

Areal pressure in grammatical evolution

An Indo-European case study

Chundra Cathcart,^{1, 2} Gerd Carling,¹ Filip Larsson,¹
Niklas Johansson¹ and Erich Round^{3, 4}

¹ Lund University / ² University of Zurich / ³ University of Queensland /

⁴ University of Surrey

This article investigates the evolutionary and spatial dynamics of typological characters in 117 Indo-European languages. We partition types of change (i.e., gain or loss) for each variant according to whether they bring about a simplification in morphosyntactic patterns that must be learned, whether they are neutral (i.e., neither simplifying nor introducing complexity) or whether they introduce a more complex pattern. We find that changes which introduce complexity show significantly less areal signal (according to a metric we devise) than changes which simplify and neutral changes, but we find no significant differences between the latter two groups. This result is compatible with a scenario where certain types of parallel change are more likely to be mediated by advergence and contact between proximate speech communities, while other developments are due purely to drift and are largely independent of intercultural contact.

Keywords: historical linguistics, evolutionary linguistics, areal linguistics, typological change, Indo-European, language contact, complexity

1. Introduction

The differential stability of typological features (alternatively characters or traits, in terminology adopted from evolutionary biology) has long been a topic of interest in linguistics. Linguists hypothesize that certain parts of a language's grammatical system will show little change over time, while other features will change frequently (Sapir 1921; Meillet 1925; Martinet 1975). With a few exceptions, quantitative approaches to the study of typological stability either focus on synchronic distributions, with no explicit way of accounting for shared behavior due to genetic inheritance, or are purely evolutionary in their scope, with geography and system-internal pressures playing no discernible role.

It is well known that languages exhibit similarity due to a number of competing forces, including shared inheritance, independent parallelism and contact (Nichols 2003; Aikhenvald 2007; Bickel 2011). The goals of this paper are twofold: we employ explicit phylogenetic and spatial models of linguistic evolution, and we use these models to investigate the evolutionary dynamics of typological features; in the case of features that are frequently in flux, we attempt to measure the possibility as to whether parallel changes come about due to areal pressure as opposed to chance or universal tendencies. Essentially, we seek to tease apart drift (cf. Sapir 1921; Yanovich 2016) and advergence (cf. Renfrew 2000; Chang et al. 2015), a phenomenon where diversifying languages, having split from a larger parent language, undergo parallel changes, conceivably due not only to interactions between precursors of change in the parent language, but areal contact as well. Our goal is to probabilistically infer information regarding how typological features change during the course of evolution of a language family with a well-documented history and largely agreed-upon prehistory (the Indo-European languages) and whether, in the case of parallel changes, these changes co-occur in spatiotemporal proximity to each other (or to locations where the feature is present or absent). Our analysis incorporates (1) a phylogenetic model of character evolution, (2) a phylogeographic model of demographic dispersal and (3) a metric for measuring spatial dispersion. This allows us to estimate how areally sensitive parallel changes to typological features are. We employ a large dataset of feature variants from several domains of morphosyntax, and we classify changes pertaining to these variants according to whether (by our account) they simplify a grammatical pattern (e.g., via the loss of a marked morphological structure, or by making morphology more transparent), render a pattern more complex (e.g., by introducing a morphosyntactic mismatch in marking, or increasing the number of elements in a clause that must be indexed by the verb) or are neutral in their complexity (e.g., by changing constituent order). We predict that changes which simplify are more areally concentrated than 'neutral' changes, our control group, and that neutral changes are more areally concentrated than changes which introduce complexity. While we do not find changes which simplify to be significantly more areally concentrated than neutral changes, we do find that changes which introduce complexity are significantly more areally dispersed than changes from the other two groups.

Our methodologies make a number of simplifying assumptions, and the procedure carried out in the paper serves largely as a proof of concept; however, it is hoped that the methods outlined in this paper will expand the possibilities of investigating interactions between genetic inheritance and areal contact in quantitative historical linguistics. A third, ancillary goal of this work is to help demystify some of the probabilistic models that have recently gained currency in our field. We hope, by making our methodology clear – and by making our code freely available – to

add to the growing community of historical linguists developing tools for collaborative use, and to further the discussion on how to adapt tools from outside the field for linguistic use.

2. Previous literature

2.1 Quantitative typology

The quantitative literature on typological stability can be roughly divided into two genres. Earlier scholarship tended to make use of basic descriptive statistics in order to characterize the stability of, preference for, or dependence between features, based largely on their synchronic distributions (cf. Nichols 1986; Dryer 1989).

More recently, researchers utilizing tools and insights from computational biology have investigated the evolution of typological characters using an explicitly phylogenetic model. These and other graphical modeling techniques (see Jordan 2004 for a general overview) allow for a greater degree of complexity in hypothesis testing; they allow us to infer what sort of underlying process is most likely to have generated the distribution of data that is observed. For instance, Dediu (2010) uses a series of Bayesian phylogenetic models to infer rates of gain and loss among typological features (from WALS, the World Atlas of Linguistic Structures; Haspelmath et al. 2005), from which indices of diachronic typological stability can be computed. Assuming fixed tree topologies for a number of linguistic stocks, the author infers gain and loss rates for different typological features and converts them into stability ranks.

However, reliance on a purely phylogenetic model fails to take into account areality, the understanding of which is inherent in a great deal of earlier work. Results of purely phylogenetic studies of typological stability have been criticized due to the fact that the methodology used has no explicit means of accounting for areal influences or system-internal typological pressures that may be at work (Wichmann 2015: 218–219); after all, genetic relationships are far from the only factors underlying typological change. Typological change often arises from areal pressures; genetically unrelated languages can converge on similar structures by virtue of inhabiting the same multilingual geographical area (cf. Gumperz & Wilson 1971; François 2011). Language contact may also create exceptions to widespread typological patterns such as word order harmonies (where OV and VO constituent order tend to co-occur with postpositions and prepositions, respectively, in languages; cf. Graham 2016).

At the same time, it should be possible to model linguistic change along both vertical (i.e., genetic transmission) and lateral (i.e., areal diffusion, contact)

dimensions, as in Meid's (1975) space-time schema. One such quantitative approach is found in Daumé (2009), which introduces a hierarchical Bayesian model designed to tease apart genetic inheritance and areal spread in typological features (from WALS), thus improving genetic trees constructed from such features. We wish to take into account areal factors that may influence linguistic evolution, and employ both phylogenetic and phylogeographic methods in order to quantify these pressures.

2.2 Phylogenetic linguistics and graphical models

A great deal of pioneering work on quantitative typology has been done using frequentist statistical tests, which are easy to carry out and intuitive to understand. Tests of this sort, along with classical linear regression, are beset by disadvantages when used to model data that violates the independence assumption. This has generally been dealt with via principled inclusion or exclusion of language samples which may bias a particular outcome (cf. Dryer 1989), though techniques of this sort may require the available data to be underused. Some exceptions in linguistics and related disciplines (e.g., Botero et al. 2014) use mixed-effects modeling in order to account for genetic skews in linguistic data.

Graphical models, which combine probability and graph theory, allow scholars to investigate relatively complex hypotheses regarding how real-world data are generated. In particular, fine-grained structures of conditional dependence between random variables can be modeled, the likes of which cannot be straightforwardly represented in more basic statistical analyses. (See Jordan 2004 for a general overview.) Models of this sort generally involve a specification where unobserved parameters are conditioned on observed data. Model fitting is usually carried out according to some type of iterative Bayesian procedure, usually a Markov chain Monte Carlo (MCMC) method, which draws samples from the posterior distributions of unobserved quantities. Alternatives to MCMC exist, such as Variational Bayesian methods, which approximate the true posterior distribution with a simpler distribution.

Some contemporary phylogenetic models used in computational biology are a type of graphical model that may be familiar to some linguists, given their extension to historical linguistics. The observations consist of trait data; reasonable prior assumptions are made about the tree parameters, and the inference procedure returns the posterior parameters (i.e., branch lengths, branching structure) which generate the data under the model specification. Historical linguists have found it interesting and fruitful to adopt such tools, which employ numerical information (either phenotypic or genetic) in the service of exploring relationships between species. We refer the interested reader to general surveys of computational cladistic

methods (e.g., Felsenstein 2004; Nichols & Warnow 2008; Yang 2014; Dunn 2015). We restrict ourselves to likelihood methods (as opposed to parsimony methods, the rough objective of which is to minimize the number of character state changes), which specify a set of parameters regarding change over the phylogeny. Maximum likelihood methods seek the values that maximize the likelihood of the observed trait data conditional on these values, while Bayesian methods place priors over the parameters of interest and return each parameter's posterior distribution (generally via a MCMC procedure, as mentioned in the previous paragraph). Likelihood approaches are preferred for a number of reasons, all of which are enumerated in the literature cited above; a particular strength is their ability to represent a wide array of character evolution scenarios.

The above methods are often used to infer a tree topology from trait data. The same models can be used to gather information about character change over a known or assumed tree topology or tree sample inferred from lexical data, which is one of the objectives in our paper.

3. Materials and methods

The methods described in the previous section provide a framework for inferring rates of typological change. It is hoped that, by taking into account both genetic and areal pressures in development, we can dispel some criticism directed at a purely phylogenetic approach to this question (which might extend to any and all phylogenetic approaches), and uncover as yet unseen patterns of linguistic change.

We seek also to include two additional design features, not fully exploited in previous analyses. The first is the use of ancestry constraints between known parent and daughter languages. This is not the norm in evolutionary linguistic work, and we believe that it is critical: if a documented language of the past is treated not as a direct ancestor of its descendant languages but as a sibling of a hypothetical ancestor, it can distort the history of a linguistic feature in that group.

