A global phylogenetic test of over 150 putative typological universals

Annemarie Verkerk (Saarland University), **Olena Shcherbakova** (Max Planck Institute for the Science of Human History), **Hannah Haynie** (University of Colorado Boulder), **Russell Gray** (Max Planck Institute for Evolutionary Anthropology), **Simon Greenhill** (Max Planck Institute for the Science of Human History) & **Hedvig Skirgård** (Max Planck Institute for Evolutionary Anthropology)

keywords: universals, correlation, phylogenetics, word order, hierarchies

Research of cross-linguistic universals has thrived since the advent of modern typology (Greenberg 1963/1966) and is still a vibrant topic today (Dunn et al. 2011, Berg 2020, Jäger & Wahle 2021). Over the course of almost six decades, the focus in universal research has shifted from asking 'what are possible human languages', to Bickel's (2007: 239) 'what's where why?', and further to positing that the idea of language universals is a myth (Evans and Levinson 2009). Despite claims to the contrary (responses to Evans and Levinson 2009, Dryer 2011, Velupillai 2012), it may be that there are no absolute universals left that are not ultimately rooted in definitional properties of human languages (i.e. 'all spoken languages have vowels', see also Bickel (2018) on how all universals might be probabilistic).

In this talk, we present further support for the 'no absolute universals' claim by using Bayesian phylogenetic methods and grammatical data from over 2000 languages to test claims about over 150 putative typological universals. The typological data are taken from Grambank (Skirgård et al. submitted); the universals are gathered from Greenberg (1963/1966) and Plank and Filimonova's (2006) massive collection, The Konstanz Universals Archive. We express published universals as expected correlations between pairs of features (such as gender and number; or word order of object and verb and word order of adposition and noun) and investigate whether these features are likely to co-evolve or not. We test for correlation between two binary features using explicit phylogenetic methods and *brms* in *R* (Bürkner 2017, R Core Team 2021). Rather than investigating individual families (as done by Dunn et al. 2011), we make use of global language trees (Jäger 2018, other global trees are in preparation). This allows us to incorporate small families and isolates, as well as quantitatively appreciate the fact that many separate families that are in a certain area probably share a past, even if we cannot currently confirm with the comparative method that they are indeed of one family.

Our analyses reveal that while some typological universals are indeed statistically robust (circa one-third), the majority (two-thirds) are not. In part this may be due to mismatches: universals often have particular formulations that we cannot always match precisely to Grambank features. However, many universals we test simply do not hold against the use of more appropriate statistical testing and increased sample size. We observe an interesting split: word order universals as well as universals that relate to hierarchies ('If a language has gender categories in the noun, it has gender categories in the pronoun', Greenberg 1963/1966) display significant amounts of 'universality', whereas other types of universals rarely do so. We explain this observation in diachronic as well as theoretical terms, considering the evidence our results bring to the status of scales in grammar (see contributions to Bornkessel-Schlesewsky et al. 2014; Christofaro & Zúñiga 2018). Overall, we propose that universality should be seen as a matter of degree and argue for a more nuanced view of the role of historical, cultural and cognitive factors in driving correlations between typological features.

References

Berg, Thomas. 2020. 'Nominal and Pronominal Gender: Putting Greenberg's Universal 43 to the Test'. *STUF - Language Typology and Universals* 73 (4): 525–74.

Bornkessel-Schlesewsky, Ina, Andrej L. Malchukov, and Marc D. Richards, eds. 2014. *Scales and Hierarchies: A Cross-Disciplinary Perspective*. Berlin: De Gruyter Mouton.

Bürkner, Paul-Christian. 2017. 'Brms: An R Package for Bayesian Multilevel Models Using Stan'. *Journal of Statistical Software* 80 (1): 1–28.

Bickel, B. 2007. 'Typology in the 21st Century: Major Current Developments'. *Linguistic Typology* 11 (1): 239–51.

