

The Bantu area: (towards clearing up) a mess

Jouni Maho

*Department of Oriental and African Languages
Göteborg University*

<jouni.maho@african.gu.se>

Introduction

During the twentieth century, several classifications of the Bantu languages have been published. Following Johnston (1919/1922), there have been various more or less authoritative statements on how to classify these languages by Doke (1945; see Cole 1961), Guthrie (1948; 1971), Cope (1971), Heine (1973), Heine, Hoff & Vossen (1977), Bastin, Coupez & Halleux (1983), SIL (see Grimes 1996ab), Bastin, Coupez & Mann (1999), Dalby (2000), and others, some of who have focused on sections of the Bantu area, for instance, Doke (1954), Nurse & Philippson (1980), Nurse & Hinnebusch (1993), and others.

Most classifications have employed some kind of coding system reflecting the classification as such. For instance, Johnston (1919/1922) not only classified the languages, but he also referred to the various entities in his classification with a coding system consisting of letters and digits signifying linguistic groupings as well as individual languages. The most well-known classification and coding system is undoubtedly that of Guthrie (1971). However, many people have noted that, while his coding system has become widely used as a referential tool when talking of individual languages, as a linguistic-genetic statement it was not so successful. For the latter reason, post-Guthrie classifications have sought to revise or replace it. Some of these have done so while retaining Guthrie's original coding system, though in a modified version.

The unfortunate consequence of this is that there now exists a veritable mess when referring to individual languages as well as larger linguistic groupings, since (1) different codes are used for identical languages and groupings, and (2) the same codes are used for different languages and groupings. What I intend to do here is to suggest that the only decent thing we

can do is to stick to Guthrie's original classification, or rather, his coding system. We can update it by adding new languages, but should deter from re-coding old languages.¹ This mean, of course, that we can and should use Guthrie's coding system only as a referential tool, and as such it is sufficient.²

Guthrie, two runner-ups and a mess

As already mentioned, the most wide-spread authority on the classification of Bantu languages is still Guthrie (1971). The coding system that Guthrie employed has become immensely popular among Bantuists. According to this, each Bantu language that Guthrie classified was given a three (occasionally four) character code, consisting of an upper-case letter indicating a regional zone (see figure 1) followed by two digits indicating language group and language (which is occasionally followed by a lower-case letter indicating dialect). For example, the code G42 refers to Zone G, Group 40, language no 2, which is Swahili. (G42d refers to the Unguja variety/dialect.)

Since Guthrie's classification is basically an areal-typological exercise, it exhibits anything from much to no linguistic-genetic validity, depending on what part of the classification one looks at. Thus it can best be characterized as a referential classification. As such, however, it has been and still is extensively used in referring to Bantu language groups, languages and/or dialects, and thereby it is also difficult to replace with any new system of codes, even though people have tried. The two most noteworthy such efforts are those of the Tervuren scholars (most recently by Bastin, Coupez & Mann 1999) and the Summer Institute of Linguistics or SIL (deducable from Grimes 1996 and Grimes & Grimes 1996). These three classifications — Guthrie's, Tervuren's and SIL's — all have wide followings, and they are the only ones used as authoritative classifications in the contemporary literature on Bantu languages.

Tervuren's and SIL's aims are to construct a Bantu classification with a linguistic-genetic validity. Thus there are differences between these three classifications. In some areas, they are in agreement. Guthrie's zone S is one such area. Not only is the entire group usually held together, but they also agree on the internal subclassification of this particular zone (largely, anyway). Other areas, like Congo-Kinshasa, represent anything from minimal to maximal disagreement. For example, the languages comprised in Guthrie's D20 and H40

¹ The entire enterprise was suggested to me by Derek Nurse & Gérard Philippson. I am greatly indebted to their invaluable (electronic) help and comments in the course of this work, for which see further below.

² Naturally, I do not suggest that genetic classifications should be abandoned. Far from it. See further below.

