Commentary

Kenny Smith*

An experimentalist's perspective on replicability in typology: commentary on "Replication and methodological robustness in quantitative typology" by Becker and Guzmán Naranjo

I use experimental methods to test hypotheses about cognitive and interactional mechanisms underlying linguistic universals. My comments pertain to analytic flexibility as experienced in experimental psychology, and the relationship between observational and experimental work in uncovering and understanding language universals.

Experimental psychology has been enjoying a "replication crisis" since at least 2010, when it became clear that many published results are not replicable (see e.g. Open Science Collaboration 2015 for a large-scale replication attempt). One possible cause of this lack of replicability is flexibility in the analytic choices made by individual researchers. Any analysis proceeds through a "garden of forking paths" (Gelman and Loken 2014) where the path chosen might be (explicitly or implicitly) data-dependent and outcome dependent, i.e. directed towards statistically-significant, surprising, publishable results. One collective response to the replication crisis has therefore been to attempt to reduce researchers' flexibility during analysis (e.g. by preregistering a planned analysis prior to collecting data).

Implicit in the garden of forking paths is the possibility that different researchers analysing the same data might follow different pathways and find different results. This has repeatedly been shown to be the case (e.g. Silberzahn et al. 2018, Botvinik-Nezer et al. 2020, Breznau et al. 2022): multiple teams of analysts analysing the same dataset to test the same hypotheses obtain different results,

*Corresponding author: Kenny Smith [kɛnɪ smɪθ], Centre for Language Evolution, School of Philosophy, Psychology & Language Sciences, The University of Edinburgh, Edinburgh, United Kingdom, E-mail: kenny.smith@ed.ac.uk

including finding significant effects in opposite directions. B&GN's finding that they cannot always reproduce the conclusions of existing analyses is therefore entirely expected. Unfortunately there are reasons to suspect these problems might be particularly acute and intractable in typology, more so than in experimental psychology.

First, quantitative typology involves complex data and complex analyses, requiring many decisions on data preprocessing and analysis; this is precisely where we should expect the greatest scope for analytic flexibility, and therefore the greatest potential for divergence and uncertainty. Second, while experimentalists can usually collect more data relatively easily, typologists cannot: their data are time-consuming to obtain (involving painstakingly-obtained expert linguistic knowledge) and inherently limited (there are only so many languages in the world). This might mean that there is unresolvable uncertainty around conclusions that can be drawn from typological datasets.

If one's goal is primarily to document constraints on cross-linguistic variation then this is obviously deeply troubling. However, if the central interest is the cognitive and interactional mechanisms responsible for those constraints – what it is about the way languages are learned, used and transmitted that leads to convergent cultural evolution on recurring constellations of linguistic features (see e.g. Haspelmath 2019, 2021) – then this uncertainty may be less problematic than it first appears, since we should in any case be running controlled experiments to test hypotheses about those mechanisms.

B&GN (Becker and Guzmán Naranjo 2025) refer to experimental approaches briefly in a footnote as "triangulation", "the combination of different empirical approaches to study the same phenomenon in order to test how robust results are across methods and to, ideally, find converging evidence". I think the value of experimental work lies not in providing some additional data from another source, but a fundamentally different KIND of data which allows us to test cognitive and interactional mechanisms hypothesised to be responsible for potential universals. Being observational, no matter how rigorously conducted, analyses of typological data cannot speak to those causal mechanisms. However, the observational data from typology is a rich source of potential hypotheses about mechanisms shaping linguistic systems, which can subsequently be tested in controlled experiments that can go beyond correlation and speak to causality.

This is a standard model in many disciplines, and will be familiar from medical research, where observational studies suggesting an association between a condition and potential risk factors lead to hypotheses about causal mechanisms which are then tested in randomized controlled trials. But the approach is quite general, and there are

already successful examples of this combined approach in typology. For instance, in a series of studies testing learning-based explanations for Greenberg's (1963) observations about word order in the noun phrase, Culbertson and colleagues find that artificial languages with cross-linguistically frequent word orders are easier to learn (e.g. Culbertson et al. 2012; Culbertson and Adger 2014), even for people whose first language violates the relevant universal (Culbertson et al. 2020; Martin et al. 2024). Pleasingly, these studies feature universals in the same domain as B&GN's first case study, where there is the least disagreement between the original and replication analysis. Other work from our group has explored the role of learning and efficient communication in accounting for fundamental design features shared by all human languages (e.g. compositionality, combinatoriality, Zipf's law of abbreviation: Kirby et al. 2015; Verhoef et al. 2014; Kanwal et al. 2017; see K. Smith 2022 for review), as well as hypotheses about mechanisms accounting for recurring diachronic processes such as regularisation (Smith et al. 2023), obligatorification (Fehér et al. 2019), and grammaticalisation (Kapron-King et al. under review, 2025).

