) (see Ehret 1998).

Most of these classifications share two results. They, the majority, see (1) the northwestern languages (those of Zones A, B, C, and parts of D and H) as being clearly distinct from the rest; and (2) thereafter, a split in the rest between western (Zones H, K, R, sometimes L and parts of M) and eastern languages. Readers should read these classifications with care, as western and eastern often go under different names and have different membership. These linguistic splits
lead to the historical interpretation that the original Bantu community first split into the northwestern languages versus the remainder, and later the remainder split into a western and an eastern group.

The fact that these classifications have similar results is linked to the fact that they share similar (lexically-based) methods. Any approach based solely on the use of words has to be suspect because words are the linguistic feature most easily borrowed, so linguists can never be completely sure whether a word is shared by two languages because it is inherited by both or borrowed in one or both. Linguists agree that lexicostatistics is but a preliminary step to a lasting classification, which should be based on groups of languages, at different levels within a family, sharing common (non-lexical) innovations. Work of this kind has scarcely begun (e.g. Nurse and Philippson (forthcoming) has a set of suggestions) and it seems likely that this is the direction of the next decade or two, work in which a role could well be played by SIL, with its large group of field-workers and field data. A final linguistically-based historical classification of the Bantu languages has to be regarded as work in progress.

Appendix

(see maps on following pages)
The Bantu languages of Africa
Groupings based on Guthrie, 1948*

* This map follows Guthrie's 1948 classification, except for the red listed of group J.

Note: Dashed lines show overlap of language areas.

Disclaimer: The representation of international and language boundaries is not necessarily authoritative. No political statement is intended of any kind in regard to language or territorial boundaries. This map is provided for general reference only. Maps are subject to change.
References


