

Implication of the Geographical Axis for international scale language research Implicações do eixo geográfico na investigação da escala linguística internacional

Alexandru Marius Tătar

PhD student Babeş-Bolyai University Cluj – Napoca
Faculty of Geography, Doctoral School of Geography, Romania
Alexandrumarius232@gmail.com
<https://orcid.org/0000-0002-8624-7260>

Artigo recebido a 21 de junho de 2023 e aprovado a 15 de setembro de 2023

Abstract

The research of the article is based on theories that analyze the links between Geography and Linguistics. A core theory in the analysis is Diamond's, 1999 idea of the different population histories of continental areas. Güldemann (2008, 2010) speculatively proposed that macro-areal aggregations of linguistic features might be influenced by large-scale geographical factors. In line with Diamond's geographical axis hypothesis, it is assumed that the way linguistic features assemble timescales and large geographical space is determined, among other variables, by two factors. These are the potential latitudinal spread and the longitudinal spread of the constraint. This paper examines the links between language and geographic space. The exposition of longitudinal degrees is a fundamental aspect of research on the linguistic axis. Regarding the second factor, spatial constriction provides the first results suggesting that linguistic diversity within language families tends to be higher along the longitude axis. If these findings can be replicated by more extensive and diverse testing, they promise to become an important methodological basis for a comprehensive theory of human history across space and time within linguistics and beyond.

Keywords: Geographical Axis, Areal linguistics, Large-scale feature distributions.

Resumo

A presente investigação baseia-se em teorias que analisam as ligações entre a Geografia e a Linguística, nomeadamente através do contributo de Diamond (1999) sobre as diferentes histórias populacionais de áreas continentais e das aportações de Güldemann (2008, 2010), que propôs que as agregações macro-reais de características linguísticas podem ser influenciadas por fatores geográficos de grande escala. De acordo com a hipótese do eixo geográfico aventada por Diamond, supõe-se que a forma como os traços linguísticos reúnem escalas temporais e grandes espaços geográficos é determinada, entre outras variáveis, por dois fatores, a saber: a potencial extensão latitudinal e a extensão longitudinal do constrangimento. Este artigo examina as relações entre a língua e o espaço geográfico. A exposição dos graus longitudinais é um aspeto fundamental da investigação sobre o eixo linguístico. Relativamente ao segundo fator, a constrição espacial fornece os primeiros resultados que sugerem que a diversidade linguística dentro das famílias linguísticas tende a ser maior ao longo do eixo da longitude. Se estes resultados puderem ser reproduzidos através de testes mais extensos e diversificados, podem vir a tornar-se uma base metodológica importante para a configuração de teoria abrangente da história humana através do espaço e do tempo, no âmbito da linguística e não só.

Palavras-chave: eixos geográficos, *Areal linguistics*¹, Distribuição de características em grande escala.

1. Introduction: areal linguistics and linguistic areas

Building on the observation that the histories of human populations have been quite different across

distinct continental areas, Diamond (1999, chapter 7) argued that in the long term, the historical dynamics of continents are decisively determined by the orientation of their geographical axis: geographical spread and linguistic characteristics are facilitated of the latitudinal axis but the geographical spread of linguistic characters along the longitudinal axis is problematic because languages of diverse structures and different families tend to share common patterns if they are spoken in geographic proximity this convergence is often explained by horizontal

¹ Nota da tradutora: por *Areal linguistics* entende-se o estudo de áreas/ regiões linguísticas cujas línguas exibem propriedades comuns, fruto da proximidade geográfica e do contacto linguístico e cultural entre si. Por se tratar de uma área dos estudos linguísticos com escassa tradição entre nós, optámos por deixar a expressão em inglês. Expresso aqui o meu agradecimento à Professora Doutora Liliana Inverno pela sua preciosa dilucidação sobre esta questão.

diffusibility, which is typically ascribed to language contact. In such a scenario, speakers of two or more languages interact and influence each other's languages, and in this interaction, more grammaticalized features tend to be more resistant to diffusion compared to features of more lexical content.

I take as a case study the worldwide distribution of nominal classification systems (grammatical gender, noun class and classifier) to show that more grammaticalized systems, such as gender, and less grammaticalized systems, such as classifiers, are almost equally widespread, but the former has spread more through the historical expansion of the language, while the latter has spread more through feature diffusion. Initial results indicate that quantitative models measuring areal diffuseness and stability of language features are likely to be affected by language expansion occurring by historical coincidence. I anticipate that the results will support studies of linguistic diversity in a more sophisticated way, with relevance to other parts of language such as phonology. This would be because climatic-ecological factors remain more homogeneous in the direction of capitalization, and therefore the conditions of human adaptation and subsistence depend on historical expansion at the level of latitudinal axis and climatic factors characteristic of the degree of latitude encountered.

These two interrelated phenomena, so-to-speak "the two sides of one coin," will be called henceforth the "latitude spread potentially" and the "longitude spread constraint".

Trying to interpret his findings on large-scale distributions of linguistic features within the European continent, Güldemann (2008, 2010) proposed to extend Diamond's geographical-axis hypothesis in two directions.

First, it should not only be relevant for the spread of cultural artefacts and ideas but also for linguistic features associated with languages whose speakers migrate and/or are in contact with both on the level of individuals and entire communities. Second, if the relevant geographical area associated with a certain feature distribution is sufficiently large, the mechanism should not only leave traces in areas of continental but also smaller sub-continental size.

In other words, the geographical-axis effect (as a statistical tendency concerning numerous individual spreads of populations and their features in

space and time) can be expected to be a major factor influencing the formation of large-scale aggregations of linguistic features. This hypothesis needs to be seen before the background of the current general discussion in areal linguistics and within this sub-discipline on the assumed role of geography.

