Introduction
Koen Bostoen, Mark van de Velde

To cite this version:
Koen Bostoen, Mark van de Velde. Introduction. The Bantu Languages, 2019. halshs-02504383
CHAPTER ONE

Introduction

Koen Bostoen & Mark Van de Velde

This second edition of The Bantu Languages consists of two parts. Part 1 contains general chapters that provide an overview of the state of the art in the study of the sound systems and morphosyntactic structures of the Bantu languages and of their classification, reconstruction and different contact situations. Part 2 contains short grammatical analyses of individual Bantu languages for which no book-size grammar is available. Together, these chapters provide a thorough introduction to the grammatical structures of the Bantu languages and to the historical evolutions that have shaped them. The focus on language structure and history means that this volume does not aim at giving an exhaustive overview of contemporary Bantu studies. Such an overview would have to include work on documentation, orthography creation, lexicography, youth languages and so on. This edition is very similar in structure and approach to the highly influential first edition (Nurse & Philippson 2003a), but its contents are almost entirely new. Some chapters from Part 1 have been thoroughly revised. Bonny Sands has renewed the chapter on the Sounds of the Bantu Languages (Chapter 3) in consultation with Ian Maddieson, the author of this chapter in the first edition. Larry Hyman has updated his chapter on Segmental Phonology (Chapter 4). David Odden and Michael Marlo have revised the chapter on Tone (Chapter 5), originally written by Charles Kisseberth and David Odden. The chapter on Derivation was renamed Word Formation (Chapter 6) after an update by Koen Bostoen in close collaboration with its primary author, Thilo C. Schadeberg. The chapter on Aspect and Tense has been revised by its first author Derek Nurse and complemented with a section on mood/ modality by Maud Devos to become a new chapter on Aspect, Tense and Mood (Chapter 7). The six other chapters of Part 1 are entirely new. For Part 2, we chose to invite chapters on a new set of languages, because sketch grammars are both extremely useful and relatively hard to get published. For some of the languages included in Part 2, such as Chimpoto N14 (Chapter 23) and Pagibete C401 (Chapter 15), hardly any other published information is available, while others, such as Ngazidja G44a (Chapter 20), have a rich literature but no reference publication that provides a coherent overview of the basic grammatical features of the language. The sketch grammars of the first edition are available on the companion website of this second edition. They have not been revised for this second edition.

The remainder of this chapter is a brief portrait of the Bantu family, starting with its delimitation, number of speakers and geographical distribution (Section 1), and some of its main typological characteristics (Section 2). It finishes with a concise history of its scholarly study, including some early attempts at external classification (Section 3).
1. The Bantu languages: delimitation, speakers and geographical distribution
The most recent inventory of the Bantu languages (Hammarström, Chapter 2, this volume) lists 555 distinct Bantu languages. It uses the same language versus dialect divisions as the 18th edition of the Ethnologue (Lewis et al. 2015). Hammarström observes that a stricter adherence to the criterion of mutual intelligibility would decrease the number of Bantu languages by about 15%, to more or less 472. In order to keep track of so many languages, Bantuists make use of a referential classification, devised by Malcolm Guthrie, in which every language is identified by means of a so-called Guthrie code, which gives an indication of the language’s geographical location (see Chapter 11, Section 3.1.1). The emergence and discovery of new languages and the extinction of others are minor factors to account for, but the difficulty in drawing a discrete line between a language and a dialect is the main reason why numbers diverge considerably in earlier inventories of Bantu languages or language varieties, e.g., 440 in Guthrie (1971), approximately 680 in Mann and Dalby (1987), 542 in Bastin et al. (1999) and 660 in Maho (2003). In the latest online version of his New Updated Guthrie List, Maho (2009) inventories 950 different varieties with a unique extended Guthrie code. Many of these are recognised as different varieties of a single language, i.e., those ending in a lower-case or upper-case letter. The number of distinct varieties without such a final letter in Maho (2009) is 631 (Harald Hammarström, pers. comm.). The 555 languages listed in Chapter 2 are represented by the letter corresponding to their Guthrie zone on the map in Figure 1.1.
Figure 1.1. The Bantu languages represented by the letter corresponding to their Guthrie zone

