A commentary: Replication and methods in typology
Ilja A. Seržant (University of Potsdam)

1. Introduction

The target paper by Laura Becker and Matías Guzmán Naranjo "Replication and methodological robustness in quantitative typology" (henceforth B&GN) is an important methodological and substantive contribution to the field. The aim of the paper is replication of data analyses with new methods and exploring how potentially more appropriate methods of statistical data analysis may affect the results reported in earlier studies and thus contribute to the replicability and robustness of the methods applied in typology. It advocates a more fine-grained way to control the two main typological biases: the *areal* and the *genealogical* bias. Instead of more coarse-grained methods for the genealogical control which are met by various types of balanced sampling (such as one language per genus, etc.), it suggests controlling for it not on the data collection level but on the level of data analysis by means of *phylogenetic regression* (cf. de Villemereuil & Nakagawa 2014) which accommodates relatedness between any two languages in a non-categorical, gradual way. It thus takes into account the degree of relatedness as well. Likewise, the authors offer a more fine-grained method to control for the areal bias in the response variable(s) than the traditional categorical method via random effects and/or already on the level of sampling. Instead, the authors put forward a method that takes into account the geographical distance (acquired from the GPS coordinates in Glottolog). Moreover, the difference in distance is not taken as a linear function but rather as a Gaussian process which allows for the areal effect to decrease faster than the distance between two languages, given the intuition that after some distance threshold the differences in distances do not play a role anymore. Introducing these controls only at the level of statistical modelling allows working with unbalanced samples and does not force the researcher to leave out datapoints solely for balancing – an approach that is heavily needed in typology because, for many linguistic phenomena, balanced samples are often difficult or even impossible to produce.

      The focus of B&GN is also on replicating data analyses with the new methods in order to identify those findings that will reveal themselves as robust across different ways of statistical testing as well as to make a methodological contribution to the methods and replicability in typology. The target studies are Dryer (2018), Seržant (2021a), Shcherbakova et al. (2023) and Berg (2020) (in the appendix). The authors explicitly do not alter or question the initial datasets of these studies, their focus being solely on replicating the original claims. While the methods used by the authors mainly support the claims made in Dryer (2018), not all the results produced in Seržant (2021a) could be corroborated. Even less results of Shcherbakova et al. (2023) found support. The B&NG even found positive evidence for some contrary claims here.

2. Seržant (2021a)

Here, the authors scrutinize the part in which Seržant (2021a) examines the areal distribution of the decay of paradigms of person-number indexes across six language families in Eurasia as well as the specific position of Slavic languages here.[1] The dataset contains 150 languages and the reconstruction of the respective paradigms into six proto-languages as found in the historical-comparative literature (Seržant 2020).

---

[1] The decay of a person-number index paradigm is measured cumulatively on the basis of three factors: the reduction of the lengths of the indexes as compared to the respective proto-language, the number of syncretic slots in the paradigm and the number of zeros.