Concerning the first issue, it is important to recognize that in the recent past, there has been an intensified debate about the concept of "linguistic area." The distribution of linguistic features in the more than 7000 languages of the world (Hammarström, 2016; Hammarström et al., 2019) reflects a scenario where some features may have emerged and spread by horizontal diffusion, whereas others are represented by vertical stability within their lineage. Generally, different feature types vary concerning their inherent stability (Nichols, 1992; Dediu & Cysouw, 2013), which may reflect their functional role and cognitive preference. In the evolutionary dynamics of language, high stability implies that a feature has high gain and low death rates (attractor feature) whereas low stability implies that a feature has high gain and loss rates (unstable feature), or low gain and high death rates (recessive feature). Due to their cognitive preference, features of high stability can be both stable in lineage and diffuse by contact, but as a rule, features bound by morphology show a tendency to higher stability in the lineage (Carling & Cathcart, 2021). Both lexicon and grammar vary concerning their inherent stability (Haspelmath & Tadmor, 2009; Dediu & Cysouw, 2013) but in general, more grammaticalized features of grammar have higher stability rates than more lexical features and more frequent grammatical and lexical features have higher stability rates than less frequent features (Thomason & Kaufman, 1988; Wilkins, 1996; Matras, 2009). Even though lexical morphemes can be borrowed at varying degrees, grammatical morphemes are very seldom borrowed (Matras & Sakel, 2007). The most frequent lexical items of basic vocabulary have high stability rates and are usually not borrowed (Greenhill et al., 2017), but a majority of the lexicon has lower stability rates and is subject to borrowing at varying degrees (Haspelmath, 2009; Carling et al., 2019).

Grammaticality can be viewed as a continuum, ranging from the most grammatical items of grammar (frequent function words of low transparency) to the least grammatical items of the lexicon (cultural and non-frequent content words of high transparency)

(Matras & Sakel, 2007). Even though stability is a property that is independent of the grammar-lexicon axis, we expect to find the most grammaticalized items in the domain of high-stable and preferred features, whereas the least grammaticalized items of the lexicon are expected to be less stable in diachrony. While the distribution of every linguistic feature is likely to be shaped by both horizontal diffusion and vertical stability, few analyses based on real data have been proposed to examine how these two pathways simultaneously shape the distribution of specific language features in languages of the world. We aim at filling this gap by providing a case study on nominal classification systems.

This can be discerned from many more theoretically oriented papers, e.g., Muysken (2000, 2008), Thomason (2001, p. 99), Dahl (2001), Stolz (2002, 2006), Bisang (2006a, 2006b), Campbell (2006), Tosco (2008), Bickel and Nichols (2012), and Muysken et al.

In several of these works, one can notice a considerable tendency towards a pessimistic look at the usefulness of this concept. At the same time, it has been realized that a potential way out of the problem is proposing better integration of the contact-induced linguistic-area concept in a more general theory of areal linguistics. In this framework, the first step would have to focus on the mere fact that/whether a geographical distribution of linguistic features can be observed and is at all somehow significant. Within areal linguistics, a linguistic area is viewed then more neutrally as a distribution of features according to a non-trivial “compact” geographical entity independent of any historical (or other) explanation. At this stage, one would first be concerned with the statistical probability that an identified distribution is diagnostic of an interesting relation between the feature and the associated area instead of mere coincidence. While this issue could be considered essential, it cannot be pursued further in the context of this paper (Daumé 2009; Lucas, Cule & Mathieson, 2009).

A linguistic area in the traditional, narrow sense, entails a second step: it is a feature distribution according to a compact geographical interaction with a specific historical scenario, namely that language contact is the central explanation for the observed distribution, rather than coincidence, universal tendencies, and in particular genealogical inheritance. Henceforth, we will call this concept a

“(linguistic) contact(-induced) area” to avoid confusion between the two senses of the linguistic area entertained here.

The second issue of the relation between linguistic areas and geographical space is equally controversial. On the one hand, there is no doubt that at least some large-scale linguistic distributions are partly determined by geographical factors, e.g., the significantly higher linguistic diversity in the tropics (Nettle, 1999; Collard & Foley 2002).

Moreover, it is intuitively clear that geographical patterns and events are among the factors which determine linguistic history in space and time, including the presence, trajectory, and speed of contact-induced diffusion, and thus steer more generally the distributional dynamics of features. There are, of course, documented case studies on this topic (Bostoen, Grollemund & Muluwa, 2013). Given this, there is no *a priori* reason why the result of some such events should not linger on for a longer period in the form of a particular spatial distribution of one or more linguistic features.

At the same time, a causal role of geography has recently been denied explicitly for contact-induced areas by Campbell (2006). Linguistic borrowings are paramount and geographical areas are merely a reflection of them.

Finally, language speakers tend to stay in similar environments when they migrate (Nichols, 1992; Gray & Jordan, 2000; Ramat, 2012; Hock & Joseph, 2019). Therefore, if gender and noun class languages spread more by language expansion, we expect to find less variance within the natural environment surrounding their location.

I assume that features spreading by language expansion should have a smaller variance of the environmental factors that facilitate migration and farming (Antunes et al., 2020). We investigate environmental factors that are less likely to vary across geographical areas (Moore et al., 2002; Pacheco Coelho et al., 2019; Antunes et al., 2020). As an example, the mean temperature varies drastically across geographical areas, which is likely to affect its variance. We thus select these three environmental factors: low variation of elevation, distance to water bodies, and rainfall. Low variation of elevation is generally more suitable for farming, as topographically complex areas largely correspond to versatile ecosystems and may pose restrictions on settlement options (Hassan, 1975; Kavanagh et al., 2018), while

Table 1
Levels of scale in the analysis of linguistic contact areas

Level	Space	Time	Sources	Scenarios
Micro	Bilingual community	20-200 years	Fieldwork data	Specific contact scenarios
Meso	Geographical region	200-One thousand years	Comparative data; historical sources	Global contact scenarios
Macro	Larger areas of the world	Deep time	Typological, genetic, archaeological data	Vague or no contact scenarios

Source: Muysken, 2008.

accessibility to an adequate source of water (either by river or rainfall) is also one of the basic conditions considered when expanding and finding new settlements.