My suggestion is therefore that, if we view the work of typologists as part of a joint endeavour with experimentalists, as it needs to be to speak to these causal mechanisms which many of us are interested in, then the pressure on typologists to single-handedly identify the real cross-linguistic universals with high certainty is reduced.

If this joint approach is widely adopted, there will be cases where we need to resolve mismatches between typological and experimental findings. I have encountered two such mismatches in my recent work: one case where I think the hypothesised mechanisms are plausible even if the typology is highly contested, and one where the natural language facts seem to be agreed upon but we cannot support the mechanism experimentally.

The first case relates to the well-known claim that languages spoken in larger, more heterogeneous communities, with more non-native speakers, tend to be morphologically simpler (e.g. Trudgill 2011); there is some quantitive evidence in support of this claim (e.g. Lupyan and Dale 2010; Sinnemäki 2020), although as we might expect in light of B&GN, different measures of morphological and social complexity and different analytic techniques produce different results (Koplenig 2019; Kauhanen et al. 2023; Shcherbakova et al. 2023). A small series of experiments have nevertheless tested some of the proposed mechanisms. In Smith (2024) I used an artificial language learning paradigm to test whether L2-like morphological simplifications made during (imperfect) learning could result in cumulative simplification of complex morphology as a language is transmitted across generations, and found that they could. Other work suggests other mechanisms that could explain the putative

correlation, including e.g. the difficulty of converging on shared conventions with many versus few interlocutors (Raviv et al. 2019). My current impression is that there are probably several plausible mechanisms with at least some experimental support by which population size and proportion of non-native speakers could influence language complexity. If that link is not evident in the cross-linguistic data, it could be that the factors identified in the experiments are outweighed by other factors in the wild, or that the natural language data cannot show the correlation with confidence.

In the second case we are testing mechanisms for unidirectionality in gram-maticalisation (Kapron-King et al. under review, 2025). Concrete concepts (e.g. terms for body parts) tend to become grammatical markers (e.g. adpositions marking spatial relationships) but not vice versa. The intuitive and apparently quite widely-held assumption is that this unidirectionality is due to an inherent asymmetry in associations between these sets of concepts, such that e.g. body-part terms evoke spatial concepts but not the reverse. We have not been able to find evidence for this asymmetry across several semantic extension experiments; while our participants reliably associate body part terms and spatial relationships that are frequently involved in grammaticalisation pathways (as documented in Heine and Kuteva 2002, e.g. "head" and "above"), these associations are quite symmetrical. We have there-fore (reluctantly!) become somewhat sceptical about association-based explanations of unidirectionality, and are exploring reanalysis-based accounts which do not rely on this asymmetry.

In principle, there may eventually be cases where we should collectively reconsider a conclusion from typology based on evidence from experiments (e.g. consistent failure to find evidence for ANY hypothesised mechanism generating the putative universal). This will be painful and any reconsideration might be more likely in practice if the methods used in experimental work are agreed as acceptable tests of mechanistic hypotheses by typologists, which may not currently be the case: I know that the compromises required in constructing controlled experiments can be unpalatable to people who are accustomed to working with natural languages in their full complexity.

To wrap up. B&GN show that well-known analytic flexibility seen in other data-driven disciplines also occurs in typology, introducing an additional and perhaps inescapable degree of uncertainty into conclusions from typological data. That is not disastrous if we view typology as part of a joint endeavour to uncover mechanisms shaping languages, where other methods (specifically, controlled experiments) are anyway required to test the critical mechanisms held to be responsible for candidate linguistic universals. This is the norm in other fields where claims about causal mechanisms are made based on complex and messy observational data.

References

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

Botvinik-Nezer, Rotem et al. 2020. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 582(7810). 84–88.

Breznau, Nate et al. 2022. Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences* 119(44). e2203150119.