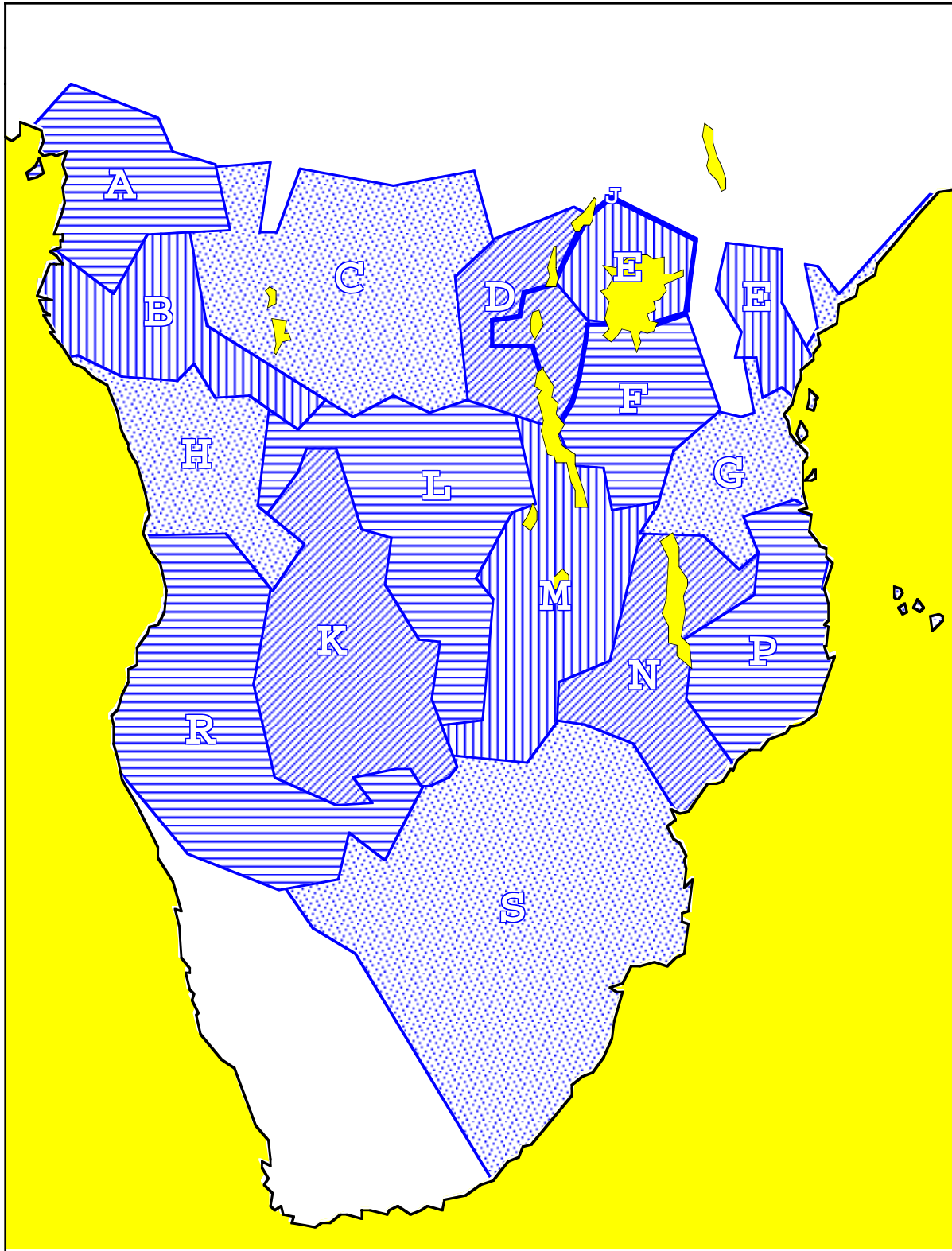


Figure 1. Malcolm Guthrie's zonal classification of the Bantu languages. Adapted.
(The J Zone does not derive from Guthrie; see text for explanations.)

groups do not constitute coherent groupings in neither SIL's nor Tervuren's classification.

One thing ties all three classifications together. They all use the same basic

<i>Guthrie</i>	<i>Tervuren</i>	<i>SIL</i>	<i>Guthrie</i>	<i>Tervuren</i>	<i>SIL</i>
A10	A10	A10	G20	G20	G20
A20	A20	A20	G30	G30	G30
A30	A30	A30	G40	G40	G40
A40	A40	A40,Non-Bantu	G50	?	G50
A50	A50	A50	G60	G60	G60
A60	A60	A70,Non-Bantu	H10	H10	H10
A70	A70	A70	H20	H20	H20
A80	A80	A80	H30	H30,K50	H30,K10
A90	A90	A90	H40	H40,K50	H40,K60
B10	B10	B10	K10	K10	K20
B20	B20	B20	K20	S30	S30
B30	B30	B30	K30	K30	K40,M60
B40	B40	B40	K40	?	K50
B50	B50	B50	L10	K50	K10
B60	B60	B60	L20	D10?,L20	L20
B70	B70	B70	L30	L30	L30
B80	B80	B80	L40	L40	L40
C10	C10	C20,C40	L50	K20	K30
C20	C20	C30	L60	L60	L50
C30	C30	C40	M10	M10	F10
C40	C40	C50	M20	?	M20
C50	C50	C60	M30	M30	M30
C60	C60	C70	M40	M40	M40
C70	C70	C70,C80	M50	M50	M50
C80	C80	C90	M60	M60	M60
D10	C60,D10	D10	N10	N10	N10
D20	D20,D30,L20	D20,D30,L20	N20	N20	N20
D30	D30	D30	N30	N30	N30
D40	D40,J40	D40,J40	N40	N40	N40
D50	D50,J50	D50,J50	P10	?	P10
D60	J60	J50,J60	P20	P20	P20
E10	J10	J10,J30	P30	P30	P30
E20	J20	J20	R10	R10	R10
E30	J30	J30	R20	R20	R20
E40	E40	E10,J30	R30	R30	R30
E50	E50,E70	E20	R40	?	R40
E60	E60	E30	S10	S10	S10
E70	E70	E40	S20	S20	S20
F10	F10	F10	S30	S30	S30
F20	F20	F20	S40	S40	N10,S40
F30	F30	F30	S50	S50	S50
G10	G10	G10	S60	?	S60

