



Check for updates

ESSAY

REVISED **Global language geography and language history: challenges and opportunities**

[version 2; peer review: 2 approved, 2 approved with reservations]

Matthias Urban^{1,2}

¹Dynamique Du Langage, Centre Nacional de la Recherche Scientifique, Lyon, France

²Universite de Lyon, Lyon, France

v2 **First published:** 07 Oct 2024, **4**:213
<https://doi.org/10.12688/openreseurope.18421.1>
Latest published: 20 Jan 2025, **4**:213
<https://doi.org/10.12688/openreseurope.18421.2>

Abstract

While it has almost become a truism of comparative linguistics that linguistic diversity is unevenly distributed across the globe, the reasons are poorly understood up to the present day. Linguists are thus in the embarrassing situation that they do not understand significant regularities in the way the objects of their study –languages– pattern. In this essay, I explore three interrelated strands of thought to create a perspective on the question that is different from those explored so far: first, I suggest that instead of looking at present-day levels of diversity statically, we should take an approach that looks into how these distributions were generated. Related to this point and in contradistinction to extant work, second, I advocate an inductive approach that departs from qualitative case studies that inform theory-building. Third, I ponder that, in contrast to the traditional focus of historical linguistics on language diversification and expansion, understanding how the ranges of languages are reduced might be the key missing piece of evidence in a global theory of language diversity and its genesis. This new perspective is also able to address the striking correlation between linguistic and biological diversity that suggest that the processes that created and maintain both are, on some level, qualitatively similar.

Plain language summary

This essay reflects on the distribution of linguistic diversity across the globe, which at present cannot be explained satisfactorily. I suggest that to understand this distribution we need to look at actual, concrete processes that have generated the ranges of individual languages, and that we can extrapolate from these to understand patterns in linguistic diversity at large.

Open Peer Review

Approval Status ✓✓??

	1	2	3	4
version 2 (revision) 20 Jan 2025		✓ view		
version 1 07 Oct 2024	✓ view	? view	? view	? view
1. Stan Brunn , University of Kentucky, Lexington, USA				
2. Guglielmo Inglese , Universita degli Studi di Torino, Turin, Italy				
3. Alexandre François , Centre National de la Recherche Scientifique, Paris, France Ecole Normale Superieure, Paris, France Australian National University School of Culture History and Language, Canberra, Australia				
4. Tuomas Väisänen , University of Helsinki, Helsinki, Finland University of Helsinki, Helsinki, Finland University of Helsinki, Helsinki, Finland				
Any reports and responses or comments on the article can be found at the end of the article.				

Keywords

Language geography, language dispersal, linguistic diversity, typology, historical linguistics



This article is included in the [European Research Council \(ERC\)](#) gateway.



This article is included in the [Horizon Europe](#) gateway.



This article is included in the [Linguistic Diversity](#) collection.

Corresponding author: Matthias Urban (matthias.urban@cnrs.fr)

Author roles: Urban M: Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Research Council (ERC) under the European Union (LANGUAGE REDUX) Grant agreement No. 101124345.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Urban M. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Urban M. **Global language geography and language history: challenges and opportunities [version 2; peer review: 2 approved, 2 approved with reservations]** Open Research Europe 2025, 4:213 <https://doi.org/10.12688/openreseurope.18421.2>

First published: 07 Oct 2024, 4:213 <https://doi.org/10.12688/openreseurope.18421.1>

REVISED Amendments from Version 1

This is a substantially revised version of the article compared to the first version. Major changes made in response to the extensive comments from the reviewers include an addition of primary language data illustrating different aspects of linguistic diversity; extended discussion of traditional multilingual settings; and clarification on data sources, as well as more minor clarifications. The manuscript has also been rewritten for greater clarity and stringency.

Any further responses from the reviewers can be found at the end of the article

Introduction: the puzzle of linguistic diversity

During the 19th and 20th century, our knowledge concerning the diversity of human languages has increased massively. We know now that there are many more distinct languages spoken around the world than what has generally been thought possible some centuries ago, namely around 7000. We also know now that the way these languages function is massively more diverse than has been thought possible as well (Evans & Levinson, 2009).

“Language diversity” can mean several related things that must be distinguished carefully. First, the term can refer to the sheer number of languages relative to the size of a study area, e.g. a country (this is sometimes called “language richness”). Second, “language diversity” can reference the number of different language *families* relative to the size of a study area (this is sometimes referred to as “phylogenetic diversity”).

These two parameters often, but not always, correlate. In Africa, for instance, the sub-Saharan belt hosts hundreds of clearly distinct languages, but they all belong to the so-called Bantu branch of the Niger-Congo family that began to diversify several thousand years ago. Orthogonally to these two dimensions, finally, the languages in a given study area may also be very similar or very diverse in how they sound and how their grammars work (this parameter is sometimes also called “structural diversity”, Nettle, 1998).

These parameters together define very distinct linguistic landscapes in different parts of the world.

Some regions have turned out to be massively diverse linguistically on all levels. A textbook case is the island of New Guinea, which alone hosts almost 1000 distinct languages which belong to a sizable number of different language families. Given the size of the island, this means that observers will in some cases encounter a different language as they just travel to the next village. A slightly situation obtains in nearby Vanuatu, where the approximately 9,400 inhabitants of the 50 villages on Banks and Torres Islands recognize 17 distinct languages. However, these form part of the Oceanic subgroup of the vast Austronesian language family, and within that group, are more or less closely related to one another. Below is the sentence *They don't know our language very well yet* in each of these 17 languages, collected by Alexandre François (2011), the linguist who has studied the “particular alchemy” (2011: 177) of Northern Vanuatu's language ecology most intensively.

Hiw	sisə	tati	jəjmə ⁹ Len	wu ⁹ Lɔɣ	k ^w e	i	nə	məŋa	ta
Lo-Toga	nihə	tat	lolməren	ɛrβɛ	k ^w ɛ	e	nə	βəɣəβəɣə	mətə
Lehali	kej	tetnɛ	ɣlal	ɣalsɛ	k ^w ɔ		n-	βap	munɣɛn
Löyöp	kiej	tɛ	ɣilal	tʃɔjmat	tʃɛk ^w ɛ		n-	βaβap	ŋim ^w ɔnien
Volow	⁹ gɪj	et	ɣɣlal	ɣalsi	tɛ ⁹ gɔ ^w ɛ		n-	ɣatɣat	njɔnɣɪn
Mwotlap	kij	et	ɣɣlal	ɣalsi	k ^w ɛtɛ		nɔ-	hɔhɔlɛ	nɔnɔnɣɪn
Lemerig	tær	ɪ	ɣɔlɔl	ʔɔrmaʔ	ʔæ.kiʔis		n-	tɛktek	mɔɣɔt
Vera'a	ⁿ dir	ɪʔ	lamai	ɛntɛɣ	ʔɪn		ɪn	tɪktɪk	mu ⁿ dɪ
Vurës	nɪr	ɣɪɪ-	ɣilal	wareɣ	tɛn		ɔ	k ^w ak ^w	namɔɣɣɪnɪn
Mwesen	nɪr	ɛtɛ	ɪɪɪ	manɰtɛ	βɪs		ɔ	ɣatɪɛ	mɔɣɔnɪn
Mota	nira	ɣate	ɣlala	mantay	t ^w k ^w ɛ		o	βaβae	naŋim ^w unina
Nume	nir	βitis	ɣil	liŋliŋi	mi		u	luwluw	namɣɪn
Dorig	nɪr	sɔwsɛ	βɪɣɪɪ	taβul	tɛ		na	ɪŋa-	ɣɪn
Koro	nɪr	tɪ	rɔŋ	taβul	wɔs.meɪɛ		ɔ	βalβalaw	namɪɣɪn
Olrat	nɪj	tɪ	rɔŋ	βɪɪɪ:	wɔs.meɪɛ			ususra:	mɔtʃ
Lakon	ɣɪ:	atɪ	rɔŋ	kɛɛ	aβɔh.male			ɛɪŋa-	nɣɪtʃ
Mwerlap	ker	ti	βalyɛar	minmin	tɪk ^w ɪtɛa		nɔ-	liŋɪ-	ɣɛan
	3PL	NOT.YET ₁	know	properly	NOT.YET ₂	[OBL]	ART	speech	POSS:1INCL.PL

A subset of languages resemble each other in their words: for instance, some have forms like *ylal*, *ylal* or *ylal* for ‘to know’ while others have completely different equivalents like *jəjmə^sLen* or *jəjmə^sLen*. In spite of such lexical differences, the structure of the sentence in each of the 17 languages is remarkably similar: the subject pronoun comes first, followed by predicate and object; there are grammatical particles expressing that the state of knowing the language has not been achieved yet not once, but twice in the sentence, and always in the same places; and the possessive pronoun ‘our’ indicates that the group referred to includes speaker and addressee in all languages. While there are also differences, this means that one can sometimes translate word by word from either one of the Banks and Torres Islands languages to any other because their grammars work in similar ways. In sum, with 17 languages, language diversity is very high on these islands; phylogenetic diversity is very low as they are all related to one another; and structural diversity likewise is low.