Additionally, we wish to project notions of contact and areality further into the past.¹ We wish to infer, when a change occurs on a branch of a tree, the likelihood that this event is a lone occurrence as opposed to one that takes place in

1. We note that some authors (e.g., Trudgill 2001) have argued that adult-induced contact is a relatively recent phenomenon, the majority of which has taken place due to societal changes that have taken place in the last two millennia. While the chronology of the Indo-European languages exceeds this period, we note that we employ an aggregate measure of the areality of a linguistic change, and it is likely that the majority of the signal will be due to changes within a comparatively recent window of time.

areal proximity to a similar event. In order to achieve this, we need an idea of the most likely geographic positions along each branch of the tree at each point in time, which we can infer using phylogeographic and spatial modeling techniques.

Our first goal is to infer rates of character birth and death, which jointly figure into a metric of viability, as outlined by Nichols (2003). Phylogenetic studies that have addressed questions regarding character evolution and the phylogenetic signal displayed by typological characters generally proceed by generating a tree topology based on these characters, then making inferences about which ones are more stable or display greater phylogenetic signal (Dunn et al. 2008). This approach stands in contrast to those that model evolution of characters on a tree whose topology has been inferred from other features, usually lexical (Dunn et al. 2011; Haynie et al. 2014). When external evidence about a language family tree's branching structure is available, we believe it prudent to make use of this information in order to provide a realistic history of character evolution within that family. Since we do not always have reliable information about the branching structure of language families, we limit our study to the Indo-European languages, whose history is relatively well known.

We employ a posterior tree sample inferred from lexical data in order to analyze the evolutionary patterns displayed by the typological feature variants in our dataset. Crucially, we make use of ancestry constraints in order to model direct descent between attested ancient or medieval and contemporary languages, where such relationships are known. With a few exceptions (e.g., Chang et al. 2015), ancestry constraints have been neglected in contemporary evolutionary linguistic literature, though they were used in early glottochronological literature in order to calibrate rates of lexical replacement (Swadesh 1952, 1955). In lexicostatistical work on Indo-European, it could be argued that ancestry constraints between attested languages (e.g., between Latin and modern Romance languages) are invalid, since they do not account for dialect or register variation between the ancient language and the possible unattested ancestor(s) of the contemporary languages. Setting that debate aside, however, it seems unrealistic to posit typological variation between ancient/medieval languages and their putative unattested sisters (though the process of establishing invariant typological characteristics such as word order for dead languages is by no means trivial (see Viti 2010; Haug 2015)). The tendency of linguistic phylogenetic research to neglect ancestry relationships is no doubt an artifact of biology, where direct ancestry between observed specimens is not frequently noted. However, in linguistics, ancestry relationships provide crucial evidence regarding directionality of change, and we seek to make use of this received wisdom rather than allow conventions from biology to supersede it.

Readers may note a glaring omission: at the time of writing, we are working solely with data from Indo-European languages. A complete account of language

contact within the history of a group of languages should treat stock-external contact. We currently lack reliable reference phylogenies for non-IE languages based on lexical data and other innovations, though we are working to expand our model to these languages. While the fact that we do not incorporate this information into our model is less than ideal, it is also fairly uncontroversial that, at least to some degree, Indo-European languages have been diversified via contact with undocumented speech varieties whose affiliation is unknown (Lass 1997) or which have left their imprint only on toponyms and other miscellaneous items (cf. Vennemann 1994). When a relatively unusual pattern emerges within a language family (e.g., VSO constituent order in Celtic, retroflexion in Indo-Aryan, ejectives in Armenian, etc., to note a few Indo-European examples), contact with another unrelated language tends to be invoked.

We hope to expand our model in the future to include contact between unrelated languages, but we are content for the time being to gain insight into the role of contact between related languages in linguistic diversification and disparification (cf. Garrett 2006), which we find highly interesting in its own right. In a number of linguistic phylogenetic studies, parallel gains on branches in the tree are treated as tree-internal contact events by default (cf. Nakhleh et al. 2005; List et al. 2014); both papers use a parsimony approach which seeks to minimize the number of gains on the tree. When investigating change over a phylogeny of languages, it seems prudent to distinguish between parallel phylogenetic gains that are due to contact and those that take place independently. It is hoped that the approach described above – i.e., making use of phylogeography – will help identify historical contact between lineages on the tree.²

4. Data

Our data consist of two major ingredients: (1) binary typological data; and (2) a tree sample for Indo-European languages. We also make use of geographic and chronological metadata pertaining to the languages in our sample.

2. We acknowledge that there are additional pressures besides geographic proximity involved in contact events, though we cannot devote much attention to them in this paper. Population size may play a role; Trudgill's (1974) gravity model, for instance, considers the influence of larger speech varieties on smaller ones (though Braunnüller 1984 provides evidence that smaller languages are resistant to the sorts of morphological simplification undergone by languages with more speakers). Since contemporary phylogenetic methods can estimate population sizes at ancestral states, future research might do well to take population size into account as a contact-related parameter.

4.1 DiACL

The typological data is taken from the Diachronic Atlas of Comparative Linguistics (DiACL) Typology Eurasia database,³ which contains 118 typological features from 149 Eurasian languages of multiple linguistic stocks (Carling 2017). The database is organized into a three-level hierarchical structure. The top level consists of GRIDS or domains of the grammar, of which there are five: alignment, nominal morphology, verbal morphology, tense and word order. Each grid is stratified according to FEATURES, which are further divided into feature VARIANTS (e.g., WORD ORDER: ORDER OF CLITIC PRONOUNS WITH RESPECT TO FINITE VERB: OBJECT VERB). Feature variants take Boolean values: NO/FALSE/ABSENT, YES/TRUE/PRESENT, or no data. 3.6% of cells are missing data. DiACL is somewhat unique in that it contains data from ancient and medieval as well as modern languages, which are crucial benchmarks against which to check the effectiveness of the models which we employ.

At the feature level, a degree of granularity is targeted which goes beyond that currently available in databases such as WALS. For example, word order is coded across clause types. Polymorphic behavior can be coded across feature variants; that is to say, two word orders that tend to be in complementary distribution can theoretically both be present in a language (though in some cases, marginally attested features are suppressed). However, in many cases, a majority rules-style coding procedure has been implemented in order to characterize the typical behavior of a language with respect to a feature variant. We restrict our analysis to the 117 Indo-European languages in the sample, and we exclude feature variants which are present in either all or none of the Indo-European languages in the database, leaving 108 variants.

There exist some interdependencies between feature variants in the database. We do not employ an explicit means of addressing this issue. Below, we evaluate the performance of our model with respect to interdependent features.

4.1.1 *Areal dynamics across the dataset*

The objective of this paper is to investigate whether changes involving different linguistic feature variants display different degrees of areal sensitivity. As such, we divide the feature variants in the dataset into three partitions. We draw inspiration from the large literature on linguistic and typological complexity (Braunmüller 1984; Miestamo et al. 2008; Lupyán & Dale 2010; Baerman et al. 2015), though we do not necessarily adhere to the explicit and rigid taxonomies developed in these papers. There are myriad sophisticated ways in which to quantify linguistic complexity; ours is a somewhat simplistic one which we attempt to ground in

3. <https://diac1.ht.lu.se/>

the notion of outsider complexity (Kusters 2008). In our view, a morphosyntactic change does not necessarily lead to COMPLEXITY by simply introducing a new morphological category (e.g., definiteness, future tense, adpositions) but by giving rise to a grammatical relation between multiple constituents in the clause that learners must contend with (e.g., whether adjectives must agree with the nouns they modify in terms of a morphological category, whether overt case must be marked on core arguments of the verb, etc.); in this sense, our understanding of complexity with respect to the dataset corresponds to the notion of syntagmatic complexity (Nichols 2009: 111–112). We take changes that reverse such a state of affairs to bring about SIMPLIFICATION. In the middle, we envision a NEUTRAL category comprising features that cannot be easily subsumed under the labels given above, largely consisting of changes to word order; features pertaining to changes in derivational and inflection morphology independent of grammatical relations (e.g., whether case and number marking is done with agglutination or fusional morphemes) is treated as neutral as well, though agglutination arguably introduces a level of transparency (and thus simplicity) that fusional morphology lacks.

The group which we predict to show the greatest areal signal includes feature variants which involve some manner of structural simplification, in that learners are exempted from acquiring a surface-level pattern signifying a relationship between multiple pieces of clausal information, or some grammatical relation. This may include gaining the feature ALIGNMENT: NOUN: PRESENT PROGRESSIVE: A=O (“in the present progressive, is the noun form for the agent the same as the noun form for the object?”) or losing the feature NOMINAL MORPHOLOGY: NOMINAL CASES: GEN/DAT (“is there a special noun form to express genitive which is not the same as dative case?”), which would point to morphological syncretism. We predict intermediate areal sensitivity among NEUTRAL features pertaining to phenomena such as word order; in general, such patterns are borrowable but may not necessarily represent the effects of second-language acquisition or other pressures pertaining to life stages that come into play in contact scenarios. We predict the lowest areal sensitivity for feature variants which introduce a degree of complexity, mirroring the variants in the first group, e.g., losing the feature ALIGNMENT: NOUN: SIMPLE PAST: A=SA (“in the simple past, is the noun form for the agent subject of a transitive clause the same as the noun form for the agent subject of an intransitive clause?”), which leads to an ergative alignment pattern. It is reasonable to hypothesize that features of this sort are the result of language-specific structuring of information driven by within-group social and cognitive needs, and could be less likely to display an areal pattern. We believe that our hypothesis is valid given the large body of evidence that many features falling under our rubric of complexity, such as redundant agreement,

pose problems for speakers and L2 learners but facilitate understanding for hearers and L1 learners (Kusters 2008: 14).⁴

