Article

Enduring constraints on grammar revealed by Bayesian spatiophylogenetic analyses

 Check for updates

Annemarie Verkerk[1] ✉, Olena Shcherbakova[2], Hannah J. Haynie[3], Hedvig Skirgård [2], Christoph Rzymski [2], Quentin D. Atkinson [4,5], Simon J. Greenhill [2,6] & Russell D. Gray [2,4] ✉

Human languages show astonishing variety, yet their diversity is constrained by recurring patterns. Linguists have long argued over the extent and causes of these grammatical 'universals'. Using Grambank—a comprehensive database of grammatical features across the world's languages—we tested 191 proposed universals with Bayesian analyses that account for both genealogical descent and geographical proximity. We find statistical support for about a third of the proposed linguistic universals. The majority of these concern word order and hierarchical universals: two types that have featured prominently in earlier work. Evolutionary analyses show that languages tend to change in ways that converge on these preferred patterns. This suggests that, despite the vast design space of possible grammars, languages do not evolve entirely at random. Shared cognitive and communicative pressures repeatedly push languages towards similar solutions.

Human languages are strikingly diverse. They vary in almost every way[1], from the sounds they use to the ordering of words and other grammar rules. However, such diversity does not preclude the existence of regular patterns and structured variation. A central goal of linguistics has been to describe the patterns and structure of human linguistic diversity and identify the constraints on that diversity[2–8]. In Hjelmslev's[2] words, the aim of linguistic typology/theory "must be to show which structures are possible, in general, and why it is just those structures, and not others, that are possible".

Some argue that languages all face the same pressures for communicating and encoding information, leading to convergence towards structurally good solutions[9–11]. For example, take verb agreement, the phenomenon where verbs are marked for the person and number features of their grammatical arguments, such as '-es' in 'Alex catch-es the ball'. Verb agreement in English is very limited; in other languages, it is far more extensive or non-existent altogether. One explanation for this cross-linguistic variation is that verb agreement 'trades off' with word order. Speakers of all languages need a way to differentiate argument

relations (to identify and mark subjects, objects and other syntactic arguments as such), and languages with subject–verb–object word order (as in English) do not 'need' verb agreement because subjects are on one side of the verb and objects on the other.

Others argue that all languages are shaped by our human cognitive capacity for online production, comprehension and acquisition[5,12–22]. Word order patterns, especially those related to the order of object and verb, have been explained in terms of efficient online processing. These have been claimed to be rooted in principles where the order of the 'head' (the most important element of a phrase that determines its type and syntactic behaviour) and its 'dependents' (other elements in the phrase) match each other across different types of phrases[13,17,23]. For example, if adpositions (heads of adpositional phrases) precede nouns (dependents in adpositional phrases) in a given language, we may expect the same 'matched' pattern where verbs (heads of verb phrases) also precede objects (dependents in verb phrases).

It is also possible that all languages are shaped by general pathways of diachronic language change[11,24–28]. For example, the association

---

[1]Department of Language Science and Technology, Saarland University, Saarbrücken, Germany. [2]Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [3]Department of Linguistics, University of Colorado Boulder, Boulder, CO, USA. [4]School of Psychology, University of Auckland, Auckland, New Zealand. [5]School of Anthropology and Museum Ethnography, University of Oxford, Oxford, UK. [6]School of Biological Sciences, University of Auckland, Auckland, New Zealand. ✉e-mail: annemarie.verkerk@uni-saarland.de; russell_gray@eva.mpg.de

between the order of adposition and noun and the order of object and verb has been explained through the common process of grammaticalization where adpositions develop from verbs[29]. Here, adpositions such as 'for' may arise from verbs meaning 'give' (as in 'Anna gave John a flower'), and if the word order is such that verbs come before objects, then these forms will be prepositions rather than postpositions ('for John' rather than 'John for').

The nature of universals and the extent to which the types of constraints mentioned above contribute to their emergence is of direct relevance for understanding the nature of human language and human cognition and, hence, a matter of some dispute across various approaches to linguistics[30,31]. Within formal approaches, universals such as those invoked in X-bar structure (simplistically, the idea that phrases of any type consist of specifiers, heads and their complements[32,33]) are seen as absolute rules of human language, tied to robust, innate grammatical constraints[5,7,8,32,34]. Evans and Levinson[1] argue against absolute universals of all types, emphasizing the diversity of the world's languages and the complex interactions of multiple potential constraints in any non-trivial generalizations about human grammars. Generative replies to Evans and Levinson[1], such as Freidin[35], Pesetsky[36] and Rizzi[37], argue that their account does not hold, among other critiques, because the generative level of analysis is deep. In contrast, the analysis in Evans and Levinson[1] and other work in functional typology remains surface-level. Hence, the generative study of universals is distinct from the typological one, both in its methods and in terms of explanations.

The current study is rooted in the field of linguistic typology, initially pioneered by Greenberg[38], which has subsequently generated a large body of research on linguistic universals (see, among other works, ref. [6]). This approach focuses on identifying patterns of grammatical feature co-occurrence that need not be exceptionless ('statistical universals'), in many cases using them to inform theories that invoke the aforementioned types of constraints (communication, cognition and language change). For example, Greenberg's[38] universal number 4 claims that "with overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional" (where SOV is subject–object–verb order), which follows the earlier example. Dryer[23] finds evidence for a large set of word order associations using a large language sample and proposes an explanation for them rooted in ease of online language processing. Similarly, Bickel et al.[39] have demonstrated that a strong cognitive preference to identify the first base-form noun phrase as the agent (the 'do-er' of an action) leads to a persistent bias against so-called ergative languages, which can explain the rarity of this pattern.

These types of explanations for universals are often grounded in the so-called competing motivations account, which can account for both variation across languages as well as provide (cognitive) grounds for the universal itself. A famous example of a formal application of this model is Aissen's[40] optimality theory account of differential object marking (DOM). DOM is common cross-linguistic behaviour where some direct objects in a language are marked (for example, with case or adpositions), while other direct objects remain unmarked. For example, in Spanish, direct objects that are definite and denote human referents are preceded by the marker '*a*' (lit. 'to'), whereas other objects are left unmarked. Aissen demonstrates that cross-linguistic variation in DOM can be explained by two interacting principles (motivations): (1) iconicity, which implies that more prominent objects (those that are definite, human or higher animates) are more likely to receive marking and (2) economy, which implies that marking should be avoided altogether. Her analysis suggests that these two constraints interact in such a way that the outcome is the observed cross-linguistic variation in differential object marking—in DOM languages, highly prominent objects always receive marking, but languages vary on the 'cut-off' regarding less prominent objects. Underlyingly, the iconicity motivation may be grounded in the communicative need to distinguish prominent objects from subjects, which also tend to be definite, human or higher animates.

Such accounts form the theoretical groundwork for why we may find universals in the first place. However, not all linguists are convinced of the importance of such constraints: Dunn et al.[41] argue that statistical word order universals arise in "an evolutionary landscape with channels and basins of attraction that are specific to linguistic lineages". They claim that word order correlations do not emerge in response to functional constraints, but rather are a consequence of particular diachronic changes unique to particular language families. A common view among generative linguists is that universal grammar does not (and should not aim to) explain statistical universals[34].

Resolving the complex problem of what grammatical relationships are 'universals' and what they mean for understanding language or cognition has been hindered by several fundamental challenges. First, the lack of a comprehensive grammatical dataset has meant previous work has tended to consider a relatively small subset of the world's ~7,000 languages, limiting statistical power and the ability to test for strong associations rigorously. Second, shared linguistic ancestry and the diffusion of features between neighbouring populations mean linguistic data do not constitute independent data points, violating the independence assumption of many statistical tests and potentially generating spurious statistical associations between features[42]. Finally, raw correlations between features (or traits) tell us little about the historical causal relationships between them.

