Bantu and Bantoid

Lutz Marten

1. Introduction

The Bantu family is the largest African language family in terms of geographic and demographic size. There are an estimated 450-550 Bantu languages, spoken by about 250 million speakers in 27 countries, from the Nigerian-Cameroonian borderland in the northwest of the Bantu-speaking area to Kenya in the northeast and to South Africa in the south (see Map 1).

[Map 1 here]

Map 1: The distribution of Bantu languages

The family includes languages with millions of speakers, such as Zulu, Shona, Nyanja, Kinyarwanda, and Swahili, but also smaller languages, many of which are poorly or not at all documented. However, overall Bantu languages are the most well described language group in Africa, and the first to attract comparative work, resulting in a reconstructed proto-language, Proto-Bantu. In terms of classification, Bantu is deeply embedded within the Niger-Congo phylum and belongs to the Benue-Congo family within it (Figure 1).
The place of Bantu within Benue-Congo, and indeed the exact membership and internal classification of Bantu, are subject to some debate, and despite the early establishment of a reconstruction for the family, a comprehensive sub-grouping of Bantu is still being addressed. Furthermore, in the linguistically complex Nigerian-Cameroonian borderland, about 150-200 Bantoid languages are spoken which are closely related to Bantu, and whose exact position with respect to “narrow” Bantu remains to be established. This, of course, has implications for what we mean by Bantu and for the properties of Proto-Bantu. Despite the antiquity of comparative-historical Bantu studies within African linguistics, there remains considerable scope for scholarly discovery.
2. Early studies of Bantu languages, the reconstruction of Proto-Bantu and 20th century developments

The earliest surviving records of Bantu languages consist of a few words noted down by early travelers in Africa, such as the Arab historian and traveler Al-Masʿudi, who visited the east African coast in the 10th century and recorded the Swahili word wafalme with the meaning ‘son of the Great Lord’ (Freeman-Grenville 1962:16). More comprehensive descriptions of Bantu languages were the result of the onset of missionary activity in Angola and the Congo in the second half of the 17th century: in 1659, the Italian missionary Giacinto Brusciotto published a grammar of Kikongo, the first known grammar of an African language, and in the following decades missionaries, travelers and explorers continued to publish sketches of different Bantu languages. In the latter half of the 18th century, Abbé Proyart (1776) noted the similarity between different Bantu languages spoken on the coast around the Congo estuary, as did the German missionary Oldendorp (2000), who recorded the languages of west African slaves in the Danish West Indies in the 1760s. In southern Africa, Lichtenstein (1808) compared Nguni and Tswana words and phrases he had learned when traveling in the Cape, with Tsonga examples collected by White (1800) in Maputo, and noted their relatedness. Meanwhile in London, the orientalist William Marsden, based on these accounts and on data supplied by his Mozambican servant, proposed that central, southern and eastern Bantu languages are genetically related (Tuckey 1818:384-90). With increasing colonial expansion in the 19th century, interest in African languages and their comparative study increased, too – a correlation which, as Said (1978) notes, is not entirely accidental: the classification and grouping of the languages of Asia and Africa by European missionaries, travelers and scholars was often exploited by colonial
policies of dividing and ruling Asian and African territory and societies. Grammatical
descriptions of Bantu languages increased, and by the middle of the century, the
material was rich enough for comparative studies such as Bleek (1862/69), who
introduced the term “Bantu” for the family, based on the common root -ntu ‘person’ and
the animate plural prefix ba-, so meaning ‘people’, as well as the numbering convention
for Bantu noun classes. Meinhof (1899, 1910, 1932) expanded this work and at the end
of the 19th century, based on a small number of geographically balanced test languages
(Duala, Herero, Kongo, Nyakyusa, Sango, Northern Sotho, Swahili, and Zulu),
proposed the reconstructed Ur-Bantu or Proto-Bantu, which, in broad outline, has
remained more or less accepted until today, and provided a main foundation and
reference point for subsequent work in Bantu linguistics. The understanding of the
position of Bantu languages within Niger-Congo resulted from comparative work on
west African languages: Westermann’s (1927) results implied that Bantu was part of the
Niger-Congo phylum, and Greenberg (1949, 1963) developed a detailed classification
in which Bantu became deeply embedded within Niger-Congo and one co-ordinate
branch within the Benue-Congo group. While this does not give cause for surprise
today, it was remarkable at the time, given the presumed importance of the Bantu
languages in Africa (cf. Greenberg 1949:315). Work on Bantu itself was subsequently
significantly influenced by Guthrie’s (1948, 1967-71) comparative Bantu studies,
further discussed below, and the work of the “Lolemi” research group at Tervuren,
inaugurated by Achille Meeussen, which resulted in numerous studies on lexical and
grammatical aspects of Bantu, including Meeussen’s (1967) Bantu grammatical
reconstruction. Over the last forty years or so, Bantu languages have also become
central to questions of theoretical linguistics, for example in the areas of phonology and
especially tone (Hyman and Kisseberth 1998), syntax (Mchombo 1993),
grammaticalization, tense–aspect, and information structure (see Nurse and Philippson 2003a).