4.2 Tree sample

We model character evolution on a tree sample which we inferred from 100-word Swadesh list lexical data from DiACL Eurasia, using Chang et al.'s (2015) modified version of BEAST version 1.8 (Drummond et al. 2012),⁵ which allowed us to impose ancestry constraints between languages which we take to be in a relationship of direct descent (e.g., Old, Middle and Modern Persian; Classical, Middle and Modern Greek, etc.). We used a restriction site character model of trait evolution and assumed gamma-distributed among-trait rate variation with 4 rate classes. We assumed uncorrelated lognormal-distributed among-branch rate variation. For the root age of the tree, we assumed a normal prior with a mean of 6500 BP and a standard deviation of 2000. We make use of clade constraints to ensure that the branching of the tree agrees with the general consensus regarding Indo-European subgrouping, based on the comparative method. Markov chain Monte Carlo inference was run for 20 million iterations. The first half of samples were discarded as burn-in; the remaining samples were thinned to 20 trees. Of these, root ages for the median 50% of trees range from 6850 to 7396 years BP. The mean posterior root age is 7204 years BP, and the median posterior root age is 7139 years BP. Of the trees in the sample, the root ages of 3 trees are in the range 5500–6500 BP, consistent with the Steppe Hypothesis of the origin of the Indo-European languages, while 4 are within the range 8000–9500, consistent with the Anatolian hypothesis; root ages for the remaining 13 trees are on the interval 6500–8000. We used BEAST to carry out a continuous phylogeographic analysis in parallel with the tree sample inference, which samples geographic locations for the root and interior nodes of the tree. We placed a weak multivariate normal prior of (0, 0) with a precision of .001 over the longitude and latitude of the root. For two trees, root locations were inferred that were not on the Eurasian landmass; we excluded these trees, leaving 18. Of the remaining trees, longitude values inferred for the root range from 19.4 to 57.6, with a mean of 38.4, median of 37.8 and standard deviation of 5, while latitude values inferred for the root range from 28 to 45.6, with a mean of 42.3, median of 43.2 and

4. While we treat features pertaining to word order as neutral, we acknowledge that certain word orders and combinations of word orders are argued to be more 'difficult' from generative as well as psycholinguistic standpoints. However, we lack a principled means of dividing neutral features among the two other groups.

5. Available at <https://github.com/whdc/ieo-beast/>

a standard deviation of 9 (Figure 1 displays observed and inferred node locations on a map). Together, these phylogenetic and phylogeographic analyses make it possible to infer which change events are consistent with a language contact scenario.

4.2.1 *Ancestry constraints*

Comparative-historical linguists have developed a number of criteria for establishing whether any two languages spoken during different periods are in a relationship of direct descent or rather descended from a common ancestor (Hoenigswald 1960, 1966). Previous work which has modeled the evolution of typological features over language phylogenies has tended to treat these ancestral languages not as the parents of their descendant languages, but rather as the siblings of an unobserved node ancestral to what are traditionally considered to be its descendants (Dediu 2010; Dunn et al. 2011). Figure 4 gives a graphical representation of this difference. We model ancestry relationships in our tree between known parent and descendant languages (e.g., Latin vs. Romance, Classical vs. Middle vs. Modern Greek). Whether or not ancestry constraints should be imposed remains somewhat controversial. Opponents may argue that Classical Greek cannot be taken as directly ancestral to Middle and Modern Greek, given the dialectal diversity across ancient dialects, or that a language like Classical Latin as found in documents of antiquity represents a literary register different from the common speech, and that it is improper to assume this ‘doculect’ to be directly ancestral to French, Italian, etc. Proponents (such as ourselves) argue that models without ancestry constraints posit an unrealistically high degree of divergence between hypothetical speech varieties that must have been mutually intelligible; allowing an ancestral language to be an outgroup has the potential to underestimate recurrent changes that have taken place in the prehistory of the ancestor but have been lost in its descendants. Furthermore, a great deal of dialectal variation consists of low-level phonological and morphological variation, whereas a consistent typological and lexical profile can generally be assumed for ancient doculects (cf. Fortson 2015).⁶ This leaves some issues unresolved, but the proper modeling of speech register in the context of phylogenetic linguistics remains a topic outside of the scope of this paper. The ancestry constraints we use are listed in the Appendix (<http://dx.doi.org/10.1075/dia.16035.cat.additional>).

6. At the same time, we acknowledge the fact that there are a handful of exceptions to this generalization, such as differentiation with respect to V2 word order among Middle English dialects (Kroch et al. 2000), though we remain uncertain as to whether such cases are the exception or the norm.

5. Model

Our procedure is bipartite, roughly speaking: first, we carry out a phylogenetic analysis to infer rates of character evolution (as well as likely character birth and death locations on the tree topology); then, we carry out a spatial model of phylogeographic diffusion to recover likely geographic locations of evolutionary events on the tree. Below, we give a conceptual overview of each step of this procedure, followed by a full, detailed specification.

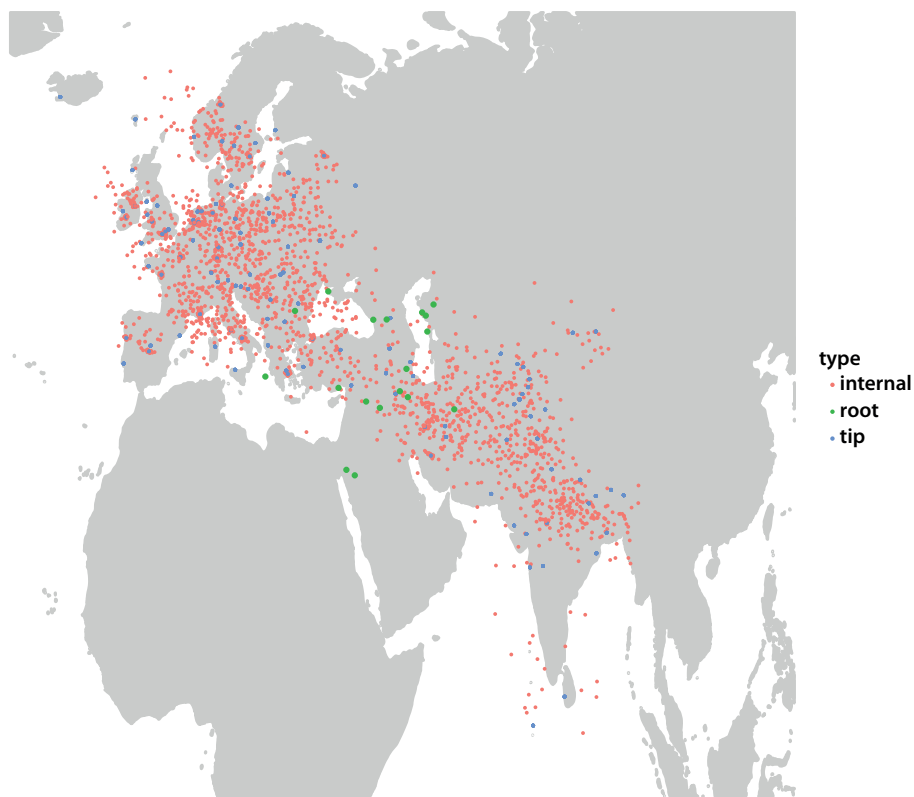


Figure 1. Locations for attested languages (blue) as well as locations inferred for internal nodes (red) and the root (green), for all 18 trees

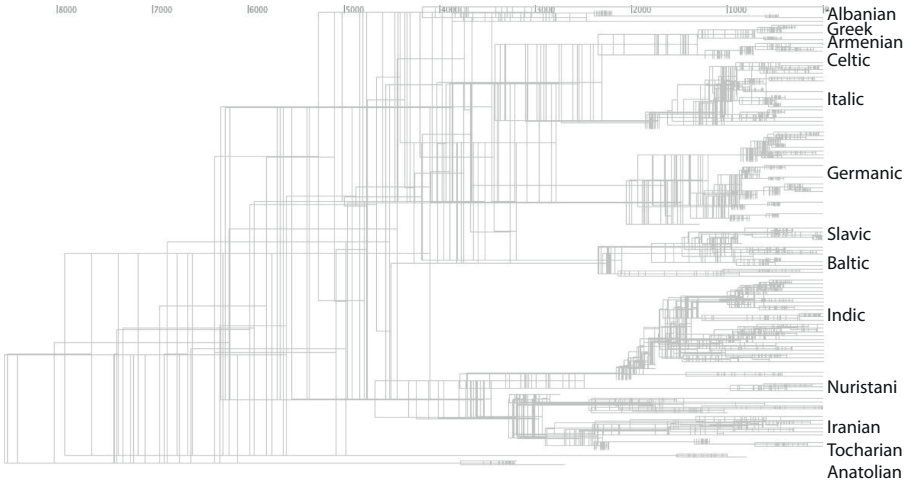


Figure 2. Indo-European languages (more detailed version available at https://github.com/chundrac/areal-pressure-gram-evo/blob/master/diacl_densitree.pdf)

5.1 Evolutionary model

Our evolutionary model works as follows: first, we infer distributions for each feature’s gain and loss rate. Subsequently, we carry out ancestral state reconstruction and character mapping to simulate change events over the phylogeny, coupling this information with that of the phylogeographic model in order to pinpoint these events in time and phylogeographic space.