Bickel, Balthasar. (2018). Estimating diachronic biases of typological traits in unknown phylogenies. Paper presented at the 12th Conference of the Association for Linguistic Typology, Canberra.

Cristofaro, Sonia, and Fernando Zúñiga, eds. 2018. *Typological Hierarchies in Synchrony and Diachrony*. Amsterdam: Benjamins.

Dryer, Matthew S. 2011. 'The Evidence for Word Order Correlations'. *Linguistic Typology* 15 (2). https://doi.org/10.1515/lity.2011.024.

Dunn, Michael, Simon J Greenhill, Stephen C Levinson, and Russell D Gray. 2011. 'Evolved Structure of Language Shows Lineage-Specific Trends in Word-Order Universals'. *Nature*, 473 (7345): 79–82.

Evans, Nicholas, and Stephen C. Levinson. 2009. 'The Myth of Language Universals: Language Diversity and Its Importance for Cognitive Science'. *Behavioral and Brain Sciences* 32 (5): 429–48.

Greenberg, Joseph H. 1963/1966. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of grammar*, ed. Joseph H. Greenberg, 2nd edition, 73-113. Cambridge, Mass: MIT Press. (Reprinted in 1966, 40-70.)

Jäger, Gerhard. (2018). Global-scale phylogenetics linguistic inference from lexical resources. *Scientific Data*, 5, 180189.

Jäger, Gerhard, and Johannes Wahle. 2021. 'Phylogenetic Typology'. *ArXiv:2103.10198 [Cs, q-Bio]*, March. http://arxiv.org/abs/2103.10198.

Plank, Frans & Filimonova, Elena. (2006). The Universals Archive: A brief introduction for prospective users. *STUF - Language Typology and Universals*, 53(1), 109-123.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Skirgård, H., Haynie, H. J., Hammarström, H., Blasi, D. E., Collins, J., Latarche, J., Lesage, J., Weber, T., Witzlack-Makarevich, A., Passmore, S., Maurits, L., Dunn, M., Reesink, G., Singer, R., Bowern, C., Epps, P., Hill, J., Vesakoski, O., Robbeets, M., Abbas, K., Auer, D., Bakker, N., Barbos, G., Borges, R., Danielsen, S., Dorenbusch, L., Dorn, E., Elliott, J., Falcone, G., Fischer, J., Ghanggo Ate, Y., Gibson, H., Göbel, H., Goodall, J., Gruner, V., Harvey, A., Hayes, R., Heer, L., Herrera Miranda, R., Hübler, N., Huntington-Rainey, B., Ivani, J., Johns, M., Just, E., Kashima, E., Kipf, C., Klingenberg, J., König, N., Koti, K., Kowalik, R., Krasnoukhova, O., Lindvall, N., Lorenzen, M., Lutzenberger, H., Martins, T., Mata German, C., Meer, S., Montoya Samamé, J., Müller, M., Muradoglu, S., Neely, K., Nickel, J., Norvik, M., Oluoch, C. A., Peacock, J., Pearey , I., Peck, N., Petit, S., Pieper, S., Poblete, M., Prestipino, D., Raabe, L., Raja, A., Reimringer, J., Rey, S., Rizaew, J., Ruppert, E., Salmon, K., Sammet, J., Schembri, R., Schlabbach, L., Schmidt, F., Skilton, A., Smith, W. D., Sousa, H., Sverredal, K., Valle, D., Vera, J., Voß, J., Witte, T., Wu, H., Yam, S., Ye 葉婧婷, J., Yong, M., Yuditha, T., Zariquiey, R., Forkel, R., Evans, N., Levinson, S. C., Haspelmath, M., Greenhill, S. J., Atkinson, Q. D. and Gray, R. D. (submitted) Grambank data reveal global patterns in the structural diversity of the world's languages.

Velupillai, Viveka. 2012. *An Introduction to Linguistic Typology*. Amsterdam: Benjamins.