

HISTORICAL LINGUISTICS

Koen Bostoën¹

Linguists and archaeologists offer complementary viewpoints on human behaviour and culture in past African communities. While historical-comparative linguistics commonly deals with the immaterial traces of the past in Africa's present-day languages, archaeology unearths the material vestiges of ancient cultures. Even if both sciences share similar core concepts, their methods, data and interpretive frameworks are profoundly different. Explaining some basic principles of historical-comparative linguistics as applied to the Bantu languages and debunking some common misconceptions are the central aims of this contribution. Due to space constraints, no detailed bibliographic references are provided throughout the text (see my earlier publications for extensive bibliographies).² Some essential readings for non-specialists are listed at the end of this chapter.

I. DIACHRONIC LINGUISTICS ON THE BASIS OF SYNCHRONIC DATA

Ideally speaking, historical linguistics is the study of distinct historical stages in the evolution of one single language or language family. This is the case in Romance, for instance, where the development of Latin into its multiple daughter languages can be empirically reconstructed. In Africa, examining language variation through time on the basis of diachronic language data is hardly ever possible, due to the lack of written documents.

The case of Kikongo, whose historical record starts in the early 17th century, is exceptional, and even not equalled by Kiswahili whose oldest surviving texts do not date further back than the mid-18th century. For most other Central African languages, written documents become at best available from the late 19th century onwards. Even today, there are still many undocumented languages, several of which are on the verge of extinction. Historical linguistics in Africa thus usually consists in the comparative study of historically-related languages. This up-stream approach, also known as

'historical-comparative linguistics', starts from extant languages and tries to reconstruct their evolution from ancestral stages through the study of current-day variation. Such inter-language variation can be phonological, morphological, syntactic, semantic or lexical.

In the case of Bantu, the hypothetical common ancestor language reconstructed on the basis of similarities observed between languages known mainly from the 19th century onwards is commonly called Proto-Bantu. This proto-language is assumed to be the best possible reflection of the ancestor language that was supposedly spoken some 4,000 to 5,000 years ago in the area from where Bantu languages started to spread through Central Africa and beyond. Bantu linguists agree to situate this homeland in the so-called Grassfields region of Cameroon, not far from the country's border with Nigeria. This zone displays the highest linguistic diversity (which means that parent languages had sufficient time to diverge locally) and is close to the area where the Benue-Congo relatives of Bantu languages are spoken.

II. REFERENTIAL VS. HISTORICAL OR GENEALOGICAL CLASSIFICATIONS

The best-known Bantu classification is no doubt Malcolm Guthrie's. In 1948, Guthrie subdivided the Bantu languages in 16 different zones labelled A, B, C, D, E, F, G, H, K, L, M, N, P, R, S and T, which he reduced to 15 in 1971 by merging the last two to one zone. Each zone is further subdivided into language groups, indicated by a decimal number, in which individual languages are indicated by a unit. Lowercase letters following certain units refer to dialects of a same language, e.g. Ciluba (L31a) and Lulua (L31b). In contrast to what is often believed, Guthrie's classification is strictly referential and was never meant to be historical: Guthrie did not rely on the 'Comparative Method' (which is the core approach of historical-comparative linguistics) or 'shared innovations', its basic principle for historical subgrouping. Shared innovations are lexical, phonological or grammatical changes that took place only once in some ancestor language from which its daughter languages inherited it and which are therefore indicative of the closer

¹ BantUGent - UGent Centre for Bantu Studies, Department of Languages and Cultures, Ghent University, Belgium.

² <http://research.flw.ugent.be/en/koen.bostoen>

relatedness between languages. By attributing a unique alpha-numeric code to each language, Guthrie wanted to facilitate comparison between the several hundred Bantu languages known at the time.

Despite its limited historical or genealogical value, Guthrie's classification remains a useful reference tool. Each one of the nearly 900 documented Bantu language varieties can be approximately situated in space thanks to its unique code. That is exactly why Jouni Maho updated Guthrie's list by adding new languages, but remained as faithful as possible to the original approach. Other scholars did propose rearrangements on historical grounds. Only one of these gained relatively wide acceptance amongst Bantu linguists, i.e. zone J proposed by the former linguistics department of the Royal Museum for Central Africa in Tervuren.

Bantu as a language family has been established ever since Bleek (1851). Its homeland is the region where the 'Narrow Bantu' languages, i.e. those conventionally classified as Bantu by Guthrie, meet the 'Wide Bantu' languages, i.e. their closest Benue-Congo relatives aka 'Bantoid'. The small 'Mbam-Bubi' subgroup, consisting of several languages of the Mbam region of central Cameroon and Bubi spoken on Bioko Island, is the genealogical junction between Narrow and Wide Bantu. The (Narrow) Bantu family further branches into five major subgroups: 'North-Western', 'Central-Western' (aka 'North Zaire' or 'Congo'), 'West-Western' (aka 'West-Coastal'), 'South-Western' and 'Eastern'. We mainly owe this robust understanding of Bantu genealogy to quantitative analyses of so-called 'basic vocabulary', such as lexicostatistics and phylogenetics. Qualitative approaches based on phonological and/or grammatical features fit less with the tree model of language divergence, and emphasize that convergence due to language contact also had a significant impact on the speciation of Bantu languages.