Inversely such big areal patterns involve greater time depths, and it is thus less likely that one can reconstruct the specific historical scenarios or even concrete events and circumstances that caused the synchronic picture. The last point is nicely captured by Muysken's (2008) holistic approach to language contact in general, which includes large contact areas.

As seen in Table 1, he introduces several levels of scale which differ in certain parameters, among them the kind of historical scenarios entertained (last column). In this context, we would venture to add dimension, namely the role of geography, which we assume to be small on the micro level and important on the macro level.

The way a linguistic feature ends up in a particular language variety that is spoken in a certain geographical location – the individual data point of a large-scale distribution pattern – is an extremely complex matter, normally without any direct relation between geographical space and feature.

The relation is instead mediated by intermediate layers, which requires one to consider:

- the feature within the linguistic system of an idiolectal variety;
- the idiolect as a member let of an abstract language;
- the language is spoken by an abstract population found in a certain location;
- the location in geographical space.

Movement of features through geographical space largely happens in two idealized ways: a) across geographically “stable” human populations through contact between them (and thus metaphorically between languages) or b) with/on geographically mobile human populations.

Such complex and indirect feature–geography interrelation, and thus the history of large distribution patterns, can only be captured by more abstract, metaphorical modelling the idea to conceptualize population features like potentially contagious viruses on a host has been distracted both outside and within linguistics – compare, e.g., Cullen (2000) & Enfield (2003, 2008).

In the present context, a yet more abstract concept of (linguistic) features as “particles in a liquid/pulp” is developed possibly even more appropriate.

What we propose here is that the distributional dynamics of linguistic features – the “particles” of the last metaphor – are not only steered by their more “active” inherent properties and/or the properties of their hosts, languages, speakers, populations) but also by the more “passive” reactive interplay of the different kinds of feature hosts with the geographical environment – the “liquid” – in which all these hosts emerge, thrive, and degrade. As introduced above, the relation between feature aggregations and one such environmental factor, the geographical axis, will be dealt with in the remainder of this paper.

2. Geographical axis effect

2.1. Values of the three environmental factors influencing the Geographical Axis model

Language contact is largely determined by complex and sparsely documented social factors (Hickey, 2010; Bower et al., 2011). Nevertheless, if the distribution of a feature is mostly influenced by feature diffusion, the feature is more likely to be found across languages from different language families located in geographic proximity. This can be explained by the fact that the diffusion would happen by language contact and diffuse from a language to its geographic neighbours, with few restrictions of family affiliation (Coupé et al., 2013).

In our case study, classifiers are expected to diffuse more than gender and noun class, thus expect that classifiers are more likely to be found across different language families in the same area. As for gender and noun class, if they expand more by language expansion, we expect that a gender or noun class language is less likely to have geographic

neighbours from different language families since languages from different families are more likely to have been pushed away and/or replaced by the family with gender.

To investigate this hypothesis, we divide the world map into 3267 grids (Derungs et al., 2018). For each grid and each feature, we count the number of language families represented by languages within the grid with the feature in question (Supplementary material 2.3). Our data show that the family density of classifier languages is indeed higher than the family density of gender languages ($w = 294,410$, $p < 0.001$) and noun class languages ($w = 171,006$, $p < 0.001$). The data also show that the family density of gender languages is significantly higher than the family density of noun class languages ($w = 117,264$, $p < 0.001$).

As a summary, we show that the geographical distribution of nominal classification systems is likely to have been influenced by the mechanisms of language expansion. Evidence from language family density, the geographical coverage of language families, and the variance of environmental factors

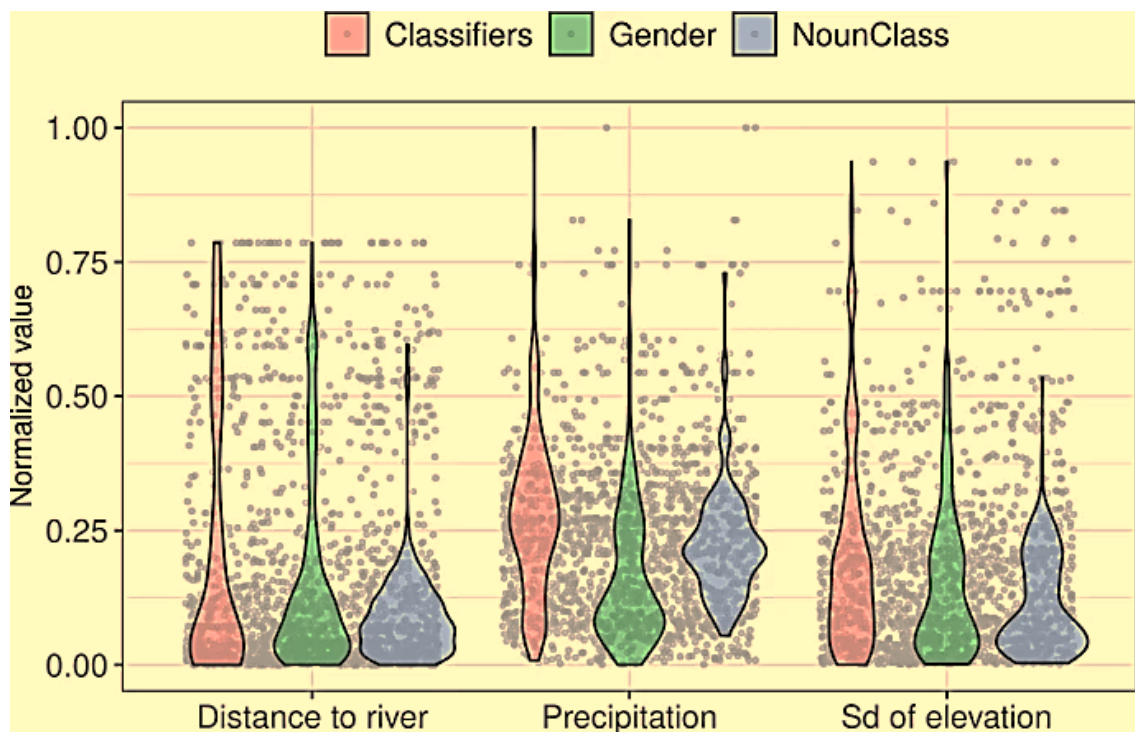


Figure 1

The values of the three environmental factors have been normalized to a scale from 0 to 1.
Source: produced by the author.



