

George Starostin
(Russian State University for the Humanities, Moscow)

Lexicostatistics in Africa and its role in the reconstruction of Proto-Niger-Congo

The lexicostatistical method is regarded today by many comparative linguists as one of the most reliable and easily formalized means not only to establish the internal classification of a given linguistic family, but also to test the very existence of such a family. This is particularly important for various hypotheses of long-range linguistic relationship, where unambiguously interpretable comparative evidence is relatively scarce, and one has to resort to rigorous analysis of the most «stable» components of the language, such as morphology and the basic lexicon.

The Niger-Congo hypothesis, along with other hypotheses on African macrofamilies, usually associated with the work of J. Greenberg, requires a general lexicostatistical «validation». Although some of its branches have already been subject to lexicostatistical analysis (such as, e.g., Bantu or Mande), there has so far been no global assessment of the validity of Niger-Congo based on percentages of matches on the Swadesh list. This is partly due to the great number of languages (and the relatively small number of specialists who would be willing to assess all of the supposed branches of Niger-Congo at once), and partly to the lack of a formally established system of regular phonetic correspondences between these branches, usually deemed necessary to offer credible cognacy judgements from which the lexicostatistical matrix is derived.

In my talk, I will demonstrate a highly preliminary lexicostatistical assessment of the Niger-Congo hypothesis, based on roughly assembled wordlists for several hundred languages that include specimens from most of its branches and run through an automatic algorithm that tries to establish «pseudo-cognacy» based on phonetic similarity of the compared lexical items. For diagnostic purposes, a large group of languages from the other hypothetical African macrofamilies (such as «Nilo-Saharan» and «Khoisan») will be included in the calculations as well. This preliminary method has proven itself useful in the formulation of preliminary hypotheses and confirmation of language relationship in certain cases (such as comparison between phonetically archaic languages, or languages separated by relatively small time depth). The resulting lexicostatistical matrix and classificatory tree will show which parts of Niger-Congo are relatively «safe» even on preliminary levels of comparison, and which ones will require additional significant comparative-historical research to confirm or disprove.

Lastly, the overall results will be compared with a more «robust» lexicostatistical classification, produced for parts of the «Nilo-Saharan» family, which is based on more meticulous «step-by-step» reconstruction of proto-wordlists for some of its branches, helping to filter out both accidental similarities and traces of recent linguistic contacts. It is to be believed that similar work, performed for all or most branches of Niger-Congo, will greatly improve its lexicostatistical potential as well, not to mention aiding the overall reconstruction of the phonological structure and semantic status of its morphemic inventory.