5.1.1 Rate inference

We assume that the features in our dataset evolve according to a two-rate continuous-time Markov (CTM) process, with a gain rate α and loss rate β . For a branch of length $t_{\text{parent}[n],n}$ connecting a node n to its direct ancestor, denoted by $\text{parent}[n]$, the probability that $\text{parent}[n]$ is in state $s(\text{parent}[n])$ and n is in state $s(n)$, $p_{s(\text{parent}[n])s(n)}(l_{\text{parent}[n],n})$, can be calculated from α and β as follows (Liggett 2010: 59–50):

$$(1) \quad \begin{array}{c|cc} & s(n) = 0 & s(n) = 1 \\ \hline s(\text{parent}[n]) = 0 & \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta} e^{-(\alpha+\beta)t_{\text{parent}[n],n}} & \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta} e^{-(\alpha+\beta)t_{\text{parent}[n],n}} \\ s(\text{parent}[n]) = 1 & \frac{\beta}{\alpha+\beta} - \frac{\beta}{\alpha+\beta} e^{-(\alpha+\beta)t_{\text{parent}[n],n}} & \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta} e^{-(\alpha+\beta)t_{\text{parent}[n],n}} \end{array}$$

The values for α and β are unknown. We wish to draw samples for these parameters from their posterior distributions via an MCMC procedure, as outlined below.



Figure 3. Indo-European language locations

The likelihood of the data observed at the tips of a tree with a fixed topology given a gain rate α and loss rate β can be efficiently computed via the PRUNING ALGORITHM (Felsenstein 2004: 253–255), which starts at the tips of the tree and recursively computes the likelihood that each interior node of the tree is in state 0 or 1 (in the case of a binary character) under the current rate parameters, as follows:

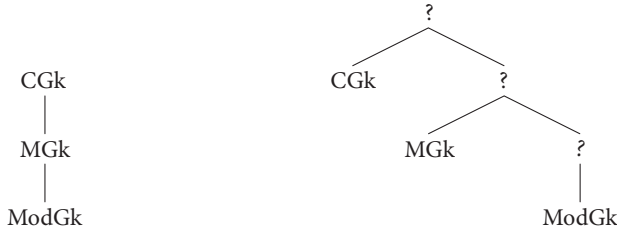


Figure 4. Tree topologies representing the relationship between Classical, Middle and Modern Greek, with and without ancestry constraints

$$(2) \ell(s(n) = i; \alpha, \beta) = \prod_{\text{child} \in \text{children}[n]} \left(\sum_{j=0}^1 p_{ij}(t_{n, \text{child}}; \alpha, \beta) \ell(\text{child} = j; \alpha, \beta) \right)$$

Above, $\text{children}[n]$ denotes the direct descendants of node n , and $\ell(\text{child} = j; \alpha, \beta)$ the likelihood of a given child having the state j under the rate parameters. For tips of the tree where data is observed, this likelihood is independent of the rate parameters, and is 1 or 0 depending on whether the tip in question exhibits the state j . In the case of missing data at the tips of the tree, the likelihood of each state is set to 1 (Felsenstein 2004: 255).

The overall likelihood of the tree is the sum of likelihoods that the root is in each state, weighted by the prior probability of each state:

$$(3) \ell(\text{tree}; \alpha, \beta) = \frac{\alpha}{\alpha + \beta} \ell(\text{root} = 1; \alpha, \beta) + \frac{\beta}{\alpha + \beta} \ell(\text{root} = 0; \alpha, \beta)$$

For each feature f , we sample from the posterior distribution of α_f and β_f via MCMC, an iterative procedure. We randomly initialize α_f and β_f computing the likelihood under α_f and β_f ($\ell(\text{tree}; \alpha_f, \beta_f)$), and new values α'_f, β'_f are drawn as follows:

$$(4) \alpha'_f \sim N(\alpha_f, \sigma_{\alpha_f}), \beta'_f \sim N(\beta_f, \sigma_{\beta_f})$$

The likelihood of the tree under α'_f, β'_f $\ell(\text{tree}; \alpha'_f, \beta'_f)$, is computed as well. The proposed values α'_f, β'_f are accepted (replacing the current values) via the Metropolis-Hastings algorithm (in this case equivalent to the Metropolis algorithm, since the proposal distribution is Gaussian and therefore symmetric), i.e., if a randomly generated number $a \sim U(0, 1)$ is less than $\min(1, \frac{\ell(\text{tree}; \alpha'_f, \beta'_f)}{\ell(\text{tree}; \alpha_f, \beta_f)})$.

Above, $\sigma_{\alpha_f}, \sigma_{\beta_f}$ denote the step sizes for the two parameters. These values are nontrivial, as they have an effect on the extent to which the space of possible states is explored. We initialize these values at .5. During the initial stages of inference, we monitor the proportion of proposals that are accepted and tune the step sizes by raising or lowering them if the acceptance rate is too high or low, respectively. We aim for an optimal acceptance rate in the 20–30% range (cf. Rosenthal 2011).

We run this procedure on three chains for 10,000 iterations, averaging over all trees in the tree sample (in order to incorporate a degree of phylogenetic uncertainty) and discarding the first half of samples as burn-in. We monitor convergence by calculating the potential scale reduction factor \hat{R} for each posterior sample (log-transformed to introduce normality), which is the square root of the ratio of the between-chain variance and the average within-chain variance, under the assumption that the posterior distribution is normal (Gelman & Rubin 1992). Values below 1.1 indicate good convergence, a threshold reached by all the rate distributions we infer.

5.2 Changes in time and space

5.2.1 Character mapping

Given fixed values or distributions over rates of evolution, evolutionary character histories can be simulated over a tree topology, first by reconstructing state probabilities at internal nodes of the tree, then by simulating birth and death events over the branches of the tree, the waiting times of which follow an exponential distribution parameterized by the evolutionary rate values inferred above. This procedure is described in various papers (e.g., Nielsen 2002; Bollback 2006).

5.2.2 Ancestral state reconstruction

Given a posterior distribution of rates α_f and β_f , we can estimate the probability that feature f is present at the root of the tree or a given interior node. This is done in the following manner. First, values for α_f and β_f are sampled with replacement from their respective posterior distributions. The likelihood of each node under these rate values is computed, again via the pruning algorithm. Subsequently, the probability that feature f is present at the root can be computed, as follows:

$$(5) \quad p(s_f(\text{root}) = 1 | \text{data, tree, } \alpha, \beta) = \frac{\alpha \cdot \ell(\text{root} = 1; \alpha, \beta)}{\alpha \cdot \ell(\text{root} = 1; \alpha, \beta) + \beta \cdot \ell(\text{root} = 0; \alpha, \beta)}$$

The state of the root can be simulated via a draw from a Bernoulli distribution parameterized by the probability given above.

Moving down the tree toward the tips, we calculate the probability that feature f is present in each internal node of the tree n on the basis of the state sampled for its parent, $\text{parent}[n]$:

$$(6) \quad p(s_f(n) = 1 | s_f(\text{parent}[n]) = i, \text{data, tree, } \alpha, \beta) = \frac{\ell_f(n = 1; \alpha, \beta) p_{i1}(t_{\text{parent}[n], n}, \alpha, \beta)}{\sum_{j=0}^1 \ell_f(n = j; \alpha, \beta) p_{ij}(t_{\text{parent}[n], n}, \alpha, \beta)}$$

This update can also be carried out for tips with missing data. The above procedure is carried out many times, yielding a distribution over change events.

5.2.3 Birth/death simulation

Once state values have been sampled at ancestral states, birth and death events can be simulated along branches as follows: for branches where the parent state is 0 and the child state is 1, a waiting time for birth can be drawn from an exponential distribution parameterized by the currently sampled value of α_f ; for branches where the parent state is 1 and the child state is 0, a waiting time for death can be drawn from an exponential distribution parameterized by the currently sampled value of β_f . Given a waiting time $w \sim \text{Exponential}(\alpha_f)$ or $\text{Exponential}(\beta_f)$, the geographic locations of these evolutionary events can be simulated on the basis of the phylogenetic information inferred for the nodes in the tree:

$$(7) \text{ longitude(event)} \approx \frac{w}{t_{\text{parent}[n], n}} (\text{longitude}(n) - \text{longitude}(\text{parent}[n]));$$

$$\text{latitude(event)} \approx \frac{w}{t_{\text{parent}[n], n}} (\text{latitude}(n) - \text{latitude}(\text{parent}[n]))$$

The time (in years) at which the inferred event takes place can easily be gleaned on the basis of the age of the parent node of the branch on which the change occurs plus the waiting time drawn. We tabulate all simulated change events and analyze the spatiotemporal behavior displayed by these events as described below.

6. Results

6.1 Evaluation metrics

We modify the character mapping procedure described above in order to evaluate the accuracy with which our evolutionary model generates the data in our dataset. Our inference procedure fits the model on the basis of posterior probability, a helpful metric for model selection, but otherwise not particularly meaningful when assessing a model on its own. We present two measures designed to assess how well our parameters generate the data in our model, inspired by evaluation metrics that are the norm in Natural Language Processing.⁷

7. The rates inferred have been inferred on the basis of all the data in our dataset; at the time of writing, we have no principled or efficient way of cross-validating our rates by holding out certain observed data points.

6.1.1 Capturing data at tips

One of our evaluation methods is inspired by phylogenetic data simulation (cf. Paradis 2014), but to our knowledge it serves as a novel means of evaluation. We first draw a state for the root on the basis of Equation 5. Moving down the tree, we simulate states for each of the nodes on the basis of Equation 6, but we exclude the likelihood term such that the state is drawn on the basis of the CTM process alone:

$$(8) \ p(s_f(n) = 1 | s_f(\text{parent}[n]) = i, \text{tree}, \alpha, \beta) = p_{i1}(t_{\text{parent}[n], n}; \alpha, \beta)$$

Finally, we sample states for the tips on the basis of the above equation; we can then compare the simulated state $s_f(\text{tip})^{\text{sim}}$ to the true state $s_f(\text{tip})^{\text{true}}$ for tips without missing data, and compute an accuracy measure $1 - |s_f(\text{tip})^{\text{sim}} - s_f(\text{tip})^{\text{true}}|$.⁸

In general, we believe that this metric will be challenging for a number of evolutionary models and will return measures of accuracy that are relatively conservative. In particular, it is likely that this metric will pose difficulties for evolutionary trees where there is a great deal of featural variation within subgroups, as the process of generating the data at the tips will become increasingly aleatory – more so than in clades where most members share the same feature value (which would be reconstructed with higher probability at a common ancestor).