In contrast, Northern Eurasian languages score relatively low on all three counts: there are relatively few different languages; these belong to a small number of distinct language families; and most, in particular those resulting from recent language spreads, are quite similar in how they sound and how their grammars work.

These are some examples of the myriad different ways in which the three parameters of diversity can combine.

Visualizing linguistic diversity in an accessible way is not easy and requires a number of decisions and qualifications. While there are parts of the world where relatively clear language

boundaries exist, this is not the case everywhere. Languages may overlap in geographical space in complex ways, either because speakers of more than one language live next to each other in the same communities, or because languages “live” next to each other in the minds of bi- or multilingual speakers (societal vs. individual bi- or multilingualism, Appel & Muysken, 2005), or both. Any representation of language on maps is inadequate to capture these distributions in their full complexity (Luebbering, 2013; Urcioli, 1995), and for most of the world’s languages, such detailed information is simply unavailable anyhow.

Figure 1 nevertheless attempts to show globally uneven levels of linguistic diversity. This map focuses on the language level. It represents languages as dots in the absence of widely available polygon data that could represent the ranges in which languages are spoken more realistically, let alone more detailed information of the distribution of languages in social space.¹ Coordinates come from Glottolog, the leading catalogue

¹ More technically, the dots represent “languoids” – a neologism introduced in Cysouw & Good (2013) for linguistic entities that may correspond to language families, languages, or dialects in traditional parlance. Boundaries between these widely familiar categories are not sharply defined – for instance, a set of dialects that are so divergent that speakers have difficulties understanding each other is not very different from a small family of closely related languages. The notion languoid allows reference to linguistic entities recognized in the literature without enforcing permanent decisions on their position on the dialect–language–language family continuum, and Glottolog allows for flexible “upgrading” or “downgrading” of the classification level as more information becomes available. The dots show linguistic entities with identifiers that are currently considered “language-level” in Glottolog (Forkel & Hammarström, 2022).

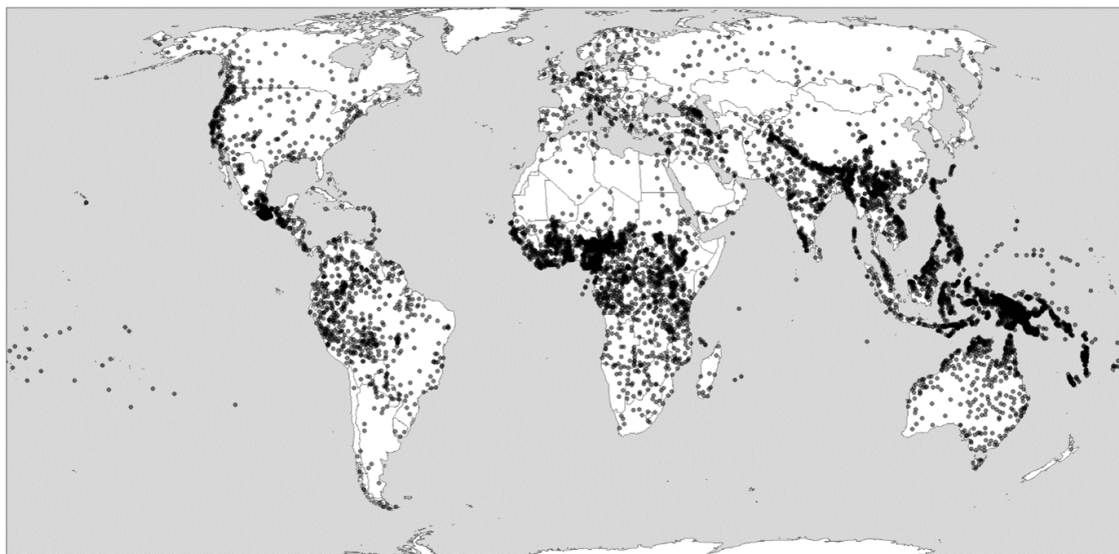


Figure 1. Global linguistic diversity (language richness). Created using the R package glottospace (Norder et al., 2022).

of the world's languages via the glottospace R package (Norder *et al.*, 2022). In the case of small-scale languages, dots typically represent the center of the area where the language is spoken; in other cases, historical or demographic information inform decisions (as when Russian is plotted at the coordinates of Moscow).

Asymmetries in diversity levels are not only observed within continent or island-sized areas such as those I have just mentioned, but scale down to variation *within* such areas: In New Guinea, it is the coastal areas in the north that host disproportionately much of the island's linguistic diversity, while the New Guinea highlands, which traverse the island longitudinally, are lower in diversity both measured in terms of individual languages and in terms of different families.¹ Figure 2 gives an impression of this.

As can be observed in Figure 3, in South America, diversity levels in greater Amazonia, while high everywhere, are particularly pronounced on the eastern margins. In the Andes themselves, and as one moves southward to Patagonia and Tierra del Fuego, language density becomes notably lower. In North America, it is only California that boasted a hyperdiverse mosaic of languages or language families, whereas to the east of the Rockies, diversity is measurably lower.

It is such observations of nested diversity clines that suggest that it is no mere coincidence of history that New Guinea, Amazonia, and California are hyperdiverse, whereas Greenland, Patagonia, and eastern north America are not to the same extent. However, up to the present day, the reasons for regionally uneven diversity levels are not understood.

As the differential patterning of linguistic diversity has come to the attention of linguists and scholars in other disciplines in the early 21st century, a striking congruence with biological

diversity was observed (Gorenflo *et al.*, 2012; Maffi, 2005; Moore *et al.*, 2002). Also this relationship is not only observed globally, but on intra-continental scales, too (though at some levels of resolution, it breaks down: Manne, 2002). For instance, in South America, the Amazonian fringe and the cloud forest ecotone at the intersection of Andes and eastern lowlands are not only hotspots of linguistic, but also biological diversity.

Languages are cultural products, shaped by the communicative behaviour of their speakers (Du Bois, 1987) and the social ecologies they are embedded in (Pakendorf *et al.*, 2021; see more extended discussion in the following sections). They are *not* species. Drawing an analogy between biological evolution and "cultural evolution" is as tempting as potentially dangerous and simplifying, especially as we have a lot of qualitative knowledge on the maintenance of linguistic diversity and their social ecologies that are specifically human (again, more extended discussion is in the following sections). Still, undeniably, languages cover geographical space in ways that are strikingly similar to the ways in which biological species do.

Like Hua *et al.* (2019), I tend to view this association as epiphenomenal (linguistic and biological diversity co-vary with climate and/or environment, which shapes the distribution of both) rather than causal (humans diversify culturally in areas with high biodiversity more readily because of the richness of available resources).

The latter scenario would be consistent with a prominent hypothesis that has been suggested to explain the uneven linguistic diversity of the world. In the next section, I turn to current perspectives on linguistic diversity, including this hypothesis of "ecological risk". I also discuss some of the issues I see with extant work, and questions that still remain open, either because studies so far have yielded contradictory answers or because they have not been a focus of attention.

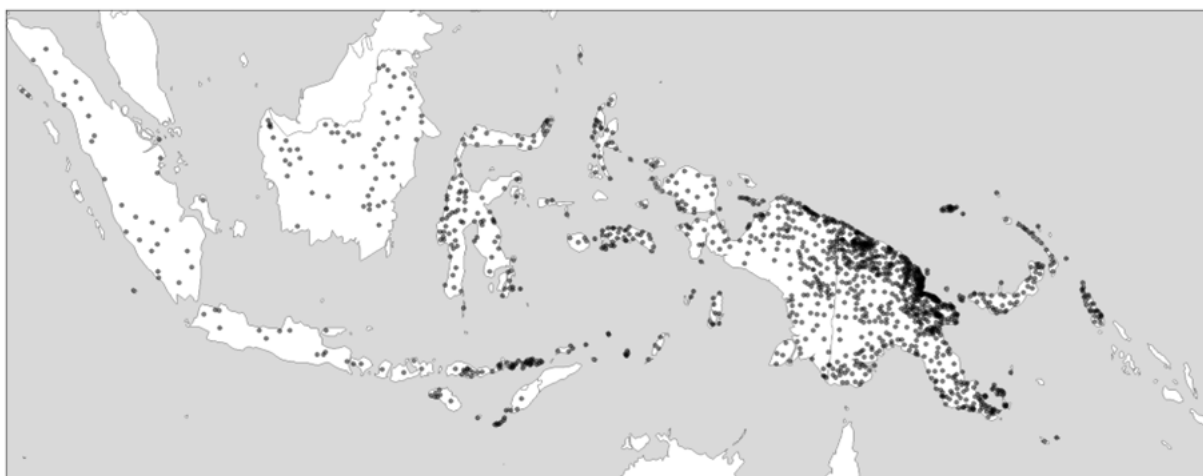


Figure 2. Linguistic diversity (language richness) in Indonesia and Papua New Guinea. Created using the R package glottospace (Norder *et al.*, 2022).