Here, we overcome these challenges by analysing a comprehensive database of grammatical features, Grambank[43], which covers more languages than previous work (Supplementary Text 1), and employing sophisticated methodologies to handle non-independence and test for co-evolution. We test 191 putative 'linguistic universals' extracted from the Universals Archive[44] (Supplementary Text 2 and 3). These are all so-called implicational universals, as Greenberg's[38] universal number 4 above (SOV ⇒ postpositions): they relate characteristics of the world's languages in an 'if X, then Y' structure. First, we apply a Bayesian generalized linear mixed effects model to evaluate the support for each hypothesis while controlling for genealogical and geographical relations. We then apply a Bayesian phylogenetic method to infer the underlying evolutionary dynamics (see Methods and Supplementary Text 5 for explanations and rationale). We see this work as rising to the challenge spelt out by Piantadosi and Gibson[45], which claims that "claims about linguistic universals should be accompanied by some measure of the strength of evidence in favour of such a universal". They propose that any hypothesized universal should be compared with the corresponding null hypothesis using a Bayes factor or similar, and point out that such measures are critical given the relevance of universals to debates on the nature of human language.

To examine differences in the reasoning behind universals we divided the generalizations from the Universals Archive into four types, reflecting recurrent themes in typological literature[10,12,46]: (1) narrow word order, (2) broad word order, (3) hierarchical universals and (4) other. Narrow word order universals link the order of words in two or more constructions in ways that generally relate to where the most important words occur, such as Greenberg's[38] universal 4 above. Broad word order universals correlate a word order feature with a morphosyntactic feature unrelated to word order, such as "non-accusative alignment may be associated with verb–initial order"[47]. Hierarchical universals are chains of implicational universals with the most frequently attested traits on the left and the rarest ones on the right, as defined within the same domain or paradigm. An example is Greenberg's[38] claim that "no language has a dual unless it has a plural"; this type of universal has also been called scalar or 'scale' in literature. The remaining universals are captured under 'other', but in practice often correlate two morphological features (see Methods and Supplementary Text 2, 7 and 8 for further details).
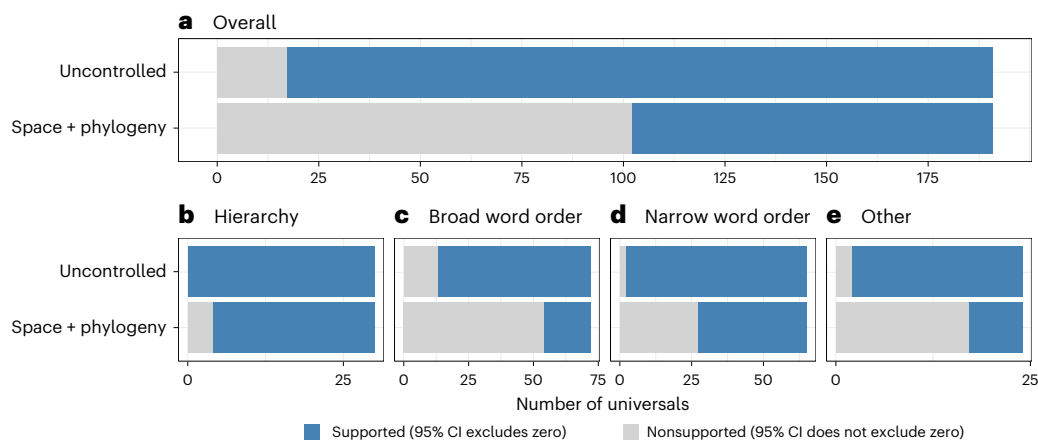
Fig. 1 | Bar chart showing the proportion of supported universals under the naive model and the spatiophylogenetic model. Support implies, for the naive model, that the 95% CI of posterior coefficient estimates does not straddle zero. For the spatiophylogenetic model (with genealogical and geographical relations controlled for), it means that the median of the 95% CI of the main fixed effect estimates does not straddle zero. Universals identified to be supported are coloured in blue, while non-supported universals are coloured in grey. **a**, The overall universals. **b–e**, The universals by subset: hierarchy (**b**), broad word order (**c**), narrow word order (**d**) and other (**e**).

Results

Phylogenetic and spatial correlation

We constructed Bayesian generalized linear mixed effects models (GLMMs) for all universals using brms[48] implemented in R[49]. We find that, without controlling for genealogical and geographical relations, the vast majority of the proposed universals are supported by our regression models—that is, the fixed effect of the predictor variable (the second part of the universal) has posterior estimates whose 95% credible interval (CI) exclude zero (Fig. 1, Supplementary Data 1 and Supplementary Table 3). In these naive models, 174 (91%) of the 191 universals are found to be supported.

However, when we do control for spatial and phylogenetic non-independence this number decreases substantially to 89 of 191 (47%). Here, we conduct the analysis over 100 phylogenetic trees[50] (Methods) and hence report on the median of posterior estimates and their 95% CIs (Supplementary Fig. 4, Supplementary Data 1 and Supplementary Table 4). We find marked differences between the strength of support for four types of universals (Fig. 1). There is strong support for hierarchical universals with 24 of 30 (80%) having posterior estimates that exclude zero. The narrow word order universals are also relatively well supported, with 36 of 65 (58%) confirmed. In contrast, there is weaker support for the broad word order (18 of 72 supported, 25%) and the 'other' universals (7 of 24, 32%).

Evolutionary dynamics

To infer the evolutionary (in the sense of diachronic) pathways behind the statistically supported universals identified in the spatiophylogenetic brms analyses, we performed co-evolution analyses using the BayesTraits program[51,52]. Again, we conducted the analyses over 100 phylogenetic trees, calculated Bayes factors (BF) by comparing the dependent and independent model and calculated the 95% high density interval (HDI) to summarize BF support for each universal (Methods and Supplementary Texts 5 and 6). We took the lower bound of the 95% HDI >10 as indicating support for the dependent model of trait co-evolution over the independent model (Fig. 2, Supplementary Data 1 and Supplementary Table 4). On this criterion, 60 of the 89 universals supported in the spatiophylogenetic model were also supported in the co-evolution analyses. We continue discussing this set of 60 universals. Across the different types of universals, we observe the same pattern as for the spatiophylogenetic correlations: the strongest evidence can be found among the hierarchical universals (evidence for all of the 24 universals supported in the spatiophylogenetic analysis,

80% of all hierarchical universals ($n = 30$)). Second, the word order universals show a more mixed pattern: for the narrow word order universals, less than half of narrow word order universals (24/36, 37% in all ($n = 65$)) and a much smaller fraction of the broad word order universals (8/18, 11% in all ($n = 72$)) are supported. Third, we find that only four of the seven 'other' universals supported in the spatiophylogenetic model hold (17% in total, $n = 24$).

Robustness

Since both studies are dependent on a global phylogeny of languages[50] (Methods), we conducted additional tests with categorical control for language family. Our results also hold with this approach (Supplementary Texts 5 and 6). Of the 89 statistically supported universals identified in the spatiophylogenetic analysis using the global phylogeny, 67 (75%) are supported using a categorical control for language family instead (Supplementary Fig. 6, Supplementary Data 1 and Supplementary Table 3). The main effect estimates of the two studies (the spatiophylogenetic brms models and the categorical control for language family brms models) are highly correlated (Spearman's $r = 0.93$ (189 degrees of freedom), $P < 0.001$, 95% CI 0.91–0.95).

Taking the universals supported in the spatiophylogenetic brms model and additionally also supported in the BayesTraits analyses, we find support for 60 (of 191, 31%) universals. The strength of the correlations between the two terms linked in the putative universal (median of the main effect estimates in brms and the median BF in BayesTraits) are high (Spearman's $r = 0.72$ (189 degrees of freedom), $P < 0.001$, 95% CI 0.65–0.78). In the remainder of the paper, we will consider these 60 as well-supported universals and discuss them further. See Supplementary Text 5 for more on the comparison of these two analyses.