We report the results of this evaluation procedure for readers' general interest. In reality, if this measure is to be useful in future work, it will have far more utility in facilitating direct comparison between two different types of evolutionary model, something that is not at issue in this paper. Accuracy levels averaged across languages for each feature (ranging from .49 to .97, with a mean of .68 and median of .66), as well as levels averaged across features for each language (ranging from .6 to .75, with a mean of .68 and median of .69), can be found in the Appendix (<http://dx.doi.org/10.1075/dia.16035.cat.additional>). In general, features with high accuracy levels tend to be fairly homogeneous at the tips of the tree; lower values are found for heterogeneous features. Accuracy across languages shows a much narrower range of values, and it is harder to come up with a clear account of why some languages show higher accuracy than others, though it is worth noting that older languages tend to display higher than younger ones. This is likely due to the fact that older languages have a great deal of influence over the value reconstructed for the root; in turn, due to their chronological proximity to the root, the values simulated for older languages are more likely to be identical to the root under a CTM process, resulting at least in part in more accurate values (while values

8. A reviewer suggests the use of the F-score, which entails partitioning the data to be evaluated into true positives, false positives and false negatives. While this is theoretically possible for us to do, we find the simple evaluation metric we describe here to be more intuitive, since the dataset to be evaluated and the dataset which serves as the basis of evaluation are of the same size.

sampled for younger languages in the tree are increasingly subject to randomness and hence lower accuracy).

6.1.2 *Capturing ancient/medieval data*

Our use of ancestry constraints serves as a crucial calibration point for the rates of change in our model. One possibility is to evaluate the accuracy of this calibration by holding out the data found in languages which are treated as directly ancestral to other languages, and seeing how well our model recapitulates the states found in these languages. To do this, we employ the evolutionary rates inferred in the procedure described above. Then, we carry out stochastic character mapping for each feature variant but treat the ancestral nodes in question as if they contain missing data. We then compare the simulated states against their true values via the metric described in the previous evaluation procedure. Average accuracy for ancient and medieval languages in the database ranges from .72 to .93 (with the value 1 representing complete accuracy), with a mean and median of .84. Values for each language are given in the Appendix (<http://dx.doi.org/10.1075/dia.16035.cat.additional>).

Old Italian has the highest average accuracy, while Classical Greek has the lowest; in general, medieval languages in a direct lineage between an ancient and modern language (e.g., Middle Greek, between Classical and Modern Greek; Middle High German, between Old and Modern High German, etc.) fare better than ancient languages. This could potentially be due to a number of factors, including the fact that these languages in particular are both typologically and chronologically close to their daughter languages (which are few in number and relatively homogeneous). It is worth noting that, for all but two categories attested for both languages, Old and Modern Italian agree. Classical and Modern Greek show a much higher degree of disagreement, so when we attempt to reconstruct the former language on the basis of data attested by the latter language, there is a greater possibility of error. A greater potential for error is also to be expected for ancient languages with a large number of diverse daughter languages, such as Latin (with an accuracy level of .73). However, Sanskrit fares relatively well (.89), despite having a similar genealogical profile. This is perhaps due to the inclusion of Prakrit, which agrees with Sanskrit to a large extent and is close enough to Sanskrit chronologically to influence the character state which is reconstructed. Were Prakrit to be omitted, Sanskrit features would likely be reconstructed with lower accuracy.

6.1.3 *Dependencies*

A number of dependencies exist between features in the DiACL dataset. In this paper, we employ a relatively hands-off approach to dealing with these dependencies; our rationale is that, given the fact that ancestral state likelihoods are computed

on the basis of the distributions of features at tips on the tree, the probability of reconstructing a combination of features that violates dependency rules is relatively unlikely.⁹

We identify a number of sets of interdependent features. Some dependencies involve attraction, i.e., in order for a feature to be valued 1, one or more additional features must have that value. Other dependencies are repellent, i.e., in order for a feature to be valued 1, one or more additional features must have the value 0. For some interdependent features, such as features pertaining to main and subordinate clause word order, at least one feature must be valued 1 (according to the coding system employed in this paper). Additionally, word order variants such as V2 should not be able to co-occur with other orders within a single grammatical system.

We tabulate illicit value combinations for these sets of interdependent variants. For instance, the variant NOMINAL MORPHOLOGY: NOMINAL CASE: DAT/GEN (“does the language have genitive morphological case as distinct from dative case?”) requires that two other variants, namely NOMINAL MORPHOLOGY: NOMINAL CASE: DAT and NOMINAL MORPHOLOGY: NOMINAL CASE: GEN, have values of 1. Similarly the variant NOMINAL MORPHOLOGY: NOMINAL CASES: OBL-CASE (“are there any cases besides agent, object, genitive, dative, vocative?”) requires that all variants pertaining to the presence of these cases be valued 1. The variant NOMINAL MORPHOLOGY: NOMINAL CASES: > 7 CASES (“are there more than 7 cases?”) ought to entail a value of 1 for OBL-CASES.

The full list of interdependent variant sets, along with licit combinations for these sets, is given in the Appendix (<http://dx.doi.org/10.1075/dia.16035.cat.additional>). For the majority of the sets identified, illicit combinations are found in no more than 10% of character mapping simulations. However, there are a small number of exceptions. Illicit combinations for sets involving the variants NOMINAL MORPHOLOGY: NOMINAL CASES: > 7 CASES and NOMINAL MORPHOLOGY: NOMINAL CASES: OBL-CASES for 38% and 41% of iterations; additionally, scores for main-clause and subordinate-clause word order variants show that V2 frequently co-occurs with other word order variants, which is not permissible under the definitions for the features in question. These values suggest that features such as NOMINAL MORPHOLOGY: NOMINAL CASES: > 7 CASES and NOMINAL MORPHOLOGY: NOMINAL CASES: OBL-CASES be excluded from future phylogenetic analyses, given their dependency on a large number of features already in the dataset, and that word order should perhaps be modeled as a multistate character. We retain these features with these caveats in mind.

9. In theory, a way to deal with the dependency problem is to bundle multiple interdependent binary features together as a single multistate feature.

6.2 Reconstructions

As automated ancestral state reconstruction gains increasing currency in historical linguistics, it has become common to evaluate the performance of ancestral state reconstruction against comparative reconstructions carried out by one or more experts (Dunn et al. 2017; Jäger & List forthcoming). Given the long history of scholarship on Indo-European morphosyntax, as well as the frequently opposing views found in the literature, a full appraisal along these lines is beyond the scope of this paper; we set this evaluation aside for future work. We give the full series of PIE reconstructions in the Appendix (<http://dx.doi.org/10.1075/dia.16035.cat.additional>) (i.e., for a given feature, the proportion of simulations for which a value 1 has been drawn, according to the procedure described in §5.2.2), and we briefly discuss key results below.

Nominative/accusative alignment is reconstructed across the scheme of features pertaining to alignment; for each feature, the variant A=Sa (“is the subject of a transitive clause marked the same as the subject of an agentive intransitive verb?”) is reconstructed with probability greater than .5, whereas alternative alignment patterns are reconstructed below this threshold. Nominative/accusative alignment was proposed for Proto-Indo-European by Delbrück (1900); in later years, the idea that PIE had active/stative alignment (promoted by Gamkrelidze & Ivanov 1995) enjoyed popularity, though this view has been reconsidered in recent years (cf. Viti 2014). Our results fail to bear on debates as to whether SOV (Delbrück 1900) or SVO (Friedrich 1975) word order should be reconstructed to Proto-Indo-European, as both word orders are chosen with roughly equal probability. All variants pertaining to definiteness are reconstructed with low probability. This result is uncontroversial, given evidence that Indo-European branches developed definiteness marking independently, by means of grammaticalization (Bauer 2007). Problematically, neuter gender is reconstructed with very low probability; this is undoubtedly due to the influence of Tocharian A and B, which lack a productive neuter gender.

At the time of writing, automated ancestral state reconstruction is still somewhat uncommon in evolutionary linguistics. We anticipate a growth in the application of these methodologies to Indo-European prehistory, which will allow scholars to compare results generated by different character evolution models, as well as develop a more refined framework for evaluating their accuracy.

A pressing question for us is whether potentially problematic reconstructions will bias the results of our spatiotemporal analysis. Since we classify gains of certain morphological categories involving case marking as introducing complexity, and expect that such changes will be more areally dispersed, a change taking place on a branch leading from PIE to Proto-Anatolian could potentially inflate the areal

dispersion which we measure, though we generally find this to not be a major problem for our analysis.¹⁰

6.3 Analysis

The birth/death simulation procedure described above allows us to estimate the location of each type of event (gain or loss) for each variant in time and space. This allows us to quantify the areal sensitivity displayed by each feature variant. We require a metric that allows us to determine how areally concentrated or dispersed parallel changes are, in order to establish a continuous measure which differentiates between changes sensitive to areality and those insensitive to it.

