Commentary

Francesca Di Garbo*

Embracing uncertainty, and the multifaceted soul of linguistic typology: commentary on "Replication and methodological robustness in quantitative typology" by Becker and Guzmán Naranjo

The newly published article by Laura Becker and Matías Guzmán Naranjo (2025) empirically demonstrates the importance of replication and replicability for assessing the robustness of research results and proposed generalizations in quantitative linguistic typology. The four case studies presented in the paper, as well as the insightful discussion that accompanies them, illustrate how to embed replication of statistical analyses into the research practices of quantitatively minded typologists, and why. The authors argue that replication practices in linguistic typology may invaluably contribute to assessing the degree of (un)certainty of our findings, and to capitalize on the fact that there is not a single best way to analyze a typological dataset. The article culminates in a call for more replication studies to be run and published in quantitative typology, as a way of consolidating typological results and the generalizations stemming from them, thus also enriching the range of possible scientific outcomes in comparative cross-linguistic research. Importantly, for replication studies to become common practice among typologists, full transparency in data sharing is needed.

Given that best practices concerning the robustness and replicability of typological results have been rarely discussed in their own terms (Corbett 2005; Haspelmath and Siegmund 2006), Becker and Guzmán Naranjo's paper is likely to have a long-term impact on the epistemological foundations of the field. This commentary broadly focuses on the scope of the paper and attempts to relate it to a wider discussion of the inherently multifaceted nature of linguistic typology.

*Corresponding author: Francesca Di Garbo [fran'ʧeska di 'garbo], Aix-Marseille University – CNRS LPL, Marseille, France, E-mail: francesca.di-garbo@univ-amu.fr

In her 1992 book *Linguistic Diversity in Space and Time*, Johanna Nichols brilliantly defines linguistic typology as a population science, on a par with biology and population genetics, whose scope ranges between micro-level observations of individuals to high-level generalizations on human linguistic behavior across time and place. Owing to this highly diversified way of approaching the description and modeling of linguistic diversity, linguistic typology has always been a very heterogeneous field of research, bringing together a diverse range of scholars with equally varied research agendas. Currently, the typologist's research profile could be said to fall into four major subgroupings, with inevitable intersections across each of them:

1. Descriptive and documentary linguists, engaged in primary data collection and community work
2. Comparative linguists interested in developing methods and tools for comparing languages and speech communities with one another
3. Comparative linguists interested in theoretical generalizations on the functioning of human languages in and beyond their context of use
4. Data scientists working with already annotated (large) typological datasets using statistical methods, sometimes borrowed from other disciplines, such as population genetics, evolutionary biology, and geography.

What brings together this heterogenous scholarly crowd is the overarching interest in describing and explaining human linguistic diversity, its distribution in space and time as well as its non-linguistic, e.g. environmental or cognitive, correlates.

Such a broadly defined domain of investigation naturally translates into a very diverse set of research tools and methodologies, none of which, in my opinion, should be conceived of as scoping above the others. Here's why.

– Primary data stemming from field-based research as well as case studies of individual languages and speech communities are vital to continuously generate evidence on the distribution of linguistic diversity and to inspire new research questions in comparative cross-linguistic research.
– When typological data collection and annotation are run from scratch, working on large (say, over 1,000 languages) datasets of languages of the world is costly in human terms, and hard to attain within the time frame of the research projects that are most typically funded at universities and research institutions from across the world.
– Exploratory studies based on sampling still provide an important, and in my opinion, even necessary testing ground to investigate the cross-linguistic frequency of lesser-known phenomena. These data can be used to formulate hypotheses about the areal and genealogical distributions of linguistic patterns.

Statistical testing can thus be a valid way to (start) investigating small to moderately large typological datasets.

– Statistical modelling using statistical bias control enables typologists to investigate typological distributions from the bottom up, and to better capture the complexities of the factors at stake when considering the distribution of languages and linguistic structures in place and time. When applicable, this approach can be very beneficial to disclose various types of patterns in the distribution of linguistic diversity, as Becker and Guzmán Naranjo clearly show.