An analysis of possible biases in past studies

The conditions under which the universals were first proposed were considered as possible explanatory factors of why some universals are well-supported in the spatiophylogenetic brms models and the BayesTraits analyses. Many of the strongest supported universals were formulated in the 1960s and 1970s, and the language samples then were often relatively small and biased (often featuring mainly Eurasian languages). We considered the size and geographical bias of the original language sample and implementation of the universal in terms of the Grambank questionnaire (Supplementary Text 9). We found that none of these had an impact on the support of universals as defined above. This shows that the limited representation of worldwide linguistic

Fig. 2 | Median natural log BF and their 95% HDI from the BayesTraits analyses showing support for co-evolutionary models. Universals with the lower bound of the 95% HDI on the distribution of BF >10 are considered supported and are coloured blue. The relationships are ranked from strongest to weakest, top to bottom, per category. The universals are given in short form, where the formula X ⇒ Y means 'if X, then Y'; the full citations, sources and all abbreviations can be found in Supplementary Table 2. Sample sizes can be found in Supplementary Data 1 and Supplementary Table 4.
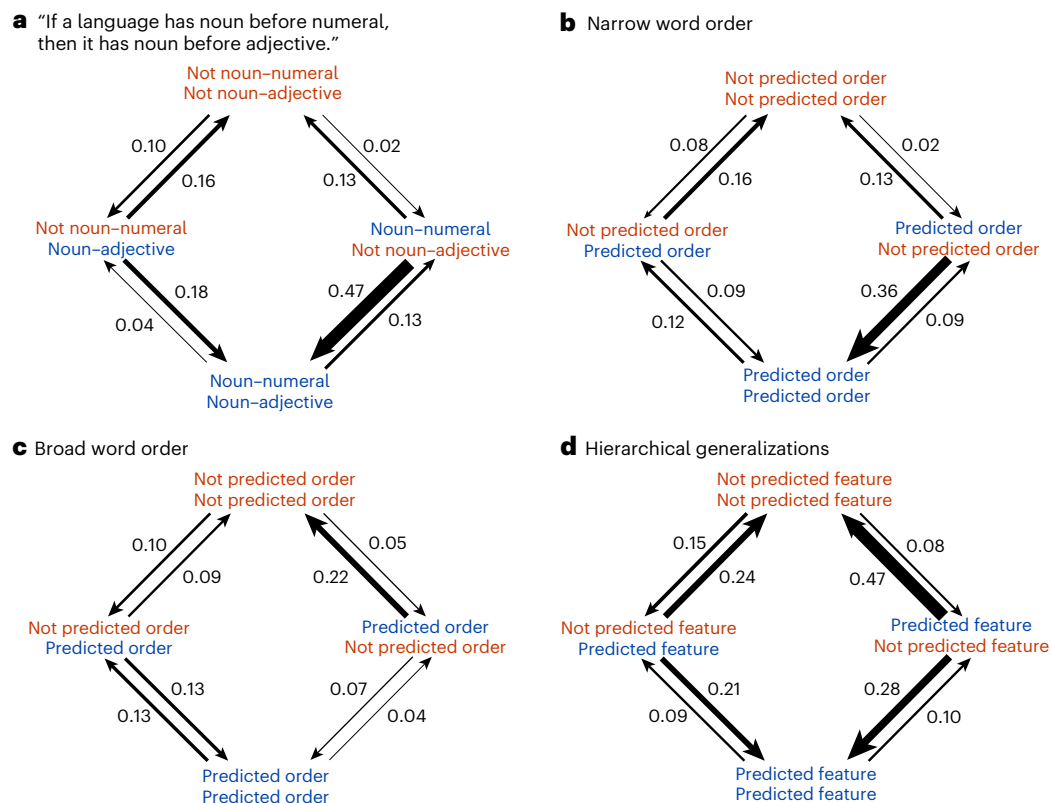
Fig. 3 | Median rates of change towards harmonic (bottom) and disharmonic states (left and right) for supported universals in BayesTraits analyses (*n* = 60). a, An illustration of the dependent model in BayesTraits for the universal 'if a language has noun before numeral, then it has noun before adjective'[12]. **b**, Median rates of change for supported narrow word order universals. **c**, Broad word order. **d**, Hierarchical universals. Blue text refers to orders or features predicted by the universal, red text refers to orders or features that the universal does not predict and is agnostic about.

diversity in twentieth century research on universals did not prevent the discovery of universals that still hold in our study, with its massively increased sample size (Supplementary Text 1) and appropriate statistical methodology.

Hierarchical universals

The tested hierarchical universals generally deal with the expression of grammatical categories through agreement, and within the (pro) nominal paradigm. An assessment of why certain universals are supported (and others are not) is provided in Supplementary Text 9. Here, we discuss that hierarchical universals have often been explained using competing motivations accounts, where 'competition' between certain principles ('motivations') such as economical behaviour (avoidance of overt grammatical marking) and salience (preferred marking of entities that matter, such as humans) results in the attested cross-linguistic distributions[9,53–55]. We think that the success of the studied hierarchical universals (80%, 24 of 30, are supported by both the spatiophylogenetic brms and BayesTraits analyses) can be explained in such an account, especially if we consider the role of diachronic change and if the universals are formulated in a specific enough manner: two unsupported ones are probably too general. Note that we do not and cannot test which explanations are 'best' in this paper, and that which follows (also on other types of universals) is left for further research. As is clear from Fig. 2 and Supplementary Figs. 6 and 7, there is a gradient in support for hierarchical universals despite them being highly supported as a category; this gradient persists across the different analyses and should be the focus of further investigation.

Many of the supported universals in this category posit the presence of a more common feature given the presence of a rarer feature within the same paradigm. Hierarchical universals hence capture the

frequency at which various features appear in the world's languages and language change in (morphological) paradigms. Exemplified in Supplementary Fig. 10 is Plank's[56] "If determiners agree within NPs, modifiers are likelier also to agree than not to agree". Agreement within noun phrases (NPs) is common in the languages of the world, with determiners being a less common agreement target than, for example, adjectives; Plank's[56] generalization is highly supported (BayesTraits median BF 200, median spatiophylogenetic brms coefficient estimate 2.01 and 95% CI 1.42–2.65). A second example is provided in Supplementary Fig. 11.