Culbertson, Jennifer & David Adger. 2014. Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences* 111(16). 5842–5847.

Culbertson, Jennifer, Julie Franck, Guillaume Braquet, Magda Barrera Navarro & Inbal Arnon. 2020. A learning bias for word order harmony: Evidence from speakers of non-harmonic languages. *Cognition* 204. 104392.

Culbertson, Jennifer, Paul Smolensky & Géraldine Legendre. 2012. Learning biases predict a word order universal. *Cognition* 122. 306–329.

Fehér, Olga, Nikolaus Ritt & Kenny Smith. 2019. Asymmetric accommodation during interaction leads to the regularisation of linguistic variants. *Journal of Memory and Language* 109. 104036.

Gelman, Andrew & Eric Loken. 2014. The statistical crisis in science. *American Scientist* 102. 460.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 73–113. Cambridge, MA: MIT Press.

Haspelmath, Martin. 2019. Can cross-linguistic regularities be explained by constraints on change? In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 1–23. Berlin: Language Science Press.

Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *Journal of Linguistics* 57. 605–633.

Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Kanwal, Jasmeen, Kenny Smith, Jennifer Culbertson & Simon Kirby. 2017. Zipf's Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication. *Cognition* 165. 45–52.

Kapron-King, Anna, Simon Kirby, Graeme Trousdale & Kenny Smith. 2025. No directional preference for grammaticalization in semantic extension game. In *Proceedings of the 46th annual Conference of the Cognitive Science Society*.

Kapron-King, Anna, Simon Kirby, Graeme Trousdale & Kenny Smith. under review. Grammatical unidirectionality is not reflected in individual preferences when performing artificial semantic extension. https://doi.org/10.31219/osf.io/c5zks.

Kauhanen, Henri, Sarah Einhaus & George Walkden. 2023. Language structure is influenced by the proportion of non-native speakers: A reply to Koplenig (2019). *Journal of Language Evolution* 8. 90–101.

Kirby, Simon, Monica Tamariz, Hannah Cornish & Kenny Smith. 2015. Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141. 87–102.

Koplenig, Alexander. 2019. Language structure is influenced by the number of speakers but seemingly not by the proportion of non-native speakers. *Royal Society Open Science* 6. 181274.

Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLoS One* 5. e8559.

Martin, Alexander, David Adger, Klaus Abels, Patrick Kanampiu & Jennifer Culbertson. 2024. A universal cognitive bias in word order: Evidence from speakers whose language goes against it. *Psychological Science* 35. 304–311.

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349. aac4716.

Raviv, Limor, Antje Meyer & Shiri Lev-Ari. 2019. Larger communities create more systematic languages. *Proceedings of the Royal Society B: Biological Sciences* 286. 20191262.

Shcherbakova, Olena, Susanne Maria Michaelis, Hannah J. Haynie, Sam Passmore, Volker Gast, Russell D. Gray, Simon J. Greenhill, Damián E. Blasi & Hedvig Skirgård. 2023. Societies of strangers do not speak less complex languages. *Science Advances* 9. eadf7704.

Silberzahn, Raphael et al. 2018. Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science* 1. 337–356.

Sinnemäki, Kaius. 2020. Linguistic system and sociolinguistic environment as competing factors in linguistic variation: A typological approach. *Journal of Historical Sociolinguistics* 6. 20191010.

Smith, Kenny. 2022. How Language learning and language use create linguistic structure. *Current Directions in Psychological Science* 31. 177–186.

Smith, Kenny. 2024. Simplifications made early in learning can reshape language complexity: An experimental test of the linguistic Niche hypothesis. In Larissa. K. Samuelson, Stefan L. Frank, Mariya Toneva, Allyson Mackey, & Eliot Hazeltinea (eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*, 1346–1352. Cognitive Science Society.

Smith, Kenny, C. Ashton & Helen Sims-Williams. 2023. The relationship between frequency and irregularity in the evolution of linguistic structure: An experimental study. In Micah Goldwater, Florencia K. Anggoro, Brett K. Hayes & Desmond C. Ong (eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*, 851–857. Cognitive Science Society.

Trudgill, Peter. 2011. *Sociolinguistic typology: Social Determinants of linguistic complexity*. Oxford: Oxford University Press.

Verhoef, Tessa, Simon Kirby & Bart de Boer. 2014. Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics* 43. 57–68.