The boundary between the Bantu languages and their closest relatives of the Bantoid family spoken in Cameroon and Nigeria is to a certain extent established by convention, rather than by a set of shared innovations that are attested in all and only the Bantu languages. According to convention, languages are considered to belong to the Bantu family if they have a Guthrie code. An area where this convention is most likely to be at odds with reality is the little-studied group of Jarawan languages of Nigeria and northern Cameroon, which lack a Guthrie code but which have been argued to be Bantu, possibly Bantu A60 (Gerhardt 1982, Blench 2015). Several studies in genealogical classification based on basic vocabulary (cf. Bastin & Piron 1999, Grollemund 2012: 349, Bostoen et al. 2015, Grollemund et al. 2015) do indeed recognise Jarawan Bantu languages (as well as certain other Bantoid languages) as being most closely related to the languages of the Mbam-Bubi group, which do have Guthrie codes (in groups A30, A40, A50 and A60), without a discrete cut-off point between Bantu and Bantoid or so-called “Narrow Bantu” and “Wide Bantu.”

As for the number of speakers of Bantu languages in Africa, Nurse and Philippson (2003b: 1) estimate that about 240 million Africans speak one or more Bantu languages, multilingualism being the rule rather than the exception in Africa. In 2003, this meant that one African out of three to four spoke a Bantu language (given a total African population of about 875 million at that time). This is more than a half of all Niger-Congo speakers, which Nurse and Philippson (2003b: 1) estimate at about 400 million. Patin et al. (2017) estimate that there are about 310 million Bantu speakers. This would correspond to about one African in four, the number of Africans in 2018 being around 1.2 billion. Of the 556 Bantu languages in the 20th edition of the Ethnologue (Simons and Fennig 2017), 529 have a population estimate, whose sum is 276,513,509 speakers (Harald Hammarström pers. comm.). This number could be extrapolated to about 290 million for the 556 Bantu languages. However, one could also argue for a higher contemporary number, taking into account that according to the 20th edition of the Ethnologue the total African population including Madagascar is 929,932,101. This estimate is based on 2039 languages out of the 2178 present in Africa that have a population estimate (from well before 2018). The proportion of this estimate with respect to the actual number of 1.2 billion Africans today is 0.775. Applied inversely to the sum of Bantu speakers estimated in the Ethnologue 20, their number would amount to about 350 million today (Harald Hammarström pers. comm.), which is even more than what Patin et al. (2017) propose without explaining on what their estimate is based.

The Bantu languages are mainly spoken between Cameroon’s South-West region (4°8’N and 9°14’E) in the North-West, southern Somalia’s Barawe (Brava) area (1°6’N and 44°1’E) in the North-East and Cape Agulhas (34°48’S and 20°E), the continent’s southernmost tip in the Western Cape province of South Africa. Their distribution area is contiguous – some very rare languages surrounded by non-Bantu languages notwithstanding – and spans twenty three countries on the African mainland. In alphabetical order, these are Angola, Botswana, Burundi, Cameroon, Central African Republic, Congo-Brazzaville, Congo-Kinshasa, Equatorial Guinea, Gabon, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, Somalia, South Africa, Southern Sudan, Swaziland, Tanzania, Uganda, Zambia and Zimbabwe. In some of these, such as Burundi, Malawi and Rwanda, Bantu languages are the only indigenous African languages. Long-standing Bantu speech communities are also found on the islands of Bioko (part of Equatorial Guinea), Mayotte (an overseas department of France) and the Comoros (see also Nurse and Philippson 2003b, Hammarström et al. 2017).
Nurse and Hinnebusch (1993: 14) report a variety of the Bantu language Swahili spoken on the small island of Nosse-Be, off the northwest coast of Madagascar, with another pocket further down the west coast of this island. In other African countries, especially those in the northern and southern borderlands, Bantu languages coexist with languages that belong to other families, such as Central Sudanic, Nilotic, Cushitic, and Omotic, which are part of the wider Nilo-Saharan or Afro-Asiatic phyla, or that are considered isolates today, such as Hadza and Sandawe (formerly considered to be Khoisan). Yet in others, such as the Central African Republic, Southern Sudan and Somalia, Bantu languages are sporadic, not to say nearly absent. All in all, Bantu is the predominant language group in Central, Eastern and Southern Africa. Other Niger-Congo languages – apart from Adamawa-Ubangi and Kordofanian – predominate in Western Sub-Saharan Africa, but have a distribution area that is about a third to a half of the Bantu area. Thanks to the vastness of the Bantu area, Niger-Congo is by far Africa’s most widespread language phylum.

