

Linguistic Diversity in the Digital Age

Exploring the Effect of Digital Literacy on Languages

Katharina Zeh, Lale Tüver

Faculty of Philological and Cultural Studies, University of Vienna, Vienna



Global linguistic diversity is declining. Does digital literacy influence this trend, particularly for minority languages?



It depends on how digital literacy is measured!

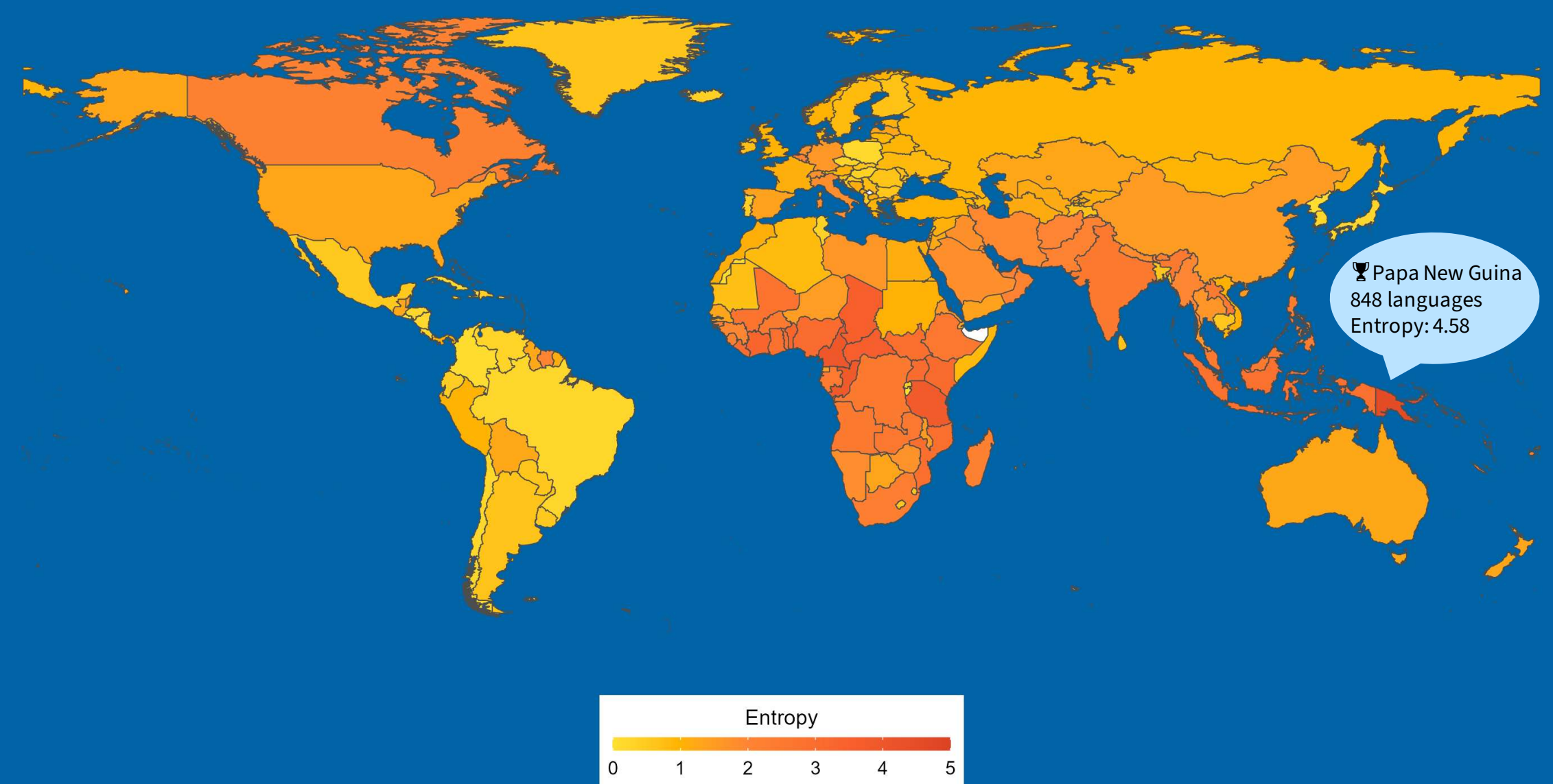


Figure Global linguistic diversity by country, measured using entropy-based metrics of language distribution.

Research Questions/ Project Goals

Global linguistic diversity has declined significantly in recent decades, raising concerns about the survival of minority languages. Previous research has primarily focused on socioeconomic factors, but the impact of digitalization remains largely underexplored.