These different ways of investigating linguistic diversity are equally important sources of evidence in the typologist's workflow. While, in most cases, the academic profile of individual scholars does not encompass all of these aspects, cumulative evidence stemming from collaborative work in and across these domains can, in its own terms, be seen as a way of corroborating typological results and verifying the validity of typological generalizations. Emphasizing the importance of this incremental workflow, from descriptive evidence to large-scale generalizations, is crucial to avoid overly relying on any of the research steps mentioned above, and, in turn, to abstain from qualifying any of these steps as superior to the others.

In this approach, the entire workflow of typological research, and not just statistical modeling, is held accountable for validating results and generalizations. This is also more reasonable in terms of the diverse academic conditions in which members of the research community operate. Still today, scholars' expertise and training vary enormously across countries, research institutions, and generations. For instance, traditionally trained typologists are usually well versed in relating descriptive data on language-specific structures to the comparative concepts that they design and use for cross-linguistic comparison. However, they tend to be less well versed in Bayesian statistics. Conversely, not all quantitative-minded typologists have received traditional linguistic training, having sometimes developed an interest for linguistic diversity through related disciplines such as data science, evolutionary biology or population genetics. Finally, advanced academic training in linguistic typology (including statistical training) is far from being equally widespread across the world and many young students in linguistics discover typology "late", through the lenses of descriptive and theoretical linguistics. While nothing is wrong with any of these scenarios, each of them inevitably bears its own consequences in terms of the type of research that typologists of different orientations produce.

Thus, given that, for all the above-mentioned reasons, not all typological investigations are amenable to statistical modeling, it is important to underscore that inferential statistics may not be the only way to validate research results in our field. Exploratory cross-linguistic research based on sampling remains, for instance, much

needed, given that existing typological datasets, even the largest and best curated ones, do not exhaust the range of linguistic phenomena that typologists may be interested in studying. Moreover, triangulation studies, aiming at capturing converging evidence by using data simultaneously coming from typological databases, corpora of language use and experimental designs, represent yet another promising way to test the validity of old and new typological generalizations, as Becker and Guzmán Naranjo also point out. Independently of the methods chosen for replicating or cross-validating earlier results, accessible and transparent data management practices are indeed essential in linguistic typology. This requires placing empirical data, and their organization, at the center of the publication process in our discipline, no matter the orientation of the studies, and irrespectively of the qualitative or quantitative nature of research designs.

In light of the above, the scholarly debate sparked by Becker and Guzmán Naranjo's paper is both welcome and enlightening, and the article is likely to have a lasting impact on our discipline. At the same time, some passages of the paper seem to imply that statistical modeling holds a certain methodological superiority over other, say more traditional, approaches to data collection and analysis in linguistic typology. For instance, the authors write: "[t]his suggests that sampling as a form of bias control (phylogenetic or contact) may not be ideal, and that statistical bias control in the form of a phylogenetic regression term and a Gaussian Process are able to represent the dependencies between languages in a sample more accurately". In a similar vein, they add "our comparisons showed that more advanced statistical techniques that can model the phylogenetic and contact relations between languages do pick up more complex patterns in the data than traditional sampling methods". These claims seem to suggest a hierarchy of reliability between typological studies based on sampling and those based on statistical bias control. These observations are, of course, very well-grounded in the analyses presented throughout the paper. However, they may also run the risk of polarizing methodological choices in typological research. To avoid this, it is essential to reiterate that not all typological investigations can, or should, be designed around statistical bias control. However, when applicable, statistical bias control offers a way to expand the empirical basis of typological generaliations and refine them where necessary.

In sum, while advocating for explicit and transparent data management practices in typological research is certainly timely, I believe that embracing and fostering the plurality of research approaches that has characterized linguistic typology since its early days remains vital to the progress of our field.

References

Becker, Laura & Matías Guzmán Naranjo, 2025. Replication and methodological robustness in quantitative typology. *Linguistic Typology*. https://doi.org/10.1515/lingty-2023-0076.

Corbett, Greville. 2005. Suppletion in personal pronouns: Theory versus practice, and the place of reproducibility in typology. *Linguistic Typology* 9(1). 1–23.

Haspelmath, Martin & Sven Siegmund. 2006. Simulating the replication of some of Greenberg's word order generalizations. *Linguistic Typology* 10(1). 74–82.

Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: Chicago University Press.