

REVISITING THE UNIFORMITARIAN HYPOTHESIS: CAN WE DETECT RECENT CHANGES IN THE TYPOLOGICAL FREQUENCIES OF SPEECH SOUNDS?

EITAN GROSSMAN^{*1}, ELAD EISEN¹, DMITRY NIKOLAEV^{1,2}, and STEVEN MORAN^{*3}

^{*}Corresponding Authors: eitan.grossman@mail.huji.ac.il, steven.moran@uzh.ch

¹Department of Linguistics, Hebrew University of Jerusalem (HUJI), Jerusalem, Israel

²Department of Linguistics, University of Stockholm, Stockholm, Sweden

³Department of Comparative Linguistics, University of Zurich, Zurich, Switzerland

The Uniformitarian Hypothesis, which in essence stresses the time-independent unity of human languages, has been interpreted in a variety of ways, e.g., as a constraint on language change or as a constraint on synchronic distributions (Janda & Joseph, 2003). One version of the Uniformitarian Hypothesis, formulated explicitly as ‘[H]uman languages have always been pretty much the same in terms of the typological distribution of the units that compose them’ (Newmeyer, 2002), is often assumed by theoretical linguists in order to infer universal (and time-independent) properties of human language directly from present-day distributions. For example, linguists have proposed cognitive explanations of word order distributions that crucially rely on Greenbergian word order typologies, e.g., Hawkins (1983). However, the possibility of inferring causes directly from present-day distributions has been called into question (Dryer, 1989; Nichols, 1992; Maslova, 2000).

Here we ask to what extent are the current cross-linguistic distributions of phonological segments the result of events of recent human history? Specifically, are cross-linguistically frequent speech sounds frequent because of language contact? The answer to this question bears on language evolution because present-day distributions may be, at least in part, due to the result of relatively recent historical events, such as colonization and globalization.

Our study is based on the comparison of three large-scale cross-linguistic databases, including BDPROTO (representing past states of human languages) (Marsico, Flavier, Verkerk, & Moran, 2018; Moran, Grossman, & Verkerk, To appear), PHOIBLE (representing contemporary distributions) (Moran & McCloy, 2019), and SEGBO (Grossman, Eisen, Nikolaev, & Moran, Submitted). SEGBO is a new database of phonological segment borrowing, which documents more than 1600 borrowed contrastive segments from 500+ borrowing languages, from 100+ families and 220+ donor languages.

To obtain a genealogically-balanced sample and robust estimates, we sample one language per phylum from both datasets 10k times and compute the relative frequencies of different phonemes in the two bootstrap samples (PHOIBLE and BDPROTO). These relative frequencies are presented in Fig. 1 for a subset of segments. To test if these shifts in frequencies are correlated with segment borrowing in general, we regressed the difference between means of bootstrap distributions of consonant frequencies on their borrowing frequency. The results are statistically significant ($p = 0.017$), pointing towards a possible role that segment borrowing has played in the evolution of segment inventories.

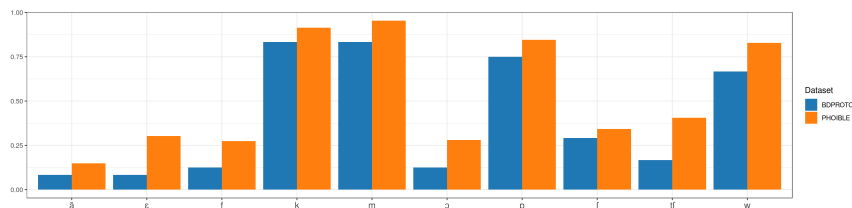


Figure 1. Mean bootstrap frequencies of a sample of segments in BDPROTO and PHOIBLE

Descriptively, we find that at a global level, some sounds are likely the result of relatively recent borrowing events, e.g., as reported by Blasi et al. (2019). In particular, the labiodental /f/ is borrowed far more frequently than any other segment (nearly 10% of the observations in the database), a fact that might point to an even later date of diffusion than previously supposed. However, different macro-areas show different patterns of borrowing. For example, languages of the Americas mostly borrowed the basic voiced stops (/b d g/), while Eurasian and Papunesian languages mostly borrowed affricates and fricatives. And African languages borrowed mostly sibilant fricatives and the voiceless bilabial stop /p/.

Each of these area-specific patterns points to a higher degree of areal specificity in pre-contact areas. Another finding is that a handful of colonial languages, which expanded in the recent past, were the major contributors of borrowed segments. Lastly, the differences in the frequencies of segments in BDPROTO and PHOIBLE correlates with the frequency of borrowing in SEGBO. Taken together, our findings point to substantive and rapid evolutionary changes in phonological segment inventories in the recent past. These findings are also relevant for linguistic theory, as they provide further evidence that present-day distributions cannot be taken as direct evidence for the naturalness of sound systems.

Acknowledgements

We thank students and colleagues at the HUJI, the University of Helsinki and the Higher School of Economics; Erich Round for a dataset on Australian languages; and Einav Levanon for his Papunesian dataset. Thanks to Swiss National Science Foundation grant/award number PCEFPI_186841 (Steven Moran, PI).

References

- Blasi, D. E., Moran, S., Moisić, S. R., Widmer, P., Dediu, D., & Bickel, B. (2019). Human sound systems are shaped by post-neolithic changes in bite configuration. *Science*, 363(6432), eaav3218.
- Dryer, M. S. (1989). Large linguistic areas and language sampling. *Studies in Language*, 13(2), 257–292.
- Grossman, E., Eisen, E., Nikolaev, D., & Moran, S. (Submitted). SegBo: A database of borrowed sounds in the world's languages.
- Hawkins, J. (1983). *Word order universals*. Academic Press.
- Janda, R., & Joseph, B. (2003). On language, change, and language change or, of history, linguistics, and historical linguistics. In B. Joseph & R. Janda (Eds.), *Handbook of historical linguistics* (pp. 3–180). Malden: Blackwell.
- Marsico, E., Flavier, S., Verkerk, A., & Moran, S. (2018). BDPROTO: A database of phonological inventories from ancient and reconstructed languages. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*.
- Maslova, E. (2000). A dynamic approach to the verification of distributional universals. *Linguistic Typology*, 4, 307–333.
- Moran, S., Grossman, E., & Verkerk, A. (To appear). Investigating diachronic trends in phonological inventories using BDPROTO. *Language Resources & Evaluation*.
- Moran, S., & McCloy, D. (Eds.). (2019). *Phoible 2.0*. Jena: Max Planck Institute for the Science of Human History.
- Newmeyer, F. J. (2002). Uniformitarian assumptions and language evolution research. In A. Wray (Ed.), *The transition to language* (pp. 359–375). Oxford: Oxford University Press.
- Nichols, J. (1992). *Linguistic diversity in space and time*. University of Chicago Press.