Research Questions

What factors contribute to the global decline in linguistic diversity?
In particular, what role does digital literacy play?

Creation of a Global Linguistic Diversity Dataset

- Extract national speaker counts from various sources
- Compute entropy values on country level

Case Study

- Analyze how digital and demographic factors influence linguistic diversity
- Measure the effect of digital literacy on languages

Datasets & Sources

Category	Source	Variables
Linguistic	Ethnologue, Joshua Project	speaker counts
Demographic	World Bank, UN	population, GDP, rural access ...
Digital Access	International Telecommunication Union (ITU)	Internet access at home, individuals using the internet, mobile subscriptions...
Digital Skills	Eurostat, ITU, World Population Review	Basic, above-basic & advanced in 5 dimensions (communication, safety, problem solving etc)

Data Preprocessing

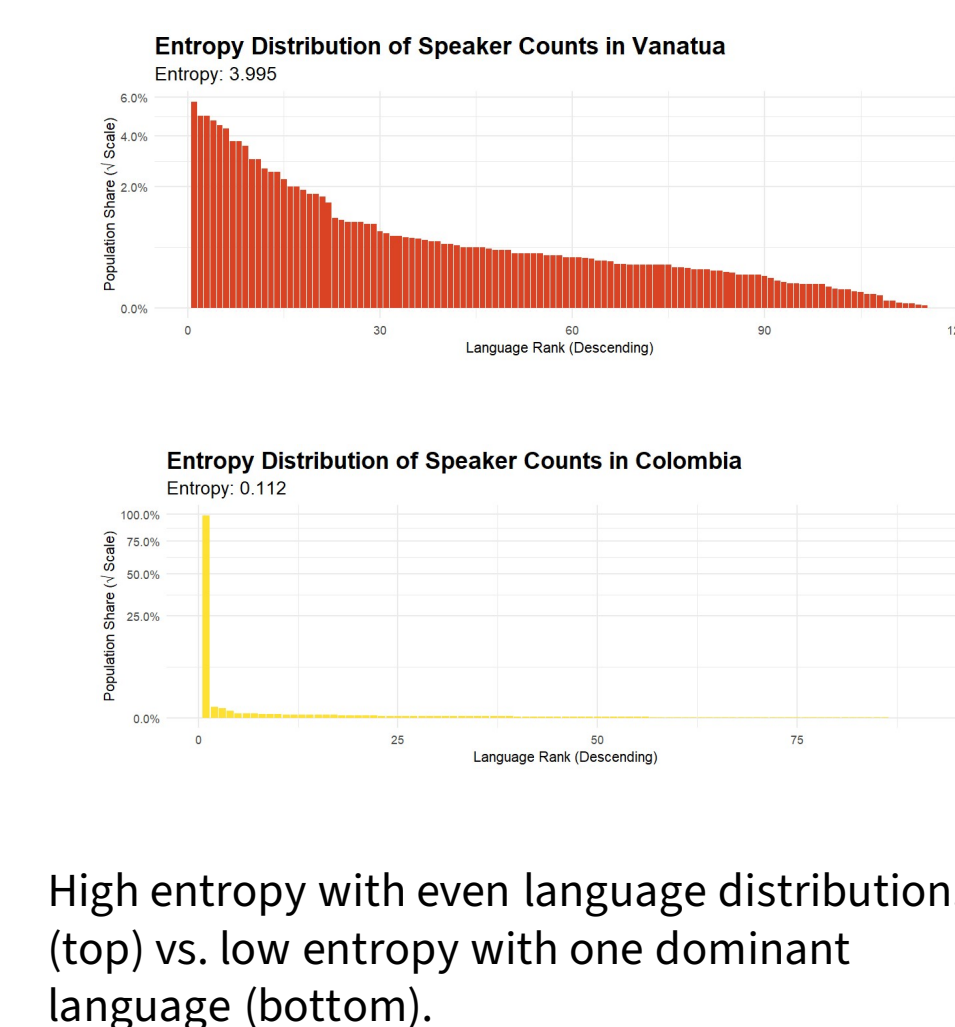
- Data Standardization**
- Removed special characters
 - Standardized number formats
 - Normalized speaker counts

- Data Enrichment**
- ISO 3166-1 Country codes
 - ISO 639 Language codes

Computation of Entropy

- **Shannon Entropy**

Entropy is a diversity metric from ecology that measures how evenly speakers are distributed across languages.