Figure 2. How Guthrie's groups correspond to Tervuren's and SIL's groups.

coding system, namely, that originally devised by Guthrie (1948; 1971). Tervuren's and SIL's classifications have both sought to revise Guthrie's original classification and coding system, and so many languages have been re-shuffled from their old places in Guthrie's classification into other groups and zones. Thus they have also been re-coded. Some groups have disappeared, while others have been created. In fact, an entire new zone has been created by the Tervuren scholars, the so-called J zone, which has been adopted also by SIL. (This corresponds, roughly, to Guthrie's D40-D60 and E10-E40.)

This means, among other things, that identical and/or similar languages and language groupings are labelled/coded differently in the different classifications, even where there would intuitively seem to be no need for it. For example, in SIL's classification, for some reason, Guthrie's C10 has become C20, Guthrie's C20 has become C30, Guthrie's C30 has become C40, and so on. Moreover, due to the re-shuffling of languages, some of Guthrie's original groups have been "emptied" completely of languages in the classifications of both Tervuren and SIL so that, for instance, D60, E50, E60 and M10, are non-existent in SIL's classification while L10 and L50 do not seem to exist in Tervuren's classification. Occasionally, Tervuren and SIL have used such "emptied" groups for other languages that they were originally used for by Guthrie. For example, Guthrie's K20 group consists of a single language, Lozi, which both Tervuren and SIL have re-classified into S30. Thus K20 is, or was, empty. Tervuren has moved Guthrie's L50 group into K20, leaving the L50 label empty(!), while SIL has moved Guthrie's K10 group into K20, and are now using K10 for some of Guthrie's L10 languages. The resulting mess when comparing the use of these codes is tremendous, and very annoying (see figure 2).

Individual languages are now variously labelled and referred to with different codes. Add to this the fact that individual scholars tend to make idiosyncratic revisions as they see fit. For example, Se(n)geju is coded E56 by Guthrie, E73 by Tervuren and E20 by SIL.³ Ngangela is coded variously as K12b (Guthrie), K19 (Tervuren) and K20 (SIL). Lunda appears as L52 (Guthrie), K22 (Tervuren), and K30 (SIL). Ntomba-Bikoro appears as C35a (Guthrie), C35 (Tervuren), C70 (SIL), and C65 in Botne (1999). Gciriku is not mentioned by Guthrie, but appears geographically withing Guthrie's K30 group. It is coded K39 by Tervuren, K70 by SIL, and K38 by Möhlig (1983). The list could be made much longer, but these suffice as examples.

The classifications: differences of opinion or just a big mess?

There would be little to argue against any revision of Guthrie's classification and coding if that revision would have the likeliness of remaining unrevised in itself. That is, if the revision is linguistic-genetically valid, it would ultimately be accepted as the standard classification, and we would all deter from using Guthrie. So the question is: is either of the two runner-up classifications linguistic-genetically valid?

The runner-up classifications differ with respect to methodology as well as goals. Tervuren's classification is strictly based on lexicostatistics. This is a

³ Note that the SIL classification only classifies "down" to groups. Thus the SIL-codes always end in zeroes.

method with many and oft-noted inherent flaws, but it continues to be used nonetheless. Despite their own trust in lexicostatistics, Tervuren's classification is not likely to remain as it is. SIL's classification is difficult to say anything about. The methodology seems somewhat obscure, since there are no explicit statements about methods, criteria, or in fact anything. The only way to arrive at SIL's classification is to deduce it from the language lists accompanying the *Ethnologue*. It looks much like a compilation, or rather conglomerate, of many people's work. Thus one cannot conclude anything specific about the potential longevity of SIL's classification.

