

Gerhard Jäger, University of Tübingen, gerhard.jaeger@uni-tuebingen.de

Tracking the change that leads to typological variation

Word-order universals and the phylogenetic comparative method

Abstract: Subject-object-verb (SOV) and subject-verb-object (SVO) word order are the most common word orders across the languages of the world, jointly covering more than 80% of the documented languages. To what degree this biased distribution reflects genuinely universal tendencies of language change, as opposed to historical contingencies, is an open question. The current distribution is shaped by at least three confounded factors: the inherent directionality of language change, differential diversification rates of language families with different typological profiles, and the properties of ancestral languages inherited by their extant descendants. Plain counting of the current distribution cannot disentangle these factors. I deploy the *phylogenetic comparative method* from computational biology and model word-order change as a continuous-time Markov chain (CTMC) running on language phylogenies inferred via Bayesian phylogenetic inference. The language sample comprises 1,567 languages: 1,443 from 107 multi-member families and 124 singleton lineages; family phylogenies were inferred from lexical data using topological constraints from Glottolog and time calibrations from a Bayesian global tree. The fitted model assigns an equilibrium probability of 65% to SOV and 28% to SVO, clearly diverging from the apparent symmetry between the two types. The dominant SOV/SVO exchange is asymmetric, with a net flux from SVO towards SOV.

Keywords: typological universals, word order, phylogenetic comparative method, continuous-time Markov chain, Bayesian phylogenetics

This invited contribution is based on the plenary lecture given on **7 March 2025** at the Annual Conference of the Linguistic Society of Germany.

Cite as: Jäger, Gerhard. 2026. Tracking the change that leads to typological variation: Word-order universals and the phylogenetic comparative method. *Zeitschrift für Sprachwissenschaft* 45(2).
<https://doi.org/10.18148/zs/2026-2008>.

Published by: publikon / dgfs (Linguistic Society of Germany) under a Creative Commons 4.0 BY license

1 Introduction

The standard approach in linguistic typology has been, since Greenberg's (1963) pioneering work, to collect samples of languages, to classify them according to structural features, and to interpret skewed distributions as evidence for universal tendencies. A canonical example is word order. Out of the six logically possible arrangements of subject (S), object (O) and verb (V) in a transitive clause, two of them – SVO and SOV – account for the overwhelming majority of the documented languages. In the World Atlas of Language Structures (Haspelmath et al. 2008), for instance, SOV is the most frequent dominant order and SVO the second most frequent.

Making the step from a skewed synchronic distribution to a claim about diachronic tendencies is not trivial though. Languages within a genealogical family are not independent data points; they all descend from a common ancestor and inherit properties of that ancestor. If, for instance, an SOV proto-language diversifies at a high rate to many present-day daughter languages which all inherit the proto-language's SOV order, this type appears inflated in the current frequency distribution. A frequently used mitigation strategy is to work with *stratified samples* of languages, e.g., by using only one language per family. It was shown, however, by Maslova (2000) that this might be insufficient to sidestep the problem. The current distribution is the product of three interlocked factors, (1) universal tendencies of the direction and rate of diachronic change, (2) differences in the diversification rate between lineages with different feature values, and (3) the feature value of the common ancestral language. Stratified sampling addresses the second issue, but not the two others. According to Maslova (2000), it is mathematically impossible to separate these factors from present-day language counts alone.

The *phylogenetic comparative method* from computational biology (see Cooper & Freckleton 2026 for an overview) is designed to address precisely those problems. It was designed in the context of evolutionary biology (see, e.g., Felsenstein 1981; Pagel et al. 2013). For a small number of unordered discrete states, this approach models evolution as a *continuous time Markov chain* (CTMC) operating on the branches of a phylogenetic tree. If the parameters of the CTMC, as well as the topology of the tree and the lengths of all branches, are known, the model assigns a probability to a vector of observations at the leaves of the tree. This enables statistical inference – if the full phylogeny and the vector of observations are known, the parameters of the CTMC can be inferred. These parameters in turn determine the stationary distribution and allow probabilistic statements about the states at the root and the intermediate nodes of the tree, as well as about the expected number of transitions between the possible states. Since the model is conditional on the tree structure, it factorizes universal properties of the diachronic process – encapsulated in the rate matrix – from historically contingent facts represented by the phylogeny. This approach is applied here to the evolution of word-order types, using Bayesian inference. Following the approach by Maurits & Griffiths (2014) and Jäger & Wahle (2021), I will

use not a single tree covering all languages under consideration, but a forest containing one tree per lineage. Like Jäger & Wahle (2021), I do not use a single phylogeny per lineage but a sample from a probability distribution over phylogenies which were obtained in a preparatory step from lexical data via Bayesian phylogenetic inference.

For the present study I used the word-order dataset of Gell-Mann & Ruhlen (2011), comprising 1,567 languages matched to unique glottocodes; full sample details are given in Section 3.1. Phylogenetic trees are inferred from lexical data obtained from the Lexibank database (List et al. 2022): 1,443 languages belong to 107 genealogical families for which posterior tree samples were inferred using MrBayes (Ronquist & Huelsenbeck 2003), while the remaining 124 are treated as independent singleton lineages. The posterior distribution over a universal rate matrix is then inferred using a combination of *No U-Turn Sampling* (NUTS) (Hoffman & Gelman 2014) and *Gibbs sampling* in the Turing.jl framework for Julia (Ge, Xu & Ghahramani 2018).

The results are inconsistent with two influential proposals regarding the long-term diachronic trends in word-order evolution. Gell-Mann & Ruhlen (2011) argue that Proto-World was SOV, and that there is a dominant directed path $SOV \rightarrow SVO$ which can only be reversed under language contact. SVO is then in bidirectional exchange with VSO/VOS. This is not supported by this study's finding. Changes $SOV \rightarrow SVO$ are inferred to be less numerous than changes in the opposite direction. Ferrer-i-Cancho (2016) proposed that transition probabilities are inversely related to the number of pairwise neighbour permutations in a permutation circle of the six possible orders. This is likewise not supported by my findings.

The main finding of this study is that in the long-term stationary distribution, SOV, with 65% is about twice as likely as SVO with 28%, with VSO a distant third at 5%. The three remaining orders jointly contribute less than 3%.

The paper is organized as follows. Section 2 reviews the typological background, the non-independence problem, and earlier quantitative approaches to word-order change. Section 3 describes the data, tree inference, and the CTMC model. Section 4 presents the estimated rate matrix, equilibrium distribution, stochastic character mapping, and ancestral-state reconstructions for selected families. Section 5 discusses implications and limitations, and Section 6 concludes. Technical details of the computational pipeline are provided in the online repository.

2 Background

2.1 Typological universals and word order

Typology is the subfield of linguistics studying the structural properties of natural languages in a comparative way. The goal is to identify linguistic features and feature combinations which

are universally present, universally absent or distributed across the languages of the world in a non-random way that is not determined by common descent. The modern debate was initiated by Greenberg (1963), who investigated a diverse sample of 30 languages, which led him to 45 universals of varying strength, many of them pertaining to word order.

Greenberg's approach distinguishes between *absolute universals*, which hold without exception across all known languages, and *statistical universals* (tendencies), which occur significantly more often than chance would predict. Word order falls almost entirely in the second category. Universal 1 states that in declarative sentences with nominal subject and object, the dominant order almost always places the subject before the object – ruling out OVS, VOS, and OSV as primary orders. Beyond this subject-first preference, Greenberg identified clusters of co-occurring properties: languages with verb-before-object (VO) order tend to have prepositions, adjectives after nouns, and relative clauses after nouns, while those with object-before-verb (OV) order tend towards the mirror image – postpositions, adjectives before nouns, relative clauses before nouns (Universals 2, 3, and many others). These correlations suggest that word order is not an arbitrary parameter but reflects a deeper typological alignment that constrains the space of possible change.

The distributional picture is strikingly skewed. In the 1,567-language sample used in the present study (Figure 1), SOV accounts for 48.5% of the languages and SVO for 39.0%, together covering 87.5% of the sample. VSO is a distant third at 8.0%, attested mainly in Semitic, Celtic, and Oceanic languages. VOS (2.6%), OVS (1.2%), and OSV (0.6%) are rare. The geographic distribution (Figure 1) reflects both broad typological tendencies and the demographic history of specific families: SOV predominates across South and Central Asia and East Africa; SVO is dominant in sub-Saharan Africa and mainland South-East Asia; VSO clusters in the Pacific and the Semitic-speaking world.

2.2 The non-independence problem

The most immediate obstacle to inferring evolutionary tendencies from synchronic distributions is the genealogical non-independence of languages. Languages within a family all descend from a common ancestor and therefore share its entire evolutionary history. If Proto-Turkic was predominantly SOV, for instance, then the dozens of extant Turkic languages – all SOV – do not represent independent data points in favour of SOV; they represent a single historical thread, diversified but not statistically replicated. Treating each language as an independent observation inflates the apparent evidence for whatever order is frequent in large, well-sampled families.