III. LANGUAGE AS AN HISTORICAL SOURCE

Our knowledge of the environmental, social, cultural, and historical phenomena underlying language change is often very limited in Africa. Its languages most often need to 'speak for themselves'. The study of language has in itself become an important method of reconstructing history to which not only linguists, but also historians and archaeologists dedicate themselves. Founded on the basic premise that vocabulary shared between speech

communities³ is a reflection of shared history, the study of widespread cultural vocabulary usually provides interesting insights on the lifestyle of past societies. This sub-discipline is also known as the 'words-and-things method' (see Ricquier, this volume, pp. 261-263) or linguistic palaeontology. To archaeologists, language data are particularly useful as a source of indirect historical evidence for those aspects of human culture which are either immaterial or whose material traces do not conserve well. Similar words with similar meanings shared by numerous languages can be inherited from a common ancestor language and spread through the dispersal of its daughter languages. They can also have been adopted through contact and spread across languages as loanwords.

To distinguish between inherited and borrowed vocabulary, linguists depend on the principle of regular sound correspondences. These are phonological similarities between languages, which cannot be the outcome of historical accident, because they are recurrent, systematic and without unexplainable exceptions. While synchronically widespread inherited terms can be reconstructed into a putative proto-language via these regular sound changes, loanwords cannot. Several Great Lakes Bantu languages, for instance, have a lexical doublet to refer to calabashes and glass bottles. These are two words that are historically related, but one of them was acquired through regular intergenerational transmission from an ancestor language, while the other was obtained from vehicular Swahili through contact-induced diffusion. The inherited word for calabash is phonologically much more heterogeneous, e.g. Sukuma *cuba*, Nyamwezi *nsòhá*, Ganda *ènsúwà*, Shi *nshùhá*. These words were subject to the regular sound changes that their language underwent since Proto-Bantu for which **-cópà* 'calabash' has been reconstructed. Such is not the case for the term for glass bottle, which they recently borrowed from Swahili resulting in much more similar loanwords: Sukuma *cupá*, Nyamwezi *cupa*, Ganda *ccúpà*, Shi *ícúpà*. In Swahili itself, the word *chupa* refers to both calabashes and glass bottles. When the latter type of containers were introduced along the East African coast, Swahili speakers called them after their traditional containers using the word for calabash which they inherited from Proto-Ban-

³ A speech community is defined here as a group of people who consider themselves to speak a same language.

tu. Swahili speaking long-distance traders subsequently introduced this new specimen of material culture and its Swahili word in several East-African communities, many of them already having a regularly inherited Bantu word for ‘calabash’.

Unlike archaeologists, linguists do not have a standard and universally accepted method for the absolute dating of language change. In the absence of diachronic language data and without the tentative association of language data to archaeological data, linguists need to limit themselves to relative dating. To do so, they rely on a number of principles for which they are indebted to archaeology: stratigraphy, geographic distribution and seriation.

Linguists refer to the concept of stratigraphy to disentangle the successive strata in the formation of a language. The grammar and lexicon of a language are transmitted through time and transformed due to the loss of old elements and the incorporation of new elements. They accumulate formative layers, which are never neatly superposed. Unlike archaeological strata, language layers are not subject to the law of superposition. There is permanent stratigraphic contamination, so to speak. It is the task of the historical linguist to order the present-day data into successive strata. The words for ‘calabash’ and ‘glass bottle’ in the example above clearly belong to two distinct strata of language history.

Linguistic geography or geolinguistics can help with the relative dating of language layers. This method deals with the geographic distribution of linguistic features. It is used for mapping loanword diffusion routes and for determining their direction of borrowing and also as a relative chronology device. ‘Linguistic isoglosses’ are the equivalent of stylistic horizons in archaeology. They mark the geographic distribution of a given linguistic feature shared by a number of languages. For example, cognate words for Kimbundu **njila** (‘bird’) are only found in a geographically restricted cluster of Bantu languages spoken in the southwestern part of the domain, while cognates for Kikongo **nuni** ‘bird’ are found throughout the Bantu domain. Such spatial distribution is interpreted as a function of time: the Kikongo word is a shared retention going back to Proto-Bantu while the Kimbundu word is a more recent shared innovation. The relative chronological interpretation of isoglosses is done according to certain areal norms which are not strict rules, but rather hermeneutic principles, e.g. the oldest form is

the most scattered one, which is preferably attested in the more outlying areas, while the younger form occurs in a group of adjacent languages, which may be large, but not as scattered as the older form. A judicious historical interpretation of isoglosses requires a basic insight into the internal classification of a language family. The relative time depth does not depend so much on the number of languages in which a feature occurs, but rather on its distribution over distinct historical subgroups. Hence, a term that is rare but scattered amongst the north-western and western Bantu languages is considered older than a synonym that is densely spread among eastern Bantu languages only.

