

August 27, 2010

**Comparing ASJP Approaches to Automated Classification:  
Correspondence-Based and Lexical-Based Trees for Mayan**

Cecil H. Brown and Eric W. Holman

Eight classificatory trees for Mayan languages of Mesoamerica produced through ASJP automation are compared to one another and to expert (human) classification. These include both the traditional lexical-based ASJP trees (LEX trees) and trees (COR trees) based on shared phonological innovations/retentions made possible by a computer program developed for identifying sound correspondences for languages of the ASJP database. Four LEX trees and four COR trees are generated by using two different databases for Mayan languages, i.e., word lists entailing 40 referential items and 100 items respectively, and by using two different tree-producing algorithms, NJ and UPGMA. The eight trees are evaluated with respect to how well each replicates consensus expert subgroups of the family. In general, LEX trees replicate consensus subgroups better than COR trees; LEX and COR trees based on 100-item lists replicate better than those based on 40 items; and both LEX and COR trees produced through use of the NJ algorithm replicate better than those produced through UPGMA. Based on expert classification and that of automated trees, a new internal classification of Mayan is proposed. This classification accepts as most plausible those subgroups of best performing LEX and COR trees that are in agreement. When LEX and COR trees are not in accord, subgroups of the best performing COR tree typically trump those of LEX trees, this based on the assumption that shared phonological innovations are more indicative of genetic relationship than is lexical similarity. Finally, consensus expert judgment on subgrouping, when documented, has precedence over all other considerations.