An example of a universal possibly rooted in the interaction between economy and salience (broadly construed as principles on what humans find important) is Croft's[54] claim "if there is a construction in which the verb agrees with some member of the relational hierarchy subject ⇒ direct object ⇒ indirect object ⇒ oblique ⇒⇒ [⇒genitive], then there are at least some constructions in which the verb agrees with members higher on that hierarchy". We tested this proposal by investigating if object indexing implies subject indexing. This is a highly supported claim (BayesTraits median BF 120, median spatiophylogenetic brms coefficient estimate 3.81 and 95% CI of 2.32–5.35), which Croft[54] explains by stating that verbal indexing is used to mark important or salient arguments, that is, arguments that are high on animacy, definiteness and case scales that line up to be important in the speaker's perspective on the event. Hierarchical universals, hence, ultimately deal with the expression of those grammatical categories that are most salient for humans across cultures (and more frequent) in comparison with those that are less salient (and less frequent). A complementary account may be a tendency of languages to cover connected regions in conceptual space[10]. They may have diachronic explanations too; see results below (Fig. 3d) on the analyses of rates of change, with a change
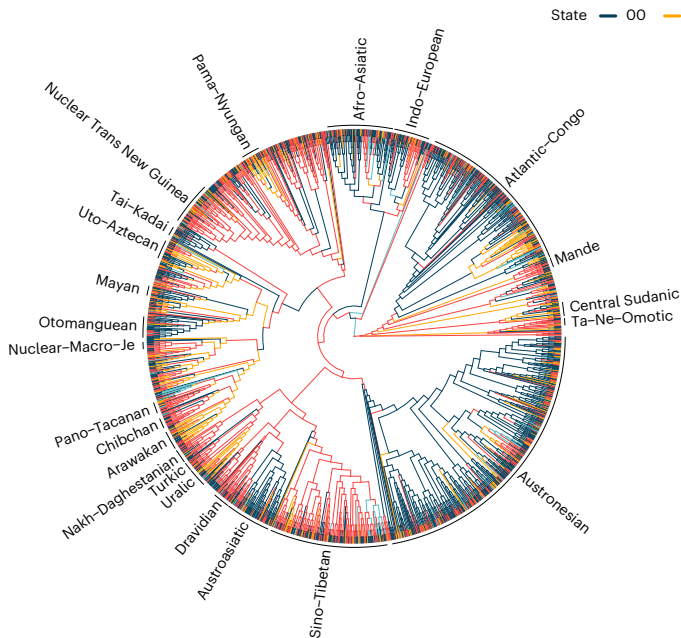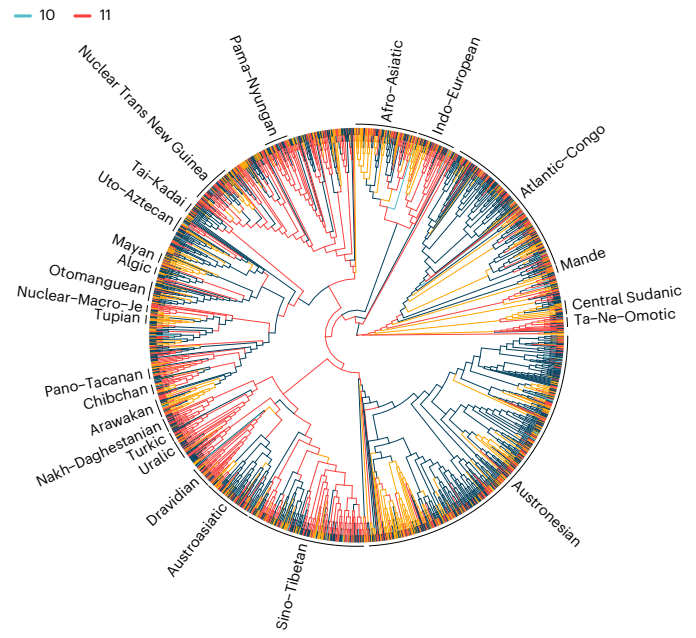
a Narrow word order: SOV ⇒ N–Adp

State ▬ 00 ▬ 01 ▬ 10 ▬ 11

b Hierarchy: case nouns ⇒ case pronouns

Fig. 4 | Ancestral state reconstruction of two highly supported universals. State 11 (red) is the prediction made by the universal (the harmonic state 4, see above). State 00 (black) is when both features are absent. State 01 (orange) and 10 (light blue) are disharmonic states (see above). See 'Ancestral state reconstruction' section in the Methods for further details. **a**, "With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional"[38]. **b**, "If there is case-inflection on nouns, there is also case-inflection on some pronouns" (attributed to Edith Moravcsik, source unknown; Supplementary Text 7).

towards the presence of both features predicted by the universal being more common than the reverse sets of changes away from this state.

Narrow word order universals

We find support for 37% (24 of 65) of the narrow word order universals (implications between one (or more) word order(s) and another word order). We find support for generalizations between the order of object and verb, the order of adposition and noun, as well as generalizations that describe the order of modifiers and other elements of the noun phrase and the head noun, in relation to other word orders, reflecting findings of previous investigations[12,13,23,38,41,57,58]. An assessment of our results in light of other papers on word order universals is given in Supplementary Text 8; on explanations, see Supplementary Text 9. A key take-away from the latter overview is that the order of adposition and noun is somehow central, as it is involved in the greatest number of supported correlations (Supplementary Fig. 9). Hawkins[12] has in fact already proposed that "Prep and Postp are more general typological indicators". There is a tendency for harmonic ordering of the noun and its various adnominal modifiers. Aside from that, VO (verb–object order) languages and OV (object–verb order) languages seemingly evolve differently as they do not engage in the same word order correlations, which is again not a new finding (see, for example, research on the final-over-final condition[59,60]). As is clear from Fig. 2 and Supplementary Fig. 7, some universals receive higher support than others, pointing to a gradience that is in line with the lineage-specificity proposed by Dunn et al.[41] and findings of other quantitative studies[57,58]. One possible interpretation of our results is that different universals have differential strengths across lineages.

We find evidence for prepositional languages to have a noun-modifier order within the noun phrase and, though less pronounced, for postpositional languages to have modifier-noun order. Such patterns have been explained in terms of strong cognitive predictions associated with 'headedness', although the relevant properties differ from author to author[12,13,23,61]. We refer to universals positing a consistent head-dependent order across different phrases or, in other words, a tendency for languages to consistently put the most important words of a phrase in the same position, as 'harmonic'[62].

The phylogenetic co-evolution analyses in BayesTraits allow us to assess whether explanations rooted in harmony hold in a diachronic context[10]. We take N–Num ⇒ N–Adj[12] as an example. The values predicted by the universal, the presence of the orders where nouns precede numerals and adjectives, N–Num and N–Adj respectively, are given state '1', and the reverse order where nouns follow numerals and adjectives, Num–N and Adj–N, are given state '0' (Supplementary Text 5). Such generalizations about states (as shown in Fig. 3a) were coded in the same way: in the BayesTraits models, state 4 (1,1) is harmonic, while state 2 (0,1) and state 3 (1,0) are disharmonic. Note that state 1 (0,0) does not concern us here, as diachronic change to state 4 always goes through state 2 or 3. Figure 3b shows that change to the harmonic state 4 is more frequent than change to disharmonic states for narrow word order universals. This applies to both 'simple' word order universals that condition exclusively on a single word order (such as N–Num ⇒ N–Adj), as well as more complex universals that have multiple word order conditions in the implication part (such as Adp–N and N–Num ⇒ N–Adj[12]), where Adp stands for adposition. Figure 4 provides an illustration of this pattern by depicting the support for two highly supported universals on the global phylogeny, showing that the harmonic state is most probable in large chunks of the global tree. We suggest that these findings may benefit future research into the why and how of narrow word order universals, especially considering the interaction between cognitive preference for global head-dependent orders and diachronic explanations rooted in grammaticalization.

Broad word order and other universals

The broad word order universals are less well supported (11%, 8 of 72) than narrow word order ones, and the rates of change towards combinations of states predicted by the universal are not faster than those towards other state combinations (Fig. 3c). Much the same applies
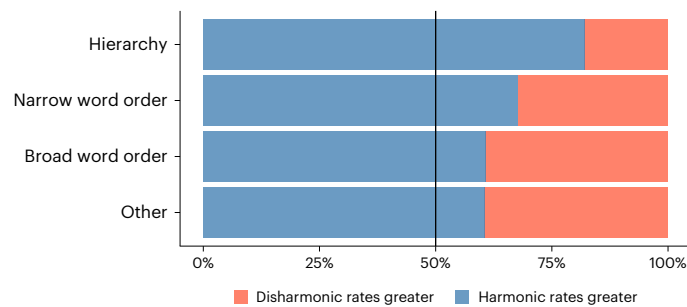
Fig. 5 | Bar plot showing the proportion of sampled rates from supported universals where the harmonic rate was greater than the disharmonic rate. Distributions in blue show the proportion of times the rates in the harmonic categories were larger than the rates in the disharmonic categories, while distributions in red show the proportion of times the rates in the disharmonic category were larger.

to other universals, only 17% (4 of 24) are supported. Several of the supported broad word order universals are involved in the resolution of argument relationships; supporting the well-studied interaction between grammatical case marking[47,63–66], the order of adposition and noun, and free word order. Other supported ones are also highly specific: across all four categories of universals, generally formulated universals fare poorly (such as Greenberg's[38] universal "if a language is exclusively suffixing, it is postpositional; if it is exclusively prefixing, it is prepositional"). These generally formulated universals make sweeping statements, for example, about all affix positions in an entire language (see above) or about different alignment systems at the same time, as in Nichol's[47] "non-accusative alignment may be associated with verb–initial order". In contrast, more specific universals refer to specific phrase types, parts of speech or function words or morphs. Many broad word order universals, however, are highly specific and nevertheless unsupported; such as Stassen's[67] "if attributive adjectives have the form of relative clauses, there is a positive correlation with SVO".