The standard response in typological work is *stratified sampling*: restrict the sample to one language per family and, where possible, draw from geographically diverse families (Dryer 1992; Haspelmath et al. 2008). Stratification reduces the inflationary effect of large families:

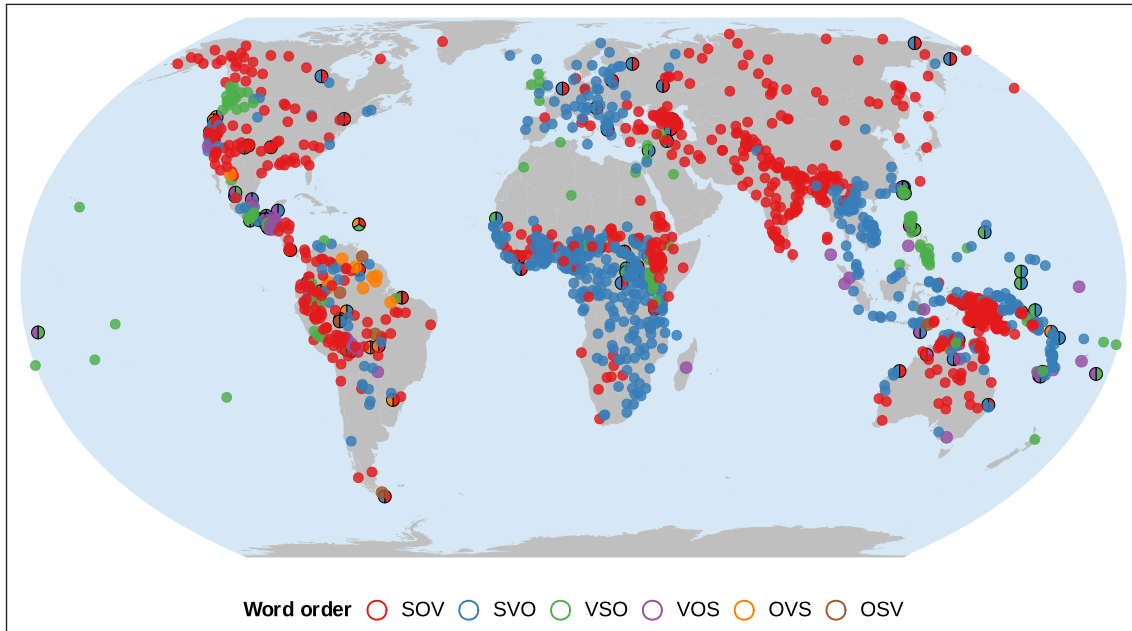


Figure 1: Geographic distribution of the six basic word-order types across the languages in the sample. Each language with a single dominant order is coloured by that order; the 104 languages with more than one attested order are omitted for clarity

a sample that includes one Indo-European language, one Dravidian language, and one Turkic language weights the SOV-heavy Eurasian macro-area less heavily than one that includes every Turkic daughter.

Maslova (2000) demonstrated, however, that stratified sampling addresses only one of three distinct sources of bias. Even a sample containing exactly one language per family conflates three independent confounds. First, if there is a universal tendency for languages to change in one direction – say, from OV to VO more often than from VO to OV – this tendency impacts the extant distribution, but the shape of that impact depends on how long the process has been running and how often families have split. Second, if unacknowledged macro-families with feature value A give rise to more daughter lineages than macro-families with value B, then A may be over-represented in any sample, independently of any tendency in change direction. Third, if the feature values of deep proto-languages were themselves non-uniformly distributed – whether through a single proto-world bottleneck or through correlated founding effects in multiple independent lineages – that ancestral configuration propagates forward in time regardless of subsequent dynamics. As Maslova (2000) proved formally, these three quantities are jointly under-determined by the synchronic distribution: many combinations of values for the three factors are, in principle, consistent with the observed counts. No amount of stratification can resolve this indeterminacy.

The conclusion is sobering: to make progress on the question of universal word-order tendencies, one needs not just a better sample but a fundamentally different kind of evidence. Precisely this is what the phylogenetic comparative method provides, by conditioning on the tree rather than marginalizing over it.

2.3 Previous approaches

Several quantitative studies have attempted to go beyond frequency counts and model word-order change directly. Gell-Mann & Ruhlen (2011) inferred SOV as a proto-world state and argued for a dominant SOV $\rightarrow$ SVO drift. All transitions in the opposite direction are, according to them, the result of contact-induced change. Ferrer-i-Cancho (2016) proposed a model where the transition probability between two states is inversely related to the number of pairwise neighbour swaps required to transform one order into the other. Maurits & Griffiths (2014) took a more principled approach, fitting a CTMC model to word-order data for seven language families (Afro-Asiatic, Austronesian, Indo-European, Niger-Congo, Nilo-Saharan, Sino-Tibetan, and Trans-New Guinea) using neighbour-joining trees. Their inferred rate matrix (see Figure 2) suggests that SOV is the most probable proto-language state in most families. The study is methodologically the closest predecessor to the present one, but it is limited in scale, considering only seven families and using trees without topology uncertainty.

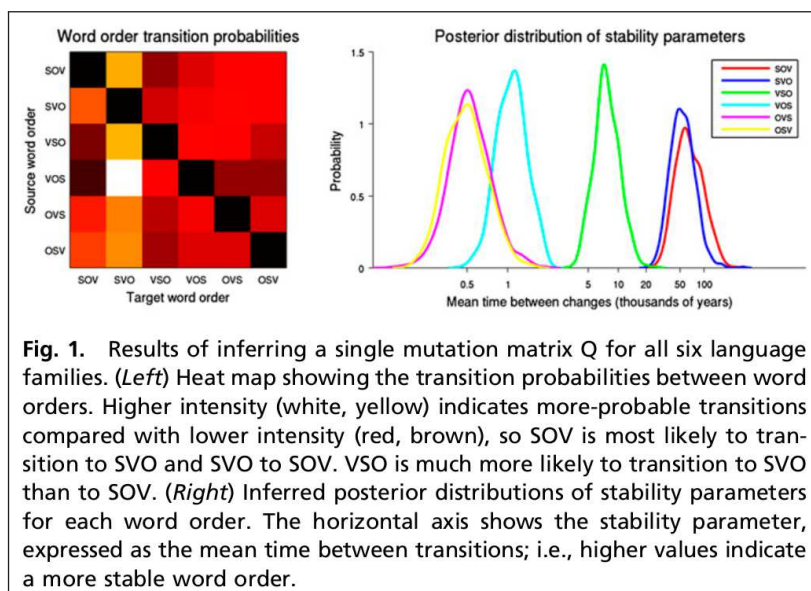


Figure 2: Estimated transition rates among major word orders from Maurits & Griffiths (2014); the original figure caption reads “six language families,” but the study in fact analysed the seven listed in Section 2.3. The figure is reproduced from the original article (© the authors) and is not covered by the CC BY licence of the present article

Dunn et al. (2011) applied the correlated-evolution test of Pagel & Meade (2006) to four families and found that word-order co-occurrence patterns differ across families, concluding that universals are better understood as family-specific tendencies. Jäger & Wahle (2021) revisited this with CTMC models for eight word-order features across 175 lineages, finding decisive support for a universal (shared-rate) model over lineage-specific alternatives and credible evidence for thirteen pairwise correlations. The Dunn et al. conclusion was argued to rest on data sparseness. The present paper builds on this foundation, shifting from pairs of binary features to the full six-state system and using the full Lexibank rather than just the Automated Similarity Judgment Program (ASJP) of Wichmann, Holman & Brown (2018).

3 Methods

3.1 Data

Typological observations come from Gell-Mann & Ruhlen’s (2011) word-order dataset, which codes the dominant word order of transitive declarative clauses as one of six basic constituent orders: SOV, SVO, VSO, VOS, OVS, and OSV, encoded as binary flags. Languages for which multiple flags are set (i.e. several co-existing orders) are retained with a uniform tip-likelihood distribution over their attested orders. The geographic distribution of the sample (Figure 1) is broad, covering all major linguistic areas, though sampling density varies with the documentation depth of individual regions and families. Phylogenetic trees are inferred from lexical data obtained from the Lexibank database (List et al. 2022), a cross-linguistic lexical repository that consolidates wordlist data from a wide range of sources. The extraction of character matrices from word lists follows the workflow of Jäger (2025), in which phonetic transcriptions are aligned with a pair-hidden Markov model and the presence or absence of sound classes in the resulting alignment columns is recorded as binary characters; the procedure is described in more detail in Section 3.3.

Genealogical family membership follows Glottolog (Hammarström et al. 2026). For each of the 107 non-singleton families, Bayesian phylogenetic inference yielded a posterior sample of 50 trees used in all downstream analyses; the inference procedure is described in Section 3.3. In total, 1,443 languages belong to one of those families. The remaining 124 languages in the sample are either isolates or are the only representative of their family (see Table 1).

Table 2 and Figure 3 summarize the distribution of word-order types in the 1,567-language sample. SOV and SVO together account for 48.5% and 39.0% of the sample respectively – 87.5% jointly. VSO adds a further 8.0%, leaving less than 5% distributed across VOS (2.6%), OVS (1.2%), and OSV (0.6%). These raw counts are, however, dominated by a small number of large families:

Table 1: Dataset summary

Subset	Languages	Lineages
Full GR sample	1,567	231
Families with posterior phylogenies	1,443	107
Singleton lineages	124	124

Table 2: Distribution of major word-order types in the 1,567-language sample. Languages coded with more than one dominant order have their weight distributed evenly across those orders. Raw percentages treat each language equally; weighted percentages weight each language by the inverse size of its lineage, so every lineage contributes equal total weight

	SOV	SVO	VSO	VOS	OVS	OSV
<i>n</i>	760	612	125	42	18	10
% (raw)	48.5	39.0	8.0	2.6	1.2	0.6
% (weighted)	66.5	22.5	5.5	2.4	1.7	1.4

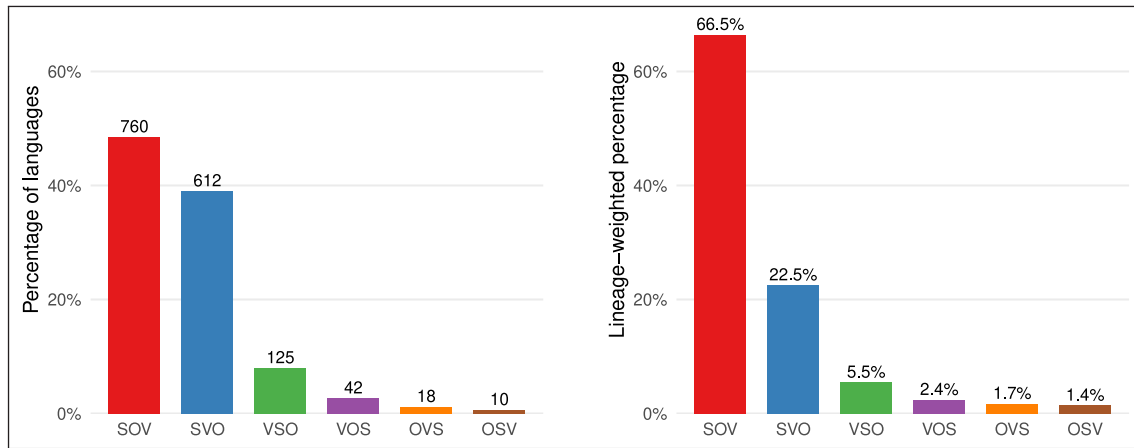


Figure 3: Left: raw frequencies of the six major word-order types in the 1,567-language sample (each language counts equally). Right: lineage-weighted frequencies (each language is weighted by the inverse size of its lineage, so every lineage contributes equal total weight). The shift from SVO towards SOV in the weighted distribution reflects the dominance of SVO in a small number of large, recently diversified families

Atlantic-Congo (216 languages, predominantly SVO) and Austronesian (209 languages, mostly SVO or VSO) alone account for a large share of the SVO total. As discussed in Section 2.2, raw counts cannot disentangle universal tendencies from family-size effects and ancestral states; that is precisely the task of the CTMC analysis that follows.