3. Internal classifications

Bantu as a language family was firmly established by the beginning of the 20th century. However, the internal sub-classification of Bantu languages remained outstanding, and despite several studies, this remains true today. An early, comprehensive study addressing the question is Johnston (1919/22). However, it was not until almost fifty years later that the most influential study of the internal classification of Bantu appeared, the geographical-referential classification by Malcolm Guthrie (1967-71). Based on comparative material of over 200 languages, Guthrie divided Bantu languages into 84 language groups (Table 1), which are assigned to fifteen geographical zones. Each zone is assigned an alphabetical letter from A in the northwest to S in the south (Map 2), and within each zone, languages or language groups are assigned a letter and a number, so that, for example, the Swahili group is G40, and within this Swahili itself G42 (see Maho 2003).

<table>
<thead>
<tr>
<th>A10 Lundu-Balong</th>
<th>D60 Ruanda-Rundi</th>
<th>L30 Luba</th>
</tr>
</thead>
<tbody>
<tr>
<td>A20 Duala</td>
<td></td>
<td>L40 Kaonde</td>
</tr>
<tr>
<td>A30 Bube-Benga</td>
<td>E10 Nyoro-Ganda</td>
<td>L50 Lunda</td>
</tr>
<tr>
<td>A40 Basa</td>
<td>E20 Haya-Jita</td>
<td>L60 Nkoya</td>
</tr>
<tr>
<td>A50 Bafia</td>
<td>E30 Masaba-Luhya</td>
<td></td>
</tr>
<tr>
<td>A60 Sanaga</td>
<td>E40 Ragoli-Kuria</td>
<td>M10 Fipa-Mambwe</td>
</tr>
<tr>
<td>A70 Yaunde-Fang</td>
<td>E50 Kikuyu-Kamba</td>
<td>M20 Nyika-Safwa</td>
</tr>
<tr>
<td>A80 Maka-Njem</td>
<td>E60 Chaga</td>
<td>M30 Konde</td>
</tr>
</tbody>
</table>
Table 1: Main groups of Bantu languages (based on Guthrie 1967-71; Maho 2003)
In addition, Guthrie proposed a set of reconstructions, to add to the forms proposed by Meinhof and his followers, as well as a long list of what he called “starred forms”. These were not meant to be historical hypotheses, as reconstructed forms are, but merely summary statements of the variation found in his data ("Common Bantu" as opposed to Proto-Bantu). Yet, the distinction is subtle, and in practice the forms often serve as reference points for historical work, and they also formed one of the bases for the large on-line database of Bantu Lexical Reconstructions BLR (BLR3). Guthrie’s zonal classification has sometimes been criticized for not corresponding well to genetically more plausible groupings. For example, a genetically justified group of interlacustrine (i.e. between the east African Great Lakes) languages (including, for example, Rwanda, Rundi, Haya and Luganda) straddles Guthrie’s zones D and E, and sometimes a new zone, zone J, is used to refer to this group. But in a sense, the criticism is beside the point. The only measurable quality of a referential classification such as Guthrie’s is whether it is used widely (which it is), while a genetic classification may result in principle in entirely different groupings (cf. Maho 2003). The problems with a truly genetic grouping of Bantu have been noted frequently (e.g. Meinhof 1932:176, 1933; Nurse 1994/95; Schadeberg 2003; Vansina 1995:179): the large amount of data needed – from all 500 or so Bantu languages – for a comprehensive comparative study, the effect of language contact and borrowing, which may obscure historical divergence, and the question whether a traditional genetic tree based model, assuming binary splits of diversion, is adequate for a situation which might better be characterized as a convergence area, or “a huge single pool of dialects” (Vansina 1995:180).

