Creating a Baseline to Evaluate Correlations Between Language and Environment

JCoLE 2022: Machine Learning and the Evolution of Language

Abstract

Previous studies have proposed a correlation between environmental features such as elevation, humidity, and temperature and language features such as ejective consonants, complex tone systems, and consonant-vowel ratio (Everett, 2013; Everett et al., 2015; Everett, 2017), though the findings are contested (Urban and Moran, 2021; Roberts, 2018). Through a series of statistical analyses, we echo the concern that the correlation is not robust and could even be coincidental. After taking into account the random effects of language family and region with mixed-effects models (Bates, 2007), we failed to replicate two of the three correlations previously proposed to be significant; the only correlation that was replicated was the case of high humidity and the presence of tone (Estimate = 176.22, SE = 58.44, $p < 0.01$). Furthermore, we found significant correlations between environmental and linguistic features where the latter is highly unlikely to be influenced by the former, such as adposition-noun order $\sim$ temperature (Estimate = -0.002, SE = 0.0005, $p < 0.01$) and the existence of numeral classifiers $\sim$ elevation (Estimate = 0.98, SE = 0.29, $p < 0.001$).

The agent-based model provides support for a more nuanced view of these relationships, where the language-environment effect may be modulated by other factors (Janssen and Dediu, 2018), such as diet (Blasi et al., 2019), community size (Raviv et al., 2019), articulator anatomy and genetic differences (Dediu and Moisik, 2019; Dediu et al., 2019, 2021; Butcher, 2018; Everett and Chen, 2021), interaction and cultural evolution (Nölle et al., 2020), language contact (Moran et al., 2021), frequency (Macklin-Cordes and Round, 2020), and other physical and social factors (Bentz et al., 2018).

To do this, we will find the correlation coefficients between the three environmental features mentioned above and a variety of language features, such as presence of certain sounds (e.g. lateral consonants) or relative order of verb and object. We will identify the distribution of these correlation coefficients, which will establish a baseline to compare the tone-humidity correlation to and allow us to test whether the correlation is significantly stronger than chance.

Keywords: Quantitative typology, language evolution, environment, mixed effects models

References

Bates, D. (2007). Linear mixed model implementation in lme4. *Manuscript, University of Wisconsin*, 15.

Bentz, C., Dediu, D., Verkerk, A., and Jäger, G. (2018). The evolution of language families is shaped by the environment beyond neutral drift. *Nature Human Behaviour*, 2(11):816–821.

Blasi, D. E., Moran, S., Moisik, S. R., Widmer, P., Dediu, D., and Bickel, B. (2019). Human sound systems are shaped by post-neolithic changes in bite configuration. *Science*, 363(6432):eaav3218.

Butcher, A. R. (2018). The special nature of australian phonologies: Why auditory constraints on human language sound systems are not universal. In *Proceedings of Meetings on Acoustics 176ASA*, volume 35, page 060004. Acoustical Society of America.

Dediu, D., Janssen, R., and Moisik, S. R. (2019). Weak biases emerging from vocal tract anatomy shape the repeated transmission of vowels. *Nature Human Behaviour*, 3(10):1107–1115.

Dediu, D. and Moisik, S. R. (2019). Biology matters: Variation in vocal tract anatomy and language. *Selected papers on theoretical and applied linguistics*, 23:19–33.

Dediu, D., Moisik, S. R., Baetsen, W., Bosman, A. M., and Waters-Rist, A. L. (2021). The vocal tract as a time machine: inferences about past speech and language from the anatomy of the speech organs. *Philosophical Transactions of the Royal Society B*, 376(1824):20200192.

Everett, C. (2013). Evidence for direct geographic influences on linguistic sounds: The case of ejectives. *PloS one*, 8(6):e65275.

Everett, C. (2017). Languages in drier climates use fewer vowels. *Frontiers in Psychology*, 8:1285.

Everett, C., Blasi, D. E., and Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences*, 112(5):1322–1327.

Everett, C. and Chen, S. (2021). Speech adapts to differences in dentition within and across populations. *Scientific reports*, 11(1):1–10.

Janssen, R. and Dediu, D. (2018). Genetic biases affecting language: What do computer models and experimental approaches suggest? In *Language, Cognition and Computational Models*, pages 256–288. Cambridge University Press.

Macklin-Cordes, J. L. and Round, E. R. (2020). Re-evaluating phoneme frequencies. *Frontiers in psychology*, page 3181.

Moran, S., Lester, N. A., and Grossman, E. (2021). Inferring recent evolutionary changes in speech sounds. *Philosophical Transactions of the Royal Society B*, 376(1824):20200198.

Nölle, J., Fusaroli, R., Mills, G. J., and Tylén, K. (2020). Language as shaped by the environment: linguistic construal in a collaborative spatial task. *Palgrave Communications*, 6(1):1–10.

Raviv, L., Meyer, A., and Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B*, 286(1907):20191262.

Roberts, S. G. (2018). Robust, causal, and incremental approaches to investigating linguistic adaptation. *Frontiers in psychology*, 9:166.

Urban, M. and Moran, S. (2021). Altitude and the distributional typology of language structure: Ejectives and beyond. *Plos one*, 16(2):e0245522.