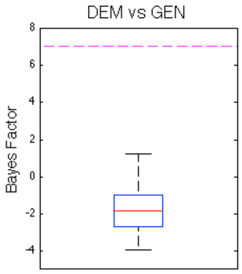
Languages do *not* show lineage-specific trends in word-order universals

In recent work, Dunn et al. [1] used a probabilistic evolutionary model of word-order language traits overlaid on a linguistic phylogeny to determine whether pairs of word-order ‘universals’ have evolved in tandem or not, concluding that coupled evolution was particular to certain language families but not others, hence ‘lineage specific.’ In this paper we show that when one carries out more extensive computer simulations refining the ‘noisy’ probabilistic estimates inherent in such methods, most trait co-variation disappears, at least in the Indo-European case. Further, for at least some traits in some language families (e.g., the Germanic languages), the possibility of non-binary, or ‘polymorphic’ traits also leads to inaccurate results. Given this, one cannot securely conclude that ‘word order universals’ are lineage specific, at least from the data and analysis presented in [1], because its results do not appear to be either reproducible or accurate.

The methodology in [1] first constructed linguistic “family trees” for four presumed independent language families, Indo-European, Austronesian, Bantu, and Uto-Aztecan, using Swadesh cognate word lists along with a Bayesian phylogenetic computer program. These trees were then overlaid with possible word-order trait pairs, such as demonstrative-noun order (DEM) and genitive-noun order (GEN). A second program simulated the evolution in each trait as it might change from the root of a language family tree forwards in time, with the output the likelihood that a value for the trait would be observed at the tree tip, given an initial trait state, the tree, and possible changes in between. This approach critically depends on inherently stochastic computer simulations using a random search technique that requires potentially billions of iterations to converge, and may not always do so [3, 4]. To partly guard against this possibility, [1] ran each test pair in each language family for 6 billion iterations, for six replicates; the median value out of 6 was reported as the ‘true’ value [2]. However, using Dunn’s original data and computer code, for the Indo-European family we re-ran each of the simulations in [2] for 10 billion iterations, 100 times, obtaining a more accurate picture of the *distribution* of likelihood values. The statistical signal for covariance largely disappears. Fig. 1 gives a representative example for DEM (demonstrative-noun order) and GEN. Dunn et al. reported that DEM-GEN *did* covary, since their reported statistical “Bayes Factor” test yielded a strongly significant test statistic, noted by the horizontal magenta line as a Bayes Factor of 7.65. Our 100-replicate run results are given by the box with a median value of –2, far below this, and strongly *not* significant.



In sum, powerful new computational tools for language analysis like these are only as secure as the assumptions and computations that ground them. In the case of compute-intensive, inherently difficult-to-replicate approaches, it is imperative that accuracy and reproducibility is confirmed before proceeding. **Words: 467**

[1] Dunn M, Greenhill SJ, Levinson SC, Gray RD. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473, 79–82.

[2] Dunn M, Greenhill SJ, Levinson SC, Gray RD. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473, Supplementary Online Materials.

[3] Meade, A. User manual for Bayes Traits and Bayes Phylogenies.

[4] Nichols, J., Warnow, T. 2008. Tutorial on computational linguistic phylogeny. Language and Linguistics Compass, 2/5 760-820.