Figure 2
The 20 largest single-feature areas for numeral systems.
Source: produced by the author.

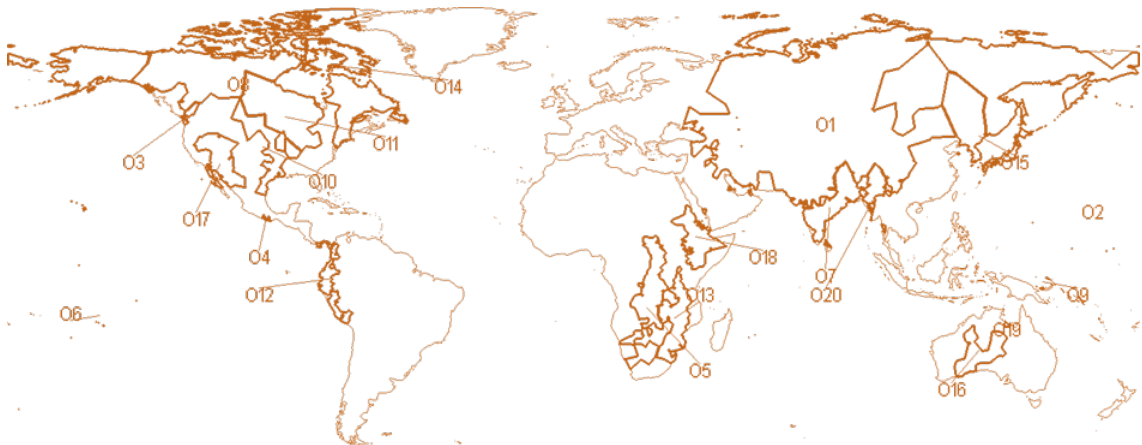


Figure 3
The 20 largest single-feature areas for transitive sentence order.
Source: produced by the author.

highlight the importance of distinguishing between the two mechanisms of language expansion and feature diffusion (Figure 1).

These two mechanisms are generally not distinguished in quantitative assessments of the horizontal stability of linguistic features; however, they can lead to similar results, while telling a drastically different story about the importance of grammaticalization in the diffusibility and the stability of the analyzed linguistic features. Our study also points out the importance of testing such an assumption for other linguistic features and other factors. I demonstrate how the effects of language expansion could be investigated in a case study of nominal classification systems. I encourage future studies to

replicate the analysis on other features related to phonology, syntax, and semantics, among others, to compare their dynamics in spreading.

I also encourage the building of evolutionary models to consider the impact of non-linguistic factors, such as language expansion, along with linguistic factors, such as grammaticalization, so that the spreading dynamics of linguistic features are modelled in a more accurately-scale feature distribution involving linguistic convergence

The hypothesis that macro-areas induced by multiple language contacts tend to have an east-west axis as a partial reflex of the latitude spread potential is at present hard to test systematically.

This is because there is no inventory of such areas that is large enough for statistical testing and would find enough agreement in the linguistic

community. For the time being, a way out of this dilemma is to investigate compact distributions of single linguistic features, irrespective of whether these overlap with other distributions, and determine whether they show any bias concerning their geographical axis.

One is on numeral systems containing data for 6837 languages (Hammarström 2010); the other one is on basic word orders in the transitive sentence representing 4653 languages (superseding the data of Dryer, 2005, Hammarström, 2007, and Lewis et al., 2013).

These two data sets were subjected to various tests one of which yielded relevant results for the geographical-axis hypothesis. The most direct evidence comes from the assessment of areas which are homogeneous concerning a certain feature value within the two linguistic domains. By joining geographical areas which have the same feature value and assigning each point on a map to its nearest language we obtain single-feature coherent areas. For details on this procedure, the reader is referred to Hammarström and Güldemann (submitted).

Figure 2 and Figure 3 – present a subset of the resulting linguistically homogeneous areas, namely the twenty largest ones (in terms of languages) for the two featured domains. Following the hypothesis, I suggest above, we can picture that the larger the areas the more their geographical shapes tend to be latitudinal rather than longitudinal. To measure the axis orientation, we take an area's East-West and North-South endpoints to get a distance East-West and a distance North-South in kilometres. An area's axis ratio is the ratio of dew/dns, whereby a value >1 means that the relevant area is more latitudinal. While there is a lot of variation in the axis ratio across the individual areas, correlating such axis ratios to geospatial size one can determine a mean of axis ratios for all areas that are geospatially larger than a certain threshold size.

2.2. Language families in large geographical space

The latitude spread potential and language family axis (Truncated line).

If the latitude spread potential is indeed a factor for large-scale linguistic distributions, it can be expected to have yet further implications. The logic of the hypothesis predicts that the East-West

trend holds for any type of historically mediated linguistic distribution which has a sufficient geographical size for environmental factors to come to bear. Although the propagation/transmission of features over space is different in genealogical language groups, henceforth called linguistic lineages, it can still be modelled in an abstract sense as clustered distributions of linguistic isoglosses forming large geographical areas.

Accordingly, Güldemann (2010: 582) hypothesized that linguistic lineages, too, may have a latitudinal rather than longitudinal axis orientation with growing geospatial expanded

This possible effect of the latitude spread potential can be tested more easily. In Güldemann and Hammarström (forthcoming), we used various data sets to test this hypothesis, because there is still no agreement in the linguistic community as to the exact genealogical composition of the world's languages.

The hypothesis on the first-order subfamilies of the two primary datasets is also analyzed (Campbell and Poser, 2008).