Seržant claims that two decay hotbeds can be identified: one in Northwestern Europe with languages such as English, Scandinavian or French and the other one in India, see Figure
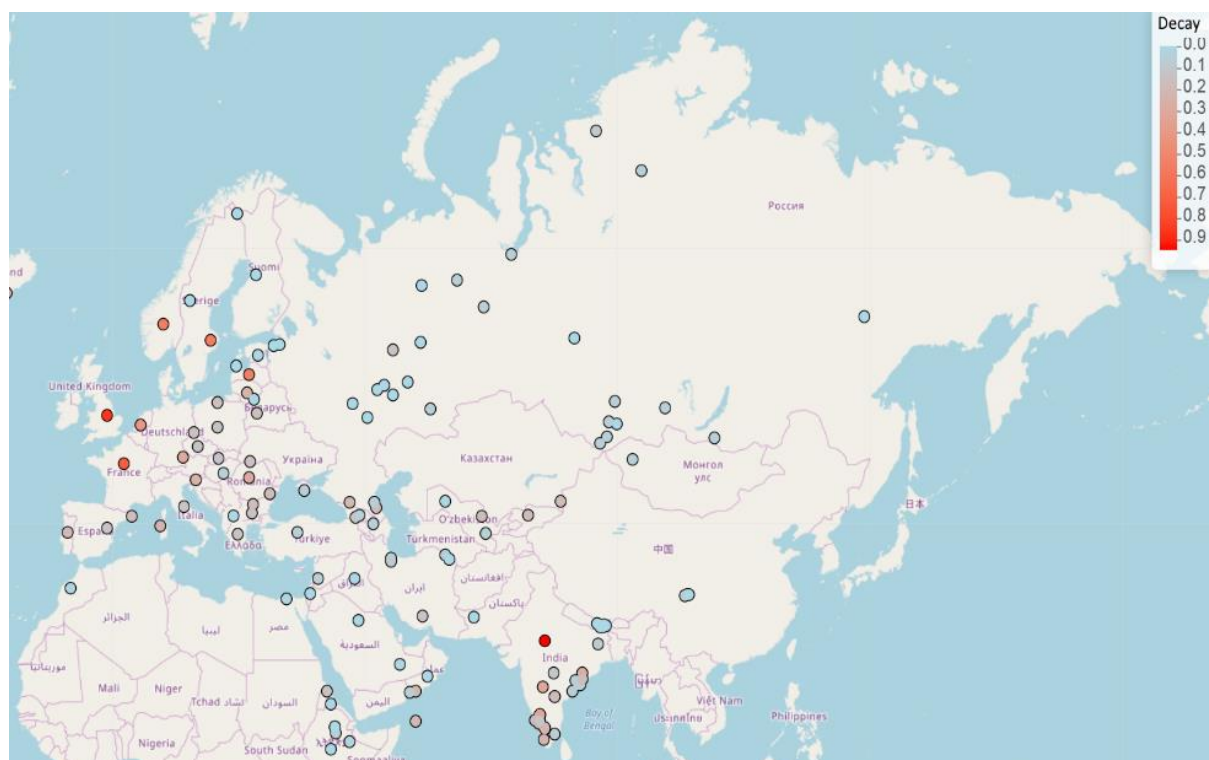
1.
Figure 1. The degree of decay factor across different languages of Eurasia

Moreover, I have also claimed that there is an East-West cline of decay in which the East is conservative, showing nearly no indication of decay while the West has languages with a quite high degree of decay. Slavic languages, in turn, inhabit the Transitional Zone. Table 1 summarizes the average decay factors for the relevant area:

Northwestern Europe	Ø 0.61	Germanic	Ø 0.49
		French	0.72
Transitional area	Ø 0.12	Greek	< 0.14
		Romanian	< 0.3
		Slavic	Ø < 0.15
Northeastern Eurasia	Ø 0.05	Turkic	Ø < 0.07
		Uralic	Ø 0.02

Table 1. Decay factors across the three areas (Ø – averaged across languages) (from Seržant 2021a: 74)

This study is somewhat different from a typical typological study – like Dryer's (2018) or Shcherbakova et al.'s (2023) – since it is not about universalist generalizations in which the areal and genealogical effects are just confounds to rule out. It is rather a study in areal and historical linguistics in which, reversely, universals are confounds. It asks the question why a specific grammatical category tends to disappear in some languages but shows no signs of decay in other languages and aims at establishing areal hotspots of decay.

While B&GN confirm the original finding that there are two hotbeds of decay (Northwestern Europe and India), they could not replicate the following results. First (i), B&GN find that the areal signal once the genealogical effect is taken into account is negligible and,

hence, there is no East-West cline of decay factors, while, second (ii), B&GN find a strong genealogical signal and suggest that the decay can be a product of inheritance alone, i.e., there is something specific to a family. Finally (iii), B&GN suggest that the distance to the hotbed explains the low decay factor.

3. Discussion

B&NG is a very important methodological contribution. They offer new and more fine-grained methods to testing areal and genealogical pressures in the data. Statistical rigor and robustness to different ways of statistical modelling diminish uncertainty in typological conclusions.

