

# Generalizability and Reproduction in Quantitative Typology: A Case Study on Sound Symbolism

Quantitative research in linguistics and other fields suffers from at least two crises: The replication crisis (Bochynska et al. 2023, Kobrock & Roettger 2023) and the generalizability crisis (Yarkoni 2020). The former relates to the reproduction and replication of established research results, which often do not hold up when tested on new data. The latter relates to our possibility of generalizing results beyond the sample to make claims about statistical universality, a common goal in linguistic typology (Bickel 2011). Both crises also affect quantitative typology, and arise for example when different kinds of bias (areal, genealogical) in the typological sample are not sufficiently accounted for (Bickel 2011, Roberts et al. 2015, Jaeger et al. 2011, Guzmán Naranjo & Becker 2021). Too often, at least one of these aspects is ignored, which impedes the generalizability of the results and leads to conclusions that potentially lack statistical validity. While those issues have found some attention in typological literature (Corbett 2005, Bisang 2011, Winter & Grice 2021, Hartmann et al. 2024), systematic approaches to replication are scarce until now (Editorial Board of Linguistic Typology 2006, Song 2007).

In the talk, I will present a replication study on a paper that investigates sound symbolism (Erben Johansson et al. 2020). The study finds a number of relationships between 12 phonetic features that are more likely than chance to occur in words for certain concepts. In order to avoid genealogical dependencies between languages, they restricted their sample to one language per language family. The authors also establish a relationship to previous studies on sound symbolism, and find a partial overlap with existing results, but also some new, previously unobserved patterns. Thankfully, the authors also share their model and code, making it possible to reproduce the results.

For the replication itself, we proceed in two steps. In a first step, we reproduce the original results and describe the challenges that arise during this process. In a second step, we modify the original model by adding a Gaussian process term for bias control of areal diffusion (Guzman Naranjo & Becker 2021). Here, we find that most of the initially claimed sound symbolism relationships do remain valid, but some evidence for areal effects is present as well. We discuss the outcome of the areal control on the original results and its implications for the replication of other studies in linguistic typology.

## References

- Bickel, B. (2011). Statistical modeling of language universals. *Linguistic Typology*, 15(2), 401–413. <https://doi.org/d10.1515/lity.2011.027>
- Bisang, W. (2011). Variation and reproducibility in linguistics. In Peter Siemund (ed.) *Linguistic universals and language variation*, 237-263. Berlin: De Gruyter. <https://doi.org/10.1515/9783110238068>
- Bochynska, A.; Keeble, L.; Halfacre, C.; Casillas, J. V; Champagne, I.; Chen, K., et al. (2023). Reproducible research practices and transparency across linguistics. *Glossa Psycholinguistics*, 2(1). <http://dx.doi.org/10.5070/G6011239>

- Corbett, G. G. (2005). Suppletion in personal pronouns: Theory versus practice, and the place of reproducibility in typology. *Linguistic Typology*, vol. 9, no. 1, pp. 1-23. <https://doi.org/10.1515/lity.2005.9.1.1>
- Editorial Board of LT (ed.) (2006). Re-doing typology. *Linguistic Typology* 10(1). 67–68. <https://doi.org/10.1515/LINGTY.2006.004>.
- Erben Johansson, N., Anikin, A., Carling, G. & Holmer, A. (2020). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology*, 24(2), 253–310. <https://doi.org/10.1515/lingty-2020-2034>
- Guzmán Naranjo, M., & Becker, L. (2021). Statistical bias control in typology. *Linguistic Typology*, 26(3), 605–670. <https://doi.org/10.1515/lingty-2021-0002>
- Hartmann, F.; Roberts, S.; Valdes, P. & Grollemund, R. (2024). Investigating environmental effects on phonology using diachronic models. *Evolutionary Human Sciences* 6. e8. <https://doi.org/10.1017/ehs.2023.33>
- Jaeger, T. F.; Graff, P.; Croft, W. & Pontillo, D. (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology*, vol. 15, no. 2, pp. 281-319. <https://doi.org/10.1515/lity.2011.021>
- Kobrock, K. & Roettger, T. B. (2023). Assessing the replication landscape in experimental linguistics. *Glossa Psycholinguistics*, 2(1). <https://doi.org/10.5070/g6011135>
- Lilienfeld, S. O. (2017). Psychology’s Replication Crisis and the Grant Culture: Righting the Ship. *Perspectives on Psychological Science*, 12(4), 660–664. <https://doi.org/10.1177/1745691616687745>
- List, J.-M.; Forkel, R.; Greenhill, S. J.; Rzymiski, C.; Englisch, J. & Gray, R. D. (2022). Lexibank, a public repository of standardized wordlists with computed phonological and lexical features. *Scientific Data*, 9(1), 1–16. <https://doi.org/10.1038/s41597-022-01432-0>
- Song, J. J. (2007). What or where can we do better? Some personal reflections on (the tenth anniversary of) *Linguistic Typology*. *Linguistic Typology* 11(1). 5–22. <https://doi.org/10.1515/LINGTY.2007.002>.
- Winter, B. & Grice, M. (2021). Independence and generalizability in linguistics. *Linguistics*, 59(5), 1251–1277. <https://doi.org/10.1515/ling-2019-0049>
- Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, 45. <https://doi.org/10.1017/s0140525x20001685>