The situation has prompted mainly two kinds of response: the detailed study of subgroups, so as to develop an overall classification from bottom up, and large scale
comparison of only a limited set of data, often with innovative methods (in addition, non-genetic classifications have been proposed as well, discussed in Section 6). Among the former are Nurse and Hinnebusch’s (1993) establishment of a Northeast-Coast Bantu group, interlacustrine Bantu languages (Schoenbrun 1990), or southern Bantu (Doke 1954; Herbert and Bailey 2002; Janson 1991/92). From this and similar work, a more comprehensive classification of Bantu overall can gradually be built up, although this requires work on many more subgroups (cf. also Möhl 1981). The second possible response is to work with a large number of languages, but with only a restricted data set. This has been done in particular with lexical data sets employing lexicostatistic, and later phylogenetic methods, in a number of studies (e.g. Heine et al. 1977; Bastin et al. 1999; Holden 2002; Holden and Gray 2006; Grollemund et al. 2015; cf. Marten 2006). Heine et al.’s (1977) results show a distinction between three main Bantu subgroups: a north-western group, a western-central group, and, as the largest group, a south-eastern group. Bastin et al.’s (1999) extensive lexicostatistical study is based on a 92-word list of basic vocabulary from 542 Bantu languages. The results of the study are, however, ambiguous. In their overall analysis, Bastin et al. (1999) conclude that, first, hierarchical, binary tree relations are not well suited for expressing their results, second, Bantu languages share histories of divergence as well as contact and convergence, and these histories are sometimes impossible to tell apart, and third, by varying the computational parameters used to analyze the data, a series of trees results which expresses the ambiguous relations between the Bantu languages of the sample better than a single tree. The computational parameters are connectivity (the degree of coherence within a given group of languages) and exclusivity (the degree of similarity with members from outside the group), as well as different thresholds of
similarity between different nodes in the tree. The resulting family of trees provides a “different perspective on linguistic relationships which are all part of the total picture” (1999:109), and so there is no single correct tree to model the relationship between different Bantu languages. Vansina (1995), commenting on pre-publication results of the study, takes the argument further, and proposes that the relationship between Bantu languages cannot be represented by the tree-model, but fits rather well a wave-model of language change which allows overlapping changes and multiple inheritance. Another approach is taken by studies using phylogenetic methods, adopted from evolutionary biology, such as Holden (2002), Rexová et al. (2006), and Grollemund et al. (2015), which allow computation over the set of possible trees compatible with the data to established the most consistent, or “consensus” tree. Grollemund et al.’s (2015) study is based on 100-word lists from 424 languages and does result in such as consensus tree with hierarchical sub-groupings, in which five main branches are identified: northwest, central-west, westwest, southwest and east Bantu (Figure 2).

A few studies have also been conducted comparing non-lexical data (Bastin 1983; Ehret 1999; Nurse and Philipsson 2003b). Nurse and Philipsson (2003b), for example,
use a variety of phonological and morphological features. They argue that lexical data are not well suited for sub-classification and the historical study of language development, since large parts of a language’s vocabulary might be borrowed, and if, as in the case of Bantu, borrowing is often from closely related languages, borrowed forms are hard or even impossible to detect. Based on their non-lexical features, Nurse and Philippson (2003b) propose a number of language groups, which can be combined into four large groups: west, forest, west-central and north-eastern savannah. No south-eastern group is postulated, in contrast to many previous studies, and Nurse and Philippson (2003b) note that no such group is supported by the features adopted.

In sum, most proposed sub-classifications of Bantu languages identify a north-western group, including more or less the Bantu languages of Cameroon, Gabon, Congo and the north-western part of the DRC, and a large eastern or south-eastern group of Bantu languages from Kenya to South Africa. In addition, two further groups are often distinguished: a south-western group to the south of the north-western group, including languages of Angola and Namibia, and a central group of languages between the eastern and the western groups. Even though no fully worked-out sub-classification of Bantu has been proposed so far – despite considerable efforts – the results which have been achieved had a great influence on the historical study of the areas where Bantu languages are spoken today, in particular with respect to the “Bantu expansion”, or “Bantu dispersals”.

