

## The relation between European colonialism and linguistic diversity

Andreas Baumann<sup>\*1</sup>, Bernd Lenzner<sup>2</sup>, Hannes Fellner<sup>1</sup>, and Franz Essl<sup>2</sup>

<sup>\*</sup>Corresponding Author: andreas.baumann@univie.ac.at

<sup>1</sup>Faculty of Philological and Cultural Studies, University of Vienna, Vienna, Austria

<sup>2</sup>Department of Botany and Biodiversity Research, University of Vienna,  
Vienna, Austria

European colonialism was shown to have entailed a global loss of biodiversity (Crosby, 2004; Yang et al., 2021; Lenzner et al., 2022). Similar effects of colonialism were discussed in the linguistic literature as well (Simons & Lewis, 2013). On the linguistic level, however, the picture seems to be complicated and clearly multicausal (Nettle & Romaine 2000). Most prominently, Mufwene (2002) has suggested colonialism to have differential effects on the linguistic ecosystem, conditioned by the intensity of colonialism in a region. The goal of this paper is to quantitatively examine to what extent the duration under colonial rule in the history of a country is associated with its present status regarding linguistic diversity. In doing so, we examine different operationalizations of linguistic diversity.

Several data resources were combined in this study in order to derive country level measures of colonialism and linguistic diversity. For each country, we used colonization beginning and end dates from COLDAT (Becker, 2019) to estimate ‘colonization duration’.<sup>1</sup> The global distribution of colonization time is displayed in Fig. 1. Note that COLDAT is restricted to European empires.

Four ways of measuring linguistic diversity were employed. First, and most straight-forward, we assessed the ‘number of languages’ currently spoken in each country based on Ethnologue (excluding extinct languages; log-transformed). Second, we computed, for each country, the ‘index of linguistic diversity’ as introduced by Harmon and Loh (2010). It is computed as one minus the average normalized endangerment level in that country. Endangerment was assessed by means of contemporary EGIDS scores in Ethnologue. Third, we computed ‘glottogenetic diversity’ as the entropy of

---

<sup>1</sup> In the case of multiple overlapping colonizers, we only considered the most extreme dates.

language families represented in a country. Glottogenetic information was taken from WALS (Dryer & Haspelmath, 2013). Fourth, we assessed the average ‘structural distance’ between languages in a country.<sup>2</sup>

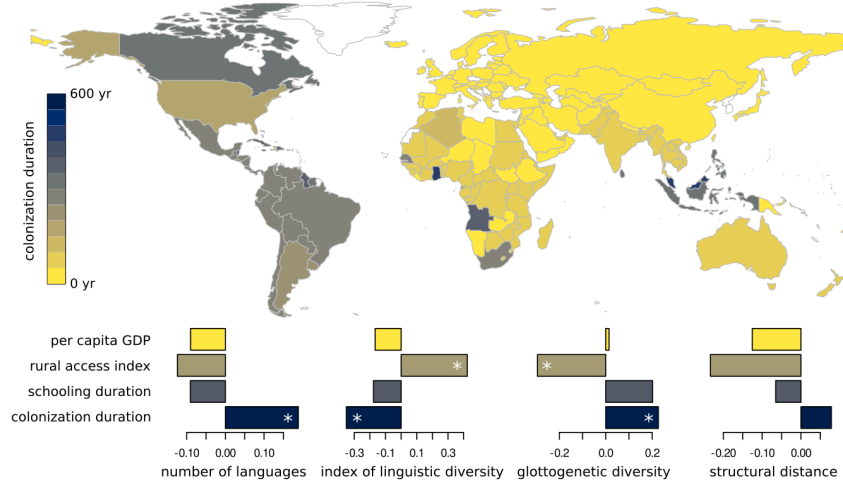


Figure 1. Top: Global distribution of colonization duration. Bottom: standardized coefficients of the linear models (\* denoting significant effects at a 95% confidence level).

We computed four generalized linear models (GLM, Poisson and quasi-binomial, resp.), one for each measure of linguistic diversity depending on country size (area). In a next step, we computed linear models of the residuals of the four GLMs, featuring colonization duration and a selection of socio-economic covariates (schooling duration, rural access index, per capita GDP; World Bank) that were shown to be relevant to linguistic diversity (Bromham et al., 2022) as predictors (checking for collinearity).

The analysis reveals that colonization duration has differential effects on the four measures of linguistic diversity in the models. More specifically, colonization is negatively related with the linguistic diversity index, but positively with glottogenetic diversity and the number of languages. There is no robust effect on structural distance. One, as we think plausible, interpretation of the results is that while colonization promotes the in-take of genetically distant languages and creolization (Blasi et al. 2017), thereby also increasing glottogenetic diversity, colonization and the implementation of a dominant *lingua franca* has simultaneously lead to an increase in the endangerment of lesser supported ambient languages.

<sup>2</sup> For this, we first computed for each language pair the fraction of non-overlapping linguistic features in Grambank (Skirgård et al. 2023) to obtain pairwise distances. Structural distance was then computed as the mean of all pairwise distances in a country, weighted by the number of features in pairwise comparisons.

## References

- Becker, B. (2019), Introducing COLDAT: The Colonial Dates Dataset, *SOCIUM/SFB1342 Working Paper Series*, 02/2019.
- Blasi, D. E., Michaelis, S. M., & Haspelmath, M. (2017). Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour*, 1(10), 723-729.
- Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., & Hua, X. (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution*, 6(2), 163-173.
- Dryer, M. S. & Haspelmath, M. (eds.) 2013. *WALS Online (v2020.3)* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7385533>
- Crosby, A. W. (2004). *Ecological imperialism: the biological expansion of Europe, 900-1900*. Cambridge University Press.
- Harmon, D., & Loh, J. (2010). The index of linguistic diversity: A new quantitative measure of trends in the status of the world's languages. *Language Documentation & Conservation*, 4, 97-151.
- Lenzner, B., Latombe, G., Schertler, A., Seebens, H., Yang, Q., Winter, M., Weigelt, P., van Kleunen, M., Pyšek, P., Pergl, J., Kreft, H., Dawson, W., Dullinger, S., & Essl, F. (2022). Naturalized alien floras still carry the legacy of European colonialism. *Nature Ecology & Evolution*, 6(11), 1723-1732.
- Mufwene, S. (2002). Colonisation, globalisation, and the future of languages in the twenty-first century. *International Journal on Multicultural Societies*, 4(2), 162-193.
- Nettle, D., & Romaine, S. (2000). *Vanishing voices: The extinction of the world's languages*. Oxford: Oxford University Press.
- Simons, G. F., & Lewis, M. (2013). The world's languages in crisis. Responses to language endangerment: In honor of Mickey Noonan. *New directions in language documentation and language revitalization*, 3-20.
- Skirgård, H., Haynie, H. J., Blasi, D. E., Hammarström, H., Collins, J., Latarche, J. J., Lesage, J., Weber, T., Witzlack-Makarevich, A., Passmore, S., Chira, A., Maurits, L., Dinnage, R., Dunn, M., Reesink, G., Singer, R., Bowern, C., Epps, P., Hill, J., Vesakoski, O., ... Gray, R. D. (2023). Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances*, 9(16), eadg6175.
- Yang, Q., Weigelt, P., Fristoe, T. S., Zhang, Z., Kreft, H., Stein, A., ... & van Kleunen, M. (2021). The global loss of floristic uniqueness. *Nature Communications*, 12(1), 7290.