Figure 3. Linguistic diversity (language richness) in South America. Created using the R package glottospace (Norder *et al.*, 2022).

Current perspectives (and their problems)

There are minimally three dimensions to the problem of explaining global linguistic diversity: first, there is the question which types of environments (if any are identifiable) are conducive to high linguistic diversity and which mitigate its development; second, there is the question if the congruence between cultural and biological diversity is a coincidence, and, if not, how it should be explained; and third, especially because languages emphatically are not species, there is the question how such factors, if they could be identified, link up with actual human linguistic and non-linguistic behavior.

In the last 25 years or so, a number of studies have appeared that address the question of linguistic diversity and its drivers. However, they have mainly focussed on the first dimension, while the actual behaviour of people that, in the long term, could lead to these distributions has received much less attention. I will return to this point later.

Early studies (Cashdan, 2001; Collard & Foley, 2002; Hillebrand, 2004; Mace & Pagel, 1995) mainly noted a latitudinal gradient to global levels of linguistic diversity (see also Gavin & Stepp, 2014).

Nettle (1996; 1998) was an influential pioneer in singling out a possible explanation, and his “ecological risk” hypothesis is still viable today: Nettle argued that language diversity in the sense of language density correlates not just with latitude, but more directly with the length of the Mean Growing Season, i.e. the number of months per year that allow, given local conditions like temperature, rainfall, etc., vegetation to grow. The length of the Mean Growing Season, in turn, is taken by Nettle as a proxy for “ecological risk”: Where it is long, ecological risk is low, and societies are hypothesized to be self-sufficient: the environment allows enough resources for most parts of the year to ensure reliable subsistence. Such conditions are said to foster the development of many small,

self-reliant speech communities that do not need to maintain outside ties to safeguard against supply shortages. Where climatic conditions are less stable, or where resources are scarcer due to a shorter Mean Growing Season, people must increasingly derisk their subsistence base. This involves the exploitation of resources on larger tracts of land, and/or the maintenance of stronger inter-community ties and larger networks – hence, so the hypothesis, fewer speech communities spread out over wider ranges².

In the meantime, the size of the data analyzed, the number of possible environmental variables considered, and the methodological sophistication of relevant work has grown enormously. Spatial autocorrelation, a typical phenomenon in spatially structured data, has been shown to confound results in research on the environmental and social conditioning factors of linguistic diversity (Cardillo *et al.*, 2015). It has also been noted that the effect of surveyed climatic variables may be non-stationary and interact in locally specific ways in generating linguistic diversity or inhibiting it (Pacheco Coelho *et al.*, 2019). As a whole, research in the drivers of language diversity, reassuringly, shows the typical signs of maturation of a field of scientific investigation.

At the same time, however, the results of relevant studies that have appeared after Nettle differ widely. We have now a whole series of articles, often in high-standing journals and using sophisticated methodology, that share the general concern of identifying ecological drivers of linguistic diversity, and that assess a wider range of environmental parameters rather than just Mean Growing Season as a proxy for ecological risk. Axelsen and Manrubia (2014) identify the presence of rivers and terrain roughness (technically, “rugosity”) as factors that are conducive to the rise of linguistically diverse landscapes – but earlier Currie and Mace (2009) only found a weak effect of rugosity (which they measured differently, however). While Sutherland (2003) found no evidence for Nettle’s ecological risk hypothesis, Hua *et al.* (2019) found the effect of geology-related variables like rugosity to be relatively negligible compared to climate and year round productivity, consistent with an account like that of Nettle. Derungs *et al.*’s (2018) results are likewise consistent with the “ecological risk” hypothesis, but they conclude that climate –latitude, precipitation, and temperature – is more relevant for linguistic diversity of food-producing (agriculturalist) societies than for hunter-gatherers, which would mean that “ecological risk” should affect different types of people in different ways.

While the question which environmental or climatic conditions foster language diversity thus remains open for the time being, a broader question is how *any* climatic and environmental factors could lead to actual linguistic and non-linguistic behavior of people that, in turn, sets into motion processes that would

generate observed patterns of diversity. As sketched above, an answer to this question is encapsulated in the “ecological risk” hypothesis. However, as far as I am aware, that ecological risk actualizes linguistic behavior with repercussions on language density, or even phylogenetic and structural diversity, has so far never been shown through detailed qualitative case studies, but only stipulated by hypothesis.

Where case studies on the ecology of linguistic diversity in the tradition of the ethnography of communication (Hymes, 1964) exist, it often turns out that it is language ideologies of one kind or another that sustain diversity in arrangements of so-called small scale multilingualism (Pakendorf *et al.*, 2021). For instance, in the Vaupes region of the Amazon, linguistic exogamy together with a linguistic ideology that deprecates language mixing creates a situation in which a range of distinct languages belonging to different language families are maintained and converge with each other grammatically, but not lexically (e.g. Epps, 2020). Also some big data quantitative studies –such as Antunes *et al.* (2020) on New Guinea– suggest that environmental factors alone are insufficient to explain observed levels of linguistic diversity, and that socio-cultural factors must be taken into account.

Similarly, linguistically highly fragmented Melanesia is characterized by common ways in which language is considered to be anchored with particular communities and places (e.g. Laycock, 1982; Sankoff, 1980). One extreme case for this Melanesian way of thinking about language and space comes again from Vanuatu: forms of speech, whether linguists would classify them as dialects or languages, are considered to be locally anchored to communities in individual villages, as are other cultural activities such as ways of preparing food, singing etc. This link between place and language is valued and actively maintained and fosters the development of locally specific speech forms, including details of pronunciation and vocabulary that come to be viewed as typical of different places – and these are precisely those highly salient features of speech that people are likely to note and associate with different languages. At the same time, there is what François (2012) calls “egalitarian multilingualism:” no language is considered per se more valuable than any other (even though, in practice, some languages are particularly important because they have disproportionately high speaker numbers). People do not mind learning the languages of others, also because marriages are typically exogamous so that multilingual households are common. Such kinship ties are complemented by commercial ones, and together they lead to dense networks of social and commercial interaction. This, of course, is fertile ground for convergence effects such as the ones reflected in the parallelism in grammar between Vanuatu languages. In sum, the particular linguistic ecology of Northern Vanuatu fosters both differentiation and fragmentation with high language density and low structural diversity: that is, exactly the configuration of linguistic diversity that can be observed.

On the other hand, other qualitative, “thick” (Geertz, 1973) descriptions of language ecologies suggest that notions like

² This should in principle be relevant both for agriculturalist and more “traditional” lifestyles involving hunting and gathering, as ultimately plants form the basis for subsistence in both cases.

“group boundary formation” (Gavin *et al.*, 2013), i.e. the essentializing use of a form of speech to identify one’s “ethnic group”, set it off against other such groups, and thus yield “ethnolinguistic groups” are far from universal. Nettle himself, who relies on the “ethnolinguistic group” as a unit of analysis, quotes Brooks (1993: 27) to the effect that in West Africa, “individuals and families change their language and modify their social and cultural practices in ways that are often perplexing to outsiders.” In other words, language is not always firmly linked to particular individuals that together would make up an “ethnolinguistic group”, but language use may be dependent on social context, roles, and local language ideologies. An influential analysis of such configurations in the Balkans is in Irvine and Gal (2000); for the ancient and present-day Central Andes in Babel (2018) and Urban (2018; To appear); and for Upland Southeast Asia and the Himalayas in Scott (2009) and Shneiderman (2010). As Evans (2018: 13) aptly puts it, “the focus on language diversity as something manifested by discrete, internally coherent entities can remove the very types of evidence we need to tackle the diversification problem.”

In sum, the qualitative literature suggests that the recruitment of languages as markers of local identity and groups, while they exist also in some “traditional” contexts, is far from universal. General, one-size-fits-all explanations for the genesis of linguistic diversity are thus problematic.