The massive spread of the Bantu languages is striking, especially in consideration of the group’s estimated age of no more than 4,000 to 5,000 years (Vansina 1995: 52, Blench 2006: 126). This time depth is quite shallow compared to the 10,000 to 12,000 years that have been proposed for the Niger-Congo phylum (Blench 2006: 126). The geographic distribution of Niger-Congo minus Bantu is much smaller than the spread zone of its tardive Bantu offshoot. Bantu languages would have gradually split off from their closest South-Bantoid relatives in the borderland of South-Eastern Nigeria and Western Cameroon, an area of high linguistic diversity within the Bantoid subgroup of Benue-Congo, one of the major Niger-Congo branches (cf. Blench 2015). Ever since Greenberg (1972), there is great unanimity to consider this area as the Bantu homeland. It is from this ancestral homeland that the concurrent dispersal of Bantu languages and Bantu-speaking people across Central, Eastern, and Southern Africa started. This phenomenon is commonly referred to as the Bantu Expansion (Oliver 1966, Bouquiaux 1980, Vansina 1995, Ehret 2001, Bostoen 2018). During their initial migration between roughly 5,000 and 1,500 years ago, Bantu speech communities not only introduced new languages in the areas where they immigrated, but also new lifestyles, in which technological innovations such as pottery making and the use of large stone tools originally played an important role, as did farming and metallurgy subsequently. Wherever early Bantu speakers settled down, they left an archaeologically visible culture (Phillipson 2005, de Maret 2013, Bostoen et al. 2015). New insights from the field of evolutionary genetics show that the Bantu Expansion was not just a spread of languages and technology through cultural contact, as was once thought (Lwanga-Lunyiigo 1976, Gramly 1978, Schepartz 1988, Vansina 1995, Robertson & Bradley 2000), but involved the actual migration of people (Pakendorf et al. 2011, Li et al. 2014, Patin et al. 2017). Moreover, Bantu-speaking newcomers strongly interacted with resident hunter-gatherers, as can still be observed in the gene pool (Destro-Bisol et al. 2004, Wood et al. 2005, Quintana-Murci et al. 2008, Verdu et al. 2013, Patin et al. 2014), and/or the languages of certain present-day Bantu speech communities (Herbert 2002, Bostoen & Sands 2012, Gunnink et al. 2015, Pakendorf et al. 2017). The driving forces behind what is the principal linguistic, cultural and demographic process in Late Holocene Africa are still a matter of debate, but it is increasingly recognised that a climate-induced crisis of the Central African rainforest around 2,500 years ago boosted the scale and pace of the Bantu Expansion (Schwartz 1992, Brncic et al. 2009, Ngomanda et al. 2009, Maley et al. 2012, Neumann et al. 2012, Oslisly et al. 2013, Bostoen et al. 2015, Grollemund et al. 2015, Hubau et al. 2015).

3. Typological characterisation
As pointed out in the conclusion of Chapter 3 and in the introduction of chapters 4 and 9, one of the most attractive features of the Bantu family is that it allows for the comparative study of linguistic variation in a huge set of closely related languages. There is a marked typological divide between the Northwestern Bantu languages and the others. The Northwestern languages are spoken close to the Proto-Bantu homeland and in a spread zone called Macro Sudan belt (Güldemann 2008) or Sudanic zone (Clements & Rialland 2008). They typically have dense tone systems, with an equipollent opposition between low and high tones, few or no tonally underspecified morphemes and a high number of floating tones. At the other end of the typological spectrum are the few Eastern languages that have lost tone. In languages with intermediate tonal density, many morphemes are tonally underspecified and receive their surface tone through the application of rules. Chapter 5 discusses many more typological differences between the tone systems of the Bantu languages, such as the nature of the tone bearing unit or the way in which rules like tone spreading and tone shift work. The high amount of floating tones in the Northwest is due to the loss of segmental material, which is itself due to the existence of maximality constraints on the size of stems (see e.g. Hyman 2004). These same constraints also explain why the verbal derivational suffixes discussed in Chapter 6 can hardly be stacked in many Northwestern languages, whereas they typically can in the East of the Bantu domain, sometimes exuberantly.