The way to determine the axis ratio of a lineage is essentially the same as that used for large-scale feature aggregations above. I used the geographical positions (centre points) of languages or individual sources for languages not listed there.

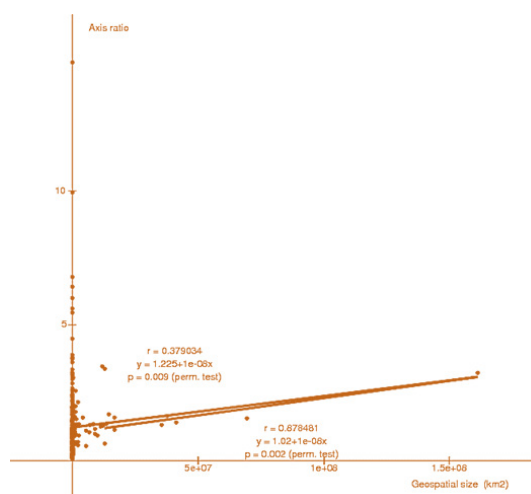


Figure 4
Size-axis ratio relationship for independent lineages according to G2 with regression lines fit the 50/10 largest ones.
Source: produced by the author.

Table 2
Relationship of axis-ratio and geospatial size across four data sets

Data set	Geometric mean of axis ratio	Linear regression for Fifty largest lineages	p	Linear regression for Ten largest lineages	p
G2	1.019	$r=0.39$	$<.01$	$r=0.89$	$<.01$
E17	1.144	$r=0.46$	$<.01$	$r=0.95$	$<.01$
G2 subfamilies	1.008	$r=0.13$	$=.12$	$r=0.99$	$<.05$
E17 subfamilies	1.084	$r=0.18$	$=.06$	$r=0.98$	$<.05$

Source: created by the author.

A lineage's axis ratio is the ratio between the East-West expansion dew (as the distance in km between the eastern and western endpoint languages of a lineage) and the North-South expansion dns as the distance in km between the northern and southern endpoint languages of the lineage.

A lineage's geospatial size for the pr is simply determined by multiplying dew and dns. When plotting the axis ratio against geospatial size, the above hypothesis is confirmed the geometric mean of the axis ratio of all lineages is commonly close to 1, that is, neutral concerning a latitudinal or longitudinal shape.

This is expected because many lineages are small and on a small geographical scale environmental factors should not make a discernible impact. However, from a certain size on, the expected linear relationship between the two dimensions emerges. Thus, for the G2 data set, as shown in Figure 3, taking only the 50 largest lineages linear regression gives a modest ($r=0.39$) but significant ($p<.01$) trend, and taking only the 10 largest ones gives a stronger ($r=0.89$) also significant ($p<0.1$) relationship.

This picture is largely replicated in the analysis of the data set and the first-order subfamilies of both G2 and E17. All results are summarized in Table 2.

Not all large lineages behave according to their geospatial size, because other more accidental contingencies like local geographical factors (e.g., geophysical barriers), histories that outplay any geographical constraints, etc. counteract the latitude spread potential.

The possibility of achieving a level playing field in terms of the geographical area of the spread of the language family is examined. I have therefore systematically studied the mean axis ratio according to variable size thresholds, improving on the arbitrary choice of the top 10/50 largest lineages, expecting to find a smoothly rising curve rather than a randomly fluctuating one. Figure 5 shows for the G2 data set the mean axis ratio of all lineages larger or equal to

the size of the nth largest lineage, as n ranges from all lineages down to one, confirming the expected trend. Again, this also holds overall for the other three data sets. while this tendency is weak for the class of small and medium-sized lineages, it nevertheless shows up when considering means.

When evaluating our conclusion that there is a trend of linguistic lineages towards a latitudinal shape which gains in prominence with increasing geospatial size, it is also important to take the three following points into account. First, the latitude bias correlates only with geospatial lineage size, in line with our hypothesis. The observed trend disappears or becomes weaker as soon as lineage size is determined by other criteria, e.g., the number of languages.

Second, we do not expect and do not find, that necessarily holds for all (sub)continental areas separately, which elsewhere is a good criterion for large-scale typological investigations (Dryer,1989). This is because other potentially opposing geographical factors can dampen or annihilate the latitude spread potential on a more local scale. Factors at issue in this respect seem to be particularly pronounced in Central and South America, in that the land masses have a strong north-south orientation (Diamond, 1999), the Andes have a strong north-south axis which may tend to steer the movement of (features associated with) human populations along longitudes, and even climate zones may be more longitudinal in South America (Ricklefs, 2001). Indeed, if evaluating the hypothesis on this geospatially large zone alone, no latitude bias of large lineages is found. With this background, it is important to acknowledge that the latitude spread potential is strong enough to show up on a worldwide scale.

Finally, considering the overall shape of the Earth's landmasses inhabited by humans one impressionistically observes that they are overall latitudinal and might thus be tempted to assume that this fact induces the global trend. We have designed a model