What unites supported universals of the remaining 'other' type is that they tend to relate patterns of morphology, such as Keenan's[68] universal "if heads of possessive constructions agree with their possessors in a given language then verbs agree with subjects in that language". This finding, along with the poor support for broad word order universals in general, may imply that there are few universals that impact both morphology and syntax, except where morphology and syntax 'connect', as in argument resolution (Supplementary Text 9).

Harmonic tendencies across types

The estimated rates of change for the narrow word order and hierarchy categories (Fig. 3) showed a significant but moderate tendency towards having higher rates towards harmonic states (Fig. 5). We tested significance using two-sided Wilcoxon signed rank tests to avoid making assumptions about normality: narrow word order ($V = 6,792$, $n = 22$ and $P < 0.001$), broad word order ($V = 1,878$, $n = 10$ and $P = 0.07$), hierarchy ($V = 3,241$, $n = 24$ and $P < 0.001$) and other ($V = 240$, $n = 4$ and $P = 0.21$). For narrow word order and hierarchical universals supported in the BayesTraits analyses, the proportion of harmonic rates > disharmonic rates exceeds the expected proportion (0.5). The 'other' category was not found to show a significant effect here, presumably owing to its small size ($n = 4$).

Discussion

In this paper, we overcome the limitations of previous studies of linguistic universals in several ways. First, our large sample of languages from across the world gives us greater statistical power to detect robust linguistic generalizations. It also avoids problems created by small samples focused on a specific region of the world or a small set of language families[42]. Supplementary Fig. 1 shows that the amount of data

available in Grambank[74] to test these hypotheses dwarfs that used to propose these hypotheses in the first place. Second, we used appropriate computational methods that explicitly control for the phylogenetic and spatial non-independence of the languages sampled and enabled us to investigate the diachronic relationships between putatively linked grammatical features[57]. In this way, we meet the challenge spelt out by Piantadosi and Gibson[45] "that claims about linguistic universals should be accompanied by some measure of the strength of evidence in favour of such a universal". Third, by investigating a large and varied number of putative universals, we can detect general patterns across distinct types of universals with differing rationales and identify common underlying properties. The limitations of this study are that we focused on just the universals listed in the Universals Archive[44] that could be tested with the data in Grambank (Supplementary Texts 1 and 2). While the statistical models and set of phylogenetic trees we use are the state of the art, it is hoped that future developments will bring further refinements of our results.

We tested 191 putative linguistic universals. The vast majority of these were statistically supported in a naive analysis that did not control for phylogenetic and spatial autocorrelation. However, less than half of these universals were supported once genealogical and geographical relations between the languages were taken into account. In other words, many proposed universals are artifacts of the non-independence of features among closely related or neighbouring languages. This result demonstrates the critical importance of fully controlling for spatial and phylogenetic autocorrelation.

We find statistical support for 60 of the 191 universals. This includes support for most hierarchical universals (80%, 24 of 30) and over a third of the tested narrow word order universals (37%, 24 of 65). The two other types of universals, broad word order universals and 'other', do not have the same level of statistical support. Of the 72 broad word order universals we tested, only 8 were found to be supported (11%); of the 24 'other' universals, just 4 were supported (17%). The fact that we find statistical support for a third (31%) of the proposed linguistic universals suggests that, while grammar is far less constrained than linguists have claimed[23,61], there are indeed some enduring constraints on grammatical variation (contra Dunn et al.[41]). Given that universals differ in strength (Fig. 2 and Supplementary Fig. 7), our results elucidate future directions in universals research, which should be aimed towards explaining gradient support in terms of linguistic, lineage-specific and areal factors. Linguistic theory has generated a multitude of explanatory accounts for universals, and the current study is not in a position to differentiate between their relative merit. Instead, our analyses can shed light on different hypotheses and may be valuable to multiple theoretical perspectives (Supplementary Text 9 and ref. 69).

One central take-away is that hierarchical and narrow word order universals are supported more often than other types of universals. Many have argued that the use of harmonic word orders is rooted in processing constraints[13,22,23]. In such an account, the supported universals can then be explained in terms of a cognitive preference for heads and (nominal) dependents to be ordered in the same direction[12,13,23]. Alternatively, in a generative framework, merge and move (internal merge) operations explain such dependencies[70] or overarching parameter settings[71]. Regardless of the approach, our results support the following long-standing caveat: not all head-dependent pairs engage in correlated behaviour, and some correlations are stronger than others; any account of narrow word orders needs to be able to capture this.

Explanations of hierarchical universals have likewise been rooted in functional and cognitive explanations[9,16,54], with generative accounts explaining these effects with theoretical mechanisms such as semantic feature geometries[72,73]. We speculate that these are so widely supported because they are highly specific, ingrained in the structure of agreement and other morphological paradigms and possibly rooted in language change[11,74].

Unlike the narrow word order universals, the broad word order universals do not rely on the concept of harmonic structures in different parts of the grammar. Likewise, the 'other' category includes a diverse set of proposed universals. Because both types of universals target such a wide range of morphosyntactic phenomena, the processes that potentially shape them are also more disparate (Supplementary Text 9). We can, however, conclude that there are few links between word order and other aspects of morphosyntax—many of the unsupported broad word order universals deal with supposed characteristics of verb–initial or verb–final languages, as well as languages with free or rigid word order. One of the few exceptions to this is universals involved with the resolution of argument relationships, which are probably involved in a functional trade off[64–66].

Each individual language could be thought of as an experiment in how to construct an effective communication system that can be learnt and used by both sender and receiver. An enormous array of combinatorial possibilities in all aspects of grammar is available to construct these systems. Dunn et al.[41] claim that constraints on these possibilities ('universals') may arise through lineage-specific processes. Figure 4 depicts the support for two highly supported universals on the global phylogeny. It reveals that, although the feature configurations associated with well-supported universals do tend to be disproportionately concentrated in specific lineages, they also evolve repeatedly in different language families and areas. Thus, despite the flexibility of language pragmatics[75], and the vast combinatorial array of possible language systems, some aspects of language do repeatedly evolve toward the same preferred regions of the 'design space' of morphosyntactic variation. This convergent evolution may reflect common cognitive and communicative pressures, opening the door to integrative accounts of language where language change is a key component, next to functional or formal constraints[26,74]. However, the precise nature of these pressures cannot be identified by large-scale comparative analyses alone. There are numerous possible explanations for these preferred outcomes. Our analyses do not distinguish between different potential causal mechanisms but do provide a restricted set of universals to investigate further from a wide set of theoretical viewpoints. Mechanistic research in the fields of cross-linguistic psycholinguistics, artificial language learning, corpus-based typology, historical linguistics and computer simulation is ideally suited to this task. Combining mechanistic and macro-evolutionary analyses is an exciting challenge for future research that will reveal the complex interplay of factors that shape both linguistic diversity and commonalities.

Methods

Universals data
We evaluated over 2,000 documented universals in the Universals Archive[44] (https://typo.uni-konstanz.de/rara/category/universals-archive/). We compared the morphosyntactic features involved in these universals with the data in Grambank version 1.0[43] to determine which of them could be analysed using features from Grambank. Then, we reformulated those universals that were testable, first by splitting complex universals that address multiple ostensibly associated linguistic features into simple implicational universals (which we refer to here as simply 'universals' or 'generalizations'). Second, we matched feature constellations from universals to Grambank questionnaire questions and combinations thereof. Since Grambank only covers morphosyntax, any universals involving phonology or semantics were excluded; many other claims were excluded because they were too specific, were explicitly diachronic or were not implicational. From 146 original universals that matched the Grambank variables and our criteria for synchronic implicational claims, we formulated 191 simple universals to test. These were classified into four types (see main text and Supplementary Text 2). For more on the Bayesian statistical analyses including checks for analysis convergence, see below and Supplementary Text 5. All data, code and output that support the findings of this study, including

Supplementary Data 1, are available via GitHub (https://github.com/SimonGreenhill/TestingLinguisticUniversals).

