

State of the art of the ASJP project

Søren Wichmann (MPI-EVA & Leiden University) and the ASJP Consortium

Since early 2007 a group of scholars, including myself, have pursued the idea of classifying the world's languages by a computer-automated lexicostatistical method (the Automated Similarity Judgment Program or ASJP). We have built up a database which currently contains data for around 2500 languages, 10% of which are standard 100-item Swadesh list and 90% of which are 40-item subsets of this list, representing the most stable items. The data have allowed for various statistical tests, including testing the basic premise of lexicostatistics and glottochronology that words change at a regular rate. Evidence from many languages have been brought forward against this assumption, but has tended to be anecdotal in nature. A systematic test across languages shows that within a tolerable margin of error the assumption is actually correct. This justifies pursuing both lexicostatistics and glottochronology. We will show some results for the world's languages in these regards, and in addition demonstrate a method for identifying areas of high genealogical diversity within families, a method which may be used as input to the inference of homelands.