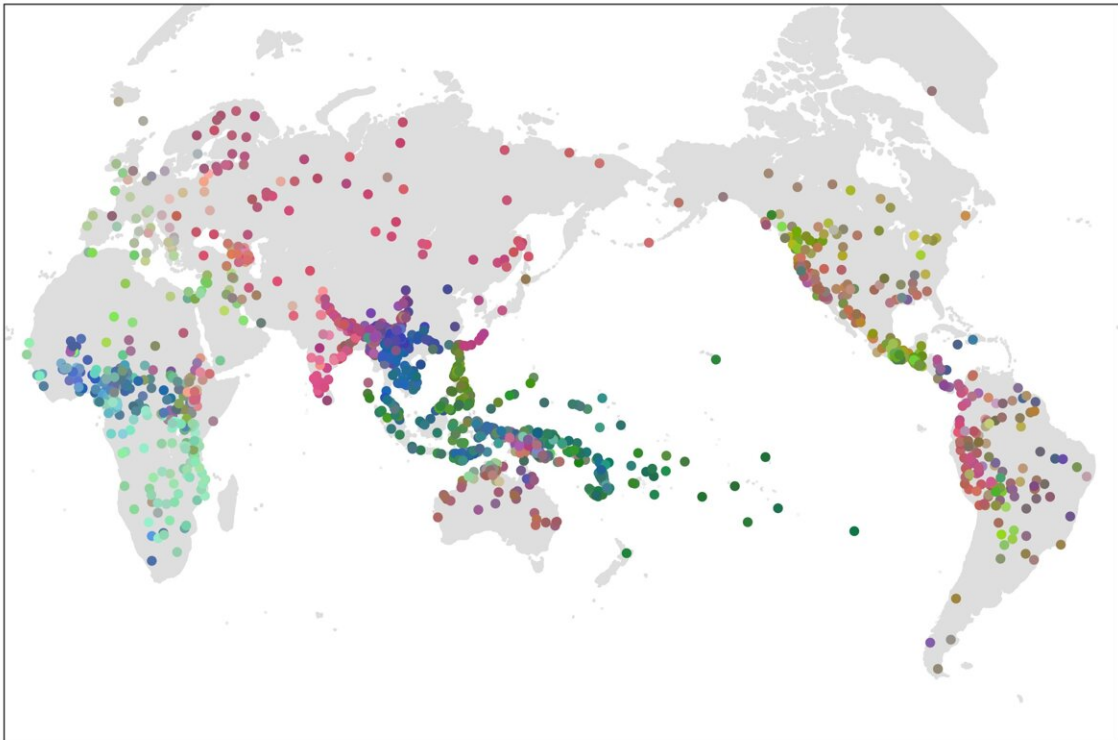


Database shows the diversity of the world's languages

April 19 2023



Grammatical similarity in the Grambank sample of languages. The color coding represents the distribution of languages according to the first three principal components of a Principal Component Analysis mapped onto RGB color space (PC1 = Red, PC2 = Green and PC3 = Blue). Similarity in color indicates similarity in grammatical structure on the first three dimensions. Credit: © MPI f. Evolutionary Anthropology

What shapes the structure of languages? In a new study, an international team of researchers reports that grammatical structure is highly flexible across languages, shaped by common ancestry, constraints on cognition and usage, and language contact.

The study used the Grambank database, which contains data on grammatical structures in more than 2,400 languages. The project was initiated by the Department of Linguistic and Cultural Evolution at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, in collaboration with a team of more than one hundred linguists from around the world. The work is published in the journal *Science Advances*.

Linguists have long been interested in [language](#) variation. What are common or universal patterns across languages? What limits the possible variation between them? [Grambank](#), the world's largest and most comprehensive database of language structure, enables researchers to answer some of these questions.

Grambank was constructed in an international collaboration between the Max Planck institutes in Leipzig and Nijmegen, the Australian National University, the University of Auckland, Harvard University, Yale University, the University of Turku, Kiel University, Uppsala University, SOAS, the Endangered Languages Documentation Program, and more than one hundred scholars from around the world. Grambank's coverage spans 215 different language families and 101 isolates from all inhabited continents.

"The design of the feature questionnaire initially required numerous revisions in order to encompass many of the diverse solutions that languages have evolved to code grammatical properties," says Hedvig Skirgård, who coordinated much of the coding and is the lead author of the study.

Limits on variation

The team settled on 195 grammatical properties, ranging from word order to whether or not a language has gendered pronouns. For instance, many languages have separate pronouns for "he" and "she," but some also have male and female versions of "I" or "you." The possible "design space" would be enormous if grammatical properties were to vary freely. Limits on variation could be related to cognitive principles rooted in memory or learning, rendering some grammatical structures more likely than others. Limits could also be related to historical "accidents," such as descent from a common language or contact with other languages.

The researchers discovered much greater flexibility in the combination of grammatical features than many theorists have assumed. "Languages are free to vary considerably in quantifiable ways, but not without limits," explains Stephen Levinson, Director emeritus of the Max Planck Institute for Psycholinguistics in Nijmegen and one of the founders of the Grambank project. "A sign of the extraordinary diversity of the 2,400 languages in our sample is that only five of them occupy the same location in design space (share the same grammatical properties)."

Languages show much greater similarity to those with a common ancestor than those they are in contact with. "Genealogy generally trumps geography," says Russell Gray, Director of the Department of Linguistic and Cultural Evolution and senior author of the study. "Nevertheless, if processes of linguistic evolution and diversification were run again from the beginning, there would still be some resemblance to what we now have. The constraints of human cognition mean that, while there is a great deal of historical contingency in the organization of grammatical structures, there are regular patterns as well."

Diversity under threat

"The extraordinary diversity of languages is one of humanity's greatest cultural endowments," concludes Levinson. "This endowment is under threat, especially in some areas such as Northern Australia, and parts of South and Northern America. Without sustained efforts to document and revitalize endangered languages, our linguistic window into [human history](#), cognition and culture will be seriously fragmented."

The Grambank database is an open-access comprehensive resource maintained by the Max Planck Society. "It puts linguistics on an even footing with genetics, archaeology and anthropology in terms of quantitative, large scale, accessible data," says Gray. "I hope it will facilitate the exploration of links between linguistic diversity and a broad array of other cultural and biological traits, ranging from religious beliefs to economic behavior, musical traditions and genetic lineages. These links with other facets of human behavior will make Grambank a key resource not only in linguistics, but in the multidisciplinary endeavor of understanding human diversity."

More information: Hedvig Skirgård et al, Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss, *Science Advances* (2023). [DOI: 10.1126/sciadv.adg6175](https://doi.org/10.1126/sciadv.adg6175). www.science.org/doi/10.1126/sciadv.adg6175

Provided by Max Planck Society

Citation: Database shows the diversity of the world's languages (2023, April 19) retrieved 27 April 2024 from <https://phys.org/news/2023-04-database-diversity-world-languages.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private

study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.