The Bantu languages are well known for their rich noun class systems. They have on average about fifteen noun classes. On top of those, most language outsiders of the Northwest also have three locative classes. The few Bantu languages that have considerably reduced or lost their noun class system are either contact languages or spoken in the North of the Bantu domain. For some of the latter, loss of noun classes has been argued to be due to contact with languages from other families, notably Central Sudanic (see Chapters 8 and 12). Another well-known characteristics of the Bantu languages is their high number of past and future tense distinctions, discussed in Chapter 7. Probably less well known are the pragmatically conditioned freedom of constituent order on the clause level in some languages (see Chapter 9), and typologically unusual word order patterns in the noun phrase (Chapter 8).

4. A brief history of the description and classification of the Bantu languages

Bantu languages started to kindle the scholarly curiosity of Europeans as early as the late 15th century, when Portuguese sailors began their voyages along the coasts of Central, Southern and Eastern Africa. Bantu words and phrases slipped into the writings of European seafarers, merchants, soldiers and missionaries. In their endeavours to spread the holy word among the peoples of Africa, missionaries had the most direct utilitarian interest in acquiring knowledge of Bantu and other African languages. It is therefore not surprising that the oldest extant Bantu language text is a Kongo translation of the catechism by the Portuguese Jesuit Mattheus Cardoso (1584–1625) from 1624 (see also Bontinck & Ndemb Nsasi 1978). The oldest Bantu (Latin-Spanish-Kongo) dictionary was compiled through close collaboration between the Kongo priest Manuel Roboredo (†1665) and several Spanish Capuchins. It was subsequently hand copied by the Flemish Capuchin Joris Van Gheel (1617–1652), and his copy from 1652 is the only one available to us (Van Gheel 1652, Van Wing & Penders 1928, De Kind et al. 2012). The oldest Bantu grammar, on Kongo too, is also a product of clerical scholarship. It was written by the Italian Capuchin priest Giacinto Brusciotto (1601–1659) – “Brusciotto” would be a misspelling of “Brugiotti” according to Pacchiarotti (2017: 8) based on Turchetta (2007) – and published in 1659. This grammar was translated into English and made available to a wider public by Guinness (1882). All three documents pertain to one and the same variety of South Kongo, i.e., the court language spoken at Mbanza Kongo, the
capital of the Kongo kingdom, the direct ancestor of present-day Sikongo H16a (Bostoen & de Schryver 2018). Detailed historical accounts of early missionary and other research into the Bantu languages are to be found in Johnston (1919: 1–14) and Doke (1935, 1959).