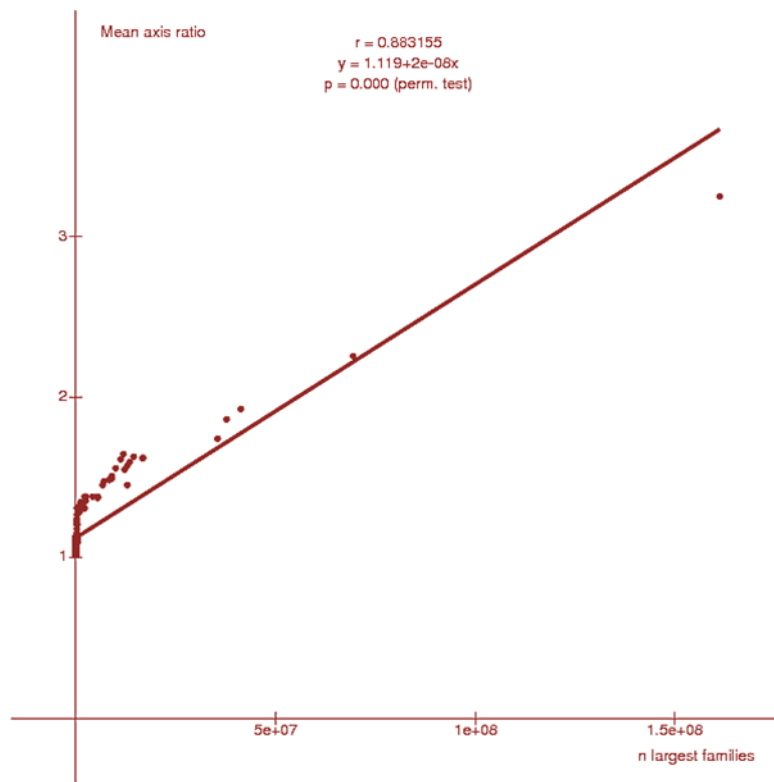


Figure 5
 Size-threshold vs. mean axis ratio relationship for independent lineages according to G2 with regression lines.
 Source: produced by the author.

of random growth of language families to control this possible geographical contingency for which the reader is referred to Güldemann and Hammarström (submitted). The result is that landmass shape alone can indeed be held responsible for a certain amount of latitude bias of lineages but importantly is not sufficient to account for the degree of bias found in the real world. We can thus conclude overall that our initial hypothesis remains valid that the latitude spread potential is a factor for the global trend of lineages to have a latitudinal rather than longitudinal axis orientation with growing geospatial size.

2.3. B. The longitude spread constraint and intra-lineage diversity

The inverse of the latitude spread potential is the longitude spread constraint, namely a propensity towards changing environmental conditions that speaking hamper rather than facilitate the long-distance spread of populations and their features. Clear North-South-South movements are possible and amply

attested historically. However, in these cases, a geographical axis effect can be expected to turn up unusually: generally speaking, populations more often confront different kinds of barriers when moving in longitudinally and are more prone to local adaptation and thus change.

To mention just one scenario directly relevant to Linguistics, greater environmental challenges encountered by a colonizing group in a new area are more likely to necessitate intensive contact with autochthonous populations, also leading potentially to considerable change in its linguistic profile, as hypothesized by Güldemann (2010, pp. 580-582).

Permed repetition of this phenomenon over a long time spans and sufficiently large space tends to steer intra-lineage linguistic diversity to be higher along the north-south axis and lower along the east–west axis.

An intuitively suggestive case in point seems to be pidgins and creoles with European lexifier languages, which are distributed around the Tropics and may structurally even cluster together

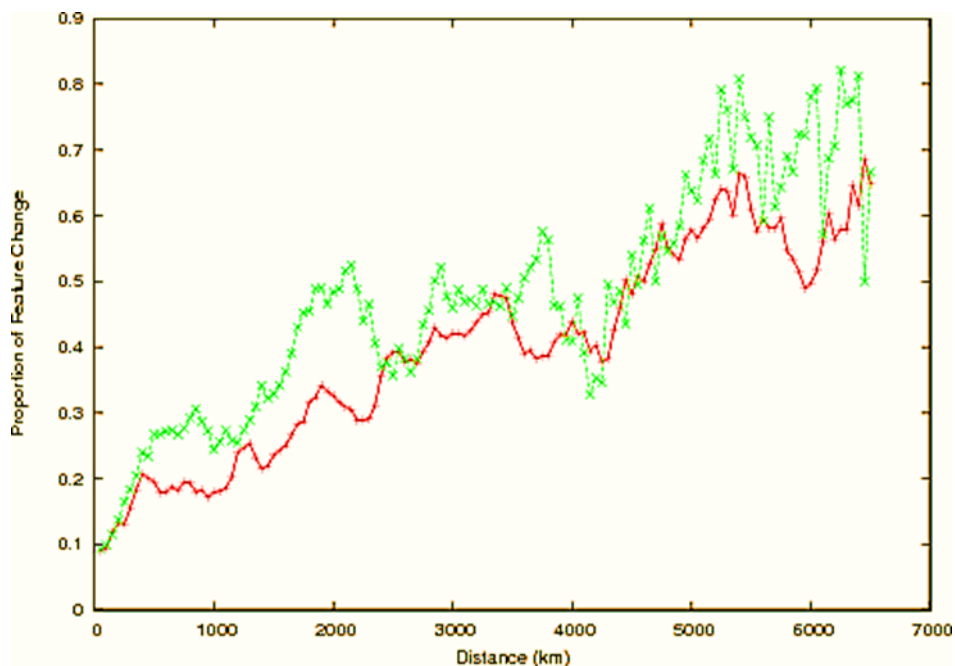


Figure 6

Proportion of change of transitive sentence order in latitudinal (red line) and longitudinal (green line) language pairs according to distance intervals.

Source: produced by the author.

irrespective of their location (Kortmann, 2013) on Anglophone linguistic variation). Mainstream historical-comparative linguistics does not treat them as members of the relevant European lineage. However, one way to model their very emergence can be potentially related to the phenomenon entertained here: the European languages drastically changed in a certain geographical, historical, and social setting, up to a point of arguably shifting away from their original genealogical alliance.

One aspect, viz. widespread demographic inequality of the colonizers vis-à-vis the other populations (Mufwene, 2001; 2008), was at least partly steered by a drastically different environment – a challenge which the colonizers could not compensate by their strong and long-lasting socio-political and economic dominance.

The above observations are still purely impressionistic, though. For a first more systematic test regarding an axis bias of intra-lineage linguistic diversity, we investigated the data on transitive sentence word order. First, I identified all pairs of genealogically related languages according to G2 (see Table 2); these numbered 713193. For the record, 62.7% of those displayed the same transitive sentence

order (which was the inverse of the situation for pairs of unrelated languages, where 65.5% of the total differed).