More specifically, in direct contradiction to the idea that ecological risk fosters a smaller number of languages with wider ranges, in traditional settings, it is precisely linguistic diversity which, sustained in multilingual landscapes, can be observed as a strategy “that maximizes alliances and protective networks through different languages” (Lüpke, 2016: 53). My point is not that ecological risk, or any other climate-related notion, does not have an influence, however indirect, on what parts of the world stabilize at what level of linguistic diversity. It is rather obvious that there *must* be some such proximate or, more likely, ultimate effect, even though it remains elusive. I merely wish to point attention to the fact that we have not, again as far as I am aware, observed climate- and environment-related factors “in action” to generate linguistic diversity in a set of diverse context (such as those sketched above) and that would be capable to explain language diversity on global scales. According to Evans (2018), who presents the most complete model for how all these factors might link up, climatic conditions *do* shape societies, their values, and their attitudes towards others, and these in turn govern linguistic behaviour in conversational praxis within one’s own social group and with outsiders. From such accounts, it becomes clear that how climate and environment shape linguistic diversity is a significant more complex problem than what much of the extant literature assumes.

At present, thus, research is in a remarkably open situation. The questions are on the table, but I think it is fair to say that answers that are robust to different analytic approaches and datasets and that are consistent with actually observed linguistic and non-linguistic behaviour have not crystallized yet. The

impasse pertains most pressingly to the first two dimensions of linguistic diversity that I have sketched in the introduction. Astonishingly, linguists do not understand the ways in which their objects of study –languages– are distributed on the largest imaginable scales, and especially why they are distributed this way. This is in contrast to the micro-levels of variation. Dialectologists, since the 19th century, have developed methods to describe and (at least to some extent) explain how features of pronunciation, grammar, and words change in geographical space, and the entire discipline of variationist sociolinguistics explores how language varies in social space. It is also in contrast to what we, as linguists, know on the distribution of features across whole languages and large regions (a field of study often called “areal typology”, in which clines and very large skewings, not dissimilar to those concerning language diversity, have been observed up to continental scales – Bickel, 2020; Dryer, 1989; Güldemann & Hammarström, 2020; Nichols, 1992; Urban *et al.*, 2019), and what we are beginning to learn on how the social ecologies in which languages are learned and spoken influence their structure and complexity (e.g. Lupyan & Dale, 2010; Trudgill, 2011). Zooming out from linguistics, the situation is also relevant to the human sciences at large, which are faced with the fact that the processes by which human societies, within and between each other, produce that key cultural products which they have used to define themselves since antiquity –languages– remains in the dark.

Dynamizing linguistic diversity

As the main contribution of this essay, in this section I explore lines of reasoning that might lead to perspectives on language diversity and its origins that is able to bring together the above-mentioned perspectives more satisfactorily. Since this is an essay, my discussion is mainly programmatic; whether the picture I will try to get into focus has merit is an open question.

I want to make three interrelated points: first, I suggest that instead of attempting to find parameters of variation (climatic, environmental, political, etc.) between regions with differential levels of linguistic diversity, a process-based approach that looks into how these distributions were generated can furnish perspectives that would be otherwise missed. In this vein, I suggest a way to dynamize the question of linguistic diversity and its drivers in a way that references the observed congruence between linguistic and biological diversity.³ Second, I ponder that, in contrast to the traditional focus of historical linguistics on language diversification and expansion, understanding how the ranges of languages contract might be the key missing piece of evidence in a global theory of language diversity and its genesis. Related to this point and in contradistinction to extant work, third, I advocate an inductive

³ This parallels how language typology has been dynamized by not looking at static synchronic features of languages, but at the historical processes of language change that shape how languages function at present (e.g. Cristofaro, 2019; Greenberg, 1978).

approach that departs from qualitative case studies which inform theory-building.

The starting point for developing my argument is the well-known distinction between spread and residual (or accretion) zones. Nichols (1992; 1997) distinguishes these as prototypical and contrasting types of language distributions in geographical space, and sketches their underlying dynamics. The dichotomy references the uneven distribution of linguistic diversity and at the same time contains elements of a theory to account for these dynamically.

As far as language geography is concerned, spread zones are dominated by relatively few language families (i.e. have low phylogenetic diversity), and at times even just contain few individual languages (i.e. have low language richness). These distributions are shaped by frequent and long-distance language expansions, which tend to completely or almost completely obliterate preexisting languages. These may include the languages that spread via earlier episodes of such expansions (“spread-over-spread” dynamics). As is expected in situations of rapid and long-range language spreads, the spreading languages are quite similar to one another as a result of the little time available for diversification through language change (i.e. low typological diversity). Thus, in spread zones, which languages and language families dominate the linguistic landscape can change drastically and quickly, but net linguistic diversity does not: it remains low on all three counts.

Residual zones, in the definition of Nichols (1992: 14–15), have the following characteristics: they contain old families (i.e. ones that are deeply diversified internally – this does not necessarily mean that these families must contain many individual languages, but that the languages belonging to its different branches are not closely related, which indicates a long time of internal differentiation). “Old families” in the relevant sense should be taken to include language isolates, which are the sole representative of lineages that are so old that relatives cannot be detected anymore. No major language expansions originate from residual zones, but they may attract intrusive languages and thus serve as a linguistic “refugium of sorts” in Nichols’s words. The arrival of these languages, however, does not lead to significant levelling of preexisting linguistic diversity in the residual zone; rather, in addition to processes of diversification through language change that take place relatively undisturbedly, they contribute further to a residual zone’s linguistic diversity.

Nichols also characterizes typical characteristics of spread and residual zones in terms of climate and environment, and explicitly mentions economic autonomy as a key condition for the genesis of residual zones. This is a clear and important point of articulation with the research on linguistic diversity that I have sketched above, in particular Nettle’s theory of ecological risk and economic autonomy or the lack thereof.

Spread and accretion zones can thus be thought of as prototypes of very different linguistic landscapes that form against

the backdrop of different economic and subsistence affordances for human societies. The crucial (prescient) contribution of these prototypes is that they are explicitly connected to different diachronic dynamics of language geography. A key relevant process is language expansion and diversification, the traditional forte of historical linguistics since the inception of the field. But what is of particular interest when it comes to explaining linguistic diversity, complementarily to the dynamics of spread zones, is the dynamics of residual zones – those parts of the world that major language spreads do not reach, or where they at least do not have an impact that would reduce linguistic diversity significantly. Indeed, “often a residual zone will be located at the periphery of a spread zone” (Nichols, 1992: 21) – consistent with the idea that the culmination point of language expansions is typically reached before these areas are affected. This is consistent with the now robust observation that language spread trajectories respond to environment (Bentz *et al.*, 2018). For instance, the thrust of the Bantu expansion reflects “a measureable preference for ... familiar savannah habitats” (Grollemund *et al.*, 2015: 13299) of the people driving it.

In Urban (2021), I provide a perspective on such dynamics through qualitative case studies on language isolates and how the ranges in which these languages are spoken have contracted in the course of attested history. Basque is a textbook example. Once spoken far into the Pyrenees and into the Ebro valley of Northern Spain (in fact, Ebro goes back etymologically to a Basque word for ‘valley’), the domain of the language has gradually shrunk, starting in antiquity and continuing up to the present. Figure 4, from Urban (2021), illustrates the process. Why is Basque spoken today in exactly that part of its former range in which we find it rather than in another? Trask (1997) explains that “the mountainous Basque terrain, with little agricultural land, no cities, few obvious resources, and harbours that faced uselessly (from the Roman point of view) onto the Atlantic, was simply too insignificant to be worth the trouble of colonization. And the same lack of Roman interest is very largely what guaranteed the unique survival of the Basque language”. Needless to say, this also means that the expansion of Latin came to a halt, or was mitigated, before reaching the Basque country.⁴ Basque is just one example of a broader patterns of neolithic Europe, where languages that likely predate the Indo-European spread are conspicuously found in peripheral regions like Basque, or on islands, in other words, at the geographical margins of Europe.

One can hypothesize similar dynamics to explain linguistic distributions and the emergence of residual zones for which we cannot rely on historical evidence. In western Mexico and Mesoamerica, language isolates and small language families are found at the edges of major agricultural spreads, suggesting a dynamics in which former, pre-spread language distributions

⁴ In Urban (2021), I also discuss the historical language geography of Burushaski, which is largely parallel to that of Basque, but adds a vertical dimension to the relevant geographical processes (see also Nichols, 2013).

