

A New Phylogenetic Classification for the Bantu Languages of the Northeastern Democratic Republic of the Congo

The Northeast of the Democratic Republic of the Congo hosts a variety of Bantu languages belonging to subgroups that are many nodes apart on the Bantu tree. West and southwest of Kisangani, we find Upper-Congo languages (Lokele, So, Mbesa, etc.), belonging to the West-Bantu group in the classification of Bastin *et al.* (1999) and the Central-Western group according to Grollemund *et al.* (2015). Southeast of Kisangani, we find Lega languages (Enya, Metoko, Songola, etc.), part of the East-Bantu group in both classifications. Northwest, Northeast and Southeast of Kisangani, we find Boan languages (Leboale, Angba, Bali, Komo, etc.) and the extraordinary “Lebonya” languages (Lengola, Bodo, Nyali) that are considered early offshoots of the Bantu tree in Bastin *et al.*, but as Central-Western by Grollemund *et al.* The discrepancy between either classification is not only due to the different methods applied, but also the volume of data included in the classification. We continue here with the phylogenetic methods of Grollemund *et al.*, but we “pump up the volume” considerably in our coverage of the languages of the northeastern DRC. We present new data for over 50 Bantu languages, collected in 2019, 2020 and 2021, which we confront to older data for the same and neighboring languages.

In order to establish our classification, we have identified cognate sets for each word. All the cognate sets were then coded into binary characters. In order to infer our tree, we have used advanced computational phylogenetic methods. We applied to our data a likelihood model of lexical evolution (allowing different rates of evolution for the words studied) implemented in a Bayesian Markov chain Monte Carlo (MCMC) approach. The results both confirm and contradict aspects of previous classifications. To name a few examples, we can identify Upper-Congo and Lega as two steady genetic units affiliated to separate branches of the Bantu tree, in line with previous observations, but our classification of “Lebonya” shakes up the discussion. We will explain why conflicting classifications appear in this specific case. We furthermore add detail, presenting a subclassification for each group.

Finally, from our new phylogenetic classification used in combination with the information on the latitudinal and longitudinal positions of the languages, we reconstruct ancestral routes of language expansion into and through the northeast of the DRC as a linguistic contribution to the region’s settlement history.

References:

- Bastin, Y., A. Coupez & M. Mann. 1999. *Continuity and Divergence in the Bantu Languages: perspectives from a lexicostatistic study*. Tervuren: Musée Royal de l’Afrique Centrale.
- Grollemund, R., S. Branford, K. Bostoen, A. Meade, C. Venditti & M. Pagel. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences (PNAS)*. <https://www.pnas.org/content/112/43/13296>.