3.2 The phylogenetic comparative method

The phylogenetic comparative method (Felsenstein 2004; Pagel et al. 2013) models the evolution of a discrete trait along a phylogenetic tree as a continuous-time Markov chain (CTMC). In the present application the trait is word order, taking one of $k = 6$ values labelled SOV, SVO, VSO, VOS, OVS, OSV (Figures 4 and 5, displaying only three states for illustrative purposes).

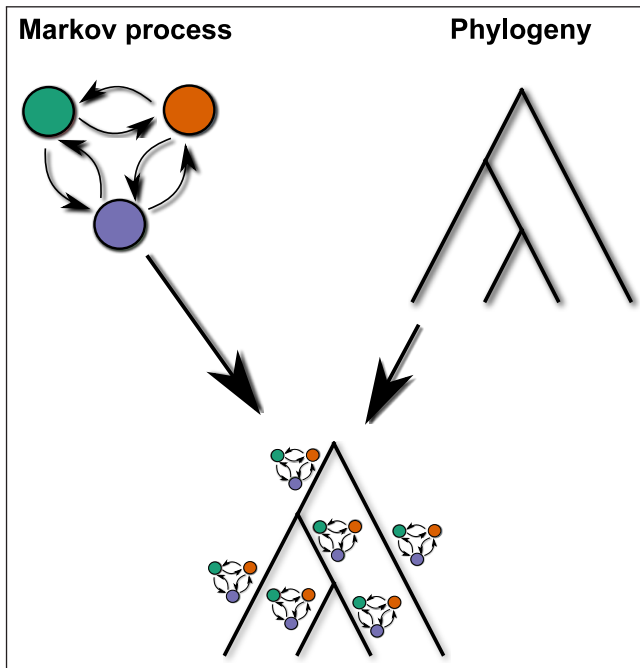


Figure 4: Schematic structure of the phylogenetic CTMC model. Independent but identically parameterized instances of a CTMC run on the branches of the phylogeny. Figure taken from Jäger & Wahle (2021)

A CTMC is governed by an instantaneous rate matrix $Q = (q_{ij})_{6 \times 6}$ (cf. Figure 5a). The off-diagonal entries $q_{ij} \geq 0$ ($i \neq j$) give the rate at which the process transitions from state i to state j ; the diagonal entries satisfy $q_{ii} = -\sum_{j \neq i} q_{ij}$, so that every row sums to zero. Given the rate matrix and a branch of length t (in kiloyears), the transition probability matrix is the matrix exponential $P(t) = e^{Qt}$, with $P_{ij}(t)$ giving the probability of moving from state i to state j over time t . When the chain is *irreducible* – that is, every state can be reached from every other state – then for large t every row of $P(t)$ converges to the same vector $\pi = (\pi_i)_{i=1}^6$, the *stationary distribution*, which is the unique probability vector satisfying $\pi Q = \mathbf{0}$ and $\sum_i \pi_i = 1$. The stationary distribution is the long-run equilibrium of the system: it gives the probability of finding the process in each word-order state after an arbitrarily long time, independently of the starting state. Crucially, π is entirely determined by the rate matrix Q and therefore captures universal tendencies rather than historical starting conditions.

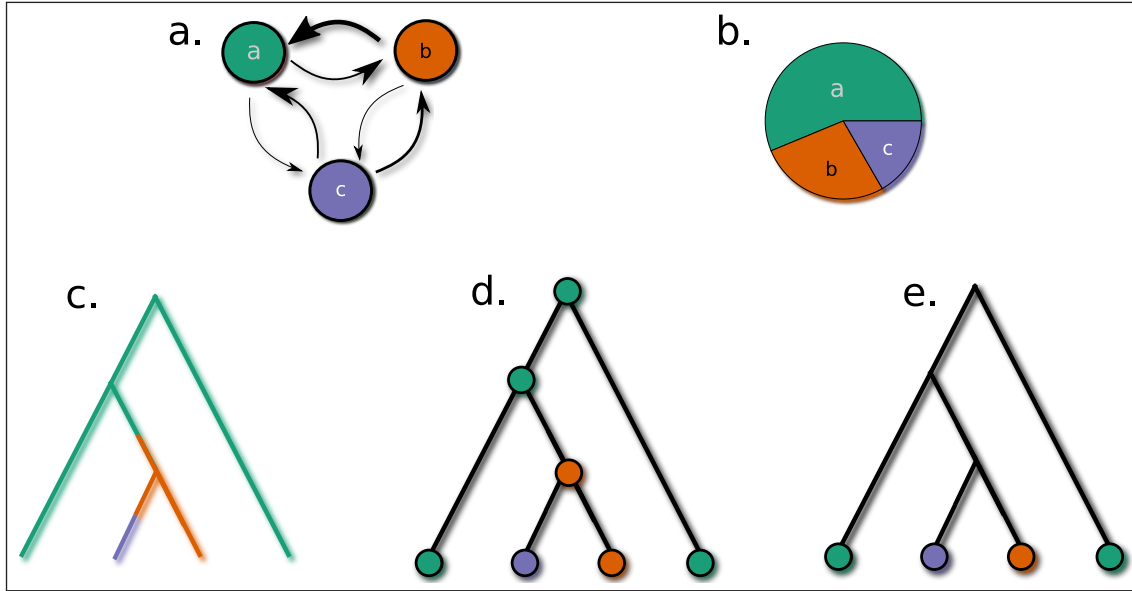


Figure 5: Panels: (a) continuous-time Markov chain; (b) equilibrium distribution; (c) fully specified history on a phylogeny; (d) marginalized over intra-branch events; (e) marginalized over states at internal nodes. Figure taken from Jäger & Wahle (2021)

The likelihood of the tip observations $y = (y_1, \dots, y_n)$ given Q and a tree T can be computed efficiently via Felsenstein's (1981) *pruning algorithm*, a dynamic programming technique. The resulting likelihood marginalizes over all possible states at the root and the intermediate nodes of the tree. The marginalization over internal states means that no prior on ancestral configurations needs to be specified; everything is inferred from the tip observations and the rate matrix. The likelihood computation assumes that the root state is drawn from the stationary distribution.

Because the phylogenies carry uncertainty, I also marginalize over a posterior sample of $K_f = 50$ trees per family. At each MCMC iteration the tree index for each family is resampled by Gibbs sampling from the full conditional distribution proportional to the Felsenstein likelihood of each candidate tree. The continuous rate parameters are sampled using the NUTS sampler. This joint sampling of rate parameters and tree indices propagates phylogenetic uncertainty into the posteriors for Q and π . The full likelihood across all 231 lineages is the product of the likelihoods of the lineages. For a singleton lineage, this simply amounts to the stationary probability of its observed state.

3.3 Phylogenetic tree inference

For each of the 107 language families I construct a binary character matrix from Lexibank lexical data using the pair-hidden Markov model (pHMM) multiple sequence alignment methodology

described in Jäger (2025).¹ Each cell of the matrix encodes the presence or absence of a sound class in an alignment column; the pHMM alignment implicitly tends to align the corresponding segments of cognate words into the same alignment column without relying on manual cognate judgements.

For each of the 210 most stable concepts from Dellert & Buch’s (2018) concept stability ranking and each language family, such a binary character matrix was constructed. As a next step, for each family these matrices were arranged according to the stability of the corresponding concepts. For the first k such matrices, k ranging from 50 to 210, the joint phylogenetic difficulty of the concatenated matrix was computed using the software *Pythia* (Haag et al. 2022). The k minimizing the difficulty was chosen for that family.

These matrices are supplied to MrBayes 3 (Ronquist & Huelsenbeck 2003) under a two-state substitution model with freely estimated stationary state frequencies with a gamma rate-variation prior and a correction for variable-site ascertainment bias (Lewis 2001). A strict molecular clock is imposed with a lognormal prior on the clock rate. Topological constraints from the Glottolog classification (Hammarström et al. 2026) are enforced as hard monophyly constraints, ensuring that inferred trees respect the established genealogical structure. The root age prior for each family is set as a uniform distribution with bounds taken from the 95% highest posterior density interval of the corresponding clade in the global language MCC tree of Bouckaert et al. (2026), as distributed with release 2.0.0 of their pipeline (Bouckaert 2025); no calibrations are applied to internal nodes. Additionally, the attestation dates of ancient languages, taken from Glottolog, were used for tip dating.

The resulting inferred root ages span from approximately 0.5 kya for recently diverged families to approximately 13 kya for the oldest groups in the sample (Afro-Asiatic; posterior mean 12.8 kya), broadly consistent with the Bouckaert estimates. Each family is analysed in two independent runs, each using one cold and three hot chains; sampling continues until the average standard deviation of split frequencies (ASDSF) falls below 0.01. Runs are then extended as long as the potential scale reduction factor (PSRF) of Gelman & Rubin (1992) exceeds 1.1 for any of the overall parameters recorded by MrBayes (tree height, tree length, clock rate and the parameters of the substitution model); this check is skipped when each of the two runs has fewer than 50 post-burn-in samples, which occurs with some small families. Control runs for these families are reported in the online supplementary material. I extract 50 posterior trees by even thinning of the pooled post-burn-in samples of the two runs. A DensiTree visualization (Bouckaert & Heled 2014) of three representative families’ posterior trees is shown in Figure 6. As can be seen, the posterior tree samples represent both phylogenetic uncertainty and uncertainty regarding branch lengths.

¹ All analyses use the corrected 2026 version of that pipeline, which fixes a sign error in pHMM training and two bugs in MSA construction.