The ultimate success of any classification is much dependent on the quantity and quality of available material on individual languages. Unfortunately, however, the Bantu area is still much un-researched. Many languages lack adequate descriptions, and for those languages for which there exist descriptive material the quantity is, with a handful of exceptions, usually very small, meaning that our view of the individual languages are one-sided, fragmentary and biased in favour of the subjective perspectives of individual scholars and schools. Thus any proposed classification depends to some degree on guess-work, or at least more so than (ideally) would be the case if more descriptive material was available. This is, perhaps somewhat simplistically, evident from the fact that all proposed classifications disagree on how to subclassify the Bantu languages in areas where little material exist, while they generally tend to agree in areas where more material does exist.

A solution?

What can we do, then? Guthrie's classification definitely needs to be updated. It has a strong areal-typological bias. His criteria for classifying any given language in a particular group is not as strict as one might have wished. Many of his groupings seem to be impressionistically based, rather than by the application of a strict methodology. Guthrie produced genetically inconsistent groupings, for instance, the sub-groupings of E60 (Chaga) and G40 (Swahili) are veritable messes.

However, the main reason for wanting to update Guthrie's classification is due to all his omissions. There are numerous Bantu languages lacking in Guthrie (1971). If we still wish to use Guthrie's coding system, then what do we do with these unclassified languages?

The solution that I propound here is one suggested to me by Derek Nurse and Gérard Philippson (see Maho, forthcoming). To put it simply: stick to Guthrie (1971) since it is already so widely used and well-established and add new languages/dialects to that without re-coding old languages. Due to the many

inconsistencies in Guthrie's classification, it can only be used as a referential classification, not a linguistic-genetic one. Any linguistic-genetically valid classification will have to await more and better data. (And preferably any such classification ought to employ a distinct coding system in order not to confuse the picture even more that it has already been.)

When deciding on new codes to previously unclassified languages, the *Bantu MapMaker* program developed by Lowe & Schadeberg (1996) employs two important principles, or rules-of-thumb.

- 1) If the addition is a language close to an already-classified language, then add a third digit to the language code of that latter language.

For example, since unclassified Ekoti, situated on the Mozambican coast, is embedded by the Emakhuwa language (P31), it has been given the code P311 (= P31+1). The closely situated Sangaji has been given the code P312 (= P31+2).

- 2) If the addition is a language with uncertain affiliation or lies close to several languages, then add a third digit to the group code.

For example, since the Swahili dialects Mgao, Makwe and Mwani lie far south from (most) other Swahili dialects, they have been coded G401, G402 and G403 (= G40+1, etc.), respectively.

To the above two principles, Derek Nurse (pc) suggested the following:

- 3) If the addition is a dialect of an already-classified language, then add an upper-case letter to the language code of the latter.

For example, Guthrie (1971) lists six Nyika (E72) dialects (or possibly languages), namely, Giryama, Kauma, Conyi, Duruma, and Rabai, coded E72a, E72b, E72c, E72d and E72e, respectively. This is not a complete list, and for consistency, three more names ought to be added, namely, Jibana, Kambe and Ribe. These have been added to Guthrie's E72-series as E72F, E72G and E72H, respectively.

A fourth principle, I have made up myself.

- 4) If the addition is a so-called restructured language (pidgin, creole, mixed), then add a upper-case letter to the group code indicating typological affinity.

Thus Kituba and Munukutuba, situated in the western Bantu area, have been

coded H10A and H10B respectively.

The main advantage with the above four principles, and which the coding principles employed in the *Bantu MapMaker* makes a point of, is that any new code can immediately be spotted as an addition. When Tervuren refers to Kweso as A93, there is no indication given that this language and its coding/classification does not derive from Guthrie, but is in fact part of the revision. Following the principles formulated above, the same language is coded A93C, that is, ending with an upper-case C, whereby anyone even dimly acquainted with Guthrie's original classification can immediately spot it as something new.

Any update of Guthrie's classification is bound to remain a half-done job. There are so many regions in the Bantu area where data is still lacking for proper statements about linguistic demography, that any linguistic-genetically validating revision, like those of Tervuren and SIL, or any referential update, as that proposed here, will most certainly be revised or updated again in the future. Presently, some 200+ languages/dialects have been added to Guthrie's 1971-classification (see Maho, forthcoming), and there is a good change more will be added as time goes on (Maho 2000). The advantage of a referential update is that there are less likely to be any re-shufflings and re-codings of languages. New languages can be added without making a mess of old codings.