We use a version of mean nearest-neighbor distance (MNN; Clark & Evans 1954) to measure the spatiotemporal dispersion of change events. Lower mean nearest-neighbor distance values indicate greater areal clumping, while higher values indicate greater dispersion. We assess the degree of CONVERGENT behavior displayed by a type of change (e.g., gain or loss of a particular variant) in the following manner: for a given change (e.g., loss of the variant NOMINAL MORPHOLOGY: NOMINAL CASE: DAT/GEN) that occurs more than once during a given simulation, we calculate the spatiotemporal distance to the nearest change of the same type, and we take the average of these distances. We define spatiotemporal distance between two change events, change_i , change_j as follows (where $d_{GC}(.,.)$ is the great-circle distance between geographic coordinates of each event and $d_t(.,.)$ is the difference between the times, in years):

$$(9) \sqrt{d_{GC}(\text{change}_i, \text{change}_j)^2 + d_t(\text{change}_i, \text{change}_j)^2}$$

Mean nearest-neighbor distance is not independent of sample size. In general, if there are a large number of objects in a fixed area, nearest-neighbor distances between them are more likely to be small. Similarly, spatiotemporal distances between locations on trees in the tree sample are highly correlated with cophenetic (i.e., genetic) distances, since the model of spatial dispersal depends highly on the genealogy of the language. At the same time, it may be possible to control for cophenetic

10. Indeed, gains for the feature NOMINAL MORPHOLOGY: GENDER / NOUN CLASS: NEUTR rank quite high according to our areal metric, described in the following section (in 14th place). However, since this change introduces complexity, and we predict such changes to exhibit low areality, this particular data point adds a bit of noise to our model (going against our prediction) rather than inflating the probability of a false positive. At the moment, we lack a principled means of excluding data points from our spatial analysis on the basis of whether their automated reconstruction accords with traditional views; we can only hope to replicate our result in future studies using more refined models of character evolution.

distance as a nuisance factor. For each spatiotemporal nearest-neighbor distance measured between two simulated events, we measure the corresponding cophenetic distance as well. We define the cophenetic distance between two locations on the tree to be the sum of the lengths of the branches connecting them to their most recent common ancestor (if they are in different lineages) or the sum of the lengths of the branches connecting them (if they are in the same lineage).¹¹ Figure 5 displays log-transformed nearest-neighbor spatiotemporal distances for each type of change averaged according to each iteration of the simulation procedure, plotted by the average of the corresponding cophenetic distances, also log-transformed. The correlation is clearly visible (Pearson's $r = .73$). We seek to factor out this association by means of principal components analysis (PCA), which transforms a set of correlated variables into a set of uncorrelated ones (known as principal components) by locating axes (analogous to the line of best fit in a linear regression model) which explain the largest portions of variance in the data. We use PCA to tease apart spatiotemporal and cophenetic distances for each dispersion metric.

We note that the data we use (i.e., simulated change events) are grouped; multiple change events are generated by the same iteration of the simulation procedure. Different sets of change events will be generated using different trees in the tree sample. The relationship between spatiotemporal and cophenetic distance, while significant across all trees in the tree sample (Pearson's $r = .31 - .60$), varies across trees in terms of the degree of correlation and line of best fit. Standard PCA has no means for accounting for group-level idiosyncrasies of this sort, so we carry out separate principal components analyses for all 18 trees in the tree sample to account for differences in group-level behavior (rather than a single principal components analysis which pools together data generated using all 18 trees in the sample).

Certain change types do not exhibit recurrent changes for every iteration of the simulation procedure, given the probabilistic nature of stochastic character mapping. For instance, the variant WORD ORDER: MAIN CLAUSES: VSO exhibits parallel

11. In general, we are wary of potentially including changes that occur within the same lineage (e.g., if a variant is gained on an ancestral branch, then lost, then gained again on a branch leading to a descendant node) for multiple reasons. First, parallel changes of this sort do not speak to a contact scenario (it is certainly possible for languages to be influenced by the language from which they are descended, particularly if the ancestral language has served as a *Dachsprache*, e.g., Latin with respect to Romance, but we do not find this to be relevant to a study of convergent evolution of the phenomena which we investigate here). Secondly, the fact that cophenetic distances between nodes in the same lineage have the potential to be shorter on average than those between nodes in different lineages could possibly create bimodality in a distribution over cophenetic distances between data points. Given the fact that we do not observe bimodality among the cophenetic distances in our sample, we surmise that our simulation procedure rarely (if at all) treats nodes in the same lineage as spatiotemporal nearest neighbors.

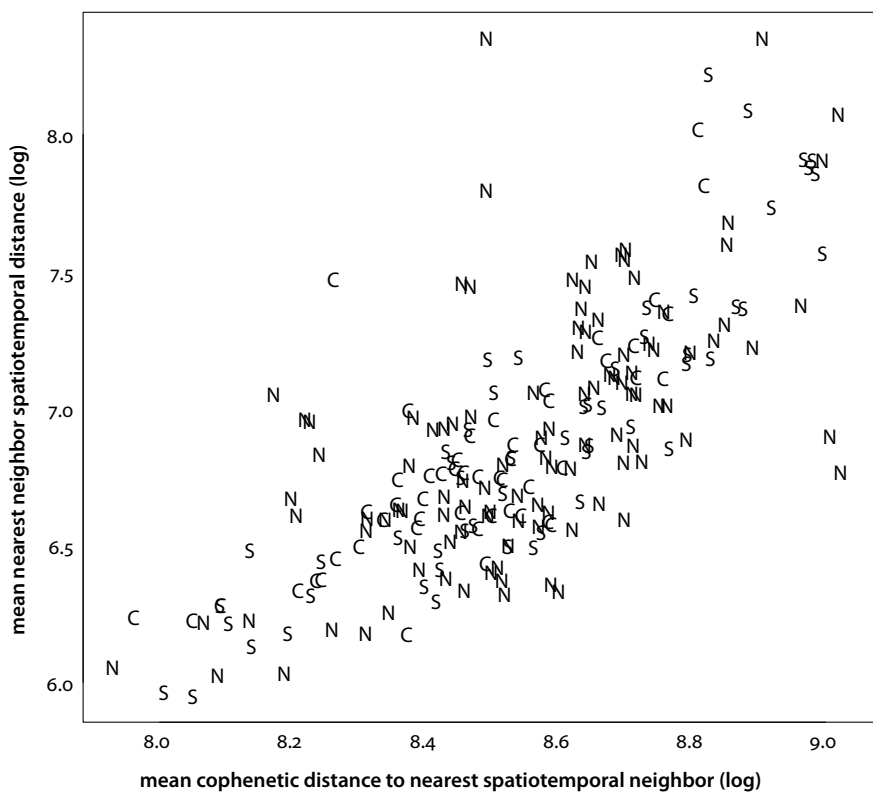


Figure 5. Mean nearest-neighbor spatiotemporal distance for each feature, plotted according to the mean of the corresponding cophenetic distances; labels indicate S(simplifying), N(eutral), C(complexifying)

losses in only 38 out of 100 iterations of the simulation procedure. This appears to be due to the fact that, in some cases, VSO word order is not reconstructed to Proto-Celtic (this result accords with the traditional comparative-historical view, given the fact that Gaulish, not in our sample, lacks VSO word order), but on occasion it is, and during a fraction of simulations it is lost in parallel on branches leading to various Celtic languages which lack main clause VSO word order. If changes which only occur more than once in a small fraction of iterations happen to display a particularly high areal concentration, the results of our linear model could be biased. We see no straightforward way to weight our areality metric according to the proportion of iterations in which a change occurs in parallel; instead, we exclude events pertaining to change types which occur in parallel in fewer than 80 iterations of the simulation procedure (leaving 185 out of 216 change types).

For each PCA, PC1 represents the correlation between spatiotemporal and cophenetic distance; high values represent data points that display high spatiotemporal and commensurate cophenetic dispersion. For all 18 analyses, this component explains 72–81% of the variance in the relationship between spatiotemporal and cophenetic distance. The remaining 19–28% of variance is explained by PC2; orthogonal to PC1, PC2 represents differences in spatiotemporal dispersion that are not explained by cophenetic distance; high values of PC2 indicate data points that are spatiotemporally concentrated but relatively cophenetically dispersed. As such, it serves as our variable of interest.

The proportion of variance explained by PC2 is above the 5% threshold above which a component in PCA is said to be meaningful (cf. Baayen 2008: 121). High values for PC2 represent data points that are more areally concentrated than would be expected given their cophenetic dispersion. We predict that changes that simplify (S) will show a greater spatiotemporal concentration (i.e., higher values of PC2) than changes that are neutral (N) and changes that introduce complexity (C), and that level N will show higher values of PC2 than level C.

We concatenate values for PC2 from the 18 principal components analyses we carried out (one for each tree in the tree sample) to create a vector variable quantifying AREALITY for the entire dataset. We fit a mixed-effects linear model which treats areality as the response to a categorical predictor with the levels (S)implifying, (N)eutral and (C)omplexifying. We include random intercepts for each TREE in the sample that is used, each GRID, each FEATURE VARIANT and each CHANGE TYPE (i.e., gain or loss); these factors have the potential to display group-level idiosyncrasies (and their inclusion is justified by a likelihood-ratio test). Since these levels are ordered, we use Helmert coding (Darlington & Hayes 2017: 283–286), a system in which the mean of each level is compared to the mean of the subsequent levels. Under our hypothesis, changes which introduce complexity exhibit the lowest areality, followed by neutral changes and changes which simplify. This yields the predefined contrasts N, S vs. C and S vs. N. Significance is assessed via a likelihood-ratio test.