4. Bantu dispersals

A particular concern of comparative Bantu has for a long time been the relation between classification and social history, and hypotheses about the spread of Bantu languages
across central, eastern and southern Africa have had considerable influence on models of African history. Because Bantu languages are quite similar, it has often been assumed that the fact that they are spoken over a wide geographic area is a comparatively recent phenomenon. This then led to the question of how Bantu languages, or their speakers, became so widely dispersed. The three interrelated questions – in which direction did the dispersal of Bantu languages take place, what were the factors involved in the process, and when did it happen – have received different answers over the last century.

In an early hypothesis, Johnston (1919/22) proposes that Bantu languages originated from Bantoid languages of Nigeria, which he called “Semi-Bantu” languages, and then were dispersed eastwards and southwards. As it turned out, this was very much along the lines of what was later shown to be the best explanation. However, in the 1960s and 1970s, the question was subject to a lively discussion between Guthrie (1962), Greenberg (1972) and the historian Oliver (1966). Greenberg argued that Bantu languages were originally spoken in the Cameroon-Nigerian borderland, the area with the highest complexity and greatest diversity of Bantu languages, and then subsequently came to be spoken throughout the rest of today’s Bantu-speaking area. Guthrie, in contrast, argued that the origin of Bantu languages, the “Bantu nucleus”, lies in the Congo basin. His main argument against Greenberg’s proposal was that in order for Bantu languages to have spread from Cameroon, they (or their speakers) would have had to have crossed the tropical rainforest of central Africa, which he considered impossible. Yet, Guthrie had to accept, reluctantly, that languages of Cameroon and Nigeria were related to Bantu, and so that speakers (or languages) had to be able to cross the rainforest, in his case from the Congo to Cameroon. But that, of course, meant that Guthrie had to concede that movement from Cameroon to the Congo was possible.
after all, since, as Greenberg pointed out, “if there is a road from New York to Chicago, there has to be a road from Chicago to New York” (1972:196). Greenberg’s proposal was further substantiated by Heine et al.’s (1977) lexicostatistical, as well as by Grollemund et al.’s (2015) phylogenetic results, and is widely accepted today. The most likely historical interpretation of the linguistic facts (cf. Ehret 1998; Vansina 1990, 1995) is that Bantu languages spread through a succession of small migrations, involving backwards and forwards movements, rather than through one large “expansion”, so that the present distribution of Bantu languages is the outcome of “many complex historical dynamics involving successive dispersals of individual languages over a time span of millennia and involving reversals as well as successes” (Vansina 1995:195). The migrations were accompanied by language shift of speakers of other languages to Bantu languages, so that the process involved both the physical movement of speakers and the dispersal of languages without speaker movements. The reasons why Bantu languages appeared attractive may be related to technical, economic or social differences. Early Bantu communities may have been larger than the original, more widely-dispersed communities they met with, and their languages may have been associated with more perceived advantages because of this. Given the frequent postulation of a main split between north-western, western and south-eastern Bantu languages (see Section 3), a possible model of the historical dispersal is an initial spread from the original Bantu-speaking area eastwards and southwards through the rain forest, with groups of Bantu speakers emerging to the south of the rain forest, reflected in the western Bantu group, and to the east of the forest, eventually leading to the further spread of Bantu languages through the drier savannah and shrub lands of eastern and southern Africa, although all these processes may have been accompanied by and
overlapped with parallel and subsequent movements. The most recent hypothesis, developed by Grollemund et al. (2015), is that early Bantu speakers took advantage of the emergence of a rainforest “corridor” resulting from periodic climate change around 500 BC, whereby for some time the rainforest receded, allowing for much easier passage through it. The time of the Bantu dispersal can further be estimated by correlation of Proto-Bantu vocabulary with archaeological evidence (Vansina 1995): Proto-Bantu vocabulary for pottery and farming indicates that Bantu languages had not dispersed before the earliest archaeological attestation of these activities in western Cameroon (around 3000 BC). On the other hand, the absence of iron-smelting vocabulary, and its distribution across different Bantu languages show that by the time the technology is first attested (around the Great Lakes ca. 800 BC) the Bantu dispersals were well under way.