A final basic archaeological concept also found in historical linguistics is seriation. Linguists usually rely on it for the sequential ordering of sound changes. Each language is subject to sound changes, which can be called regular to the extent that they affect all words sharing a given phonological environment. The chronological sequencing of sound changes is primarily used for the historical classification of languages through the principle of shared innovations. If closely related languages share a historical change (whether lexical, phonological or grammatical), there is a good chance that this innovation only happened once, i.e. in their most common recent ancestor – although independent convergent change can never be entirely excluded. Once one has an idea of the internal classification of a language group and the relative chronology of sound changes, seriation is also a helpful dating device for loanwords. The earlier foreign words are borrowed, the more sound changes they have in common with regularly inherited words and the better they are phonologically integrated, making it difficult to identify them as borrowed vocabulary.

CONCLUSIONS

The interaction between African archaeology and linguistics has been severely criticized in the past, among other things due to a lack of critical evaluation of underlying concepts and methods. Although this appreciation is certainly not undeserved, this should not refrain us from interdisciplinary collaboration. No discipline is capable of solving on its own the many complex riddles of African history. Sound archaeological-linguistic teamwork requires in the first place a good understanding of each other’s concepts, methods and evidence, to which I have tried to contribute in this chapter. A second

fundamental issue is the importance of direct collaboration between scholars of different disciplines who perfectly command their own body of evidence and are able to make a judicious assessment of its historical significance instead of leaving this task to scholars who only master one method or none at all. Finally, it is crucial that archaeologists and linguists mutually benefit from their specific advantages, e.g. absolute dating in the case of archaeology or the possibility to reconstruct vocabulary referring to immaterial or poorly preserved material aspects of human life in the case of historical linguistics.

REFERENCES

- Bastin, Y. & Piron, P. 1999. 'Classifications lexicostatistiques: bantou, bantou et bantoïde. De l'intérêt des "groupes flottants"'. In J.M. Hombert & L.M. Hyman (eds.), *Bantu historical linguistics: theoretical and empirical perspectives*. Stanford: CSLI Publications, pp. 149-164.
- Bastin, Y., Coupeuz, A. & Mann, M. 1999. *Continuity and Divergence in the Bantu Languages : Perspectives from a Lexicostatistic Study*. Tervuren: RMCA (series « Annales de Sciences humaines, série in-8° », no. 162).
- Bleek, W. 1851. *De nominum generibus linguarum Africae Australiae*. Bonnae: Formis Caroli Georgii.
- Blench, R. 2006. *Archaeology, Language and the African Past*. Lanham: Altamira Press.
- Bostoen, K., Clist, B., Doumenge, C., Grollemund, R., Hombert, J.-M., Koni Muluwa, J. & Maley, J. 2015. 'Middle to Late Holocene Paleoclimatic Change and the Early Bantu Expansion in the Rain Forests of West Central-Africa'. *Current Anthropology* 56: 354-384.
- Dimmendaal, G. 2011. *Historical Linguistics and the Comparative Study of African Languages*. Amsterdam/Philadelphia: John Benjamins.
- Grollemund, R., S. Branford, K. Bostoen, A. Meade, C. Venditti and M. Pagel. 2015. 'Bantu expansion shows that habitat alters the route and pace of human dispersals'. *Proceedings of the National Academy of Sciences of the United States of America*. <http://dx.doi.org/10.1073/pnas.1503793112>
- Heine, B. & Nurse, D. (eds). 2000. *African languages: an introduction*. Cambridge/New York: Cambridge University Press.
- Heine, B. and Nurse, D. (eds). 2004. *Les Langues africaines*. Paris: Karthala.
- Hombert, J.-M. & Hyman, L.M. (eds.). 1999. *Bantu Historical Linguistics: Theoretical and empirical perspectives*. Stanford, CA: CSLI.
- Maho, J.F. 2009. NUGL Online: The online version of the New Updated Guthrie List, a referential classification of the Bantu languages (4 June 2009). Online file: <http://goto.glocalnet.net/mahopapers/nuglonline.pdf>
- Nurse, D. & Philippson, G. (eds). 2003. *The Bantu languages*. ('Routledge language family series', no. 4). London/New York: Routledge.
- Nurse, D. and G. Philippson. 2003. 'Towards a historical classification of the Bantu languages'. In D. Nurse & G. Philippson (eds.), *The Bantu Languages*. London/New York: Routledge, pp. 164-181.
- Schadeberg, T.C. 2003. 'Historical linguistics'. In D. Nurse & G. Philippson (eds.), *The Bantu Languages*. London/New York: Routledge, pp. 143-163.
- Vansina, J. 1995. 'New Linguistic Evidence and the Bantu Expansion'. *Journal of African History* 36: 173-195.