We find that changes which introduce complexity (C) are significantly less areal than neutral (N) and simplifying (S) changes ($\beta_{S, N \text{ vs. } C} = .02$, $LR = 4.47$, $p = .03$; $\beta_{S \text{ vs. } N} = .02$, $LR = 0.83$, $p = .36$); however, we find no significant differences in areality between levels N and S. Coefficients and p -values cited above reflect simulations that were generated with the random number seed 0. We ran the simulation procedure 20 times using different random number seeds and carried out the same linear mixed-effects models described above. The directions of the effects were the same as above for 19 out of 20 random number seeds, and significance was achieved for the contrast S, N vs. C for 16 out of 20 random number seeds. We meta-analyze the p -values from the models for all 20 random number seeds via Fisher's combined

probability test (S, N vs. C: $\chi^2(40) = 166.2, p < .001$; S vs. N: $\chi^2(40) = 39.4, p = .49$). Overall, our results indicate that, according to our sample, simplifying and neutral changes form a natural class to the exclusion of changes which introduce complexity, with the latter group showing lower areal concentration among parallel changes.

7. Discussion

These results suggest that, in terms of convergent development, changes involving structural simplification as well as those pertaining to headedness and constituent order (which make up a large part of the neutral group) form a natural class to the exclusion of changes which introduce complexity, with the former group showing a greater areal signal. This suggests that the development of particular word orders, as well as the loss of verb agreement (as found in several North Germanic languages), may have been mediated by contact between speakers of Indo-European populations (as well as additional populations); in contrast, the development of marked alignment patterns such as nominative-accusative alignment or various types of verbal agreement seem to show more of a drift-like development, independent of areal interaction (according to our model). For instance, changes pertaining to the development of split ergativity (e.g., gaining the feature ALIGNMENT: COMPARE PROG-PAST: PAST_A=PROG_O “does the subject of a transitive verb in the simple past bear the same case form as the object of a verb in the present progressive?”) do not display much in the way of areal signal, according to our metric (ranking 142 out of 214 change types, with 1 being most areal).

Our results show that typological features are characterized by a great deal of variability in the degree of spatiotemporal concentration that their evolutionary patterns display. We have attempted to aggregate the overall areal sensitivity displayed by changes pertaining to these features. On a variant-by-variant basis, our results may or may not accord with views in the literature on linguistic areas and language contact. In some cases, they are compatible with other results concerning multilingualism and morphosyntactic change. For instance, we see that changes involving loss of nominal and pronominal case are well attested within the top quartile of all changes, ranked according to their areal sensitivity (e.g., loss of NOMINAL MORPHOLOGY: NOMINAL CASES: VOCATIVE, ranked 8; loss of NOMINAL MORPHOLOGY: NOMINAL CASES: GEN/DAT, ranked 16; etc.). In our view, this is compatible with views suggesting a link between so-called ‘foreigner-directed speech’ and loss of morphological case (Bentz & Winter 2013). It is conceivable that this dynamic could work toward the attrition of case in two languages that originally have it, particularly if a third substrate group (the sort that our model

cannot currently detect) were involved as well.¹² However, we also find that certain changes that involve gains of nominal case are found, albeit more sparsely, in the top quartile as well.

While nearest-neighbor distances have the potential to summarize and aggregate the areal signal displayed by linguistic evolution, some mismatches with received wisdom can be expected, given the nature of the metric. This can be seen in the case of V2 word order, considered by many to be an areal feature (Nichols 1998: 253), though our results do not show gains in V2 word order to be particularly concentrated for either metric in main or subordinate clauses (ranking 157 and 193, respectively). This is undoubtedly due to outliers like Kashmiri, which has developed V2 in isolation, and Icelandic, which developed V2 in subordinate clauses at a distance from Yiddish, the only other Germanic language with this feature. Loss of V2 main clause word order, however, is classified as highly areal (with a rank of 4); it is lost between Old French and Modern French dialects, Old and Modern Provençal, Old and Middle English, and Middle and Modern Welsh. It is likely that the relative proximity between the first two events on the one hand and the last two events on the other are responsible for this measurement.

VSO is another order thought to be areal, given its synchronic distribution; within Indo-European, it is confined to Insular Celtic. However, VSO does not exhibit great areal sensitivity according to the parallel gain or gain/presence MNN metrics (ranking 146 for subordinate clauses and 187 for main clauses); this is likely a consequence of the fact that modern Welsh developed VSO main clause word order at a relatively late date and hence at a fair spatiotemporal distance from Irish and Scottish Gaelic. The development of VSO in modern Welsh cannot be a chance occurrence, but if our model is valid, the areal-looking pattern found among Celtic languages is perhaps not due to contact alone; indeed, many of the causes of the shift from V2 to VSO in Welsh main clause word order may largely be due to intricate system-internal structural factors (cf. Willis 1998).

We note some final mismatches between our findings and conventional wisdom. The variants TENSE: FUTURE: FUT.AUX (“is there a future formed by an auxiliary?”) and NOMINAL MORPHOLOGY: DEFINITENESS: N-DEF (“is there a suffix for definiteness on the noun?”) are two prominent correlates of certain areal groups, particularly the Balkan *Sprachbund* (Schaller 1975). However, gains of these variants are not treated as particularly areal according to our metric; the changes’ ranks (89 and 106, respectively) are outside of the top quartile. This is undoubtedly due to the fact that these variants are gained in parallel outside of well-established areal

12. It is worth noting that many case studies have spurred a large amount of debate as to whether the imposition of one language’s feature onto another has taken place, or whether the two languages have undergone joint convergence induced by a third (cf. Hock 1996 [1993]).

groups, with future-tense auxiliaries coming about frequently and suffixal definiteness marking being widespread in North Germanic as well (though likely due to genetic inheritance rather than contact, since it is attested in Old Norse). Accordingly, our mean nearest-neighbor distance metric is insensitive to the fact that certain data points cluster together, if additional data points are relatively dispersed (it is worth noting that our analysis, which is restricted to Indo-European languages, excludes Turkish, a non-Indo-European member of the Balkan *Sprachbund*, and this may affect our results). It may be more appropriate to employ a distance-based metric concerned with whether a subset of evolutionary events are spatiotemporally concentrated, rather than all of them, though currently we can think of no principled way of implementing such a measure.

Our model at its current stage uses only morphosyntactic data, and thus it is ill-equipped to test some of the broader claims regarding structural (e.g., phonological as well as morphosyntactic) versus lexical effects of language contact at different life stages (Ross 2013), and our lack of reliable demographic and sociopolitical information makes it difficult to engage with notions of source versus recipient language agentivity and directionality (Winford 2005), but it is hoped that future adaptations of this methodology can be brought to bear on some questions from the larger literature on correlates and limitations of contact.

8. Conclusion

We have introduced a novel methodology for investigating the diachronic areal sensitivity of linguistic features, utilizing an explicit phylogenetic and geospatial model. We found that parallel changes which introduce linguistic complexity are less areally concentrated than parallel changes which involve linguistic simplification or are neutral, or do not affect the syntagmatic complexity of the grammatical system in either direction. We believe that this methodology presents a promising direction for research that seeks to investigate both the genetic or vertical and areal or lateral dynamics of linguistic change. Future use of this methodology may be extended to other types of linguistic features and other language families. Additionally, this method is compatible with a parsimony approach to character evolution, though parsimony lacks an explicit means of locating mutations on branches of a tree.

It should be stressed that what is presented here is not an explicit areal-phylogenetic model of evolution, and we have refrained from referring to parallel change *DUE TO* contact rather than *COMPATIBLE WITH* a contact scenario. A truly areal-phylogenetic model would need to have a generative story regarding, for each change, whether it is due to an areal or genetic factors, which in turn would have some associated parameters. This may be a distinction that requires greater

attention in the future. It may be possible to use hidden rates models (Beaulieu & O'Meara 2014) to tease apart 'genetic-looking' and 'areal-looking' change, at least to some extent, and it may ultimately be possible to explicitly model changes in evolutionary behavior as a function of the phylogeographic location of various locations on the tree. What we have presented has the advantage of diagnosing potentially contact-induced behavior among related languages. We have inferred all changes via a genetic model but have used phylogeographic information to diagnose which developments are more areally clustered together. With this model, we have taken an important step toward a fuller and more inclusive understanding of linguistic change.

Appendices

The appendices can be found online at: <https://doi.org/10.1075/dia.16035.cat.additional>

Acknowledgements

We thank Anne Goergens, Arthur Holmer, Sandra Cronhamn, Rob Verhoeven, Rob Farren, Josien Greidanus, Anastasia Karlsson and Karina Vamling for help with data entry and processing. We thank Will Chang and three anonymous reviewers for helpful discussion. All errors are naturally our own. This research was supported by the Crafoord Foundation (grant no. 20120794, 2012 [Språk- och kulturatlas for Sydamerika]) and the Marcus and Amalia Wallenberg Foundation (grant no. MAW 2012.0095, 2013–2015 [Lund Digital Information System of Contextualised Typology]). Code is available at <https://github.com/chundrac/areal-pressure-gram-evo>.