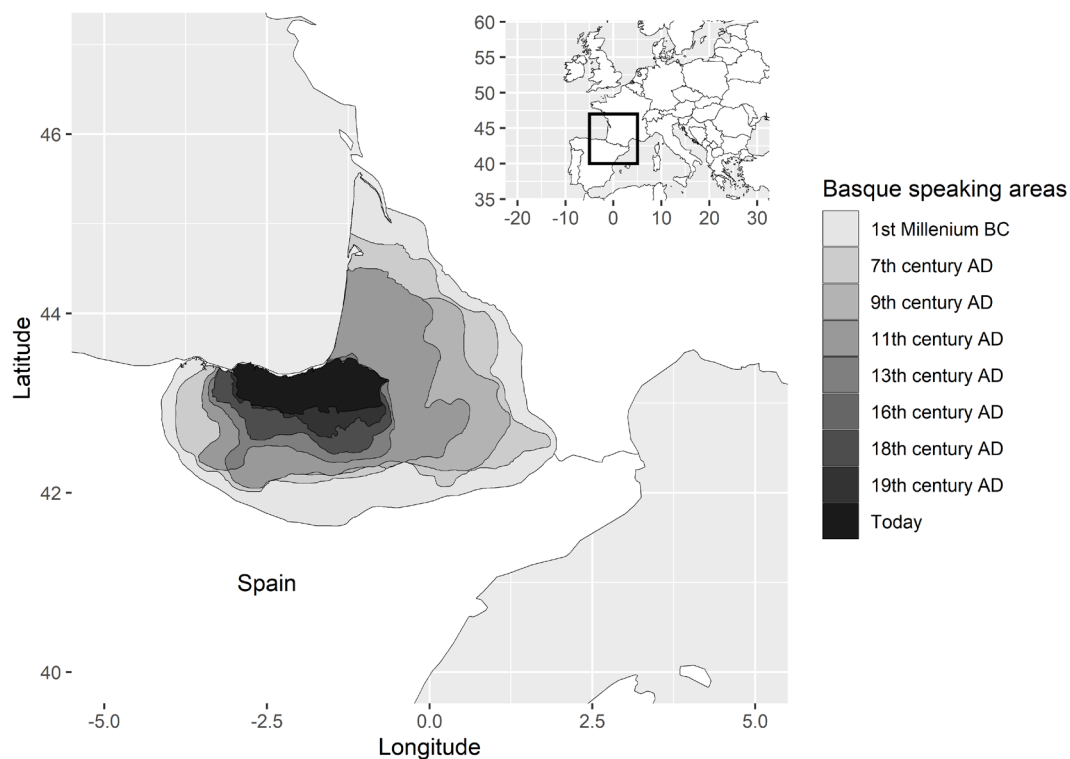


Figure 4. Historical changes in the geographical extent of the Basque language, from Urban (2021).

were reduced to geographically and economically marginal regions. Similar processes are not restricted to the deep past. They can be observed about 1500 years later in the context of colonial regimes. For instance, on the Pacific coast and in the Andean highlands of South America, Spanish colonial administrators removed Indigenous people from the agriculturally most productive lands and resettled them to less fertile regions and into mission towns. They also show themselves in the context of government-backed settler colonialism of the kind that drove the US westward expansion. Neither are such processes restricted to demographic changes introduced by agriculturalist, state-level imperial societies. Evans and McConvell (1998) provide a model for explaining a significant language spread in hunter-gatherer contexts, informed by deep-rooted Australian cultural practices.

The advantage of this process-based, range reduction oriented approach to studying linguistic diversity is that its net results are compatible with, and in certain cases predicted by, qualitative local language dynamics and ecologies. Vanuatu as described by François (2012) is one such example. Another concerns the Caucasus as a prototypical accretion zone. Nichols (2013) describes the traditional language dynamics of the Caucasus as one of asymmetric and gendered multilingualism which is embedded in and dynamized by the traditional subsistence patterns in this mountain area. Languages in the highland villages are typically community-based and the vehicles of communication for inward-facing “societies of intimates”

(Thurston, 1989). They are not or only very rarely learned by outsiders. The men of these communities, however, spend time in the lowlands to visit markets, and often stay the whole winter months in the lowlands, where herds would still find pasture. As such they are under pressure to learn languages of the lowlands, but lowlanders are under no pressure to learn highland languages. As a result, the general language dynamics of the Caucasus is one in which languages would constantly encroach the territory in an uphill direction, building up additional diversity without ousting that which already exists. As a result, the oldest layers of the diverse linguistic landscape of the Caucasus would be found at the highest altitudes, according to this model. Over the *longue durée* the Caucasus should accrete linguistic diversity both on the levels of language richness and phylogenetic diversity.

Finally, there is the question of the curious congruence between linguistic and biological diversity. The model also has advantages here. While comparisons between language diversity and species diversity have frequently been made, this has, to the best of my knowledge, only concerned the static situation at present, just like investigations of linguistic diversity have mainly been static (though see Gavin *et al.*, 2013; Gavin *et al.*, 2017; Pacheco Coelho *et al.*, 2019). However, there is a point of articulation between both when conceived of in dynamic terms. This point of articulation is the specific way how the geographical ranges of species and languages shrink and contract as they are pushed out of their former ranges.

This may happen by invasive species or anthropogenic factors in the case of species, and in the case of languages by the expansion of a language or a language family that comes to be spoken in regions that previously had had other languages (as we have seen for Basque).

Traditionally, biologists have thought that when the geographical range of a species contracts, this would likely begin in the peripheries of the region, which typically offer only less-than-optimal habitats and where the density of populations is less even and dense. Hence, in peripheral regions individuals would be more vulnerable to disruptive factors, while core populations would be less so and therefore persist longer. However, Channell and Lomolino (2000) have shown that the locales where species survive the longest typically is situated exactly at the periphery of the larger, original range. For instance, the Tasmanian tiger (*Thylacinus cynocephalus*) originally occurred throughout New Guinea and Australia, and received its name from its last refugium, the island of Tasmania at the southeasternmost periphery of the original range. The characteristics of these refugia as described by Channell and Lomolino (2000) are “those along the edge of the range, on an isolated and undisturbed island, or at high elevations”, a type of location that we have encountered before – in language dynamics.

Conclusion

Here, I have presented an overview on the puzzle of global language diversity. Highly unevenly distributed across different regions of the world, language diversity is integrated with and sustained by societies and their respective views on language, language diversity and what role they should play. I have presented a model for understanding language diversity that, in contrast to most extant work, is based on qualitative case studies of how the range of languages contract in the wake of language expansion. This model makes references to environmental variables that promote or inhibit the social, cultural, political, and economic dynamics that are associated

with language spread and thus leaves room for human agency. Furthermore, the model takes serious the fact that languages are cultural, not biological products, and does not require a brute evolutionary view on cultural and linguistic diversity. At the same time, it still opens a perspective on the dynamics of linguistic diversity that can be related meaningfully to the patently similar dynamics of biological diversity.

I acknowledge that there is a lot that is not yet understood, and that loose threads remain. One point that I wish to highlight here is the general applicability. Here, I have suggested that high-diversity zones arise because large-scale language expansion processes culminate before they are reached, and that they do so because of less favorable environments for speakers of spreading languages. However, accretion zones are also found in California, with a climate that provides suitable conditions for a reliable food and subsistence base year round – this is consistent with Nettle’s ecological risk hypothesis, but not necessarily the range reduction-based dynamic model that I have sketched here. Eventually, it may be the case that we must reckon with non-stationary effects of different environments on language diversity levels (Pacheco Coelho *et al.*, 2019) and the way they shape ideologies that sustain highly diverse linguistic landscapes. In other words, we might have to distinguish several types of residual zones (as is now done for spread zones: Nichols, 2015), created by different diachronic cultural and linguistic dynamics. This would also be consistent with the “non-stationary” nature of language ideologies, which express different attitudes towards multilingualism and which involve different roles of language, language variation, and linguistic differences in constituting ethnic or social identities.

Ethics and consent

Ethical approval and consent were not required.

Data availability statement

No data are associated with this article.

References

- Antunes N, Schievenhövel W, d’Errico F, *et al.*: **Quantitative methods demonstrate that environment alone is an insufficient predictor of present-day language distributions in New Guinea.** *PLoS One*. 2020; **15**(10): e0239359.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Appel R, Muysken P: **Language contact and bilingualism.** Amsterdam: Amsterdam University Press, 2005.
[Reference Source](#)
- Axelsen JB, Manrubia S: **River density and landscape roughness are universal determinants of linguistic diversity.** *Proc Biol Sci*. 2014; **281**(1784): 20133029.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Babel AM: **Between the Andes and the Amazon: language and social**

meaning in Bolivia. Tucson: University of Arizona Press, 2018.
[Publisher Full Text](#)

Bentz C, Dediou D, Verkerk A, *et al.*: **The evolution of language families is shaped by the environment beyond neutral drift.** *Nat Hum Behav*. 2018; **2**(11): 816–821.

[PubMed Abstract](#) | [Publisher Full Text](#)

Bickel B: **Large and ancient linguistic areas.** In: Mily Crevels and Pieter Muysken (eds.): *Language dispersal, diversification, and contact: a global perspective.* Oxford: Oxford University Press, 2020; 78–100.