Grammatical data
To obtain grammatical data to test these proposed correlations we used Grambank version 1.0[43], a database containing morphosyntactic data on 2,430 languages (Supplementary Text 1). Grambank contains information on 195 typological features, which we mapped against the universals from the Universals Archive.

Spatial data
To control for spatial diffusion, we obtained location data as longitude–latitude pairs from the Glottolog database version 4.3[76].

Phylogenetic trees
To control for phylogeny, we used the posterior sample of trees (n = 902, downsampled to 100) from a recently released global phylogeny of 6,635 languages[50].

All matching and creation of datasets was handled in R[49]. For each universal, we used the maximum number of available languages given missing data in Grambank, the Bouckaert et al. phylogenetic trees[50] or in Glottolog. Ultimately, we sampled between 329 and 2,226 languages, with mean 1,653 and median 1,679 for our analyses (Supplementary Fig. 1 and Supplementary Tables 3 and 4).

Phylogenetic and spatial correlation
To test whether the proposed universals were indeed correlated after controlling for genealogical and geographical relations, we constructed GLMMs using the package brms[48] in R[47]. This package allowed us to fit Bayesian multilevel models in R using the probabilistic programming language STAN[77]. We chose brms because it allowed us to include categorical (binary) variables as well as random effects that control for genealogical and spatial relations. Both the response variable (the condition part of the universal) and the main fixed effect (the result part of the universal) are binary. Random effects included phylogenetic distance, spatial distance and macro-area. Phylogenetic and spatial distance were calculated from covariance matrices using trees from ref. 50 for the former, and longitude and latitude from Glottolog version 4.3[76] for the latter. Language areas ('macro-areas') were likewise taken from Glottolog version 4.3[76]. Analyses were conducted on 100 phylogenies taken from Bouckaert et al.'s[50] posterior sample, hence we report on the medians of the main fixed effect (the result part of the universal) and its 95% CI. Our approach is similar to that of Guzmán Naranjo and Becker[78], who used a correlation matrix in their Bayesian GLMM to deal with genealogical autocorrelation; in contrast, they deal with spatial autocorrelation through Gaussian processes on macro-areas.

Robustness
To test whether the supported universals using the spatiophylogenetic GLMMs hold without the maximum clade credibility tree from Bouckaert et al.[50], we constructed another set of GLMMs with brms[48] in R[49], which used language family membership as a control instead of the full phylogeny. For these analyses, data on language families was taken from Glottolog version 4.3[76]. Language family membership is a categorical variable construed as a random effect, and both intercepts and slopes were estimated[79]. Only languages from families with five or more members were included.

Evolutionary dynamics
To model the evolutionary dynamics of the universals that were supported in the spatiophylogenetic GLMMs, we used BayesTraits[51,52]. We mapped our grammatical data onto the Bouckaert et al.[50] phylogenies using a continuous-time Markov model of trait evolution implemented as the 'discrete' model in BayesTraits. We seeded the mean and variance

of the gamma prior using a uniform hyperprior between 0 and 10. Correspondingly, each analysis ran for 30,300,000 iterations (the first 300,000 generations were discarded as burn-in), sampling every 30,000 generations. To better explore parameter space among Bouckaert et al.'s[50] set of posterior phylogenetic trees, we randomly sampled 100 trees (from Bouckaert et al.'s 902 posterior tree sample) and forced BayesTraits to spend an equal amount of time on each of those 100 trees (50,000 iterations). Model fit was assessed using log marginal likelihoods, estimated using stepping stone sampling[80]. Model fit of the dependent and independent model of trait evolution were then compared using natural log BF[81]

$$\text{log Bayes factors} = \text{log marginal Lh dependent model}$$
$$- \text{log marginal Lh indep. model}$$

A positive Bayes factor indicates that the dependent model performs better than the independent model, providing evidence for co-evolution of both features. Given that we are testing for multiple correlations we use a conservative cut-off. We estimate the 95% HDI of the 100 BFs and take the lower bound of the 95% HDI >10 to constitute significant evidence for co-evolution of two features.

To identify whether these models showed a tendency towards evolving into harmonic states versus disharmonic states, we extracted rate estimates from these analyses. We categorized a subset of the rates in the Q matrix as either leading to a harmonic state (both features present) or to a disharmonic state (one feature absent and one present). For example, rate q_{34} is the rate at which languages evolve from having feature 1 without feature 2 (1,0) into having both features (1,1) and is therefore harmonic. The opposite rate, q_{43}, is the rate at which languages evolve from having both features (1,1) to having just the first feature (1,0) and is therefore disharmonic. For each of these pairs of rates, in each generation of the Markov chain Monte Carlo (MCMC) chain, we asked if the harmonic rate was larger than the disharmonic rate, that is, if there was a stronger tendency towards harmonic state patterns.

Ancestral state reconstruction
To visualize the universals on the phylogeny, we recoded the states of the paired features into a four-state variable (absent + absent = 'A', absent + present = 'B', present + absent = 'C' and present + present = 'D'). We then fitted a hidden rates model using the corHMM package version 2.8[82,83] in R[49] to estimate the marginal maximum likelihood node state probabilities. We modelled the history of these features using an 'all rates different' model such that transitions between each state had their own rate. We specified the root probabilities following FitzJohn et al.[84] and corrected for ascertainment bias following Lewis[85]. Feature manipulation and visualization was done using R packages treeio[86] and ggtree[87].

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
All the data that support the findings of this study are available via GitHub at https://github.com/SimonGreenhill/TestingLinguistic Universals.

Code availability
All code used is available via GitHub at https://github.com/Simon Greenhill/TestingLinguisticUniversals.

