Linguistic Typology 2009 vol.13 N1, pages 169-181

Adding typology to lexicostatistics: A combined approach to language classification

Bakker D., Müller A., Velupillai V., Wichmann S., Brown C., Brown P., Egorov D., Mailhammer R., Grant A., Holman E.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

The ASJP project aims at establishing relationships between languages on the basis of the Swadesh word list. For this purpose, lists have been collected and phonologically transcribed for almost 3,500 languages. Using a method based on the algorithm proposed by Levenshtein (Cybernetics and Control Theory 10: 707-710, 1966), a custom-made computer program calculates the distances between all pairs of languages in the database. Standard software is used to express the relationships between languages graphically. The current article compares the results of our lexicon-based approach with the results of a similar exercise that takes the typological variables contained in the WALS database as a point of departure. We establish that the latter approach leads to even better results than the lexicon-based one. The best result in terms of correspondence with some well-established genetic and areal classifications, however, is attained when the lexical and typological methods are combined, especially if we select both the most stable Swadesh items and the most stable WALS variables. © Walter de Gruyter.

http://dx.doi.org/10.1515/LITY.2009.009

Keywords

Family tree, Inflection, Lexicon, Lexicostatistics, Linguistic areas, Linguistic atlas, Phylogenetic classification, Syntax, Time-stability