

Automated reconstruction of Proto-Bantu using a statistical model of phoneme evolution

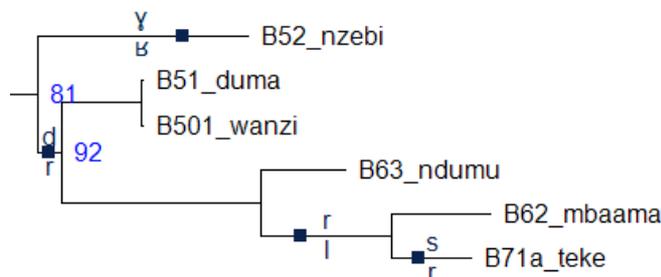
Rebecca Grollemund, Simon Branford & Mark Pagel
Evolutionary Biology Group, University of Reading

The comparative method is widely used by linguists to study the relatedness between languages and also to reconstruct proto-languages. Based on the comparison of similar words (cognates) from a group of languages and understanding of the regularity of sound changes, linguists use the Comparative Method to reconstruct proto-languages.

Here we present an automated likelihood-based statistical model that captures phonological aspects of language change in order to reconstruct proto-words. The model can detect sound changes characterizing language evolution and reconstruct the proto-phoneme for each site in the alignment.

To produce an alignment we start with the current phonological representations for a set of words in the basic vocabulary. For each of these word sets, we further split into cognate sets and these are then aligned phonemically. To this alignment we apply a Bayesian Markov chain Monte Carlo approach that can simultaneously estimate the phylogenetic tree and the matrix for the regular sound changes. This allows us to propose the most probable ancestral state for each site at any node of the tree.

We applied our model to the Bantu languages spoken in Africa (243 languages) for 100 words belonging to the basic vocabulary. Our model detects the regular sound changes that occur in Bantu (such as *d to l or *t to d).



Example of the classification of B50-60-70 languages with regular sound changes

By applying the same model we have produced a reconstruction of proto-Bantu and proto languages for several internal nodes (such as proto-Mbam-Bubi, proto north-Western, proto-Western, etc.). We will present the proto-words reconstructed and we will compare them with the ones reconstructed by Guthrie (1967-71) or Meeussen (1969).

References

GUTHRIE, M. 1967-1971. Comparative Bantu: an Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages, 4 vol, London, Gregg International Publishers Ltd.
MEEUSSEN, A. E. 1969. Bantu Lexical Reconstructions. Tervuren: MRAC.