References

1. Evans, N. & Levinson, S. C. The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–492 (2009).
2. Hjelmslev, L. *Language: an Introduction* (Univ. of Wisconsin Press, 1970).
3. Comrie, B. *Language Universals and Linguistic Typology* (The Univ. of Chicago Press, 1989).
4. Shopen, T. *Language Typology and Syntactic Description* (Cambridge Univ. Press, 2007).
5. Baker, M. C. *The Atoms of Language* (Oxford Univ. Press, 2001).
6. Bickel, B. Typology in the 21st century: major current developments. *Linguist. Typology* **11**, 239–251 (2007).
7. Cinque, G. Cognition, universal grammar, and typological generalizations. *Lingua* **130**, 50–65 (2013).
8. Roberts, I. Parameter Hierarchies and Universal Grammar (Oxford Univ. Press, 2019).
9. Keenan, E. L. & Comrie, B. Noun phrase accessibility and universal grammar. *Linguist. Inq.* **8**, 63–99 (1977).
10. Croft, W. Typology and Universals (Cambridge Univ. Press, 2003).
11. Bybee, J. L. in *The Oxford Handbook of Linguistic Typology* (ed. Song, J. .J.) 131–147 (Oxford Univ. Press, 2010).
12. Hawkins, J. A. *Word Order Universals* (Academic, 1983).
13. Hawkins, J. A. *A Performance Theory of Order and Constituency* (Cambridge Univ., 1994).
14. Hawkins, J. A. *Efficiency and Complexity in Grammars* (Oxford Univ. Press, 2004).
15. Hawkins, J. A. *Cross-Linguistic Variation and Efficiency* (Oxford Univ. Press, 2014).
16. Haspelmath, M. The best-supported language universals refer to scalar patterns deriving from processing cost. *Behav. Brain Sci.* **32**, 457–458 (2009).
17. Futrell, R., Levy, R. P. & Gibson, E. Dependency locality as an explanatory principle for word order. *Language* **96**, 371–412 (2020).
18. Hawkins, J. A. Implicational universals as predictors of language acquisition. *Linguistics* **25**, 453–473 (1987).
19. Gass, S. M. Language universals and second-language acquisition. *Lang. Learn.* **39**, 497–534 (1989).
20. Skear, P. M. Language universals and child language acquisition. *J. Integr. Sch. Arts Sci. Hiroshima Univ.* **30**, 83–109 (2004).
21. Lightfoot, D. & Havenhill, J. in *Variable Properties in Language: Their Nature and Acquisition* (eds Lightfoot, D. & Havenhill, J.) 1-12 (Georgetown Univ. Press, 2019).
22. Culbertson, J., Franck, J., Braquet, G., Barrera Navarro, M. & Arnon, I. A learning bias for word order harmony: evidence from speakers of non-harmonic languages. *Cognition* **204**, 104392 (2020).
23. Dryer, M. S. The Greenbergian word order correlations. *Language* **68**, 81–138 (1992).
24. Aristar, A. R. On diachronic sources and synchronic pattern: an investigation into the origin of linguistic universals. *Language* **67**, 1–33 (1991).
25. Bybee, J. in *Explaining Language Universals* (ed. Hawkins, J. A.) 350–379 (Blackwell, 1988).
26. Bybee, J. *Language Change* (Cambridge Univ. Press, 2015).
27. Collins, J. in *Explanation in Typology: Diachronic Sources, Functional Motivations and the Nature of the Evidence* (eds Schmidtke-Bode, K., Levshina, N., Michaelis, S. M. & Seržant, I. A.) 47-61 (Language Science Press, 2019).
28. Good, J. *Linguistic Universals and Language Change* (Oxford Univ. Press, 2008).
29. Dryer, M. S. in *Explanation in Typology: Diachronic Sources, Functional Motivations and the Nature of the Evidence* (eds Schmidtke-Bode, K., Levshina, N., Michaelis, S. & Seržant, I. A.) 63–95 (Language Science Press, 2019).
30. Haspelmath, M. in *The Limits of Syntactic Variation* (ed. Biberauer, T.) 75–107 (John Benjamins, 2008).
31. Baker, M. C. The interplay between universal grammar, universals, and lineage specificity. *Linguist. Typology* **15**, 473–482 (2011).

32. Chomsky, N. *Lectures on Government and Binding: the Pisa Lectures* (Foris Publications, 1981).

33. Alexiadou, A. & Lohndal, T. in *A Companion to Chomsky* (eds Allot, N., Lohnda, T. & Rey, G.) 25–51 (Wiley Blackwell, 2021).

34. Newmeyer, F. J. *Possible and Probable Languages: a Generative Perspective on Linguistic Typology* (Oxford Univ. Press, 2005).

35. Freidin, R. A note on methodology in linguistics. *Behav. Brain Sci.* **32**, 454–455 (2009).

36. Pesetsky, D. Against taking linguistic diversity at 'face value'. *Behav. Brain Sci.* **32**, 464–465 (2009).

37. Rizzi, L. The discovery of language invariance and variation, and its relevance for the cognitive sciences. *Behav. Brain Sci.* **32**, 467–468 (2009).

38. Greenberg, J. H. in *Universals of Language* (ed. Greenberg, J. H.) 73–113 (MIT Press, 1963).

39. Bickel, B., Witzlack-Makarevich, A., Choudhary, K. K., Schlesewsky, M. & Bornkessel-Schlesewsky, I. The neurophysiology of language processing shapes the evolution of grammar: evidence from case marking. *PLOS ONE* **10**, e0132819 (2015).

40. Aissen, J. Differential object marking: iconicity vs. economy. *Nat. Lang. Linguist. Theory* **21**, 435–483 (2003).

41. Dunn, M., Greenhill, S. J., Levinson, S. C. & Gray, R. D. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* **473**, 79–82 (2011).

42. Gavin, M. C. et al. Toward a mechanistic understanding of linguistic diversity. *BioScience* **63**, 524–535 (2013).

43. Skirgård, H. et al. Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Sci. Adv.* **9**, eadg6175 (2023).

44. Plank, F. & Filimonova, E. The Universals Archive: a brief introduction for prospective users. *STUF - Lang. Typology Univers.* **53**, 109–123 (2006).

45. Piantadosi, S. T. & Gibson, E. Quantitative standards for absolute linguistic universals. *Cogn. Sci.* **38**, 736–756 (2014).

46. Gildea, S. & Zúñiga, F. Referential hierarchies: a new look at some historical and typological patterns. *Linguistics* **54**, 483–529 (2016).

47. Nichols, J. *Linguistic Diversity in Space and Time* (Univ. of Chicago Press, 1992).

48. Bürkner, P.-C. brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).

49. R Core Team. A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2023); https://www.R-project.org/

50. Bouckaert, R. et al. Global language diversification is linked to socio-ecology and threat status. Preprint at *SocArXiv* https://doi.org/10.31235/osf.io/f8tr6 (2022).

51. Pagel, M. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. B* **255**, 37–45 (1994).

52. Pagel, M. & Meade, A. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Naturalist* **167**, 808–825 (2006).

53. Haiman, J. Iconic and economic motivation. *Language* **59**, 781 (1983).

54. Croft, W. in *Agreement in Natural Language: Approaches, Theories, Descriptions* (eds Barlow, M. & Ferguson, C. A.) 159–179 (CSLI Publications, 1988).

55. Haspelmath, M. Frequency vs. iconicity in explaining grammatical asymmetries. *Cogn. Linguist.* **19**, 1–33 (2008).

56. Plank, F. What agrees with what in what, generally speaking? *EUROTYP Working Pap.* **VII**, 20 (1994).

57. Jäger, G. & Wahle, J. Phylogenetic typology. *Front. Psychol.* **12**, 682132 (2021).

58. Hartung, K., Jäger, G., Gröttrup, S. & Georges, M. Typological word order correlations with logistic brownian motion. In *Proc. 4th Workshop on Research in Computational Linguistic Typology and Multilingual NLP* (eds Vylomova, E. et al.) 22–26 (Association for Computational Linguistics, 2022).

59. Sheehan, M., Biberauer, T., Roberts, I., Holmberg, A. *The Final-over-Final Condition: A Syntactic Universal* (MIT Press, 2017).

60. Jing, Y., Blasi, D. E. & Bickel, B. Dependency length minimization and its limits: a possible role for a probabilistic version of the final-over-final condition. *Language* **98**, 397–418 (2022).

61. Lehmann, W. P. A structural principle of language and its implications. *Language* **49**, 47 (1973).

62. Greenberg, J. H. in *Approaches to Language Typology* (eds Shibatani, M. & Bynon, T.) 145–166 (Oxford Univ. Press, 1995).

63. Kiparsky, P. in *Parameters of Morphosyntactic Change* (eds Van Kemenade, A. & Vincent, N.) 460–494 (Cambridge Univ. Press, 1997).

64. Sinnemäki, K. in *Measuring Grammatical Complexity* (eds Newmeyer, F. J. & Preston, L. B.) 179–201 (Oxford Univ. Press, 2014).

65. Levshina, N. Token-based typology and word order entropy: a study based on universal dependencies. *Linguist. Typology* **23**, 533–572 (2019).

66. Shcherbakova, O. et al. The evolutionary dynamics of how languages signal who does what to whom. *Sci. Rep.* **14**, 7259 (2024).

67. Stassen, L. in *Meaning and Grammar. Cross-Linguistic Perspectives* (eds Kefer, M. & van der Auwera, J.) 179–201 (Mouton de Gruyter, 1992).

68. Keenan, E. L. The functional principle: generalizing the notion of 'subject of'. *Chic. Linguistic Soc.* **10**, 298–309 (1974).