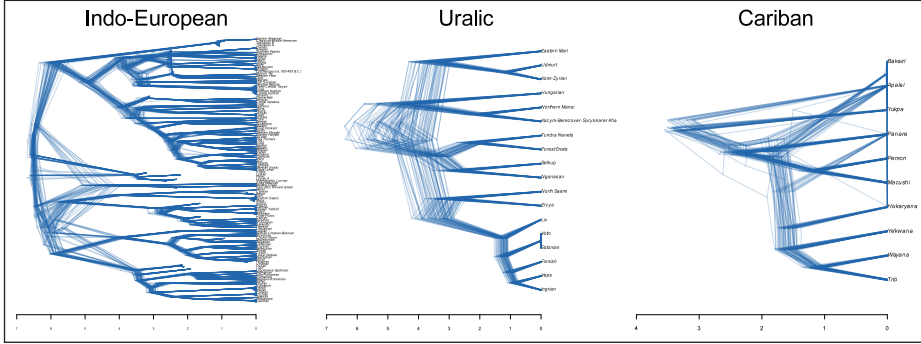


Figure 6: DensiTree visualizations of the 50-tree posterior sample for three language families: Indo-European (121 languages), Uralic (18), and Cariban (10). Overlapping tree topologies reveal uncertainty in branching order and branch lengths

3.4 MCMC inference and ancestral state reconstruction

I implemented the universal CTMC model in the programming language Julia, using the Turing.jl probabilistic programming framework. The prior decomposes into a total rate scale and a sparse allocation across the 30 off-diagonal entries. The total rate λ follows a log-normal prior, $\lambda \sim \mathcal{LN}(-4, 0.75)$, with median ≈ 0.018 transitions per kiloyear and a 95% prior interval of approximately $[0.004, 0.08]$ per kiloyear. This order of magnitude was determined *a priori* by estimating the minimal number of transitions via maximum parsimony. The allocation vector $\mathbf{d}$ across the 30 rate entries follows a symmetric Dirichlet prior, $\mathbf{d} \sim \text{Dir}(0.3 \cdot \mathbf{1}_{30})$; individual off-diagonal rates are then $q_{ij} = 30\lambda d_{ij}$, so that the mean rate per entry equals λ when the allocation is uniform. The concentration parameter $\alpha = 0.3 < 1$ induces sparsity: most of the total rate is *a priori* concentrated on a small subset of transitions, with an expected effective number of active rates of approximately 8. This is a weakly informative prior that places most mass on plausible evolutionary rates while remaining diffuse relative to the data.

To assess prior sensitivity, I re-ran the analysis under a range of alternative hyperparameter configurations, varying the Dirichlet concentration $\alpha \in \{0.1, 0.3, 0.5, 1.0\}$ and the log-normal mean $\mu \in \{-4, -5\}$. Configurations with $\alpha = 0.1$ produced pathological posteriors and were excluded; across all six converged configurations, the inferred equilibrium and waiting-time distributions are highly stable. Full details and figures are provided in the online supplementary material accompanying the deposited code and data (see the Data and code availability statement).

Posterior inference uses a combination of the No-U-Turn Sampler (NUTS) for the continuous rate parameters and Gibbs sampling for the 107 discrete tree-index variables, one per family. At each Gibbs step the tree index k_f for family f is drawn from the discrete full conditional $p(k_f | Q, \mathbf{y}_f) \propto p(\mathbf{y}_f | Q, T_{f,k_f})$, computed exactly from the Felsenstein likelihoods of the 50 candidate trees. I ran

four independent chains, each for 1,500 iterations with a 500-iteration warm-up period, yielding 4,000 post-warm-up samples in total. Convergence is assessed via the PSRF; all parameters satisfy $\text{PSRF} < 1.05$.

Expected transition counts are estimated by the stochastic character mapping (SIMMAP) approach of Bollback (2006): for each posterior draw of Q and tree index, I simulate a complete evolutionary history consistent with the tip observations and accumulate the state-change counts along every branch. The posterior mean transition count matrix summarizes which pairs of word orders exchange most frequently and, importantly, whether the exchange is symmetric or directional. Ancestral state probabilities at internal nodes are obtained as marginal posteriors from the Felsenstein pruning likelihoods, averaged over the joint posterior of Q and the family phylogeny.

4 Results

4.1 Transition rates

The negative values along the diagonal in Figure 7 indicate outgoing rates, i.e., the expected number of transitions out of a state. It can be seen that SOV and SVO are highly stable, with outgoing rates close to 0. The other four states are distinctly less stable, with OSV being the most volatile order. Among the off-diagonal entries, the transitions $\text{OSV} \rightarrow \text{SOV}$, $\text{VOS} \rightarrow \text{SVO}$, $\text{OSV} \rightarrow \text{VSO}$, $\text{VOS} \rightarrow \text{VSO}$, and $\text{OSV} \leftrightarrow \text{OVS}$ have the highest values, with high degree of uncertainty. It is remarkable that the rate $\text{SVO} \rightarrow \text{SOV}$, while small with 0.035 transitions per kiloyear, is about four times as high as the inverse rate $\text{SOV} \rightarrow \text{SVO}$ with just 0.009 transitions per kiloyear. This is in direct contradiction to Gell-Mann & Ruhlen’s (2011) suggested unidirectional preference for $\text{SOV} \rightarrow \text{SVO}$.

The expected time a language spends in a state is the inverse of the absolute value of the outgoing rate. The posterior distributions of these waiting times are shown in Figure 8 and summarized in Table 3.

4.2 Equilibrium distribution

The posterior distribution of the stationary probabilities is given in Table 4 and visualized in Figure 9. They are qualitatively similar to the weighted frequency distribution. SOV dominates with $\approx 65\%$, more than twice as likely as SVO with $\approx 28\%$. VSO accounts for about 5%, and the three orders with object before subject jointly for about 3%.

Figure 10 displays the expected number of transitions within a period of 5 kiloyears, starting from the stationary distribution. The dynamics are dominated by transitions between SOV and

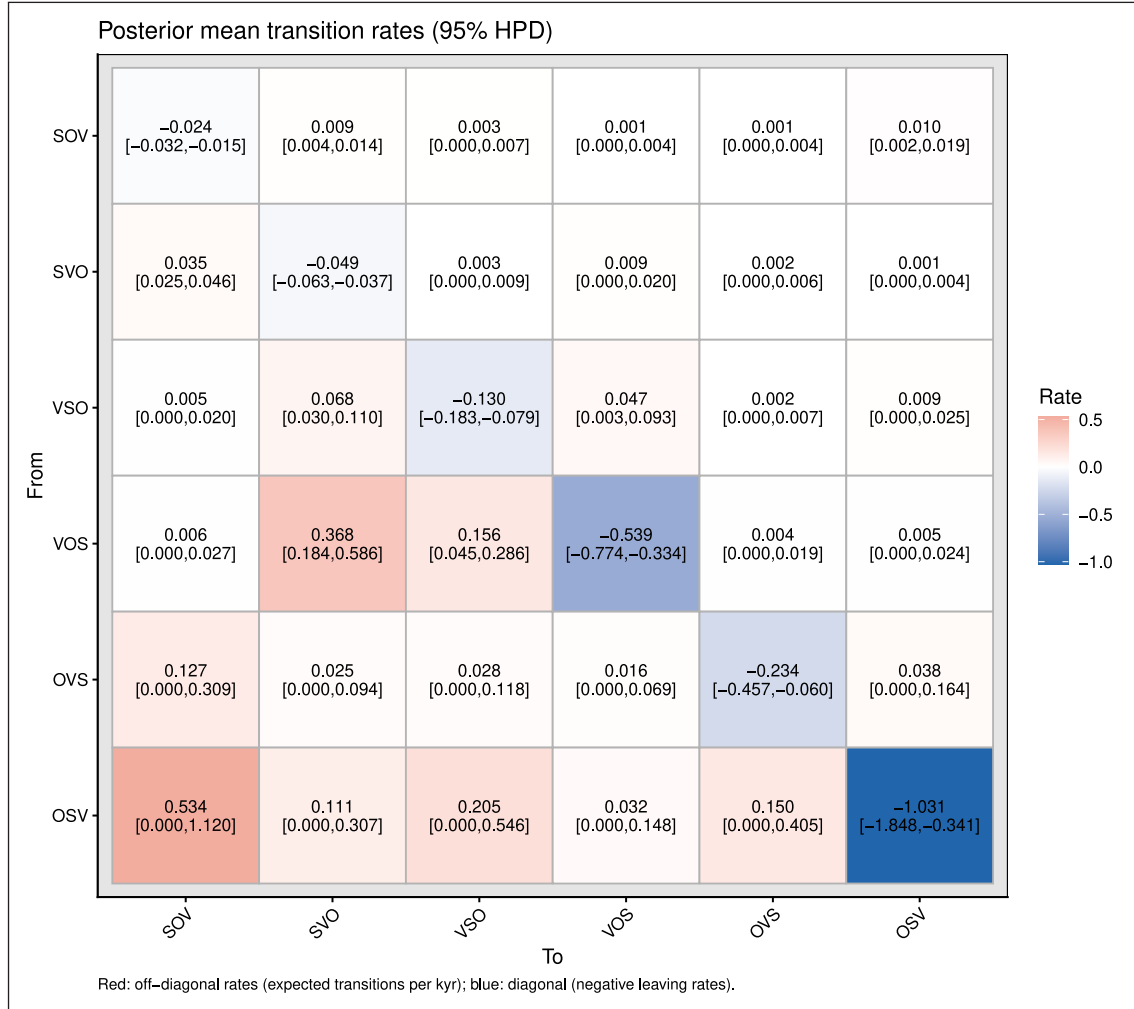


Figure 7: Diagram of the inferred transition-rates for the six major word-order types. Numbers in brackets give the 95% highest posterior density (HPD) interval, i.e. the Bayesian credible interval containing 95% of the posterior probability mass

SVO. Additionally there is a substantial expected flow from SOV to VSO and from VSO to SOV. All other transitions are expected to be very rare.

4.3 Reconstructed change history

Given the posterior distributions of family trees and of the rate matrix, it is possible to simulate the posterior distribution of entire histories of word-order changes across all families. The expected number of transitions between the different states in the history of the 107 families in the sample is given in Table 5.