[Publisher Full Text](#)

Brooks GE: **Landlords and strangers: ecology, society, and trade in Western Africa, 1000-1630.** Boulder: Westview, 1993.
[Publisher Full Text](#)

- Cardillo M, Bronham L, Greenhill SJ: **Links between language diversity and species richness can be confounded by spatial autocorrelation.** *Proc Biol Sci.* 2015; **282**(1809): 20142986.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cashdan E: **Ethnic diversity and its environmental determinants: effects of climate, pathogens, and habitat diversity.** *Am Anthropol.* 2001; **103**(24): 968–991.
[Publisher Full Text](#)
- Channell R, Lomolino MV: **Dynamic biogeography and conservation of endangered species.** *Nature.* 2000; **403**(6765): 84–86.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Collard IF, Foley RA: **Latitudinal patterns and environmental determinants of recent human cultural diversity: do humans follow biogeographical rules?** *Evol Ecol Res.* 2002; **4**: 371–383.
[Reference Source](#)
- Cristofaro S: **Taking diachronic evidence seriously: result-oriented vs. source-oriented explanations of typological universals.** In: Karsten Schmidtk-Bode, Natalia Levshina, Susanne Maria Michaelis, and Ilja Seržant (eds.): *Explanation in typology: diachronic sources, functional motivations and the nature of the evidence.* Language Science Press, 2019; 25–46.
[Publisher Full Text](#)
- Currie TE, Mace R: **Political complexity predicts the spread of ethnolinguistic groups.** *Proc Natl Acad Sci U S A.* 2009; **106**(18): 7339–7344.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cysouw M, Good J: **Langvoid, Doculect and Glossonym: formalizing the notion “language”.** *Lang Doc Conserv.* 2013; **7**: 331–359.
[Reference Source](#)
- Derungs C, Köhl M, Weibel R, *et al.*: **Environmental factors drive language density more in food-producing than in hunter-gatherer populations.** *Proc Biol Sci.* 2018; **285**(1885): 20172851.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dryer MS: **Large linguistic areas and language sampling.** *Stud Lang.* 1989; **13**(2): 257–292.
[Publisher Full Text](#)
- Du Bois JW: **The discourse basis of ergativity.** *Language.* 1987; **63**(4): 805–855.
[Publisher Full Text](#)
- Epps P: **Amazonian linguistic diversity and its sociocultural correlates.** In: Mily Crevels and Pieter Muysken (eds.): *Language dispersal, diversification, and contact: a global perspective.* Oxford: Oxford University Press, 2020; 275–290.
[Publisher Full Text](#)
- Evans N: **The dynamics of language diversity.** In: Rajend Mesthrie and David Bradley (eds.): *The dynamics of language: plenary and focus lectures from the 20th International Congress of Linguists.* Cape Town, July 2018. Claremont: UCT Press, 2018; 12–35.
[Reference Source](#)
- Evans N, Levinson SC: **The myth of language universals: language diversity and its importance for Cognitive Science.** *Behav Brain Sci.* 2009; **32**(5): 429–448.
[Publisher Full Text](#)
- Evans N, McConvell P: **The engima of Pama-Nyungan expansion in Australia.** In: Roger Blench and Matthew Spriggs (eds.): *Archaeology and language II: correlating archaeological and linguistic hypotheses.* London/New York: Routledge, 1998; 174–191.
[Reference Source](#)
- Forkel R, Hammarström H: **Glottocodes: identifiers linking families, languages and dialects to comprehensive reference information.** *Semant Web.* 2022; **13**(6): 917–924.
- François A: **Social ecology and language history in the Northern Vanuatu linkage: a tale of divergence and convergence.** *J Hist Linguist.* 2011; **1**(2): 175–246.
[Publisher Full Text](#)
- François A: **The dynamics of linguistic diversity: egalitarian multilingualism and power imbalance among northern Vanuatu languages.** *Int J Sociol Lang.* 2012; **214**: 85–110.
[Publisher Full Text](#)
- Gavin MC, Botero CA, Bowern C, *et al.*: **Toward a mechanistic understanding of linguistic diversity.** *BioScience.* 2013; **63**(7): 524–535.
[Publisher Full Text](#)
- Gavin MC, Rangel TF, Bowern C, *et al.*: **Process-based modelling shows how climate and demography shape language diversity.** *Glob Ecol Biogeogr.* 2017; **26**(5): 584–591.
[Publisher Full Text](#)
- Gavin MC, Stepp JR: **Rapoport's rule revisited: geographical distributions of human languages.** *PLoS One.* 2014; **9**(9): e107623.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Geertz C: **The interpretation of cultures.** New York: Basic books, 1973.
[Reference Source](#)
- Gorenflo LJ, Romaine S, Mittermeier RA, *et al.*: **Co-occurrence of linguistic and biological diversity in biodiversity hotspots and high biodiversity wilderness areas.** *Proc Natl Acad Sci U S A.* 2012; **109**(21): 8032–8037.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Greenberg JH: **Diachrony, synchrony and language universals.** In: Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.): *Universals of human language, method and theory.* Stanford: Stanford University Press, 1978; **1**: 61–92.
- Grollemund R, Branford S, Bostoen K, *et al.*: **Bantu expansion shows that habitat alters the route and pace of human dispersals.** *Proc Natl Acad Sci U S A.* 2015; **112**(43): 13296–13301.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Güldemann T, Hammarström H: **Geographical axis effects in large-scale linguistic distributions.** In: Mily Crevels and Pieter Muysken (eds.): *Language dispersal, diversification, and contact: a global perspective.* Oxford: Oxford University Press, 2020; 58–77.
[Publisher Full Text](#)
- Hillebrand H: **On the generality of the latitudinal diversity gradient.** *Am Nat.* 2004; **163**(2): 192–211.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hua X, Greenhill SJ, Cardillo M, *et al.*: **The ecological drivers of variation in global language diversity.** *Nat Commun.* 2019; **10**(1): 2047.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hymes D: **Introduction: toward ethnographies of communication.** *Am Anthropol.* 1964; **66**(6): 1–34.
[Publisher Full Text](#)
- Irvine JT, Gal S: **Language ideology and linguistic differentiation.** In: Paul V. Kroskrity (ed.): *Regimes of language: ideologies, politics, and identities.* Santa Fe: School of American Research Press, 2000; 35–84.
- Laycock DC: **Melanesian linguistic diversity: a Melanesian choice?** In: R.J. May and Hank Nelson (eds.): *Melanesia: beyond diversity.* Canberra: Research School of Pacific Studies, Australian National University, 1982; **1**: 33–8.
- Luebbing CR: **Displaying the geography of language: the cartography of language maps.** *Linguist J.* 2013; **7**(1): 39–67.
- Lupyan G, Dale R: **Language structure is partly determined by social structure.** *PLoS One.* 2010; **5**(1): e8559.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lüpke F: **Uncovering small-scale multilingualism.** *Crit Multiling Stud.* 2016; **4**(2): 35–74.
[Reference Source](#)
- Mace R, Pagel M: **A latitudinal gradient in the density of human languages in North America.** *Proc R Soc B Biol Sci.* 1995; **261**(1630): 117–121.
[Publisher Full Text](#)
- Maffi L: **Linguistic, cultural, and biological diversity.** *Annu Rev Anthropol.* 2005; **34**: 599–617.
[Publisher Full Text](#)
- Manne LL: **Nothing has yet lasted forever: current and threatened levels of biological and cultural diversity.** *Evol Ecol Res.* 2002; **5**(4): 517–527.
[Reference Source](#)
- Moore JL, Manne L, Brooks T, *et al.*: **The distribution of cultural and biological diversity in Africa.** *Proc Biol Sci.* 2002; **269**(1501): 1645–1653.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nettle D: **Language diversity in West Africa: an ecological approach.** *J Anthropol Archaeol.* 1996; **15**(4): 403–438.
[Publisher Full Text](#)
- Nettle D: **Explaining global patterns of language diversity.** *J Anthropol Archaeol.* 1998; **17**(4): 354–374.
[Publisher Full Text](#)
- Nichols J: **Linguistic diversity in space and time.** Chicago/London: University of Chicago Press, 1992.
[Publisher Full Text](#)
- Nichols J: **Modeling ancient population structures and movement in linguistics.** *Annu Rev Anthropol.* 1997; **26**: 359–384.
[Publisher Full Text](#)
- Nichols J: **The vertical archipelago: adding the third dimension to linguistic geography.** In: Peter Auer, Martin Hilpert, Anja Stukenbrock, and Benedikt Szmrecsanyi (eds.): *Space in language and linguistics: geographical, interactional, and cognitive perspectives.* Berlin/Boston: Walter de Gruyter, 2013; 38–60.
[Publisher Full Text](#)
- Nichols J: **Types of spread zones: open and closed, horizontal and vertical.** In: Rik De Busser and Randy J. LaPolla (eds.): *Language structure and environment: social, cultural, and natural Factors.* Amsterdam/Philadelphia: John Benjamins, 2015; 261–286.
[Publisher Full Text](#)
- Norder S, Becker L, Skirgård H, *et al.*: **glottospace: R package for language mapping and geospatial analysis of linguistic and cultural data.** *J Open Source Softw.* 2022; **7**(77): 4303.
[Publisher Full Text](#)
- Pacheco Coelho MT, Pereira EB, Haynie HJ, *et al.*: **Drivers of geographical patterns of North American language diversity.** *Proc Biol Sci.* 2019; **286**(1899): 20190242.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pakendorf B, Dobrushina N, Khanina O: **A typology of small-scale multilingualism.** *Int J Bilingual.* 2021; **25**(4): 835–859.
[Publisher Full Text](#)
- Sankoff G: **The social life of language.** Philadelphia: University of Pennsylvania

Press, 1980.