The birth of Bantu linguistics as a scientific discipline is commonly attributed to the Rhenan (German) philologist Wilhelm Heinrich Bleek (1827–1875), who established Bantu as a family and gave it the name it still has today. As pointed out by Fodor (1980: 127–128), the unity of the Bantu languages was actually recognised more than a decade earlier by the American scholar H. E. Hale (1817–1896), who collected African vocabularies during his voyage around the globe as part of the United States Exploring Expedition commanded by Charles Wilkes (1838–1842). Hale divided what would become known as the Bantu languages into two distinct branches, i.e., the Congo-Makua and the Caffrarian languages, but did not propose a common label for the group. It was Bleek himself who introduced the label Bantu in a book volume published in 1958. Further elaborating on ideas that he had started to develop in his doctoral thesis (Bleek 1851), he proposed to use that name for designating those African languages which are “prefix-pronominal” (cf. Chrétien 1985: 46), in contrast to the “suffix-pronominal” or “sex-denoting” languages, including what he called the “Hottentot” and “Bushman” languages according to the parlance of that time, i.e., those known as “KhoiSan” today. Koelle (1854) had also noticed that concord prefixes equally occurred in many languages of West Africa. Bleek defined “prefix-pronominal” languages as those “in which the pronouns were originally identical with the derivative prefixes of the nouns” and situated them in the “Tropical Regions of Africa, and probably also of the Islands in the Indian Ocean and in the Pacific” along with two families in “the African or Continental Section of this Class,” i.e., “the Bántu and the Gor Family” (Bleek 1858: 35), already envisioning there what would later become Niger-Congo. Bleek situated his “Gor” family in sub-Saharan Western Africa. Bantu, on the other hand, had both a South-African and West-African division. The South-African division comprised languages still considered to be Bantu today, subdivided in three distinct branches: South-Eastern, including all known languages of Southern Africa (except KhoiSan), North-Eastern, including languages spoken along the Eastern Coast of Africa and South-Western, including those spoken along the Western Coast of Central and Southern Africa, such as Herero R31. The West-African division, however, occupied parts of the territory of the “Gor” family and included languages such as “[t]he Otshi dialect of Ashanti-land, and the Bollum and Timneh of Sierra Leone” (Bleek 1858: 36), which are classified today in other branches of the Niger-Congo phylum. Hence, Bleek’s original definition of Bantu was territorially much wider than ours today. Moreover, it was not really genealogically founded. Bleek’s answer to the question heading this section was rather typologically oriented and took the feature of pronominal agreement through prefixes as its point of departure. In his Comparative Grammar of South African Languages, Bleek (1862, 1869) was also the first to come up with a noun class prefix system, whose structure and numbering system are still used by current-day Bantuists – and scholars of Niger-Congo more generally – some changes notwithstanding (cf. Katamba 2003: 104).

The scholar who truly defined Bantu as a genealogical unity, i.e., a family of languages descending from a common ancestor that can be reconstructed through the establishment of regular sound correspondences among modern-day languages, was the Prussian (German) philologist Carl Meinhof (1857–1944). Meinhof postulated the existence of Ur-Bantu or Proto-Bantu and stated that “Die Gesetze des Ur-Bantu sind nur aus den heute gesprochenen Bantusprachen zu erschliessen. Da sie aber in allen Bantusprachen ihre Spuren hinterlassen haben, ist ihre Kenntnis unerlässlich für die Erforschung der einzelnen Sprachen” (Meinhof 1899: 7) (“The laws of Proto-Bantu can only be deduced from Bantu languages spoken today.
However, as they have left their traces in all Bantu languages, their understanding is imperative for the study of the individual languages,” our translation]. Meinhof is not really explicit on the territorial spread of the Bantu languages, but from the map at the end of his 1899 treatise, it is obvious that his idea of the family’s distribution was much narrower than Bleek’s. Excluding Western Africa, Meinhof’s conception comes close to our own current conception of the Bantu area, except for his northern extensions into the Darfur and Kordofan regions. In his comparative work aiming at the reconstruction of Proto-Bantu, Meinhof also incorporated North-Western Bantu languages, such as Duala A24 from Cameroon.

Johnston (1919: 15) conceived the geographical delimitation of the Bantu languages along the same lines as Meinhof, i.e., “the whole of the southern third of Africa, with the exception of very small areas in the south-west (still inhabited sparsely by Hottentot and Bushman tribes) and a few patches of the inner Congo basin.” He was only hesitant about “[t]he northern boundary of the Bantu field,” which he considered to be “still a little uncertain and not easy to delineate geographically. It may be said to start on the west coast of Africa in the Bight of Biafra (due north of the island of Fernando Po), at the mouth of the Rio del Rey in the southern portion of the Bakasi peninsula, which flanks the estuary of the Old Calabar river.” He situated the north-western extremity of the Bantu area in the borderland between present-day Nigeria and Cameroon, but recognised that the languages spoken there miss some of the distinctive characteristics of a typical Bantu language:

There is no mistaking a Bantu language for a member of any other African speech family. A momentary glance at the numerals, at a dozen word-roots with their prefixes or suffixes, determines the fact whether it is or is not a member of the Bantu family. The phonology also is as a rule distinctive, though appearances may be deceptive in the case of a few languages of the north-western part of the Bantu field. The semi-Bantu languages on this north-west borderland have a vocabulary which contains a greater or smaller amount of Bantu roots, and farther north and west there are other language families which display obvious resemblances and affinities with what may have been the Bantu mother tongue; but outside the Bantu family there is no known speech group in Africa which displays all the characteristic features of Bantu word-construction and syntax and at the same time shows unmistakable affinity in word-roots.

