The deep structure of Austronesian: A Bayesian reassessment (talk)

Isaac R. K. Stead

Submission for workshop "Recent advances in computational historical linguistics: New methods and results"

The current consensus in Austronesian (AN) historical linguistics is that the primary-level subgrouping of the family comprises ten branches. Nine of these are Formosan while the tenth, Malayo-Polynesian (MP), subsumes all languages spoken outside Taiwan (Blust 1999, 2009).

However, in the last decade, a number of competing proposals have challenged this view. Most recently, Chen et al. (in press) argue that the currently accepted East Formosan primary branch and MP are sister branches descended from a single protolanguage, which they term Proto-Coastal Formosan. Their proposal is based on a shared morphological innovation (transitive usage of the PAn voice affix *ma-) and shared phonological mergers and (as well as archaeological evidence). Ross (2009, 2012) argued for an alternative subgrouping, where Rukai, Tsou, and Puyuma remain independent primary branches and Malayo-Polynesian and the other Formosan branches are subsumed within a primary-level subgroup he termed Nuclear Austronesian. This argument was based on the claim that members of the proposed Nuclear Austronesian branch share the putatively innovative feature of Proto-AN root and relative clause voice affixes becoming homophonous (Chen 2017). Ho and Yang (2000) proposed a subgrouping with Rukai-Tsouic, Saisiyat-Atayalic, and Siraya-Amis-Bunun-Kavalan-MP primary level subgroups, based on shared phonological innovations. Finally, and most controversially, Sagart (2006, 2010, 2019) argued that Tai-Kadai is a subgroup of Austronesian, rather than an independent language family.

In this paper, we compare the above hypotheses to the results of a Bayesian phylogenetic analysis performed on 190 items of basic vocabulary sampled from all the primary subgroups of Austronesian according to the current consensus (155 doculects). A broad sample of the Tai-Kadai family (34 doculects) was included as an outgroup to root the tree and assist estimation of model parameters. These data have been coded for cognacy by expert linguists over a period of more than a decade (Greenhill, Blust & Gray 2008). The analysis was performed using BEAST2 software (Bouckaert et al. 2019). Here, we demonstrate that our results support Chen et al.'s Proto-Coastal Formosan hypothesis. Furthermore, we discuss the historical population movement scenarios suggested by the tree topology and dates produced by our analysis, and how they align with the latest results in the archaeology of Taiwan (Kuo 2019).

Finally, this paper demonstrates the utility of Bayesian techniques as a means of testing and calibrating subgrouping arguments made using traditional methods. This is because the subgrouping proposals that we have tested use different types of evidence (phonological, morphosyntactic), but our analysis evaluates all proposals according to a standard set of data. These results will be of great interest to both traditional and computational practitioners in historical linguistics.

References

Bouckaert R., Vaughan T.G., Barido-Sottani J., Duchêne S., Fourment M., Gavryushkina A., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS computational biology*, 15(4), e1006650.

Blust, R. (1999). Subgrouping, Circularity and Extinction: Some issues in Austronesian comparative linguistics. In E. Zeitoun and P. J.-K. Li (eds.), *Selected papers from the Eighth International Conference on Austronesian Linguistics*, 1-54. Taipei, Taiwan: Academia Sinica.

Blust, R. (2009). *The Austronesian Languages*. (Pacific Linguistics, 602.) Canberra: Research School of Pacific and Asian Studies, Australian National University. Xxviii+824pp.

Chen, V. (2017). A re-examination of the Philippine-type voice system and its implications for Austronesian primary-level subgrouping. PhD dissertation, University of Hawai‘i.

Chen, V., Kuo, J., Gallego M.K., & Stead I.R.K. (In press). Is Malayo-Polynesian a primary branch of Austronesian? A view from morphosyntax. *Diachronica* 39(1).

Greenhill, S.J., Blust. R, & Gray, R.D. (2008). The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexomics. *Evolutionary Bioinformatics*, 4:271–283.

Ho, D.-A. & Yang, H.-F. (2000). Prologue: Austronesian and Formosan languages. In L. Huang (ed.), *Formosan languages grammar series.* (in Chinese), 1–25. Taipei: Yuanliu.

Kuo, S-C. (2019). *New Frontiers in the Neolithic Archaeology of Taiwan (5600–1800 BP)*. Springer Singapore.

Ross, M. (2012). In Defense of Nuclear Austronesian (and Against Tsouic). *Language and Linguistics* 13(6). 1253–1330.

Ross, M. (2009). Proto Austronesian verbal morphology: a reappraisal. In A. Adelaar & A. Pawley (eds.), *Austronesian historical linguistics and culture history*: A festschrift for Robert Blust, 295–326. Canberra: Pacific Linguistics.