[Publisher Full Text](#)

Scott JC: **The art of not being governed: an anarchist history of upland Southeast Asia**. New Haven/London: Yale University Press, 2009.

[Publisher Full Text](#)

Shneiderman S: **Are the central Himalayas in Zomia? Some scholarly and political considerations across time and space**. *J Glob Hist*. 2010; 5(2): 289–312.

[Publisher Full Text](#)

Sutherland WJ: **Parallel extinction risk and global distribution of languages and species**. *Nature*. 2003; 423(6937): 276–279.

[PubMed Abstract](#) | [Publisher Full Text](#)

Thurston WR: **How exoteric languages build a lexicon: esoterogeny in West New Britain**. In: Ray Harlow and Robin Hooper (eds.): *VICAL 1: Oceanic Languages. Papers from the Fifth International Conference on Austronesian Linguistics*. Auckland: Linguistic Society of New Zealand, 1989; 555–579.

Trask RL: **The history of Basque**. Abingdon/New York: Routledge, 1997.

[Publisher Full Text](#)

Trudgill P: **Sociolinguistic typology: social determinants of linguistic complexity**. Oxford: Oxford University Press, 2011.

[Reference Source](#)

Urban M: **Language ecologies and dynamics in the ancient Central Andes**. In: Matthias Urban (ed.): *The Oxford Guide to the Languages of the Central Andes*. Oxford: Oxford University Press. To appear.

Urban M: **Sprachlandschaften. Über die Rolle von Sprache in der Beziehung zwischen Mensch und Umwelt**. Würzburg: Königshausen & Neumann, 2018.

Urban M: **The geography and development of language isolates**. *R Soc Open Sci*. 2021; 8(4): 202232.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Urban M, Reyes-Centeno H, Bellamy K, et al.: **The areal typology of western Middle and South America: towards a comprehensive view**. *Linguistics*. 2019; 57(6): 1403–1463.

[Publisher Full Text](#)

Urcioli B: **Language and borders**. *Ann Rev Anthropol*. 1995; 24(1): 525–546.

[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:    

Version 2

Reviewer Report 22 January 2025

<https://doi.org/10.21956/openreseurope.20844.r49963>

© 2025 Inglese G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Guglielmo Inglese 

Universita degli Studi di Torino, Turin, Piedmont, Italy

I am pleased with the revised version of the paper and have no further comments to make.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: historical linguistics and linguistic typology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 30 December 2024

<https://doi.org/10.21956/openreseurope.19907.r46601>

© 2024 Väisänen T. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Tuomas Väisänen 

¹ Department of Geosciences and Geography, University of Helsinki, Helsinki, Uusimaa, Finland

² Helsinki Institute of Sustainability Science, University of Helsinki, Helsinki, Uusimaa, Finland

³ Helsinki Institute of Urban Regional Studies, University of Helsinki, Helsinki, Uusimaa, Finland

I welcome and enjoy any manuscript, such as this one, that bridges the gap between geography and linguistics. The article describes how linguistic diversity and its geographical distribution have

been studied in the field of linguistics over the past few decades, and proposes combining several theories (climatic, environmental, economic, language ideological) that attempt to explain why the diversity of languages varies geographically across the globe. The author manages to make a convincing case for his proposal, while pointing out remaining areas which still need to be worked on. In summary, I quite like the manuscript and find it makes a valuable contribution.

However, I have two minor conceptual issues and one minor structural issue. The remaining comments are even more minor technical stuff regarding typographical mistakes and textual clarity. First, I find it a bit confusing the author uses both "linguistic diversity" and "language diversity" while ostensibly referring to the same phenomena. However, if they are different, the distinction between these two concepts should be explained to the reader clearly. Now, it is not mentioned in the manuscript. Second, the author does not state which type of "linguistic diversity" he means, which is puzzling as the introduction lays out the different meanings the concept has (number of languages, number of language families, structural diversity). This is crucial information as it is the central concept, and it is operationalized throughout the manuscript. My take after reading the manuscript is that the author is mostly talking about the number of different languages, unless explicitly mentioning language families, but this should be explicitly stated. Third, the "Current perspectives" and "Dynamizing linguistic diversity" sections would benefit from sub-sectioning (and potentially some minor restructuring) as they are long and the reader has forgotten the three main perspectives or points by the end of the section.

Overall, the article is already in rather good shape, and my recommendation is a minor revision for the above-mentioned reasons. Although, I must confess that I am a geographer first and foremost, so I am leaning on the other reviewers' remarks on the author's grasp on the theoretical literature on historical linguistics.

Below are my detailed comments:

Introduction

- I have a comment on (or perhaps a slight issue with) the author's delineation of three types of understanding "language diversity". I argue there is a fourth type of language diversity, that draws on methods commonly used in ecology to quantify the diversity of species (see work of Peukert, 2013 [Ref 1]; Hiippala T, et al., 2019 [Ref 2]; Hiippala T, et al., 2020[Ref 3]). In fact, I think this fourth type could be integrated with the first or second examples provided (the "language richness" and "language family richness" examples), as it is related to quantification of linguistic diversity (that is, the diversity of languages). However, this additional understanding of language diversity goes a bit further than just the sheer number of languages in a given study area, and is focused on several perspectives simultaneously: richness, dominance, evenness, and diversity.
 - These perspectives can be quantified (or estimated) by a wide selection of metrics, each with their own benefits and drawbacks. In this context, richness is in its simplest form the number of languages, as the author states, but not only that. Dominance describes whether the given sample is dominated by one or more languages. Evenness describes the between-class similarity of the distribution of languages, that is, if all languages are spoken by as many individuals. Diversity tries to incorporate both richness and evenness into one metric.
- Some unclear sentences and unnecessary self-referential language in the paragraph discussing Figure 3 (starting p3)

- Last paragraph of Introduction on page 5. Please provide a reference to this "particular interpretation of a prominent hypothesis". Furthermore, I don't see how the starting sentence of this paragraph is connected to the rest of the sentences in the paragraph.

Current perspectives

- General comment: I suggest some slight restructuring and adding subsections that correspond to the three dimensions (or are they perspectives, as the section title leads the reader to believe?) laid out here. This would help the reader follow the arguments made better.
- In the starting paragraph: what is/are "lowbeds linguistic landscapes"?
- Can there be one sentence paragraphs in an ORE article? I am referring to the paragraph starting with "Early studies (Cashdan, 2001....". If not, I would integrate this somewhere else or flesh it out.
- The sentence "As a whole, research, reassuringly, shows the typical signs of maturation of a field of scientific investigation". Please rephrase this sentence. I take it, you mean this specific subfield of research has matured, but now you're referring to research in general. A similar thing happens in the last paragraph of the "Current perspectives" section in the beginning of the paragraph.
- The use of "largest imaginable scale" is a bit vague and over the top here. I presume you mean "globally", please clarify.
- Please rephrase the last sentence of this paragraph, it is very long and difficult to understand.

Dynamizing linguistic diversity

- General comment on this section: I think this section would improve if you could divide it into subsections that relate to the three interrelated points you want to make, especially as it is the section containing the main contribution by the author's own words. Now, it is a bit hard to follow how you address these points.
- Last sentence of the first paragraph. Please elaborate how your approach departs from the case studies.
- Paragraph starting with "Nichols also characterizes..." on page 8 requires a reference to the work(s) of Nichols being referred to here.
- On page 9, there is a partially missing subordinate clause in the sentence starting with "As a result, the oldest layers of the diverse linguistic landscape..."