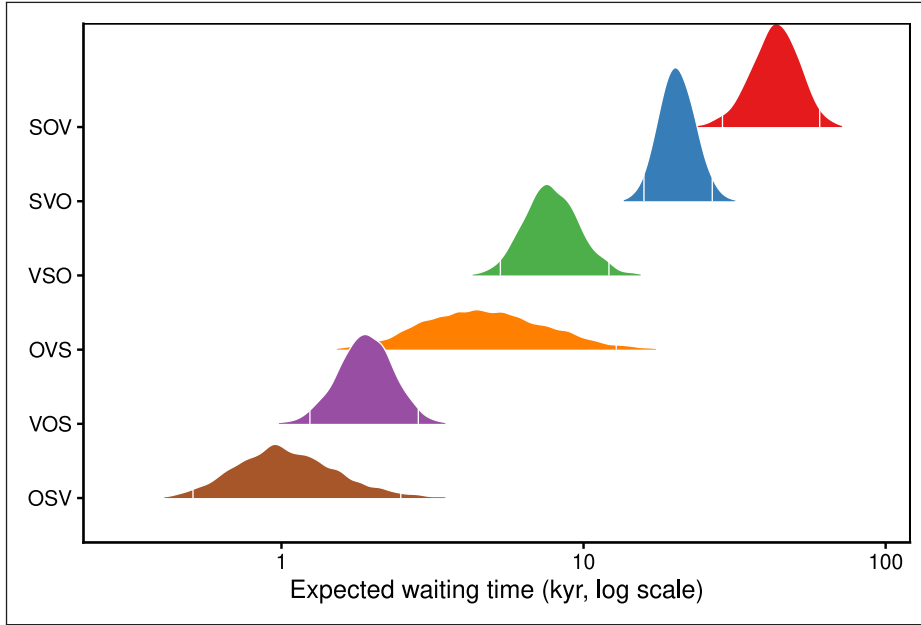


Figure 8: Posterior distributions of expected waiting times for each word-order type (log scale). Vertical lines mark the 95% HPD interval. SOV is by far the most stable state; OSV and VOS are the most transient

Table 3: Posterior expected waiting times (kiloyears) for each word-order type. Values are posterior means with 95% highest posterior density (HPD) intervals

	SOV	SVO	VSO	VOS	OVS	OSV
Mean (kyr)	43.7	20.6	8.0	1.9	5.4	1.1
95% HPD (lo)	27.6	15.8	5.1	1.2	1.6	0.4
95% HPD (hi)	59.0	26.5	11.7	2.8	10.8	2.2

Table 4: Posterior equilibrium distribution π for each word-order type. Values are posterior means with 95% highest posterior density (HPD) intervals; all values in %

	SOV	SVO	VSO	VOS	OVS	OSV
Mean	64.5	28.0	4.6	1.1	1.1	0.7
95% HPD (lo)	59.1	23.3	2.9	0.5	0.3	0.3
95% HPD (hi)	69.8	33.0	6.6	1.7	2.0	1.3

The most remarkable observation about these numbers is perhaps that they are so low. In total, the expected number of transitions is in the low hundreds (mean: 198.3, with a 95% HPD between 179 and 218). This is a small number to draw any definite conclusions from. In other words, the

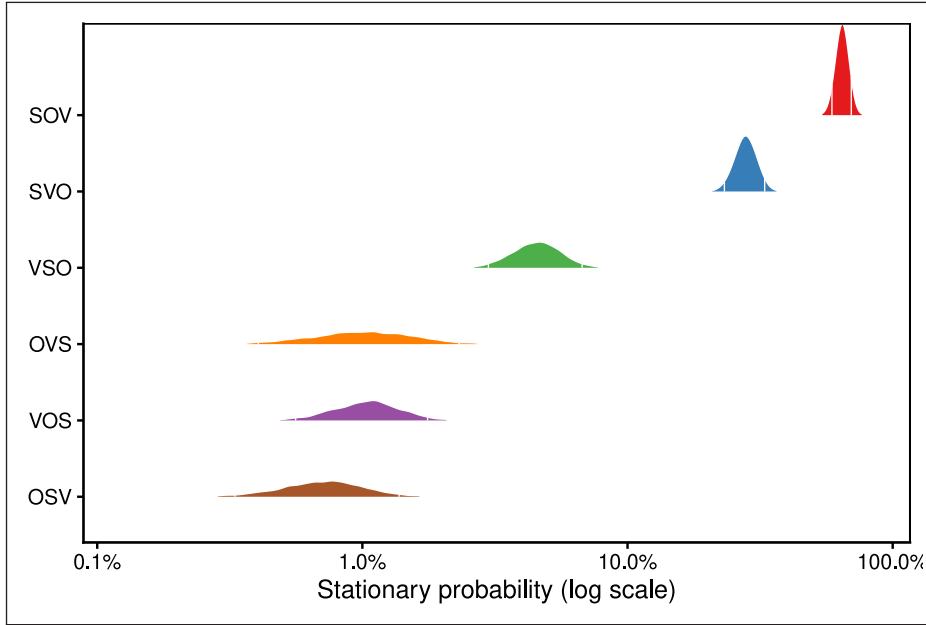


Figure 9: Posterior distribution of the stationary probabilities π_i for each of the six word-order types. Vertical lines mark the 95% HPD interval. SOV is the dominant long-run attractor with a posterior mean of 65%

existing languages – even if we had complete documentation for all of them – might actually be too small a sample of the forces shaping word-order dynamics to clearly identify causal patterns.

The number clearly standing out in Table 5 is the high number for $\text{VOS} \rightarrow \text{SVO}$. Arguably this is an effect of Proto-Austronesian being reconstructed as VOS with high probability (see Section 4.4), leading to a high number of transitions to SVO among the extant Austronesian languages.

The average reconstructed state probabilities at the proto-languages of the lineages are shown in Figure 11. Its overall shape is similar to the stationary distribution, though with somewhat less SVO and more VSO at the roots. This indicates that the inflated SVO numbers we find among the extant languages are in fact the result of fairly recent within-family diversification.

4.4 Case studies in ancestral state reconstruction

In this section I will briefly discuss the simulations for the posterior distribution of states among a few selected families.

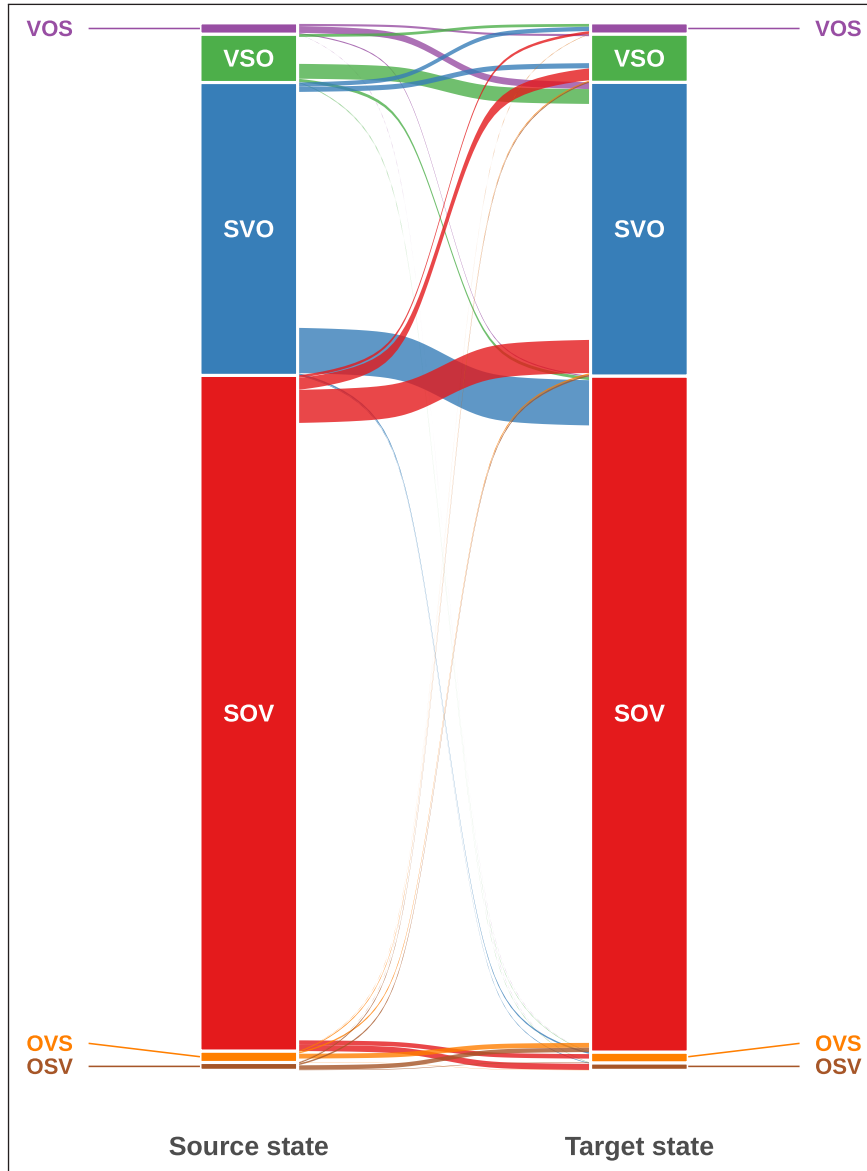


Figure 10: Expected flow between word-order types within 5 kyr, derived from the posterior mean rate matrix. Arrow widths are proportional to the expected probability mass transferred between states

4.4.1 Austronesian

The model reconstructs proto-Austronesian with highest posterior probability on VOS ($\approx 73\%$), followed by VSO ($\approx 27\%$); all verb-medial orders, including the SVO that dominates the extant sample, receive negligible probability at the root (Figure 12). This is broadly consistent with the comparative consensus, which, on the basis of the Formosan languages – the earliest-branching

Table 5: Expected number of transitions summed across all 107 families (posterior mean over 200 SIMMAP iterations). Each cell shows the mean with the 95% HPD interval in brackets. Rows = source state, columns = target state