REFERENCES

Bastin, Yvonne, André Coupez & Michael Mann. 1999. *Continuity and divergence in the Bantu languages: perspectives from a lexicostatistical study*. Annales sciences humaines, vol 162. Tervuren: Musée Royale du Afrique Centrale.

Bastin, Yvonne, André Coupez & B. de Halleux. 1983. Classification lexicostatistique des langues bantoues (214 relevés). *Bulletin des séances de l'Académie Royale des Sciences d'Outre-Mer*, 37(2), pp 173-199.

Botne, Robert D. 1999. Future and distal -ka-'s: Proto-Bantu or nascent form(s)? In: *Bantu historical linguistics*, CSLI lecture notes, no 99, pp 473-515. Edited by Jean-Marie Hombert & Larry M. Hyman. Stanford: Center for the Study of Language and Information (CSLI).

Cole, Desmond Thorne. 1961. Doke's classification of Bantu languages. In: *Contributions to the history of Bantu linguistics*, pp 80-96, by Clement M. Doke & Desmond T. Cole. Johannesburg: Witwatersrand University Press.

Cope, Anthony T. 1971. A consolidated classification of the Bantu languages.

African studies, 30, pp 213-236.

Dalby, David. 2000. "Inner"-Bantu. Observatoire Linguistique/Linguasphere Observatory. (Web-document available at <www.linguasphere.org>.)

Doke, Desmond Thorne. 1945 [repr. 1967]. *Bantu: modern grammatical, phonetical and lexicographical studies since 1860*. London: Dawsons of Pall Mall.

—————. 1954. *The southern Bantu languages*. Handbooks in African languages. Oxford, New York & Cape Town: Oxford University Press for the International African Institute.

Ehret, Christopher. 1999. Subclassifying Bantu: the evidence of stem morpheme innovation. In: *Bantu historical linguistics*, CSLI lecture notes, no 99, pp 43-147. Edited by Jean-Marie Hombert & Larry M. Hyman. Stanford: Center for the Study of Language and Information (CSLI).

Grimes, Barbara F. 1996. (Ed.) *Ethnologue: the languages of the world*. Thirteenth edition. Dallas: Summer Institute of Linguistics (SIL).

Grimes, Joseph E. & Barbara F. Grimes. 1996. (Eds.) *Ethnologue: language family index, to the thirteenth edition of the Ethnologue*. Dallas: Summer Institute of Linguistics (SIL).

Guthrie, Malcolm. 1948. *The classification of the Bantu languages*. London: Oxford University Press for the International African Institute.

Guthrie, Malcolm. 1971. *Comparative Bantu*, vol 2. London: Gregg Press.

Heine, Bernd. 1973. Zur genetischen Gliederung der Bantu-Sprachen. *Afrika und Übersee*, 56, pp 164-185.

Heine, Bernd, Hans Hoff & Rainer Vossen. 1977. Neuere Ergebnisse zur Territorial-Geschichte des Bantu. In: *Zur Sprachgeschichte und Ethnohistorie in Afrika*, pp 57-72. Edited by Wilhelm J. G. Möhlig, Franz Rottland & Bernd Heine. Berlin: Dietrich Reimer Verlag.

Johnston, [Sir] Harry Hamilton. 1919/1922. *A comparative study of the Bantu and Semi-Bantu languages*, 2 vols. Oxford: Clarendon Press.

Lowe, John B. & Thilo C. Schadeberg. 1996. *Bantu MapMaker*, v 3.1 (computer program). Berkeley & Leiden.

Maho, Jouni. 2000. Indices to Bantu languages. Manuscript. Göteborg: Department of Oriental and African Languages, Göteborg University.

———. Forthcoming. Appendix: an update of Guthrie's (referential) classification of the Bantu languages. In: *The Bantu languages*. Edited by Derek Nurse & Gérard Philippson. London: Curzon Press.

Möhlig, Wilhelm J. G. 1983. Kavangosprachen. In *Lexikon der Afrikanistik*, pp 124-125. Edited by Herrman Jungraithmayr & Wilhelm J. G. Möhlig. Berlin: Dietrich Reimer Verlag.

Nurse, Derek & Thomas J. Hinnebusch. 1993. *Swahili and Sabaki: a linguistic history*. With an addendum by Gérard Philippson. University of California publications in linguistics, vol 121. Los Angeles & Berkeley: University of California Press.

Nurse, Derek & Gérard Philippson. 1980. Bantu languages of East Africa: a lexicostatistical survey. In: *Languages in Tanzania*, pp 26-67. Edited by Edgar C. Polomé & C. P. Hill. London: Oxford University Press for the International African Institute.