Conclusion

- The first sentence is very long. Please consider dividing it into two or more sentences.
- Another paragraph of only one sentence here. Please flesh it out, or integrate somewhere else.
- The last sentence has a missing space between words.

References

1. Peukert H: Measuring language diversity in urban ecosystems. **2**: 75-96 [Publisher Full Text](#)
2. Hiippala T, Hausmann A, Tenkanen H, Toivonen T: Exploring the linguistic landscape of geotagged social media content in urban environments. *Digital Scholarship in the Humanities*. 2019; **34** (2): 290-309 [Publisher Full Text](#)
3. Hiippala T, Väisänen T, Toivonen T, Järv O: Mapping the languages of Twitter in Finland: *Neuphilologische Mitteilungen*. 2020; **121** (1): 12-44 [Publisher Full Text](#)

Is the topic of the essay discussed accurately in the context of the current literature?

Yes

Is the work clearly and cogently presented?

Partly

Is the argument persuasive and supported by appropriate evidence?

Yes

Does the essay contribute to the cultural, historical, social understanding of the field?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Urban geography, linguistic diversity, geoinformatics, regional science

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 02 December 2024

<https://doi.org/10.21956/openreseurope.19907.r45031>

© 2024 François A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Alexandre François

¹ LaTTiCe, Centre National de la Recherche Scientifique, Paris, France

² Ecole Normale Supérieure, Paris, Île-de-France, France

³ Linguistics, Australian National University School of Culture History and Language, Canberra, Australian Capital Territory, Australia

The paper “Global language geography and language history: challenges and opportunities” raises the question of linguistic diversity and of the uneven distribution of languages around the world. It discusses what different factors can account for it – including geographical, ecological, economical, ideological factors. This text is rather rich in references, taken from different disciplines relevant to this discussion. Its argument is sound and convincing.

The section on geographical and ecological factors – around the notion of “ecological risk” – is insufficiently clear about its subject. We understand that access to food and resources is an important key, but key to what exactly? to increasing human density in general (regardless of how many languages are spoken)? to favouring more numerous, smaller-size communities? Do these economic factors favour linguistic *density* (number of languages per capita)? or do they go as far as fostering linguistic *diversity* strictly speaking – i.e. the structural distance between neighbouring

languages? I wish the author brought more clarity to the discussion, so the reader can see exactly what would be direct causes vs. indirect factors explaining language diversity.

The discussion about language ideologies would gain from incorporating the notion of "egalitarian multilingualism": this attitude among neighbouring groups has been described in the literature (e.g. on Vanuatu, and Melanesia more generally) as an important factor in the maintenance or even the increase of linguistic diversity. In many parts of the world, language is taken as emblematic of place, an attitude which has been shown to intensify the divergence of languages over time. This results in a social ecology where smaller language communities are favoured, bringing about a tightly fragmented linguistic landscape (cf. studies on Vanuatu, or on Australia).

The passage on spread vs. residual zones is an excellent synthesis of J. Nichols's reflection on different language ecologies.

On the formal side, the current version of the text still has many typos, as well as some stylistic imperfections (excessive hypotaxis; repetitions; some unclear passages), and minor errors (e.g. "ephemeral" instead of "epiphenomenal"; "southwesternmost" for "southeasternmost"...). Yet this could be easily corrected by the author.

Overall, this paper is a solid synthesis of the last decades of reflection on an important topic, namely the factors underlying the distribution of languages around the planet. While it says little about the languages themselves (e.g. which aspects tend to change vs. which tend to be stable over time?), it does bring together the viewpoints of various disciplines, to construct an inspired reflection that bridges linguistics and human geography.

Is the topic of the essay discussed accurately in the context of the current literature?

Yes

Is the work clearly and cogently presented?

Partly

Is the argument persuasive and supported by appropriate evidence?

Yes

Does the essay contribute to the cultural, historical, social understanding of the field?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: historical linguistics; linguistic typology; social ecology and language ideologies

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 22 Dec 2024

Matthias Urban

Thank you for reading the article and for your extensive and thoughtful comments! I have taken them into account in revising the manuscript. I have clarified that indeed language density is what is thought to depend on "ecological risk", this indeed wasn't fully clear. You will note that I make quite extensive reference to Vanuatu with reference to your work in the revised version. This serves multiple purposes: on the one hand, it provides an opportunity for actual linguistic exemplification requested by reviewer #1; it provides an opportunity to give an example of a local language ecology and its linguistic outcomes; and it provides an opportunity to introduce "egalitarian multilingualism", as you request. Finally, I have rewritten the text relatively extensively to make it less prolix and easier to follow the line of thought, and I have emended the wrong choices of words you noted (thank you for your close reading!).

Competing Interests: No competing interests were disclosed.

Reviewer Report 28 November 2024

<https://doi.org/10.21956/openreseurope.19907.r45382>

© 2024 Inglese G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Guglielmo Inglese**

Universita degli Studi di Torino, Turin, Piedmont, Italy

The paper offers a fresh and stimulating overview current approaches to explaining linguistic diversity and the distribution of languages worldwide. It explores a number of shortcomings in existing research that links linguistic diversity to environmental factors and proposes, as an alternative, a focus on the historical process of language spread and recess that result in synchronic language diversity patterns. The paper is overall well written and provides an exhaustive and informed survey of the topic, while also laying a clear ground for future research. Its argumentation is nicely supported by a number of specific examples (e.g. the situation of Basque and that of the languages of the Caucasus). I have a few suggestions on points that deserve further improvement. First, a definition of what counts as a language is missing: this is both an empirical issue (in the sense that is not clear on which data figures 1 to 3 are based) as well as a more theoretical one. Even a working definition, such as that provided by Gil 2016 (but other definitions may be good as well, this is just one example), could make the point clearer. Second, it might be worth stressing that a shift in focus to processes in explaining linguistic diversity finds a parallel in current trends in source-oriented typology (I'm thinking of a number of papers by Sonia Cristofaro, e.g. Cristofaro 2019 (Ref 3) among others). These represent different ways in which historical thinking may reshape our understanding of linguistic diversity. Finally, the brief mention of the connection between the study of worldwide linguistic diversity with

dialectology/variationist sociolinguistics might be expanded upon (see the classic [Trudgill 2011](#) and the discussion in Inglese & Ballarè 2023).

References

1. Inglese G, Ballarè S: 1 Analyzing language variation: Where sociolinguistics and linguistic typology meet. 2023. 1-28 [Publisher Full Text](#)
2. Gil D: Describing languoids: When incommensurability meets the language-dialect continuum. *Linguistic Typology*. 2016; **20** (2): 439-462 [Publisher Full Text](#)
3. S, Cristofaro: Taking diachronic evidence seriously: Result-oriented vs. source-oriented explanations of typological universals. <https://zenodo.org/records/2583806>. 2019. [Publisher Full Text](#)

Is the topic of the essay discussed accurately in the context of the current literature?

Yes

Is the work clearly and cogently presented?

Yes

Is the argument persuasive and supported by appropriate evidence?

Yes

Does the essay contribute to the cultural, historical, social understanding of the field?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: historical linguistics and linguistic typology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 22 Dec 2024

Matthias Urban

Thank you for reading and commenting on the article! I have added relatively extensive discussion of where the data come from (viz. Glottolog) and the policies for assignment of coordinates. Extended discussion of the notion of "languoid" that is relevant here as well is in note [1]. A reference to "sociolinguistic typology" à la Trudgill is added to the main text, and a reference to dynamic diachronic approaches to typology in note [3].

Competing Interests: No competing interests were disclosed.

Reviewer Report 05 November 2024

<https://doi.org/10.21956/openreseurope.19907.r45380>

© 2024 Brunn S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Stan Brunn

University of Kentucky, Lexington, Kentucky, USA

An original, insightful and well referenced and written statement exploring linguistic diversity in new contexts: environmental and climate variable AS WELL as different continental and regional settings. I would ask the author to address mapping the location of some groups to show the difficulties in mapping the "degrees" of diversity in some historical and contemporary contexts. This focus would again address the problems of not only mapping "where a language is spoken" but also how "slippery" it is to map locations across a linguistic space, but also over time. Second, the paper could/would be stronger IF some examples of words or phrases were used by certain groups over time or in a place. A most thought-provoking contribution.

Is the topic of the essay discussed accurately in the context of the current literature?

Yes

Is the work clearly and cogently presented?

No

Is the argument persuasive and supported by appropriate evidence?

Yes

Does the essay contribute to the cultural, historical, social understanding of the field?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: the intersections of social, cultural, political and environmental geographies.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 22 Dec 2024

Matthias Urban

Thank you for reading and commenting on this article. I have included the additional information you requested in the revised version, including sample linguistic data from Vanuatu that illustrate different aspects of linguistic diversity

Competing Interests: No competing interests were disclosed.