	SOV	SVO	VSO	VOS	OVS	OSV
SOV	—	22.9 [17, 29]	5.1 [2, 9]	0.9 [0, 3]	1.4 [0, 4]	5.6 [3, 9]
SVO	38.2 [31, 46]	—	3.7 [0, 9]	4.4 [1, 9]	1.1 [0, 3]	0.2 [0, 1]
VSO	2.4 [0, 6]	25.4 [14, 37]	—	8.1 [3, 15]	0.1 [0, 1]	1.2 [1, 2]
VOS	0.6 [0, 3]	40.9 [25, 57]	13.1 [3, 21]	—	0.0 [0, 0]	0.1 [0, 1]
OVS	4.4 [1, 10]	1.1 [0, 3]	0.6 [0, 3]	0.0 [0, 0]	—	0.3 [0, 2]
OSV	8.7 [0, 16]	2.2 [1, 5]	3.5 [0, 9]	0.1 [0, 1]	1.9 [0, 4]	—

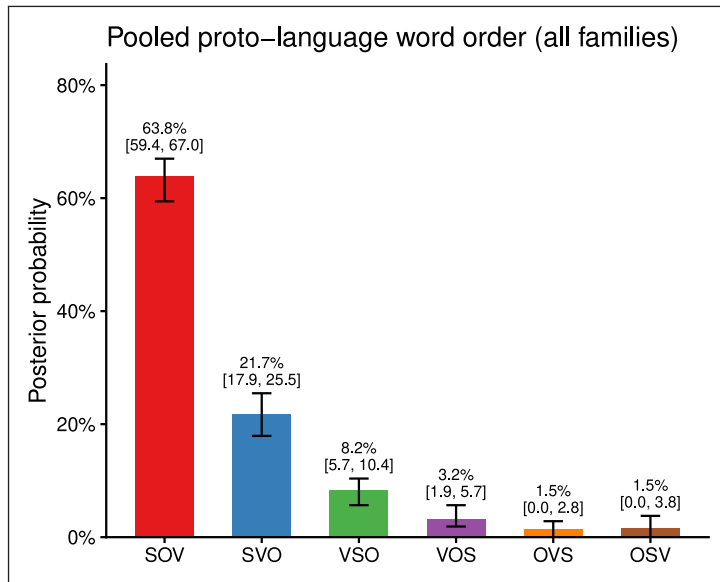


Figure 11: Inferred word-order type of proto-languages: pooled posterior root-state distribution averaged over the 107 families with phylogenies

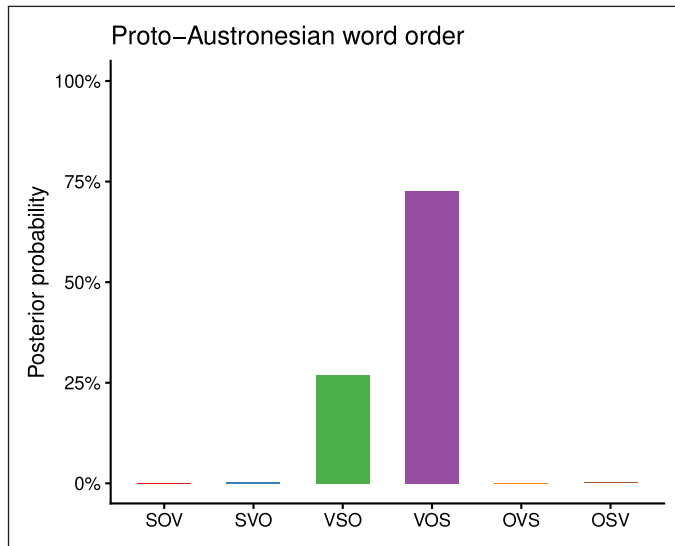


Figure 12: Marginal posterior root-state distribution for Austronesian: probability of each word-order type at the proto-Austronesian root, averaged over the joint posterior of the rate matrix and the family phylogeny

Austronesian subgroup – places the proto-language as verb-initial (Blust 2009). Despite the large number of Malayo-Polynesian SVO languages in the present sample (126 out of 209 Austronesian languages), the model correctly identifies the Formosan signal: the MrBayes posterior trees assign a deep, long branch to the Formosan clade, so the few VOS and VSO Formosan tips carry disproportionate phylogenetic weight at the root and counterbalance the numerically dominant SVO signal from Malayo-Polynesia. The Austronesian case thus illustrates a key advantage of phylogenetic over frequency-based reconstruction: by weighting tip evidence by evolutionary distance, the model down-weights the contribution of a large, shallow radiation (Malayo-Polynesia) relative to a small, deep one (Formosa), recovering the verb-initial root even though verb-medial languages dominate the sample.

4.4.2 Nilotic

The model reconstructs proto-Nilotic with highest posterior probability on VSO ($\approx 88\%$); all other orders, including SVO, receive only marginal support (Figure 13). This does not match the deep ancestor proposed by Givón (1976), who argued on the basis of internal evidence from Southern Nilotic that the family’s ultimate ancestor was SOV; it does, however, match the VSO stage that Givón posits as intermediate, on the way from SOV to the SVO pattern found in the modern Western branch. The model thus recovers this intermediate VSO stage rather than the deeper SOV root.

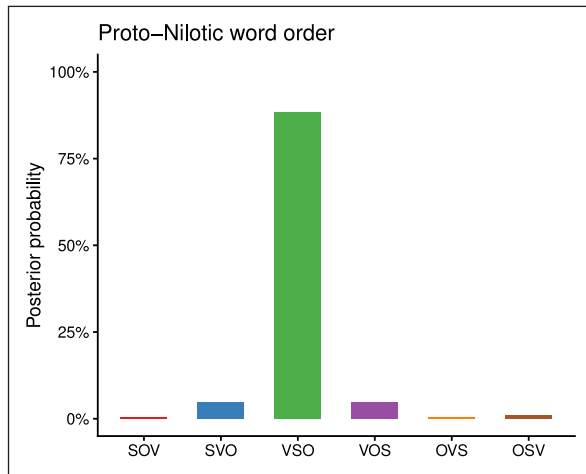


Figure 13: Marginal posterior root-state distribution for Nilotic

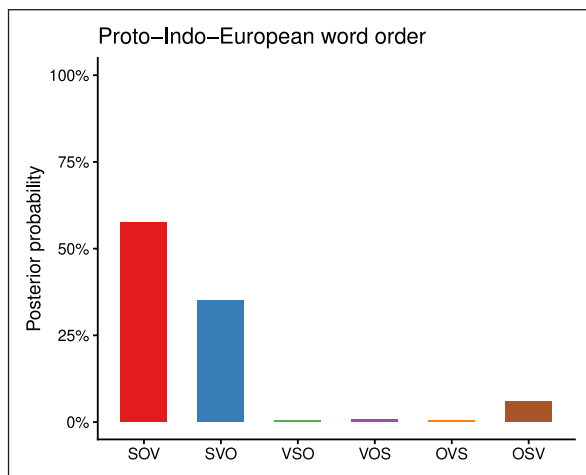


Figure 14: Marginal posterior root-state distribution for Indo-European

4.4.3 Indo-European

The model reconstructs proto-Indo-European with highest posterior probability on SOV ($\approx 58\%$), followed by SVO ($\approx 35\%$) and OSV ($\approx 6\%$); all other orders receive negligible probability (Figure 14). The dominant SOV signal is consistent with the classical reconstruction: Lehmann (1974: 30) argued on the basis of comparative syntax that “the fundamental order of sentences in PIE is OV”, a consensus that has remained influential in Indo-European studies. The large secondary SVO component reflects the typological weight of the modern European languages in the sample – the Romance languages, the Slavic languages, and most Germanic languages – which are predominantly SVO. The model is thus pulled between the deep SOV signal from

archaic branches (Sanskrit, Avestan, Old Persian, Hittite, Latin) and the SVO signal from the numerically dominant modern European branches, producing a bimodal reconstruction.

4.4.4 Uralic

The model reconstructs proto-Uralic with SVO marginally ahead of SOV (ca. 56% versus ca. 42%; Figure 15), the two orders together accounting for almost all of the posterior mass. This runs counter to the comparative-historical consensus, which places the ancestral order as SOV: the deepest-branching subgroup, Samoyedic (Nganasan, Selkup, Tundra Nenets), is uniformly SOV, as are the Ob-Ugric languages (Khanty and Mansi), pointing towards SOV as the ancestral state (Abondolo 1998). Most of the SVO languages in the present sample – among them Finnish, Estonian, Komi-Zyrian, and North Saami – belong to western branches (Finnic, Saamic, Permic) whose shift away from SOV is standardly attributed to prolonged contact with Germanic and Baltic Indo-European languages; Hungarian, the one Ugric language in the sample coded SVO, has likewise moved away from the ancestral SOV order. Unlike the Austronesian case, where a deep Formosan branch lets a few verb-initial tips outweigh a large verb-medial radiation, the Uralic SOV signal is carried by branches that are not deep enough to counterbalance the numerically dominant western SVO languages, so the root reconstruction is pulled towards SVO. The result should accordingly be read with caution: it arguably reflects the heavy sampling of contact-influenced western branches rather than a genuine SVO proto-Uralic, and it illustrates that phylogenetic weighting mitigates but does not eliminate tip-frequency bias when the deep branches are sparse. Maurits & Griffiths (2014), who also analysed Uralic, inferred a predominantly SOV proto-language; the present analysis does not reproduce that finding, a discrepancy plausibly traceable to differences in taxon sampling and tree inference.

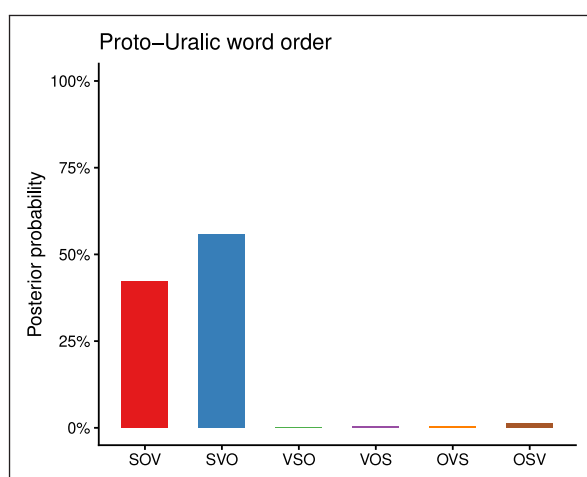


Figure 15: Marginal posterior root-state distribution for the Uralic family. SVO is marginally ahead of SOV (ca. 56% versus ca. 42%)

4.5 Proto-world reconstruction

The individual family reconstructions and the equilibrium analysis together motivate a natural question: can we say anything about the word-order state of the common ancestor of all human languages – the hypothetical proto-world? This question lies at the heart of Gell-Mann & Ruhlen (2011), but their argument is informal and takes neither probabilities nor time depth into account.