5. Bantu and Bantoid

For the sub-classification of Bantu languages, as well as for the understanding of Bantu languages within the wider Niger-Congo phylum, the relation between Bantu and its closest neighbors within the Benue-Congo group of Niger-Congo is of central importance. Yet, many questions remain at present unanswered about this relation. Geographically, the question concerns the northwest of the Bantu area: Bantu languages border on Khoisan (and Germanic) languages in the south, and on Nilo-Saharan, Cushitic and Ubangian languages in the north and northeast, and the difference between these languages and Bantu are on the whole quite clear. However, in the northwest of the Bantu area Bantu languages are spoken in the neighborhood of closely related Bantoid languages, which are felt by many observers, including such early studies as
Koelle (1854) and Johnston (1919/22) to be similar to Bantu in terms of vocabulary or structure, but yet not similar enough to be obviously part of “narrow Bantu”. There are some 200 Bantoid (excluding narrow Bantu) languages spoken in western Cameroon and eastern Nigeria (Blench 2006). Some disagreement exists as to the linguistic grouping of the languages, and their classification is still in progress. Lewis (2009) includes the following thirteen groups (with the number of languages in the group provided in brackets): Beboid (14), Dakoid (5), Ekoid (8), Fam (1), Jarawan (15), Mambiloid (13), Mamfe (3), Mbam (13), Mbe (1), Ndemli (1), Tikar (1), Tivoid (17), and Wide Grassfields (67), in addition to six unclassified languages. However, the precise membership and internal structure of these groups is often still not definitely established (e.g. Williamson and Blench 2000; Piron 1997). Grassfields languages are sometimes called “Grassfields Bantu” (e.g. Watters 2003), and the term “Ekoid Bantu” has been used as well (e.g. Crabb 1965), showing the perceived closeness to narrow Bantu, while the Mbam group is also called “Mbam Bantu” and includes languages of the Bantu A40 and A60 groups. Not only geographically, but also in terms of structural characteristics, Bantoid languages lie between proper Benue-Congo and proper Bantu. Often several lexical items can be related to Proto-Bantu forms, and many Bantoid languages have (traces of) noun classes similar to Bantu noun classes. However, it is often difficult to distinguish between genuine shared Bantu properties, and those which Bantu and Bantoid languages share because they are wider Niger-Congo features. The problems related to establishing a classification of Bantoid languages, and their relation to north-western Bantu languages, are similar to those facing attempts to sub-classify Bantu languages (Section 3): conflicting isoglosses and potentially high levels of borrowing which obscures linguistic inheritance. In addition, for a number of languages,
often small and undescribed, there are not enough data to judge their relationship with other languages in a complex linguistic area with a high degree of linguistic variation. Finally, presumably due to widespread processes of phonological attrition, many complex lexical roots in Bantu correspond to short, monosyllabic roots in Bantoid, which makes comparison more difficult. Responses to these problems are, again, similar to Bantu sub-classification: The construction of small groups so as to develop larger relationships bottom-up, the use of both lexical and non-lexical data, and the use of lexicostatistical methods, and of alternative models of language relation (see section 9.6.6). Watters (1989) provides an overview and synthesis of earlier classifications and distinguishes a northern and a southern Bantoid group, the letter including narrow Bantu. A comprehensive lexicostatistical study of Bantoid languages is provided in Piron (1997). As with other lexicostatistical studies (e.g. Bastin et al. 1999, discussed above), the results are not easy to interpret, and Piron’s study results in a number of different trees, depending on the particular analytical parameters employed. One tree, based on branch average distance, is given in Figure 3.
Figure 3: A lexicostatistical classification of Bantoid (after Piron 1997:625)

The classification identifies three different Bantu groups, but separates southeast Bantu from other south Bantoid languages (including central and northwest Bantu), and central Bantu from northwest and non-Bantu south Bantoid. Grollemund et al. (fcmg.) is a more recent phylogenetic study of Bantoid based on 100-word lists of 197 languages. It distinguishes eleven main groups (Bantu, Jarawan, Mamb-Bubi, Grassfields, Beboid, Mambiloid, Tivoid, Jukunoid, Dakoid, Bendi and Ekoid) and two isolates (Tikar and Kenyang). As in previous work, a group of Bantu A languages (A31, A40, A60 and possibly A50), the Mbm-Bubi languages, are outside of narrow Bantu and are grouped closer to the Jarawan group. More controversially, Jukunoid is classified as part of Bantoid, rather than as a higher grouping in Benue-Congo. A simplified representation of the classification is given in Figure 4.