69. Nichols, J. What, if anything, is typology? *Linguist. Typology* **11**, 231–238 (2007).

70. Cinque, G. Deriving Greenberg's universal 20 and its exceptions. *Linguist. Inq.* **36**, 315–332 (2005).

71. Biberauer, T. & Roberts, I. Changing EPP parameters in the history of English: accounting for variation and change. *Engl. Lang. Linguist.* **9**, 5–46 (2005).

72. Harley, H. & Ritter, E. Person and number in pronouns: a feature-geometric analysis. *Language* **78**, 482–526 (2002).

73. Toosarvandani, M. The interpretation and grammatical representation of animacy. *Language* **99**, 760–808 (2023).

74. Kiparsky, P. in *Linguistic Universals and Language Change* (ed. Good, J.) 23–53 (Oxford Univ. Press, 2008).

75. Levinson, S. C. The Dark Matter of Pragmatics: Known Unknowns. Elements in Pragmatics (Cambridge Univ. Press, 2024).

76. Hammarström, H., Forkel, R., Haspelmath, M. & Bank, S. Glottolog 4.3. *Zenodo* https://doi.org/10.5281/ZENODO.4061162 (2020).

77. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Soft.* **76**, 1–32 (2017).

78. Guzmán Naranjo, M. & Becker, L. Statistical bias control in typology. *Linguist. Typology* **26**, 605–670 (2022).

79. Sinnemäki, K. Word order in zero-marking languages. *Stud. Lang.* **34**, 869–912 (2010).

80. Xie, W., Lewis, P. O., Fan, Y., Kuo, L. & Chen, M. H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst. Biol.* **60**, 150–160 (2011).

81. Raftery, A. E. in *Markov Chain Monte Carlo in Practice* (eds Gilks, W. R., Richardson, S. & Spiegelhalter, D. J.) 163–188 (Springer, 1996).

82. Beaulieu, J., O'Meara, B. C. & Donoghue, M. J. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Syst. Biol.* **62**, 725–737 (2013).

83. Beaulieu, J., O'Meara, B. C., Oliver, J. C. & Boyko, C. corHMM: Hidden Markov Models of Character Evolution (R package, 2022).

84. FitzJohn, R. G., Maddison, W. P. & Otto, S. P. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* **58**, 595–611 (2009).

85. Lewis, P. O. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925 (2001).

86. Wang, L.-G. et al. Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Mol. Biol. Evol.* **37**, 599–603 (2020).

87. Xu, S. et al. Ggtree: a serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta* **1**, e56 (2022).

Acknowledgements

Author contributions

R.D.G., S.J.G., H.S. and A.V. initiated the study. R.D.G., H.S. and Q.D.A. provided resources. A.V., O.S., H.S. and H.J.H. assembled the data. A.V., O.S., C.R. and S.J.G. performed data curation and analyses. A.V., O.S., H.S. and H.J.H. validated results. A.V. wrote the original draft. R.D.G., Q.D.A., S.J.G., H.S., H.J.H., O.S. and A.V. wrote the paper.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-025-02325-z.

Correspondence and requests for materials should be addressed to Annemarie Verkerk or Russell D. Gray.

Peer review information *Nature Human Behaviour* thanks Gabriel Hassler, Ljiljana Progovac, Michelle Sheehan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature portfolio

Corresponding author(s): *Double-anonymous peer review submissions: write DAPR and your manuscript number here instead of author names.*

Last updated by author(s): *YYYY-MM-DD*

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection | No code was used to create primary data, of course, but (base) R (version: 4.3.0) was used to manipulate datafiles, such as merge datafiles, import information from Grambank, create appropriate datafiles for BayesTraits, etc.

Data that was taken from existing datasets:
Glottolog: version 4.3
Grambank: version 1.0

Data analysis	All code for analysis we used is open source. For R we used packages: brms (R version: 4.3.2, package version: 2.20.3) which depends on STAN (R version: 4.3.2, package version: 2.33.0); geoR (R version: 4.3.2, package version: 1.9-2); fields (R version: 4.3.2, package version: 15.2); ape (R version: 4.3.2, package version: 5.7-1); corHMM (R version: 4.4.0, package version: 2.8); treeio (R version: 4.4.0, package version: 1.24.3); ggtree (R version: 4.4.0, package version: 3.11.0); lme4 (R version: 4.3.0, package version: 1.1-34); lmerTest (R version: 4.3.0, package version: 3.1-3) We used BayesTraits version 4.1.2 (compiled with OpenMP support).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
  - Accession codes, unique identifiers, or web links for publicly available datasets
  - A description of any restrictions on data availability
  - For clinical datasets or third party data, please ensure that the statement adheres to our policy

All the data that support the findings of this study are available via GitHub at https://github.com/SimonGreenhill/TestingLinguisticUniversals

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Our study is closest to an evolutionary study design. We use a comprehensive new global grammatical database (Grambank) to test 191 proposed grammatical universals. The data in Grambank ultimately rely on linguistic analyses made by experts on the language and typologists. Using Bayesian phylogenetic methods we test these putative universals while controlling for spatial and phylogenetic non-independence. Here we conduct four distinct experiments: 1) uncontrolled GLMM models; 2) spatiophylogenetic GLMM models 3) GLMM models with a categorial control for language family; 4) BayesTraits/overt evolutionary modeling without correcting for spatial autocorrelation.
Research sample	Here we sample human languages and universals. Regarding the latter, we evaluated over 2,000 documented universals in the Universals Archive (Plank and Filimonova 2006, https://typo.uni-konstanz.de/rara/category/universals-archive/). We compared the morphosyntactic features involved in these universals to the data in Grambank version 1.0 (Skirgård et al. 2023) to determine which

of them could be analyzed using features from Grambank. Then we reformulated those universals that were testable, first by splitting complex universals that address multiple ostensibly associated linguistic features into simple implicational generalizations (which we refer to here as simply 'universals' or 'generalizations'). Second, we matched feature constellations from generalizations to Grambank questionnaire questions and combinations thereof. Since Grambank only covers morphosyntax, any generalizations involving phonology or semantics were excluded; many other claims were excluded because they were too specific, were explicitly diachronic, or were not implicational. From 146 original universals that matched the Grambank variables and our criteria for synchronic implicational claims, we formulated 191 simple universals to test. Regarding languages, we review the language sample in SI Sect. 1. As for sample coverage per family and per area, we refer to the Supplementary Information of Skirgård et al. (2023). The overall sampling in Grambank across areas can be assessed in Figure S2 of their paper, which shows the coverage over macro-areas (as defined in Glottolog, Hammarström et al. 2022). The coverage over macro-areas is even, all areas have between 24% and 34% coverage. Grambank version 1.0 covers between 13% and 68% of the 15 largest language families as listed on Glottolog (Hammarström et al. 2022). Coverage for Austroasiatic, Isolates (as a group), and Pama-Nyungan is fair, while Indo-European and Tai-Kadai languages are underrepresented.

Sampling strategy	For both universals and languages, a maximally exhaustive sampling strategy was chosen. We include all universals that are in the Universals Archive (see above) and that are testable using Grambank. We include all languages that have the necessary information in Grambank to test the universals on.
Data collection	See above. The core authors responsible for data collection are Hannah Haynie, Olena Shcherbakova, Hedvig Skirgård, and Annemarie Verkerk.
Timing	Data collection was performed between June 2019 and June 2021, intermittedly given involvements in other projects by all parties involved.
Data exclusions	See above on excluding universals that could not be studied using Grambank data.
Non-participation	We did not have any (human) participants.
Randomization	Randomization is irrelevant for this research.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	Antibodies
☒	Eukaryotic cell lines
☒	Palaeontology and archaeology
☒	Animals and other organisms
☒	Clinical data
☒	Dual use research of concern
☒	Plants

Methods

n/a	Involved in the study
☒	ChIP-seq
☒	Flow cytometry
☒	MRI-based neuroimaging

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A