Commentary on the paper by Bostoen et al.

Pierpaolo Di Carlo, University of Florence

Jeff Good, University at Buffalo

The paper by Bostoen et al. is a landmark attempt to synthesize results from a range of disciplines in order to refine our understanding of the Bantu expansion. It comes at an important moment in the field of linguistics where new quantitative methods are reviving disciplinary interest in questions of language and prehistory (see, e.g., Wichmann and Good 2014). We find particularly impressive the range of data types considered and are not aware of previous work drawing on such diverse strands of evidence.

Having attempted similar work ourselves, though on a much smaller scale (Di Carlo and Good 2014), we are well aware of the difficulties involved in efforts of this kind and immediately acknowledge that some degree of simplification is required in order to effectively conduct interdisciplinary investigation. Nevertheless, we believe that the presentation of the linguistic results, and, to some extent, the genetic ones, is oversimplified and rendered with more certainty than the evidence warrants. We focus on two concerns here: (i) the methods employed to arrive at a classification of Bantu languages and (ii) the assumed model of the relationship between languages and speaker communities.

The authors base the linguistic aspects of their synthesis on the results of the application of computational methods originally developed within the field of evolutionary biology to linguistic data. The particular phylogenetic methods that they employ have become commonly used in linguistic work in recent years (see, e.g., Dunn 2014), and even applied to domains associated with cultural anthropology (Jordan et al. 2009). They represent a powerful means to discover
tree-like patterns within large comparative data sets. However, the nature of the input data and the theoretical assumptions built into the algorithms necessarily limit the scope of the inferences that one can make from the results of their application. Here, it is important to note first that Bostoen et al.’s classificatory tree for Bantu is based solely on the comparison of 92 items of so-called “core vocabulary” from 542 Bantu languages. The Indo-European genealogical tree, by contrast, was developed using both lexical and grammatical data. Moreover, they have chosen to distill the data into a single tree-based representation, when it is well known that “the Bantu languages have the remarkable ability to act much more like a dialect continuum than as discrete and impermeable languages (Schadeberg 2003:158)”. In such a context, contact relationships are at least as important as genealogical ones for uncovering prehistoric patterns of change. It is, therefore, surprising that they have chosen to use a tree-generating algorithm, rather than a network-generating one (see, e.g., Heggarty et al. 2010), which would depict linguistic relationships in the Bantu area in a way which explicitly recognizes the significance of language contact. In our view, these concerns mean that the conclusions reached by examination of the subgrouping presented in Bostoen et al. can only be viewed as preliminary hypotheses. Our criticisms along these lines can hardly be considered new. Almost two decades ago, Nurse (1997), for instance, presented a detailed review of key aspects of some of the issues discussed above.

Perhaps more problematic, however, is the authors’ uncritical adoption of a model of language spread wherein branches of a phylogenetic tree are equated with migration paths (and splits) of speech communities, following an assumption that demic diffusion was the dominant mechanism of the Bantu spread. While actual demographic movements must have played a role in this process, they can hardly account for the whole of it. Moreover, such a view glosses over
culturally significant aspects of population expansion (e.g., involving differential rates of integration of non-Bantu speaking women than men in early, assumedly expanding, Bantu-speaking communities) for which there is some genetic evidence (Pakendorf et al. 2011:72). Our impression is that, here, the authors have unambiguously crossed the boundary between simplification and oversimplification in adopting a model based on the tree metaphor in its simplest form, thereby missing the chance to increase its explanatory power via the inclusion of ethnographic insights. In particular, a failure to pay attention to the pervasiveness of multilingualism throughout sub-Saharan Africa, which is so woven into the fabric of daily life that it seems impossible to consider it to be a recent phenomenon (Whiteley 1971:1), is an especially significant gap. How can a model of language diversification built around prehistoric population dispersion be rectified with the idea that the migrating communities were likely multilingual and in constant contact with nearby groups? This is not simply an incidental complication to their model, in our view, but a fundamental one.

In sum, we find ourselves deeply impressed by the level of synthesis attempted here. At the same time, however, we are anxious to see how a more complex and culturally informed linguistic and genetic picture could be integrated into the authors’ historical framework.

References