(Johnston 1919: 17)

For a language to be qualified as Bantu according to Johnston, it thus had to fulfil at least two conditions: (1) possess a sufficient number of distinctive word-roots cognate with word-roots found elsewhere in the family, and (2) manifest certain characteristic phonological, morphological and syntactic features, such as simple vowel systems and open syllables, agglutination, invariable word-roots, noun class system, the absence of sex-based gender distinctions, pronominal agreement, verb extensions, decimal numeration and the use of prepositions rather than postpositions (Johnston 1919: 18–20). The 12 so-called “propositions” laid down by Johnston to “define the special or peculiar features of the Bantu languages” were actually a critical reassessment of the 12 structural parameters originally proposed by Lepsius (1880). These were long considered “an authoritative outline of Bantu criteria” and also quoted and used by Cust (1880) and Werner (1919) in their reference sketches of African and Bantu languages, respectively (Guthrie 1948: 9).

Johnston (1919) qualified a language only fulfilling one of the two conditions mentioned above as “semi-Bantu.” Although some of these “semi-Bantu” are what we consider today as “Bantoid” or “Wide Bantu” languages (cf. infra), Johnston (1919: 17–18) also identified
“semi-Bantu” languages elsewhere in the current-day Niger-Congo area, very much in line with Bleek’s West-African division of Bantu:

Curiously enough, there are languages in southern Kordofan, in Nigeria, at the back of the Gold Coast, or in the Sierra Leone region, the syntax or construction of which frequently recalls the Bantu idiosyncrasy; but the word-roots of the vocabulary would be found wholly dissimilar. Or there are others, again, in West Central Africa that exhibit a decided likeness to Bantu in their word-roots, yet in syntax and word construction are quite unlike the Bantu.

Johnston (1919: 17–18)

Johnston’s distinction between Bantu and semi-Bantu languages persisted in the comparative work of Malcolm Guthrie (1948, 1971). Building on the Bantu scholarship discussed above, Guthrie adhered for his referential classification to two principal (1–2) and two subsidiary (3–4) criteria to define what a proper Bantu language is: (1) A system of grammatical genders (or noun classes), usually at least five, corresponding to four more features which we cannot recall here for reasons of space; (2) a vocabulary, part of which can be related by fixed rules to a set of hypothetical common roots; (3) a set of invariable cores, or radicals, from which almost all words are formed by an agglutinative process, these radicals having five more features which we also cannot recall here for reasons of space; (4) a balanced vowel system in the radicals, consisting of one open vowel ‘a’ with an equal number of back and front vowels (Guthrie 1948: 11–12). Guthrie distinguished between two categories of “languages which are incompletely Bantu,” viz. “Bantoid” and “Sub-Bantu” languages. In Guthrie’s view, Bantoid languages are those spoken in Cameroon and south-eastern Nigeria, which “have a system of grammatical genders and agreements operated by means of prefixes,” but “show little or no relationship of vocabulary with full Bantu languages” and also “do not display even the rudiments of the structural features laid down in the third criterion; moreover their vowel system is frequently complicated” (Guthrie 1948: 19). Sub-Bantu languages, on the other hand, are those responding to all criteria set out by Guthrie except the first, i.e., a system of grammatical genders. They still manifest traces of the noun classes and grammatical agreement, but these systems have become very fragmentary, if not completely defunct. Most of Guthrie’s Sub-Bantu languages, such as the Congolese language Bira D32, occur in the northern Bantu borderland. Others are vehicular languages spoken further south, such as Lingala C36d. While Guthrie excluded Bantoid languages from his referential classification, he did include Sub-Bantu languages. As has been said, this slightly arbitrary delimitation of the Bantu family at its North-Western border is still in place and is used to define the subject of this book.

ACKNOWLEDGEMENTS

We wish to thank Sara Pacchiarotti for her critical reading of an earlier draft of this introduction and for signalling us the mention of H. E. Hale in Fodor (1980). The usual disclaimers apply.

REFERENCES