Then I classified the related language pairs according to whether they were more distant from each other on a latitudinal or longitudinal axis [henceforth just “latitudinal” and “longitudinal” (language) pairs]. We then determined the proportion of language pairs that disagree in the linguistic feature at regular distance intervals of 50 km in both sets of pairs. For example, a pair of languages having a latitude distance of 75 km and a longitude distance of 45 km is put in the set of longitudinal language pairs within the distance interval of 50-100 km.

In this interval, there happened to be 4537 longitudinal pairs and 4813 latitudinal pairs with a slight difference between them in terms of feature change, namely a proportion of 9.7% vs. 9.3%, respectively.

The overall results are shown in Figure 6. Up to 400 kms latitudinal and longitudinal language pairs hardly differ concerning feature change. From then on, the longitudinal pairs consistently and by a margin show more feature disagreement than the latitudinal pairs at the same distance level, with an exception at a point near 1300 km. From ca. 2500 kms on there

is again no clear pattern of difference between the two sets of language pairs, although some tops and dips may be due to individual language families.

After 5500 kms both the latitudinal and longitudinal pairs reach a random behavior of feature change found also with unrelated languages.

It is important to reflect on what is to be expected realistically under the above longitude constraint hypothesis. On short distances, where environmental factors overall do not differ latitudinal and longitudinal, language pairs can be assumed to behave similarly. In the same vein, language pairs that are geographically extremely far apart from one another tend to be separated by a different background so that any common heritage is likely to have vanished regardless of the pairs' axis configuration. In other words, pairs of related languages should behave like random language pairs at some distance level. So, it is at a certain distance interval that one can expect that longitudinal language pairs display a feature change more often than latitudinal ones. The finding of our admittedly very restricted and preliminary test is that this is indeed the case between 400 and 3000 km and is compatible with our expectations, indicating that our hypothesis deserves further exploration.

3. Conclusions

We have argued above that various kinds of linguistic data are compatible with the hypothesis that human migration or exchange and the accompanying spread of linguistic features over large distances are facilitated along latitude axes (tested for linguistic lineages and contact areas) and hampered along longitude axes (tested for contact areas). Thus, both latitudes spread potential and latitude spread constraint influence the expansion, sedimentation, and retention of linguistic features over long periods. This adds evidence to Diamond's idea that bio-geographical factors contribute to determining human history.

However, all tests reported above are only the first steps which need and can be refined to replicate our empirically still restricted findings. Concerning language families, the above test should be repeated with a genealogical classification of the world's languages that is increasingly consensual by being permanently and collectively updated using strict

and consistent criteria. For areas of language contact, both the range of areas and linguistic features need to be extended. Testing family-internal diversity and its geographical patterning in the future requires even more data breadth and methodological sophistication. First, a larger feature set using databases of the magnitude of the World Atlas of Language Structures (WALS) but with a higher density is necessary, as evidenced also in the approach.

Moreover, not only feature diversity but also intra-family phylogenetic structure should be investigated concerning graphical axis effects.

Despite the preliminary nature of the results presented above, this study may help to reinstate Geography (beyond the universally acknowledged aspect of distance) as an important factor for the dynamics of areal linguistics on the macro level. However, direct "mechanic" correlations between Geography and linguistic distributions cannot be expected. As is evident, e.g., from the "minimal size factor" concerning geographical axis effects, the patterns emerging from investigations of this type need instead to be embedded realistically in the relevant discipline and the nature of its phenomena and data.

Moreover, trying to explain large-scale feature aggregations with the help of geography can only provide one piece of an extraordinarily complex puzzle. Any synchronic pattern is the result of a complex and long-term interplay of many distinct factors; these may conflict, and their significance shift from the historical period to the historical period.

We also do not assume any form of extreme environmental determinism in the sense that geographical factors could not be outranked by other factors determining human behavior and history. Geographical factors have lost some of their impact on human population dynamics along the historical trajectory of our species in favour of other, notably sociocultural factors.

The basic idea behind our discussion has been pronounced most prominently outside Linguistics and indeed should be independent of it. One should therefore expect geographical axis effects in other non-linguistic aspects of human populations, too, such as Cultural Anthropology, Physical and Molecular Anthropology, Archaeology, etc. In some of these fields, some attempts to test the basic idea have indeed been made, e.g., by Turchin, Adams, and

Hall (2006) on the East-West orientation of large historical empires and modern states, and by Laitin, Moortgat and Robinson (2012) on increased retention of cultural diversity on a north-south axis. These studies turned out to be far more complex in lacking a sufficiently large database and/or having to control for several additional factors. Thus, linguistic features may prove to be particularly suitable for testing hypotheses like Diamond's.