Following Maurits & Griffiths (2014), I tried a model-based answer placing proto-world as the root of a star tree whose daughters are all the lineages in the sample. For each phylogenetic family f , the branch from proto-world to the family root has length $T - R_f$, where T is the assumed age of proto-world and R_f is the posterior mean root age of family f inferred from the MrBayes analysis; for each singleton lineage g , the branch length is simply T . The Felsenstein likelihood at the proto-world root is then computed in the usual way: by marginalizing over the family root state using the transition probability matrix $P(T - R_f) = e^{Q(T - R_f)}$, and combining with the singleton likelihoods $P(T)_{r,s_g}$. The posterior over proto-world states is obtained by weighting with the stationary distribution π (the natural root prior for an ergodic CTMC) and normalizing. The minimal permitted value of T is $\lceil \max_f R_f \rceil$, i.e. the ceiling of the oldest family root in the sample – 13 kya, corresponding to Afro-Asiatic.

I chose a lower bound of 40 kya reflecting the timing of the human dispersal out of Africa: genomic evidence places this event at approximately 40–60 kya (Mallick et al. 2016), so any proto-world uniting all present-day language families cannot be younger than this. Figure 16 shows the posterior probability of each word-order type at the proto-world root as a function of T , swept from 40 to 200 kya and averaged over 400 thinned posterior draws of the rate matrix. At $T = 40$ kya the reconstruction is dominated by the family-root posteriors: since SVO is under-represented among the proto-languages in comparison to the stationary distribution, a young proto-world would need to have a more-than-stationary probability of SOV to explain this pattern. As T increases and the branches from proto-world to the family roots lengthen, the Markov process mixes and the proto-world posterior gradually converges towards the stationary distribution; at $T = 200$ kya, the SOV probability stabilizes at approximately 60–65%, broadly consistent with the inferred equilibrium of 65%. At all values of T from 40 to 200 kya, SOV carries the highest posterior probability, with SVO second from 45 kya on.

The result is therefore robust across the full plausible range of proto-world time depths. Whether proto-world is placed at 40 kya or deeper, the reconstruction consistently favours SOV as the most probable ancestral state. This broadly corroborates the first claim of Gell-Mann & Ruhlen (2011), though on an entirely different evidentiary basis: the result follows directly from the inferred rate structure rather than from a typological majority vote. Notably, the high posterior probability of SOV as the state of proto-world is mostly due to the fact that this is the most

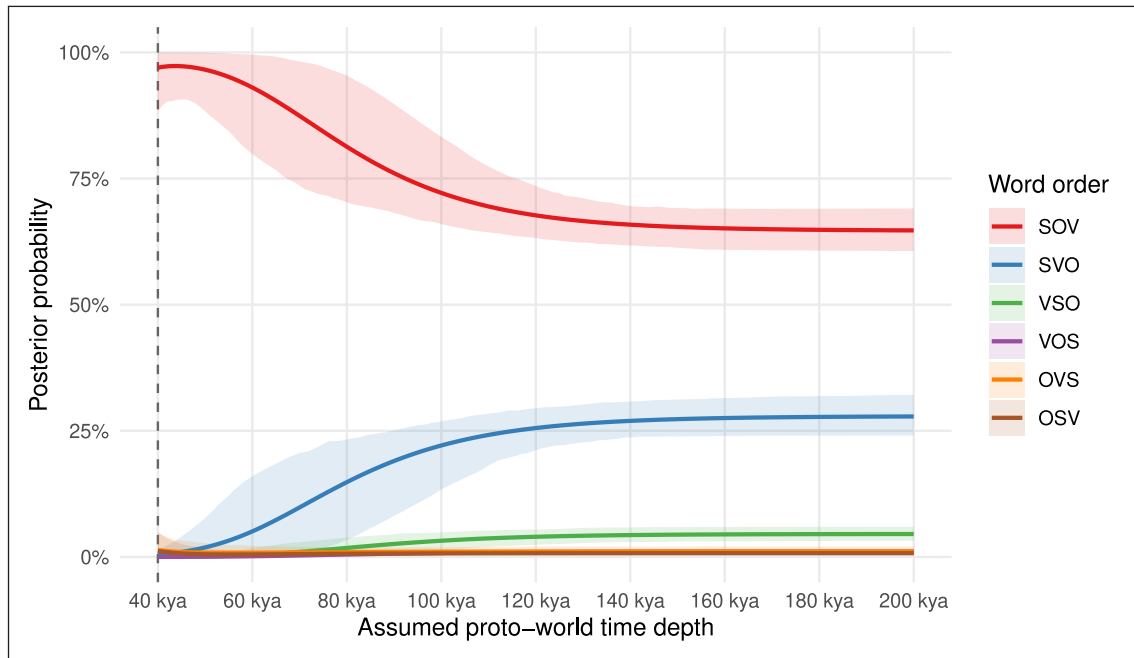


Figure 16: Posterior probability of each word-order type at the proto-world root as a function of the assumed proto-world age T (kiloyears before present). Lines show posterior means; shaded bands show the range between the 5th and 95th percentiles across 400 draws from the posterior distribution of the rate matrix, reflecting uncertainty in the inferred transition rates. The dashed vertical line marks 40 kya, the lower bound of the estimated timing of the human dispersal out of Africa

probable state in the stationary distribution; it does not follow from reconstruction in the narrow sense, which is not reliably possible at the time spans involved.

Another caveat is that the star-tree model treats all lineages as independent given proto-world, ignoring any macro-family structure above the family level. Allowing for such higher-level structure would reduce the effective number of independent data points and widen the credible intervals; the directional conclusion is unlikely to change, but the expressed confidence should be treated as an upper bound.

5 Discussion

The results converge on a consistent and interpretable picture. SOV is the long-run attractor of word-order evolution across the world's languages, with an inferred stationary probability of approximately 65%. The dominant evolutionary exchange involves the SOV/SVO/VSO triangle, and the direction of net flux within that triangle is consistently SOV-ward: SVO feeds into SOV

more often than SOV feeds into SVO, and VSO likewise feeds into SVO more often than the reverse. The rarer orders – VOS, OVS, OSV – are metastable states that arise sporadically and tend to revert, with VOS routing predominantly through SVO on its way back towards SOV.

5.1 No directional SOV → SVO trend

Gell-Mann & Ruhlen (2011) advance two separable claims: that proto-world was SOV, and that the dominant evolutionary pathway runs SOV → SVO. The second claim is the empirically stronger one, and the present analysis rejects it directly. The SIMMAP results show SVO → SOV transitions (38.2 expected) occurring substantially more often than SOV → SVO (22.9), so the net flux runs *towards* SOV, not away from it. The inferred equilibrium of 65% SOV is itself inconsistent with any sustained directional drift away from SOV: if SOV → SVO were the dominant pathway, the equilibrium would be far lower than 65%. The first claim – SOV as the proto-world state – is addressed directly by the star-tree reconstruction of Section 4.5: the model favours SOV at the proto-world root across the full range of plausible proto-world ages, in agreement with Gell-Mann & Ruhlen (2011), though the evidentiary basis is entirely different.

5.2 SVO over-representation as a possible demographic artefact

If the equilibrium is 65% SOV, why does the raw sample contain 39.0% SVO languages? A plausible explanation is the recent demographic expansion of two major SVO families: Atlantic-Congo (over 1,500 languages worldwide, 216 in the sample, predominantly SVO) and Austronesian (over 1,200 languages worldwide, 209 in the sample, mostly SVO or VSO), both of which diversified explosively within the last 5,000 years.

5.3 Failure of the permutation-circle model

The adjacency-based transition model of Ferrer-i-Cancho (2016) predicts that transition rates should be inversely related to distance on the permutation circle. The data clearly do not support this prediction. VOS → SVO (40.9) is the single largest entry in Table 5, yet VOS and SVO are not adjacent on the circle. SOV ↔ VSO is likewise very active despite SOV and VSO differing by two transpositions. By contrast, VOS and OVS are adjacent to each other but show effectively no mutual exchange (0.0 transitions in both directions). The permutation-circle model has formal elegance but no empirical support in the rate structure inferred here.

5.4 Agnosticism about the causes of SOV salience

The analysis identifies SOV as the long-run attractor of word-order evolution but is silent on *why* SOV is privileged over the other orders. Addressing this question requires a different kind

of evidence. Experimental and psycholinguistic research suggests that a preference for object-before-verb order may have roots in general cognition or perceptual structure: participants asked to communicate events nonverbally through gesture consistently adopt SOV order regardless of the word order of their native language (Goldin-Meadow et al. 2008), and similar SOV biases arise in newly emerging sign languages (Meir et al. 2010). These findings point to processing or conceptual-structure factors that operate independently of any particular linguistic convention. Whether such biases are strong enough to generate the full magnitude of the inferred equilibrium advantage – 65% for SOV versus 28% for SVO – and how they interact with demographic history and language contact, remain open questions that the present model cannot resolve.

5.5 Limitations

The most serious limitation is the assumption that a single rate matrix governs word-order change uniformly across all families. Second, the model assumes that all Glottolog lineages are stochastically independent. This is arguably a too strong assumption. It ignores phylogenetic structure above the level of Glottolog families. Since Glottolog pursues a highly conservative policy for accepting families, relevant phylogenetic structure might be missed. Third, the OSV state is represented by only seven languages coded as OSV alone and by six more coded with OSV alongside another order. The posterior OSV exit rate is consequently uncertain: the posterior mean total exit rate for OSV is approximately 1.0 per kyr (95% CI 0.3–1.8 per kyr), corresponding to a mean waiting time of 1.1 kyr (95% HPD 0.4–2.2 kyr; Table 3), with credible intervals spanning roughly one order of magnitude. The broad conclusion that OSV is among the less stable orders can be maintained, but precise quantitative claims are not warranted. Fourth, while the dominant SOV ↔ SVO pair produces sizeable transition counts (23 and 38 expected transitions for the two directions respectively), the rarer pairings involve only a handful of inferred events and the corresponding rates carry substantial uncertainty. The model estimates 30 free parameters, and many are constrained by only a small number of reconstructed transitions. Fifth, the treatment of word order as a discrete, time-invariant property abstracts away from synchronic variability and the gradient of diachronic change (Levshina et al. 2023). Sixth, the model abstracts away from language contact. Areal diffusion is a well-documented driver of word-order change – the shift from SOV to SVO in western Uralic languages under sustained Indo-European contact is a canonical example – yet the CTMC framework treats lineages as evolving independently. Contact-induced change is conflated with internally motivated change in the estimated rate matrix, and the degree to which the inferred rates reflect genuinely universal pressures versus pervasive areal influence cannot be assessed within the present model. These limitations are shared by all quantitative studies in this area and do not undermine the main conclusions, but they should be kept in mind when interpreting the precise numerical results.

