

# GRAMMATICAL COMPLEXITY IS ONLY WEAKLY INFLUENCED BY THE SOCIOLINGUISTIC ENVIRONMENT

OLENA SHCHERBAKOVA <sup>\*1</sup>, SUSANNE MARIA MICHAELIS<sup>2</sup>, HANNAH J.  
HAYNIE<sup>3</sup>, SIMON J. GREENHILL<sup>4</sup>, DAMIÁN E. BLASI <sup>5, 2, 6</sup>, RUSSELL D.  
GRAY<sup>2,7</sup>, VOLKER GAST<sup>8</sup>, and HEDVIG SKIRGÅRD<sup>2</sup>

\*Corresponding Author: [shcherbakova@shh.mpg.de](mailto:shcherbakova@shh.mpg.de)

<sup>1</sup>Max Planck Institute for the Science of Human History, Jena, Germany

<sup>2</sup>Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

<sup>3</sup>University of Colorado, Boulder, CO, USA

<sup>4</sup>School of Biological Sciences, University of Auckland, Auckland, New Zealand

<sup>5</sup>Harvard University, Cambridge, MA, USA

<sup>6</sup>Yale University, New Haven, CT, USA

<sup>7</sup>School of Psychology, University of Auckland, Auckland, New Zealand

<sup>8</sup>Friedrich-Schiller University of Jena, Jena, Germany

Recent studies claim that the social environment influences the evolution of language structures. In particular, grammatical complexity has been proposed to be lower in communities with looser social networks, higher numbers of L1 speakers, and higher proportions of L2 speakers (among others, Kusters 2003, Trudgill 2011, Lupyán & Dale 2010, Sinnemäki & Di Garbo 2018). The explanation for these relationships relies on the assumption that larger communities are exposed to more contact than smaller ones. Specifically, due to substantial proportions of L2 speakers in large communities, the more complex features are not always transmitted to further generations, and hence languages become simpler over time. Here we test these claims on a global scale using two metrics of grammatical complexity while controlling for phylogenetic and spatial non-independence.

In previous studies, metrics of grammatical complexity often capture two different phenomena (cf. Lupyán & Dale 2010): the amount of phonologically fused marking ("boundness") and the number of semantic distinctions ("informativity"). In our study, we separate these into two different metrics to evaluate potential differences in their behavior.

We use the typological information from Grambank (The Grambank Consortium 2022) to construct metrics for each of these concepts and test whether variation in the metric scores depends on the number of L1 speakers, the proportion of L2

speakers, the status of the language (official/not official) obtained from Ethnologue (Eberhard et al. 2021), number of linguistic neighbors, and population density available in Bromham et al. (2022). The status of the language has not been previously used in investigating these questions, but we include it in this study to glean the information on the function of the language in a community. In line with previous research, we expect the two dimensions of grammatical complexity, boundness and informativity, to be negatively correlated with all social variables, except for population density which serves as a proxy for social network density and can be expected to positively correlate with complexity dimensions (c.f. Trudgill 2011).

To explore relationships between language and social structures, we adopt a spatiophylogenetic modeling technique introduced in Dinnage et al. (2020), a Bayesian approach that uses an Integrated Nested Laplace Approximation (INLA) (Rue et al. 2009, Martins et al. 2013). The models fit social variables as fixed effects and two structured random effects to estimate the phylogenetic and spatial influence on variation in metric scores. This method allows us to not only control for phylogenetic and spatial non-independence of languages but also to estimate the effects of phylogeny and geographical distance between languages on variation in grammatical complexity and evaluate if adding the social variables improves the explanatory power of the model above and beyond the phylogenetic and geographic similarity. We compare the models combining sets of random and fixed effects based on obtained WAIC values (Watanabe 2010).

Contrary to prior studies, our Bayesian spatiophylogenetic modeling results do not support a hypothesis that morphological complexity (boundness) is strongly influenced by the number of L1 speakers, the proportion of L2 speakers, the number of linguistic neighbors, or population density. In contrast to previous studies, we disentangle boundness from informativity in our analyses, and our findings can be explained by the use of an extensive dataset and our rigorous control for phylogenetic and spatial non-independence of languages. As a result, the claimed link between the grammatical complexity and social variables does not appear to be strong and is found only weakly between informativity scores and some of the social variables, such as a positive effect of L1 speaker population and the official language status on informativity. This indicates that the evolution of these complexity dimensions is better explained from the perspective of inheritance and areal diffusion rather than constraints imposed by different social environments.

## **Acknowledgements**

The authors are thankful to Sam Passmore for his advice on the methods.

## References

- Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., & Hua, X. (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution*, 6(2), 163–173.
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2020). *Ethnologue: Languages of the World* (24th ed.). SIL International. [www.ethnologue.com](http://www.ethnologue.com)
- Dinnage, R., Skeels, A., & Cardillo, M. (2020). Spatiophylogenetic modelling of extinction risk reveals evolutionary distinctiveness and brief flowering period as threats in a hotspot plant genus. *Proceedings of the Royal Society B*, 287(1926), 20192817.
- Kusters, W. (2003). *Linguistic complexity*. Netherlands Graduate School of Linguistics.
- Lupyan, G., & Dale, R. (2010). Language Structure Is Partly Determined by Social Structure. *PLOS ONE*, 5(1), e8559. <https://doi.org/10.1371/journal.pone.0008559>
- Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis*, 67, 68–83.
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series b (Statistical Methodology)*, 71(2), 319–392.
- 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.
- The Grambank Consortium (eds.). (2022). *Grambank 1.0*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://grambank.clld.org>
- Trudgill, P. (2011). *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. OUP Oxford.
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.