Main groupings of the phylogenetic classification of Bantoid of Grollemund et al. (fcmg.)
Even though more data are likely to change the picture of the classification of Bantoid languages, it is clear that the situation is complex and has considerable potential impact on the classification and reconstruction of narrow Bantu: one implication from current work on Bantoid is that what is currently termed “Bantu” might not be a coherent genetic unit, and that current Bantu reconstructions do not take sufficient account of north-western Bantu languages. On the other hand, as with Bantu, the complexity of the relation between Bantoid languages has also led to exploration of non-genetic approaches to classification.

6. Non-genetic classifications

The challenges encountered in establishing robust subgroups of Bantu and Bantoid languages has led to the application of different and more complex methods for the study of linguistic relationships, including lexicostatistical and phylogenetic methods, as well as the use of non-lexical data, such as morphological innovations. These approaches share the assumption that ultimately Bantu languages can be classified in some tree-like arrangement, reflecting shared linguistic innovations. However, an alternative response has been to regard the interrelation of Bantu/Bantoid languages not as an empirical challenge, but a conceptual one, and that a potential genetic classification of Bantu needs to be supplemented, or even be replaced, by an alternative classification. The most widely used of these alternatives are typological classifications. In fact, typological criteria were often mixed with genetic criteria in the history of Bantu classification without sharply distinguishing between the two. Guthrie’s (1948:11f.) criteria for the identification of Bantu languages, as a famous example, are mainly typological, including presence of a noun class system encoded by prefixes, CVC...
lexical roots and a symmetric vowel system, but they also include the genetic criterion of reconstructible vocabulary (Gerhardt 1981). A range of comparative studies have been undertaken by members of the “Lolemi” school of Bantu linguistics in Tervuren. Although not specifically typological in theoretical outlook, cross-Bantu studies of, for example, relative clauses (Nsuka-Nkutsi 1982) or locative classes (Grégoire 1975) provide comprehensive case studies of similarities and variation across Bantu. The striking overall typological similarity of Bantu languages, and the distribution of different typological features across different subgroups of languages, is reminiscent of historical diffusion rather than historical diversification, reflecting a long equilibrium rather than punctuation in the terms of Dixon (1997). Alternatively, in areal terms, the Bantu area can be thought of as a spread zone (Nichols 1992), in which linguistic features diffuse, resulting in overall similarity, and a criss-crossing of similarities and differences across many different varieties, without clearly distinguishing groups of varieties. This idea is developed, for example, by Möhlig (1981), who proposes several historical Bantu convergence zones (“stratificational nuclei”), and more recently by Güldemann (2011) who proposes a Bantu spread zone (including mainly central and south-eastern Bantu languages) and a wider Macro-Sudan belt, which also includes north-western Bantu languages at its periphery. An areal approach provides a new perspective on Bantoid languages as well: located between the Bantu spread zone on the one hand, and the typologically quite distinct Kwa languages, Bantoid languages show similarities with both, constituting a typological “buffer zone” between the two (Good 2013). A central aspect of all these explanations is language contact. A different contact-based approach to Bantu classification is to compare languages with different roles in intra-Bantu contact situations, where it appears that convergence is particularly
driven by languages with high numbers of second language speakers (cf. Marten et al. 2007). The adoption of new models of explanation, as well as continuing work within more established paradigms, shows that, despite the relative antiquity of the family, Bantu and Bantoid classification remains a challenging and rewarding field of study.

References


Foster and C. Renfrew (eds.), Phylogenetic Methods and the Prehistory of Languages. Cambridge: McDonald Institute for Archaeological Research, 43-55.


White, W. M. (1800). Journal of a voyage performed in the Lion Extra Indiaman, from Madras to Columbo and Da Lagoa Bay, on the eastern coast of Africa (where the ship was condemned), in the year 1798: with some account of the manners and customs of the inhabitants of Da Lagoa Bay and a vocabulary of the language. London: Stockdale.