I am sympathetic with the scepticism of B&NG which strives at more cautious conclusions along Ockham's razor. In fact, I agree with the conclusions made in B&NG and appreciate the more advanced statistical modelling they use. The difference between Seržant (2021a) and B&GN is more in how statistical results are interpreted and which interpretational options given the statistical results are chosen and, accordingly, in what is actually modelled and explored.

First, there is not much of a controversy between B&NG's suggestion (iii) that the distance to the hotbed determines the decay factor and the original claim that there is a West-East cline to that hotbed. Saying that only the distance to the hotbed matters with B&NG is probably just a more accurate way of claiming a West-East cline. It takes into account the fact that the second hotbed is in India, i.e., in the very East, as well as the fact the cline is not linear.

B&GN suggest that (iii) the decay factor is only dependent on the distance from the hotbed and that Slavic languages have retained person-number indexes so faithfully from Proto-Indo-European due to their mere distance from the Northwestern hotbed and, by Ockham razor, not additionally by the contact with the conservative languages in the East such as Turkic or Uralic languages which faithfully preserve their indexes from the respective proto-languages. Indeed, in contrast to the development of new features, it is methodologically much more difficult to prove retention of inherited features due to language contact against the null hypothesis of retention through drift that is not externally conditioned. One would need a universal decay baseline here to compare the specific Slavic decay with. What is more, in another study, I have argued that the retention of indexes – once these have emerged – is universally preferred and languages do not tend to lose these (Seržant 2021b). Any retention of universally preferred features is even more likely to be independent of language contact than retention of non-universal features.

Having said this, the specific decay factors of Slavic languages remain unexplained by B&GN. While they are right in that the distance to the hotbed explains that the Slavic decay factor is low, the specific relative value of the Slavic decay is not really explained. In contrast to the very conservative languages in the East (all Turkic and Uralic languages), Slavic languages do show a minor degree of decay and pattern thus with other languages of the Transitional zone such as Greek which also exhibits some minor degree of decay (Table 1). Moreover, decay values of Slavic languages are closer to Uralic and Turkic languages as well as to the languages of the Transitional Zone than they are to the languages of Northwestern Europe. At the same time, there seems to be no significant difference in geographical distances: the Northwestern European languages are geographically as close to West and South Slavic (compare German with Czech, Slovak or Slovene) as are some Uralic and Turkic languages to East and South Slavic. Yet, I also claim that what matters is not only the geographical proximity per se but also the specific contact configuration. Historically, Slavic languages had more intensive contacts with the languages of the Transitional Zone (e.g. Greek, Romanian) as well as with Uralic and Turkic languages than with the languages of the hotbed Northwestern Zone such as German, Scandinavian, French or English. Interestingly, the intensity of these contacts

is correlated with the decay scores better than the pure geographical proximity – a proxy for the areal effect in B&GN. Thus, Slavic decay is closer to the Transitional Zone, Uralic and Turkic and more distant from the Northwestern Zone. B&GN model the decay factor as non-linearly dependent from the distance from the hotbed, i.e., the decay factor mathematically decreases disproportionally faster than the increasing distance from the hotbed. This makes sense. However, this is not an explanation but just a descriptive model. The question still remains why the areal pressure radiates non-linearly. Contact configuration would be precisely the explanation here for the non-linear dependency from the pure distance. In turn, geographical proximity is only an imprecise proxy for contact intensity.

Third (iii), should a strong genealogical signal always be taken literally to mean that inheritance alone is sufficient, i.e., that there would be something specific to a family? B&NG offer an excellent way of estimating the genealogical effect. They find that the genealogical effect is the only factor minimally needed to account for the decay while the areal and, thus, the contact effect does not play out in their proximity-based areal modelling once the genealogical effect is taken into account. While controlling for the genealogical effect is undoubtedly important for typological research, especially for the research on universals, it is not entirely clear what its role would be in exploring the decay of a grammatical category (or an emergence of a category for that matter). The genealogical effect taken literally would mean that a particular family would be genealogically (but not areally) predisposed to lose a specific category. But how such predisposition for a decay radiating via genealogical nodes and clades would work given that there were no signs of decay in the proto-language, i.e., in Proto-Indo-European? And, vice versa, how a genealogically driven predisposition for retention in Uralic or Turkic would work? Such mechanism might be the degree of migration of the speakers of a family: a heavily migrating family such as Indo-European would of course be likely to experience more loss due to intensive contact effects (cf. *simplification* in Trudgill 2011) and a more isolating family would be more prone to retention. However, crucially, this explanation is not inheritance-driven but amounts to language contact and thus to areal pressure.