Bibliography

- Antunes, N., Schiefenhövel, W., d'Errico, F., Banks, W E. & Vanhaeren, M. (2020). Quantitative methods demonstrate that environment alone is an insufficient predictor of present-day language distributions in New Guinea. *PLoS ONE* 15: e0239359 p.1-16
- Boston, Koen, Grollemund, Rebecca and Muluwa, Joseph K. (2013). Climate-induced vegetation dynamics and the Bantu expansion: evidence from Bantu names for pioneer trees (*Elaeis guineensis*, *Canarium schweinfurthii*, and *Musanga cecropioides*). *Comptes Rendus Geoscience* 345, 336-349.
- Bowern C. et al. (2011). Does lateral transmission obscure inheritance in hunter-gatherer languages? *PLoS ONE* 6: e25195, pp. 1-4.
- Campbell, Lyle (2006). Areal linguistics: closer scrutiny. In McMahon Matras and Vincent (eds.), *Linguistic Areas Convergence in Historical and Typological Perspective*, Edited by: Yaron Matras, April McMahon, Nigel Vincent, Palgrave Macmillan, 1-31.
- Carling G. (ed.) (2019). *The Mouton Atlas of languages and cultures*: vol. 1. Berlin: Mouton /De Gruyter.
- Carling G. & Cathcart C. (2021). Reconstructing the evolution of Indo-European grammar. *Language* 97, 561-598.
- Campbell, Lyle & Poser, William J. (2008). *Language Classification. History and Method*. Cambridge: Cambridge University Press.
- Cullen, Ben S. (2000). *Contagious ideas: on evolution, culture, archaeology, and cultural virus theory*. Oxford: Oxbow.
- Coupé C., Hombert, J. M., Marsico, E., Pellegrino, François (2013). Investigations into determinants of the diversity of the world's languages. In G. Peng & Shi F. (eds.) *Eastward flows the great river: Festschrift in honor of Professor William S-Y. Wang on his 80th birthday*. (pp. 75-108). Hong Kong: City University of Hong Kong Press.
- Dahl, Östen (2001). Principles of areal typology. In Martin Haspelmath et al. (eds.), *Language typology and language universals: an international handbook* (p.1456-1470), 2 vols. Berlin/ New York: Walter de Gruyter, vol. 2.
- Daumé, Hal, III. (2009). Non-parametric Bayesian areal linguistics. In *Non-Parametric Bayesian Areal Linguistics Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (NAACL '09). (pp. 593-601). Morristown, NJ, USA: Association for Computational Linguistics.
- Derungs, C., Köhl, M., Weibel, R. & Bickel, B. (2018). Environmental factors drive language density more in food-producing than in hunter-gatherer populations. *Proc R Soc B: Biol Sci*: 20172851. <https://doi.org/10.1098/rspb>.
- Diamond, Jared M. (1999, [1997]). *Guns, germs, and steel: the fates of human societies*. New York/ London: W. W. Norton.
- Dryer, MS. & Haspelmath, M. (eds.) (2013). *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dediu, D. & Cysouw, M. (2013) Some structural aspects of language are more stable than others: a comparison of seven methods. *PLoS ONE* 8:1-10. DOI: 10.1371/journal.pone.0055009
- Güldemann, Tom (2008). The Macro-Sudan belt: towards identifying a linguistic area in northern sub-Saharan Africa. In Heine and Nurse (eds.) *A linguistic geography of Africa*, Publisher: Cambridge: Cambridge University Press, 151-185.
- Güldemann, Tom. (2010). Sprachraum and geography: linguistic macro-areas in Africa. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: an international handbook of linguistic variation*, volume 2: Language mapping. *Handbooks of Linguistics and Communication Science* 30,2. Berlin: Mouton de Gruyter, 561-585, Maps 2901-2914.
- Greenhill, S.J., Wu, C-H, Hua, X., Dunn, M., Levinson, S.C., Gray, R. (2017). Evolutionary dynamics of language systems. *Proc Natl Acad Sci USA* 114: E8822-E8829
- Hammarström, H. (2016) Linguistic diversity and language evolution. *J Lang Evol* 1:19-29.
- Hammarström, Harald & Güldemann, Tom (forthcoming.) (2014). Quantifying geographical determinants of large-scale distributions of linguistic features. *Language Dynamics and Change*. E-Book ISBN: 9789004281523, Publisher: Brill, p. 67-95
- Haspelmath M. (2009). Lexical borrowing: concepts and issues. In M. Haspelmath & U. Tadmor (eds.) *Loanwords in the world's languages: a comparative handbook*. (pp. 35-54). Berlin: Mouton De Gruyter.
- Hassan, F.A (1975). *Determination of the size, density, and growth rate of hunting-gathering populations*. Berlin: Mouton De Gruyter.
- Hammarström, H. (2016). Linguistic diversity and language evolution. *J Lang Evol* 1:19-29.

- Hickey, R. (2010). Language contact: reconsideration and reassessment. In R. Hickey (ed.) *The handbook of language contact* (pp. 1-28). Oxford: Wiley-Blackwell.
- Kortmann, Bernd. (2013). How powerful is geography as an explanatory factor in morphosyntactic variation? Areal features in the Anglophone world. In Auer et al. (eds.), *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives*. Publisher Walter de Gruyter, p.165-194.
- Matras, Y. & Sakel, J. (eds.) (2007). *Grammatical borrowing in cross-linguistic perspective*. Berlin: Mouton De Gruyter.
- Muysken, Pieter (2000). From linguistic areas to areal linguistics: a research proposal. In Gilbert, Dicky, John Nerbonne, and Jos Schaeken (eds.), *Languages in contact*. (pp.263-275). Amsterdam/ Atlanta: Rodopi.
- Muysken, Pieter (2008). Introduction. In Pieter Muysken (ed.), *From linguistic areas to areal linguistics*. (pp. 1-23). Amsterdam: John Benjamins.
- Moore, J.L., Manne, L., Brooks, T., Neil, D.B., Davies, R., Rahbek, C., Williams, P. & Balmford, A. (2002) The distribution of cultural and biological diversity in Africa. *Proc R Soc Lond Ser B: Biol Sci* 269:1645-1653.
- Mufwene, Salikoko S. (2001). *The ecology of language*. Cambridge University Press.
- Mufwene, Salikoko S. (2008). *Language evolution: contact, competition, and change*. London/ New York: Continuum International.
- Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Pacheco Coelho M.T. et al. (2019). Drivers of geographical patterns of North American language diversity. *Proc R Soc B* 286:20190242
- Ricklefs, R. E. (2001). *The economy of nature*. New York: W. H. Freeman.
- Stolz, Thomas. (2002). No Sprachbund beyond this line! On the age-old discussion of how to define a linguistic area. In Ramat Paolo and Thomas Stolz (eds.), *Mediterranean languages: papers from the MEDTYP Workshop*. (pp. 259-281). Tirrenia, June 2000. Bochum: Brockmeyer.
- Stolz, Thomas (2006). All or nothing. In McMahon Matras & Vincent (eds.), *Linguistic Areas*, palgrave macmillan, 32-50.
- Thomason, S.G., Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Wilkins, D.P. (1996). Natural tendencies of semantic change and the search for cognates. In M. Durie & M. Ross (eds.), *The comparative method reviewed: regularity and irregularity in language change*. (pp. 264-304) Oxford: Oxford University Press.