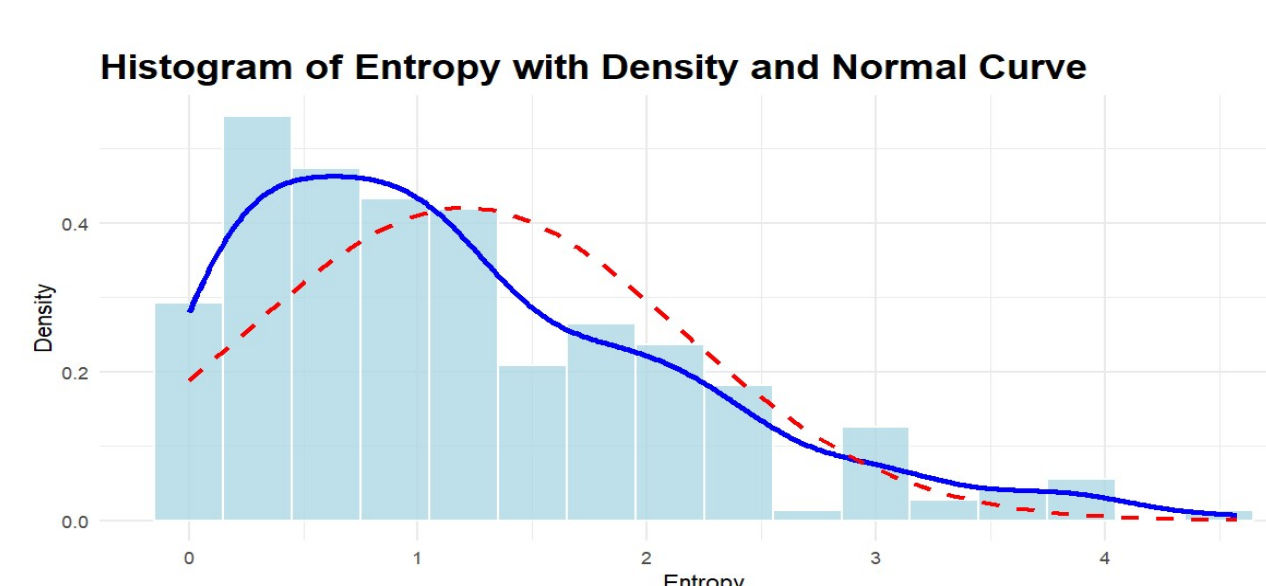


High entropy with even language distribution. (top) vs. low entropy with one dominant language (bottom).

Exploratory Data Analysis

Entropy Distribution (dependent variable)

- mean = 1.50
- standard deviation = 3.30
- observations: 214
- shape: right-skewed with long tail



Lowest Entropy Values (≈ 0.00)

Falkland Islands
Jersey
Niue, Pitcairn, Saint Helena, etc

Top Entropy Values

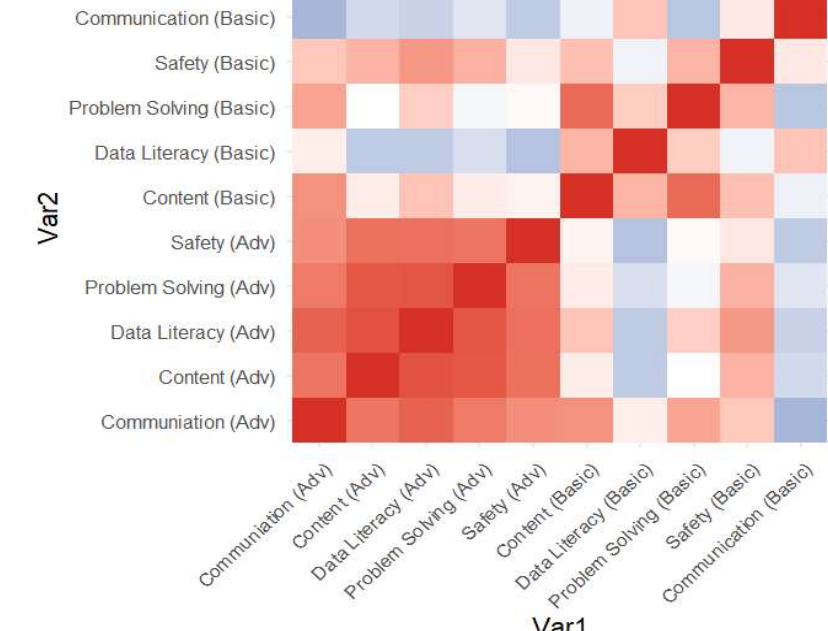
Papua New Guinea – 4.58
Vanuatu – 4.00
Cameroon – 3.98