More generally, it seems important to distinguish between areal, genealogical and universal *pressures* and the areal or genealogical *signals*. Pressures are the specific mechanisms that affect and shape languages in the respective way. Yet, signals are just statistical signals that are found if the data looked at is structured accordingly areally, i.e. along geographical proximity, or genealogically, i.e. along the distance in tree nodes. Signals are intersecting proxies for the pressures.

I suggest that the effect of the genealogical pressure might generally be overestimated while the universal and areal pressures are the strongest pressures shaping languages. In any event, the genealogical signal artificially downplays the effects of the universal and areal pressures in the data for the following reasons (not accommodated in the current typological testing).

First, genealogical signal partly covers geographical proximity and thus the areal pressure. B&NG model genealogical proximity not as a categorical variable but rather as a gradual variable. This is a much more accurate way of establishing biases in the data since, at least impressionistically, remotely related languages such as German and Hindi will hardly maintain the same inherited similarities in the grammar while this is much more likely in two more closely related languages such as German and Dutch. Thus, the degree of relatedness of two languages often correlates with their geographical proximity in both ways, compare e.g. Dutch and German vs. Dutch and Hindi. Koile et al. (2021) show that geographical proximity is a good proxy for the genealogical proximity, on the basis of the Andic subfamily of Daghestanian,[2] and, thus, the reverse is also true. Samples which sample many languages per

---

[2] "For all phylogenies suggested in the literature on Andic, we find that the correlation with geographic distances is above random, indicating that geographic distance is, in this case, a viable predictor of linguistic differentiation." (Koile et al. 2021: 4).

family like Seržant (2020) might be even more vulnerable for overestimation of the genealogical signal over the areal signal, while the phenomenon being essentially an areal one.

Secondly, genealogical pressure is logically incompatible with innovations. Taken literally, genealogical pressure for an innovation would have meant that there might be some inherited predestination in a family to a loss or to emergence of a phenomenon, - a scenario that is unlikely. Genealogical pressure is only about inheritance, i.e., only about retention of traits. In case of innovations, significant genealogical signals only reflect the fact that the spread of innovations is likely to more strongly affect genealogically closely related languages. This is because structural similarity prior to contact facilitates diffusibility of patterns (Haig 2001; Matras 2007: 34; Epps et al. 2014). Genealogy may, therefore, channel contact-induced diffusion of features. For example, genealogy has been shown to channel sound change across dialects (Heeringa & Nerbonne 2001). What is more, not only shared innovations are likely to be found in closely related languages but also shared inheritance itself does not right away exclude an effect of language contact and, thus, of an areal pressure. Language contact may as well exercise a conserving pressure for some inherited traits to be retained (see, *inter alia*, Seržant 2021; Seržant et al. 2022). Such effects would statistically increase the genealogical signal in modelling but would essentially be due to areal pressures.

Since languages continuously change (cf. Hopper 1987), logically, inheritance is expected to dramatically diminish when moving up the genealogical tree, unless the inherited phenomenon is universally preferred or there is an areal, language-contact-based pressure for it to be retained. But then, again, in these cases, the genealogical signal is due to universal and/or areal pressures and is not due to a genuinely genealogical pressure. Skirgård et al. (2023: 3) claim that the genealogical pressure is "consistently greater than that of space". However, in view of what has been said above, it seems that more fine-grained testing for this claim would be needed. The genealogical signal is boosted by the areal pressure with geographically proximate languages like Dutch and German. Moreover, Skirgård et al. (2023: 3) did not seem to have controlled for the universal pressures of specific features, although it is very likely that universally preferred features would show high stability across genealogical nodes and thus artificially boost the genealogical signal as well.

