

On the evolution of linguistic disparity

Angela, M. Chira¹, Simon, J. Greenhill^{1,2}, Mary Walworth¹

¹Department of Linguistic and Cultural Evolution, Max Planck Institute of Evolutionary Anthropology; Leipzig, Germany.

²Australian Research Council's Centre of Excellence for the Dynamics of Language, College of Asia and the Pacific, Australian National University; Canberra 2600, ACT, Australia.

Keywords: linguistic disparity; morphosyntactic change; phonological change; Oceanic linguistics

The pattern of languages around the world is very different. Some regions have high language richness, i.e., a large number of languages, or high diversity. These languages can belong to the same family, or conversely, they can be distantly related – in which case, the region would show high phylogenetic diversity. Yet other regions show high differences between local languages, a concept we refer to as disparity. Explaining broad scale patterns of language diversification and disparification are topics of considerable interest in linguistics, but formalizing these concepts is not always straightforward (e.g., Nichols 1992, Nettle 1999).

Phylogenetic comparative methods commonly used in evolutionary biology can be a powerful tool for quantifying and understanding drivers of linguistic disparity in a standardized framework. Bayesian phylogenetic models have been applied extensively for disentangling how languages are related (e.g., Gray et al. 2009, Bowerman & Atkinson 2012, Zhang et al. 2019). Bayesian methods can also be leveraged to understand the extent and timing at which lineages accumulate differences in various language aspects of interest. For example, these models can show us whether lineages record substantial changes and reach maximal disparity early in their family history – akin to the biological concept of radiation or rapid proliferation of forms when populations move into unexplored terrain (Schluter 2002). Conversely, sizable changes can occur late in the history of a particular family, for example, as a result of recurrent contact with other language families. Language disparification can occur independently of language splitting, and patterns can also differ across different language aspects.

Here we present the evolution of structural disparity among the Oceanic subgroup of the large Austronesian family. These languages offer a great test case for formalizing the concept of disparity using computational methods, as relatedness is well established via the comparative method (Lynch et al. 2002, Ross et al. 2016) and well-resolved Bayesian phylogenies have been accepted (Gray et al. 2009). Further, comprehensive lexical, phonological, and morphosyntactic datasets are available (Greenhill et al. 2008, List et al. 2021). We used a new Bayesian Oceanic phylogeny with extensive cover in Bayesian models of continuous character change for six morphosyntactic indices and four aspects of sound inventories. Morphosyntactic indices are computed using structural data from 195 typological features from the upcoming Grambank database. Phonological data were obtained by automated inference from wordlists available in Lexibank (List et al. 2021). Bayesian models reconstruct the magnitude and speed of changes in language aspects of interest at each node within the Oceanic phylogeny, thus describing the evolution of structural disparity throughout the group's history.

Our results show that linguistic disparity tends to accumulate late in Oceanic history, in contradiction with a scenario of an early family-wide radiation in forms. Rather, increased disparity emerges with time via a mix of constant language change punctuated with bursts of change within language subgroups. These localized episodes of substantial changes show the importance of extrinsic factors – such as contact with Papuan lineages – in driving language divergence. Notably, patterns for disparity accumulation can be decoupled both between and within phonological and morphosyntactic features, showing that not all aspects of language structure respond equally to potential drivers of change. Language disparification thus seems to follow a pattern of mosaic (or modular) evolution, where changes in various language aspects can occur independently from each other.

References

- Bowern, C., & Atkinson, Q. Computational phylogenetics and the internal structure of Pama-Nyungan. *Language*, 817-845 (2012).
- Gray, R. D., Drummond, A. J., & Greenhill, S. J. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*, 323(5913), 479-483 (2009).
- Greenhill, S.J., Blust, R., & Gray, R.D. The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexomics. *Evolutionary Bioinformatics*, 4:271–283 (2008).
- List, J. M., Forkel, R., Greenhill, S. J., Rzymiski, C., Englisch, J., & Gray, R. D. Lexibank: A public repository of standardized wordlists with computed phonological and lexical features (2021).
- Lynch J, Ross M, & Crowley T. *The Oceanic Languages*. Routledge (2002).
- Nettle, D. *Linguistic Diversity*. Oxford University Press, Oxford (1999).
- Nichols, J. *Linguistic diversity in space and time*. University of Chicago Press (1992).
- Ross M, Pawley A, & Osmond M. *The lexicon of Proto Oceanic*. Canberra: Asia-Pacific Linguistics (2016).
- Schluter, D. *The ecology of adaptive radiation*. Oxford University Press, Oxford (2000).
- Zhang, M., Yan, S., Pan, W., & Jin, L. Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature*, 569(7754), 112-115 (2019).