6 Conclusion

Applying the phylogenetic comparative method to cross-linguistic word-order typology yields three principal findings. First, SOV is the genuine long-run attractor of word-order evolution: the posterior equilibrium assigns 65% of probability mass to SOV (95% CI: 59–70%), a result that controls for phylogenetic non-independence, tree uncertainty, and genealogical sampling artefacts. Second, there is no universal directional pressure from SOV towards SVO; rather, the net flux within the dominant SOV $\leftrightarrow$ SVO exchange runs in the opposite direction, with SVO $\rightarrow$ SOV transitions substantially more frequent than SOV $\rightarrow$ SVO. The current overrepresentation of SVO in typological databases is arguably a demographic artefact of the recent explosive diversification of a small number of SVO families, not a signature of universal grammar. Third, the permutation-circle model of Ferrer-i-Cancho (2016) is not supported by the inferred rate structure: transition frequency does not correlate with adjacency distance on the circle.

These results validate the PCM as a principled approach to typological universals research. Of the three confounds that Maslova (2000) identified as jointly underdetermining the synchronic distribution, the present method directly addresses the first two: universal tendencies in the direction and rate of change are estimated from the rate matrix, conditioning on phylogenetic structure and marginalizing over tree uncertainty. The third confound – the distribution of proto-language states – is partially addressed through marginalizing over ancestral states. The absence of reliable information about the phylogenetic structure above the family level remains a limitation.

The present analysis largely follows Maurits & Griffiths' (2014) pioneering work but extends it both in empirical scope and in the arguably more principled way of obtaining phylogenetic family trees.

The most important direction for future work is the joint modelling of vertical transfer via phylogenetic inheritance and horizontal transfer via language contact within a single statistical model, enabling these two factors to be separated.

Acknowledgements: This research was supported by the Volkswagen Foundation project *Phylomilia – Phylogenetic linguistic inference from acoustic speech data*, by the DFG Research Unit FOR 2237 *Words, Bones, Genes, Tools: Tracking Linguistic, Cultural and Biological Trajectories of the Human Past*, and by the Akademienprogramm project *GloBasLex – Global Basic Lexicon* (Heidelberg Academy of Sciences and Humanities). Claude Code (Anthropic) was used to assist with programming the analysis pipeline and with copy-editing the manuscript; the author takes full responsibility for the content.

Data and code availability: All analysis code, the processed word-order dataset, and the main CTMC posterior sample are available at <https://codeberg.org/profgerhard/wordorder-pcm>. The lexical character matrices, posterior tree samples, and sensitivity posteriors are archived at <https://doi.org/10.57754/FDAT.b0abr-rx961>.

References

- Abondolo, Daniel. 1998. Introduction. In Daniel Abondolo (ed.), *The Uralic languages*, 1–42. London: Routledge.
- Blust, Robert. 2009. *The Austronesian languages*. Vol. 602 (Pacific Linguistics). Canberra: Pacific Linguistics.
- Bollback, Jonathan P. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7(1). 88. <https://doi.org/10.1186/1471-2105-7-88>.
- Bouckaert, Remco. 2025. *Global language tree pipeline (version 2.0.0)*. <https://github.com/rbouckaert/global-language-tree-pipeline/releases/tag/v2.0.0>.
- Bouckaert, Remco, David Redding, Artur Trebski, Oliver Sheehan, Thanos Kyritsis, Russell D. Gray, Kate E. Jones & Quentin D. Atkinson. 2026. Global language diversification is linked to socio-ecology and threat status. SocArXiv. https://doi.org/10.31235/osf.io/f8tr6_v2.
- Bouckaert, Remco R. & Joseph Heled. 2014. DensiTree 2: Seeing trees through the forest. *BioRxiv*. <https://doi.org/10.1101/012401>.
- Cooper, Natalie & Robert P. Freckleton. 2026. *Phylogenetic comparative methods: a practical introduction*. <https://doi.org/10.5281/zenodo.19222262>.
- Dellert, Johannes & Armin Buch. 2018. A new approach to concept basicness and stability as a window to the robustness of concept list rankings. *Language Dynamics and Change* 8(2). 157–181. <https://doi.org/10.1163/22105832-00802001>.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1). 81–138. <https://doi.org/10.2307/416370>.
- Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473(7345). 79–82. <https://doi.org/10.1038/nature09923>.
- Felsenstein, Joseph. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17(6). 368–376. <https://doi.org/10.1007/BF01734359>.
- Felsenstein, Joseph. 2004. *Inferring phylogenies*. Sunderland, MA: Sinauer Associates.
- Ferrer-i-Cancho, Ramon. 2016. Kauffman’s adjacent possible in word order evolution. In S.G. Roberts et al. (eds.), *The evolution of language: Proceedings of the 11th international conference (EVOLANG 11)*. <https://arxiv.org/abs/1512.05582>.
- Ge, Hong, Kai Xu & Zoubin Ghahramani. 2018. Turing: a language for flexible probabilistic inference. In *International conference on artificial intelligence and statistics*, 1682–1690.
- Gell-Mann, Murray & Merritt Ruhlen. 2011. The origin and evolution of word order. *Proceedings of the National Academy of Sciences* 108(42). 17290–17295. <https://doi.org/10.1073/pnas.1113716108>.
- Gelman, Andrew & Donald B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4). 457–472. <https://doi.org/10.1214/ss/1177011136>.

- Givón, Talmy. 1976. On the SOV reconstruction of Southern Nilotic: internal evidence from Toposa. In Larry M. Hyman, Leon C. Jacobson & Russell G. Schuh (eds.), *Papers in African linguistics in honor of Wm. E. Welmers*, 73–94. Studies in African Linguistics, Supplement 6. Los Angeles: African Studies Center.
- Goldin-Meadow, Susan, Wing Chee So, Asli Özyürek & Carolyn Mylander. 2008. The natural order of events: how speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences* 105(27). 9163–9168. <https://doi.org/10.1073/pnas.0710060105>.
- Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of language*, 73–113. Cambridge, MA: MIT Press.
- Haag, Julia, Dimitri Höhler, Ben Bettisworth & Alexandros Stamatakis. 2022. From easy to hopeless—predicting the difficulty of phylogenetic analyses. *Molecular Biology and Evolution* 39(12). msac254. <https://doi.org/10.1093/molbev/msac254>.
- Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2026. *Glottolog 5.3*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://doi.org/https://doi.org/10.5281/zenodo.18840935>.
- Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie. 2008. *The World Atlas of Language Structures Online*. Max Planck Digital Library, Munich. <http://wals.info/>.
- Hoffman, Matthew D. & Andrew Gelman. 2014. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 15(47). 1593–1623.
- Jäger, Gerhard. 2025. Beyond cognacy. In *Proceedings of the 7th workshop on research in computational linguistic typology and multilingual NLP (SIGTYP 2025)*, 52–60. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.sigtyp-1.6>. Updated version (2026) correcting a sign error in pHMM training and two bugs in MSA construction: <https://arxiv.org/abs/2507.03005v2>.
- Jäger, Gerhard & Johannes Wahle. 2021. Phylogenetic typology. *Frontiers in Psychology* 12. <https://doi.org/10.3389/fpsyg.2021.682132>.
- Lehmann, Winfred P. 1974. *Proto-Indo-European syntax*. Austin: University of Texas Press.
- Levshina, Natalia, Savithry Namboodiripad, Marc Allasonnière-Tang, Mathew Kramer, Luigi Talamo, Annemarie Verkerk, Sasha Wilmoth, Gabriela Garrido Rodriguez, Timothy Michael Gupton, Evan Kidd, Zoey Liu, Chiara Naccarato, Rachel Nordlinger, Anastasia Panova & Natalia Stoyanova. 2023. Why we need a gradient approach to word order. *Linguistics* 61(4). 825–883. <https://doi.org/10.1515/ling-2021-0098>.
- Lewis, Paul O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50(6). 913–925. <https://doi.org/10.1080/106351501753462876>.
- List, Johann-Mattis, Robert Forkel, Simon J. Greenhill, Christoph Rzymiski, Johannes Englisch & Russell D. Gray. 2022. Lexibank analysed. *Scientific Data* 9. 316. <https://doi.org/https://doi.org/10.5281/zenodo.7836668>.
- Mallick, Swapan et al. 2016. The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature* 538. 201–206. <https://doi.org/10.1038/nature18964>.
- Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4(3). 307–333. <https://doi.org/10.1515/lity.2000.4.3.307>.

- Maurits, Luke & Thomas L. Griffiths. 2014. Tracing the roots of syntax with Bayesian phylogenetics. *Proceedings of the National Academy of Sciences* 111(37). 13576–13581. <https://doi.org/10.1073/pnas.1319042111>.
- Meir, Irit, Wendy Sandler, Carol Padden & Mark Aronoff. 2010. Emerging sign languages. In Marc Marschark & Patricia Elizabeth Spencer (eds.), *Oxford handbook of deaf studies, language, and education*, vol. 2, 267–280. Oxford University Press.
- Pagel, Mark & Andrew Meade. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *The American Naturalist* 167(6). 808–825. <https://doi.org/10.1086/503444>.
- Pagel, Mark, Quentin D. Atkinson, Andreea S. Calude & Andrew Meade. 2013. Ultraconserved words point to deep language ancestry across Eurasia. *Proceedings of the National Academy of Sciences* 110(21). 8471–8476. <https://doi.org/10.1073/pnas.1218726110>.
- Ronquist, Fredrik & John P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12). 1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>.
- Wichmann, Søren, Eric W. Holman & Cecil H. Brown. 2018. *The ASJP Database (version 18)*. <https://doi.org/10.5281/zenodo.1193217>.