B&GN model areal effects as Gaussian process and thus accommodate the intuition that contact effects may only be strong in close proximity while these effects decrease rapidly with an increasing distance. I am wondering whether areal effects and language contact are not two different – albeit related – phenomena. While transfer of phenomena over language contact may happen only between two (or more) immediately neighboring languages and are thus limited by distance, areal effects represent accumulation effects emerging from an intricate series of multiple contact effects. One may thus wonder how do we know that areal effects expand linearly or non-linearly as a Gaussian process since at least theoretically such mediated expansion of features might radiate quite far away, even across an entire macroarea.

I conclude that the genealogical and areal signals in the data are not directly translatable into the respective pressures. For example, the genealogical pressure may be estimated if geographical proximity as well as universal pressures are controlled for. Genealogical pressure is logically very unlikely to affect innovations (emergence and loss).

On a more general note, replicability becomes a more difficult issue when interpretations of statistical outcomes come in. And, as a note of caution, non-replicability does not only cast doubts on the results but, alternatively, also on the appropriateness of the modelling chosen for the given research question.

References

Berg, Thomas. 2020. Nominal and pronominal gender: Putting Greenberg's Universal 43 to the test. *Language Typology and Universals* 73(4). 525–574.

de Villemereuil, Pierre & Nakagawa, Shinichi. 2014. General quantitative genetic methods for comparative biology. In Garamszegi, László Zsolt (ed.), *Modern phylogenetic comparative methods and their application in evolutionary biology*, 287–303. Berlin: Springer.

Dryer, Matthew. 2018. On the order of demonstrative, numeral, adjective, and noun. *Language* 94(4). 798–833.

Epps, Patience, John Huehnergard and Na'ama Pat-El. 2013. Introduction. Contact Among Genetically Related Languages. *Journal of Language Contact* 6. 209–219.

Haig, Geoffrey. 2001. Linguistic diffusion in present-day east Anatolia: From top to bottom. In: Alexandra Y. Aikhenvald & Robert M.W. Dixon (eds.), *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. 195-224. Oxford: Oxford University Press.

Heeringa Wilbert & John Nerbonne. 2001. Dialect areas and dialect continua. *Language Variation and Change* 13(3). 375–400.

Hopper, Paul. 1987. Emergent Grammar. *Berkeley Linguistic Society* 13. 139-157.

Koile, Ezequiel, Ilia Chechuro, George Moroz, Michael Danie. 2021. Geography and language divergence: The case of Andic languages. *PLoS ONE* 17(5). e0265460.

Shcherbakova, Olena, Volker Gast, Damián Blasi, Hedvig Skirgård, Russell Gray and Simon Greenhill. 2023. A quantitative global test of the complexity trade-off hypothesis: The case of nominal and verbal grammatical marking. *Linguistics Vanguard* 9(s1). 155–167.

Seržant, Ilja A. 2020. Dataset for the Paper 'Slavic Morphosyntax is Primarily Determined by its Geographic Location and Contact Configuration', Scando-Slavica [Data set]. *Zenodo.* http://doi.org/10.5281/zenodo.4277593

Seržant, Ilja. 2021a. Slavic morphosyntax is primarily determined by its geographic location and contact configuration. *Scando-Slavica* 67(1). 65–90.

Seržant, Ilja A. 2021b. Cyclic changes in verbal person-number indexes are unlikely. *Folia Linguistica Historica* 42(1). 49-86.

Seržant, Ilja A., Björn Wiemer, Eleni Bužarovska, Martina Ivanová, Maxim Makartsev, Stefan Savić, Dmitri Sitchinava, Karolína Skwarska, Mladen Uhlik. 2022. Areal and diachronic trends in argument flagging across Slavic. In: Eystein Dahl (ed.), *Alignment and Alignment Change in the Indo-European Family*. Oxford: OUP.

Trudgill, Peter. 2011. *Sociolinguistic typology. Social determinants of linguistic complexity.* Oxford: OUP.