Modeling Approach

Generalized Linear Model (GLM) with Gamma distribution

- Suitable for skewed, positive, and continuous data
- Applies a log-link function
- Model fit evaluated using deviance and Pearson's residuals
- Predictor variables checked for multicollinearity

Correlation Matrix of Digital Skills Predictors



Variables

Predictors

- Demographic: GDP, rural %, school life expectancy
- Digital Access: Internet use, home access, mobile subscriptions
- Digital Skills: basic & above-basic in 5 dimensions

Response Entropy

Results

Demographic

Significant negative effect of years of schooling on language diversity

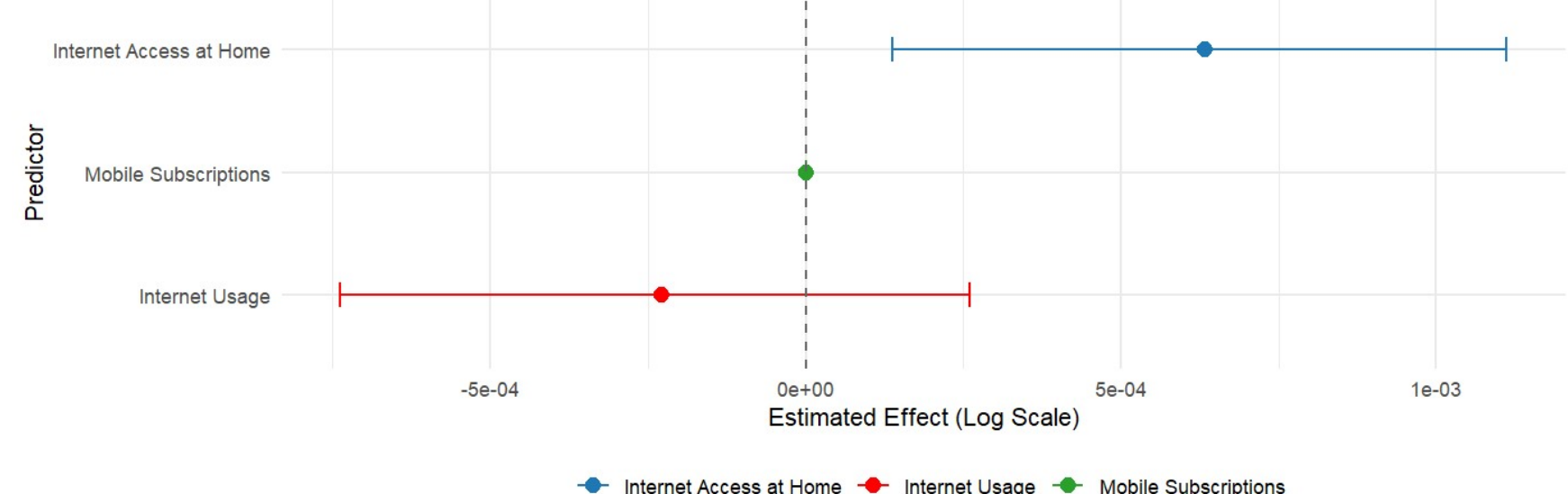
Digital Access

Significant positive effect of internet at home on language diversity

Digital Skills

No significant effects (could be caused by limited data and small sample sizes)

Effect Estimates of Digital Access Predictors



Conclusions

- Our findings reveal a **complex relationship between digital literacy, digital access, and linguistic diversity.**
- **Digital factors play a critical role in language loss and require further investigation in future research** on language diversity and endangerment.
- Despite limitations in sample sizes and available datasets, **future research can build on the findings of this project.**

References

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. <https://doi.org/10.1038/s41559-021-01604>.
Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2024). *Ethnologue: Languages of the world* (27th ed.). SIL International. <http://www.ethnologue.com>
Law, N., Woo, D., de la Torre, J., & Wong, G. (2018). *A global framework of reference on digital literacy skills for indicator 4.4.2*. UNESCO Institute for Statistics. <https://unesdoc.unesco.org/ark:/48223/pf0000265403>

Kandler, A., & Unger, R. (2023). Modeling language shift. In E. Schöll & S. H. L. Klapp (Eds.), *Diffusive spreading in nature, technology and society* (pp. 365-387). Springer. https://doi.org/10.1007/978-3-031-05946-9_18
Pagel, M. (2017). Darwinian perspectives on the evolution of human languages. *Psychonomic Bulletin & Review*, 24(1), 151-157. <https://doi.org/10.3758/s13423-016-1072-z>
Tuomisto, H. (2010). A diversity of beta diversities: Straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography*, 33(1), 2-22. <https://doi.org/10.1111/j.1600-0587.2009.05880.x>