

PHYLOGENETIC EXPLORATION OF LANGUAGE COMPLEXITY IN AUSTRONESIAN, BANTU, AND INDO- EUROPEAN LANGUAGE FAMILIES

OLENA SHCHERBAKOVA^{*1}, HEDVIG SKIRGÅRD², and SIMON J. GREENHILL^{1,2}

^{*}Corresponding Author: shcherbakova@shh.mpg.de

¹Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Jena 07743, Germany

²Australian Research Council Center of Excellence for the Dynamics of Language, Australian National University, Canberra, ACT 0200, Australia

While language complexity has received attention from sociolinguistic, psycholinguistic, and computational perspectives, the processes of simplification and complexification over time remain challenging to examine and explain. One strand of research focuses on complexity ‘tradeoffs’ and ‘local complexity’ asking whether complexification in one grammatical domain necessitates simplification in another so that all languages are ‘equi-complex’ (Miestamo, 2009, Sinnemäki, 2008). The tradeoffs may or may not occur between different language systems, such as phonetics and morphology (Shosted, 2006), morphology and syntax (Dahl, 2009, Sinnemäki, 2008), morphosyntax and vocabulary size (Real et al., 2018), and morphosyntax and semantics (Bisang, 2009).

However, most studies agree that local complexity across different domains varies between languages and there is little evidence for tradeoffs. Instead, there appear to be evidence of considerable variation in the causes and effects of complexity. For example, measures of complexity do not respond to extra-linguistic factors in the same way; Sinnemäki and Di Garbo (2018) show that verbal inflectional synthesis negatively correlates with the population size of speech communities, while nominal complexity of grammatical gender does not. Linguistic paradigms appear to be more challenging to transmit within larger communities (Nettle, 2012, Real et al., 2018) as they are harder to acquire by L2 learners, resulting in creoles being paradigmatically simpler (Good, 2012).

Here we aim to quantify and explore these issues using a large global dataset, Grambank (Skirgård et al. in review), comprised of 195 grammatical features from over 1,800 languages. To measure nominal and verbal complexity, we selected features signaling the adherence of the languages to the principle of distinctiveness (Sinnemäki, 2009). While the presence of these features facilitates comprehension for hearers providing additional grammatical information, they simultaneously contribute to the complexity from the speakers' perspective impeding the ease of production and requiring more efforts to articulate (Mufwene, 2012). Our metric of nominal complexity encompasses the presence of marking for such categories like number, gender, possessiveness, and case. Conversely, our metric of verbal complexity accounts for marking of arguments, overt signaling of tenses, aspect, and other markers on verbs. Finally, we measure the paradigmatic complexity of function words based on the distinctions existing between articles, pronouns, demonstratives, classifiers, and adpositions, which so far have been overlooked in most complexity studies.

We use these metrics to quantify language complexity along three different axes: nominal, verbal, and paradigmatic complexity. We apply cutting-edge phylogenetic methods to model change in these measures of complexity and explore how they have diversified, changed, and traded-off over time in the Austronesian, Bantu, and Indo-European language families. We then undertake a path analysis while controlling for phylogenetic non-independence to disentangle the evolutionary relationships between the measured complexity types (van der Bijl, 2018).

Our results suggest that, first, the paradigmatic complexity of function words does not interact with the nominal and verbal complexity. In contrast, nominal and verbal complexity display a weak positive correlation in Bantu and Indo-European languages. However, the correlation in the Indo-European family disappears once creole languages are removed from the analysis. Furthermore, the investigation of causal mechanisms via path analysis on a global dataset indicates that nominal complexity tends to influence verbal complexity. Our results suggest that the processes of simplification and complexification show no overall global trends but prove to be dependent on the domain and language family.

References

- Bisang, W. (2009). On the evolution of complexity: sometimes less is more in East and mainland Southeast Asia. In G. Sampson, D. Gil and P. Trudgill (Eds.), *Language Complexity as an Evolving Variable* (pp. 34-49). Oxford: Oxford University Press.
- Dahl, Ö. (2009). Testing the assumption of complexity invariance: The case of Elfdalian and Swedish. In G. Sampson, D. Gil and P. Trudgill (Eds.), *Language Complexity as an Evolving Variable* (pp. 50-63). Oxford: Oxford University Press.
- Good, J. (2015). Paradigmatic complexity in pidgins and creoles. *Word Structure*, 8(2), 184-227.
- Miestamo, M. (2009). Implicational hierarchies and grammatical complexity. In G. Sampson, D. Gil and P. Trudgill (Eds.), *Language Complexity as an Evolving Variable* (pp. 81-97). Oxford: Oxford University Press.
- Mufwene, S. S. (2012). The emergence of complexity in language: An evolutionary perspective. In A. Massip-Bonet and A. Bastardas-Boada (Eds.), *Complexity perspectives on language, communication and society* (pp. 197-218). Springer Verlag.
- Nettle, D. (2012). Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1597), 1829-1836.
- Real, F., Chater N., & Christiansen, M. H. (2018). Simpler grammar, larger vocabulary: How population size affects language. *Proceedings of the Royal Society B: Biological Sciences*, 285(1871), 20172586.
- Shosted, R. K. (2006). Correlating complexity: A typological approach. *Linguistic Typology*, 10, 1-40.
- Sinnemäki, K. (2008). Complexity trade-offs in core argument marking. In M. Miestamo, K. Sinnemäki and F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 67-88). Amsterdam, Philadelphia: Benjamins.
- Sinnemäki, K. (2009). Complexity in core argument marking and population size. In G. Sampson, D. Gil and P. Trudgill (Eds.), *Language Complexity as an Evolving Variable* (pp. 126-140). Oxford: Oxford University Press.
- Sinnemäki, K., & Di Garbo, F. (2018). Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: a typological study of verbal and nominal complexity. *Frontiers in psychology*, 9, 1141.
- Skirgård et al. (in review). Grambank, Jena: Max Planck Institute for the Science of Human History.
- Trudgill, P. (2011). *Sociolinguistic typology: social determinants of linguistic complexity*. Oxford: Oxford University Press.
- van der Bijl, Wouter. (2018). phylopath: Easy phylogenetic path analysis in R. *PeerJ* 6:e4718 <https://doi.org/10.7717/peerj.4718>