References

- Aikhenvald, Alexandra Y. 2007. Grammars in contact: A cross-linguistic perspective. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Grammars in contact*, 1–66. Oxford: Oxford University Press.
- Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511801686
- Baerman, Matthew, Dunstan Brown & Greville Corbett (eds.). 2015. *Understanding and measuring morphological complexity*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198723769.001.0001
- Bauer, Brigitte. 2007. The definite article in Indo-European: Emergence of a new grammatical category? In Elisabeth Stark, Elisabeth Leiss & Werner Abraham (eds.), *Nominal determination: Typology, context constraints, and historical emergence*, 103–139. Amsterdam: John Benjamins. doi:10.1075/slcs.89.07bau

- Beaulieu, Jeremy M. & Brian C. O'Meara. 2014. Hidden Markov models for studying the evolution of binary morphological characters. In László Zsolt Garamszegi (ed.), *Modern phylogenetic comparative methods and their application in evolutionary biology: Concepts and practice*, 395–408. Heidelberg: Springer.
- Bentz, Christian & Bodo Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3(1). 1–27.
- Bickel, Balthasar. 2011. Statistical modeling of language universals. *Linguistic Typology* 15. 401–413. doi:10.1515/lity.2011.027
- Bollback, Jonathan P. 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7. 88. doi:10.1186/1471-2105-7-88
- Botero, Carlos A., Beth Gardner, Kathryn R. Kirby, Joseph Bulbulia, Michael C. Gavin & Russell D. Gray. 2014. The ecology of religious beliefs. *Proceedings of the National Academy of Sciences* 111(47). 16784–16789. doi:10.1073/pnas.1408701111
- Braunmüller, Kurt. 1984. Morphologische Undurchsichtigkeit – ein Charakteristikum kleiner Sprachen. *Kopenhagener Beiträge zur Germanistischen Linguistik* 22. 48–68.
- Carling, Gerd (ed.). 2017. *Diachronic atlas of comparative linguistics online*. Lund: Lund University, <https://diac.l.lu.se/> (30 July, 2016).
- Chang, William, Chundra Cathcart, David Hall & Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European Steppe Hypothesis. *Language* 91(1). 194–244. doi:10.1353/lan.2015.0005
- Clark, Philip J. & Francis C. Evans. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* 35(4). 445–453. doi:10.2307/1931034
- Darlington, Richard B. & Andrew F. Hayes. 2017. *Regression analysis and linear models: Concepts, applications, and implementation*. London: Guilford Press.
- Daumé, Hal. 2009. Non-parametric Bayesian areal linguistics. In *Human language technologies: The 2009 annual conference of the North American chapter of the ACL*, 593–601. Boulder, CO: Association for Computational Linguistics.
- Dediu, Dan. 2010. A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. *Proceedings of the Royal Society of London B* 278(1704). 474–479. doi:10.1098/rspb.2010.1595
- Delbrück, Berthold. 1900. *Vergleichende Syntax der indogermanischen Sprachen*, vol. 3. Strassburg: Karl J. Trübner.
- Drummond, A. J., M. A. Suchard, D. Xie & A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29(8). 1969–1973. doi:10.1093/molbev/msso75
- Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13(2). 257–292. doi:10.1075/sl.13.2.03dry
- Dunn, Michael. 2015. Language phylogenies. In Claire Bower & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 190–211. Oxford: Routledge.
- Dunn, Michael, Tonya Kim Dewey, Carlee Arnett, Thórhallur Eythórsson & Jóhanna Barðdal. 2017. Dative sickness: A phylogenetic analysis of argument structure evolution in Germanic. *Language* 93(1). e1–e22.
- 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. doi:10.1038/nature09923
- Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink & Angela Terrill. 2008. Structural phylogeny in historical linguistics: Methodological explorations applied in Island Melanesia. *Language* 84(4). 710–759. doi:10.1353/lan.0.0069

- Felsenstein, Joseph. 2004. *Inferring phylogenies*. Sunderland, MA: Sinauer Associates.
- Fortson, Benjamin W. 2015. Indo-European: Methods and problems. In Claire Bowerman & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 645–656. Oxford: Routledge.
- François, Alexandre. 2011. Social ecology and language history in the northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics* 1(2). 175–246. doi:10.1075/jhl.1.2.03fra
- Friedrich, Paul. 1975. *Proto-Indo-European syntax: The order of meaningful elements* (Journal of Indo-European Studies Monograph 1). Butte, MT: Montana College of Mineral Science and Technology.
- Gamkrelidze, Tamaz & Vyacheslav I. Ivanov. 1995. *Indo-European and the Indo-Europeans* (translated by Johanna Nichols). Berlin: Mouton de Gruyter. doi:10.1515/9783110815030
- Garrett, Andrew. 2006. Convergence in the formation of Indo-European subgroups: Phylogeny and chronology. In Peter Forster & Colin Renfrew (eds.), *Phylogenetic methods and the pre-history of languages*, 139–151. Cambridge: McDonald Institute for Archaeological Research.
- Gelman, Andrew & Donald B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7. 457–511. doi:10.1214/ss/1177011136
- Graham, Christopher. 2016. *Geographical correlates of rare word orders: A computational approach to quantitative typology and language contact*. University of California, Davis dissertation.
- Gumperz, John J. & Robert Wilson. 1971. Convergence and creolization. In Dell Hymes (ed.), *Pidginization and creolization of languages*, 151–168. Cambridge: Cambridge University Press.
- Haspelmath, Martin, Matthew Dryer, David Gil & Bernard Comrie. 2005. *The world atlas of language structures*. Oxford: Oxford University Press.
- Haug, Dag. 2015. Treebanks in historical linguistic research. In Carlotta Viti (ed.), *Perspectives on historical syntax*, 187–202. Amsterdam: John Benjamins Publishing Company.
- Haynie, Hannah, Claire Bowerman & Hannah LaPalombara. 2014. Sound symbolism in the languages of Australia. *PLoS One* 9(4). e92852. doi:10.1371/journal.pone.0092852
- Hock, Hans Henrich. 1996 [1993]. Subversion or convergence? The issue of pre-Vedic retroflexion reexamined. *Studies in the Linguistic Sciences* 23(2). 73–115.
- Hoenigswald, Henry M. 1960. *Language change and linguistic reconstruction*. Chicago: University of Chicago Press.
- Hoenigswald, Henry M. 1966. Criteria for the subgrouping of languages. In Henrik Birnbaum & Jaan Puhvel (eds.), *Ancient Indo-European dialects*, 1–12. Berkeley, CA: University of California Press.
- Jäger, Gerhard & Johann-Mattis List. Forthcoming. Using ancestral state reconstruction methods for onomasiological reconstruction in multilingual word lists. *Language Dynamics and Change* 8.
- Jordan, Michael. 2004. Graphical models. *Statistical Science* 19(1). 140–155. doi:10.1214/088342304000000026
- Kroch, Anthony, Ann Taylor & Donald Ringe. 2000. The Middle English verb-second constraint: A case study in language contact and language change. In Susan C. Herring, Pieter Th. van Reenen & Lene Schøsler (eds.), *Textual parameters in older languages*, 353–392. Amsterdam: John Benjamins.
- Kusters, Wouter. 2008. Complexity in linguistic theory, language learning and language change. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 3–22. Amsterdam: John Benjamins. doi:10.1075/slcs.94.03kus

- Lass, Roger. 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620928
- Liggett, Thomas M. 2010. *Continuous time Markov processes: An introduction*, vol. 113 Graduate Studies in Mathematics. Providence, RI: American Mathematical Society.
- List, Johann-Mattis, Shijulal Nelson-Sathi, William Martin & Hans Geisler. 2014. Using phylogenetic networks to model Chinese dialect history. In Søren Wichmann & Jeff Good (eds.), *Quantifying language dynamics*, 125–154. Leiden: Brill. doi:10.1163/9789004281523_006
- Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLoS One* 5. 1–10. doi:10.1371/journal.pone.0008559
- Martinet, André. 1975. *Évolution des langues et reconstruction*. Paris: Presses universitaires de France.
- Meid, Wolfgang. 1975. Probleme der räumlichen und zeitlichen Gliederung des Indogermanischen. In Helmut Rix (ed.), *Flexion und Wortbildung*, 204–19. Wiesbaden: Reichert.
- Meillet, Antoine. 1925. *La méthode comparative en linguistique historique*. Oslo: H. Aschehoug & Co.
- Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson (eds.). 2008. *Complexity in linguistic theory, language learning and language change*. Amsterdam: John Benjamins.
- Nakhleh, Luay, Donald Ringe & Tandy Warnow. 2005. Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language* 81(2). 382–420. doi:10.1353/lan.2005.0078
- Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. *Language* 62. 56–119. doi:10.1353/lan.1986.0014
- Nichols, Johanna. 1998. The Eurasian spread zone and the Indo-European dispersal. In Roger M. Blench & Matthew Spriggs (eds.), *Archaeology and language II: Correlating archaeological and linguistic hypotheses*, 220–266. London: Routledge. doi:10.4324/9780203202913_chapter_10
- Nichols, Johanna. 2003. Diversity and stability in languages. In Brian D. Joseph & Richard D. Janda (eds.), *The Oxford handbook of historical linguistics*, 283–310. Oxford: Blackwell. doi:10.1002/9780470756393.ch5
- Nichols, Johanna. 2009. Linguistic complexity: A comprehensive definition and survey. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 110–125. Oxford: Oxford University Press.
- Nichols, Johanna & Tandy Warnow. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2(5). 760–820. doi:10.1111/j.1749-818X.2008.00082.x
- Nielsen, Rasmus. 2002. Mapping mutations on phylogenies. *Systematic Biology* 51(5). 729–739. doi:10.1080/10635150290102393
- Paradis, Emmanuel. 2014. Simulation of phylogenetic data. In László Zsolt Garamszegi (ed.), *Modern phylogenetic comparative methods and their application in evolutionary biology: Concepts and practice*, 335–350. Heidelberg: Springer.
- Renfrew, Colin. 2000. At the edge of knowability: Towards a prehistory of languages. *Cambridge Archaeological Journal* 10(1). 7–34. doi:10.1017/S0959774300000019
- Rosenthal, Jeffrey S. 2011. Optimal proposal distributions and adaptive MCMC. In Steve Brooks, Andrew Gelman, Galin L. Jones & Xiao-Li Meng (eds.), *Handbook of Markov chain Monte Carlo*, 93–112. Boca Raton, FL: Chapman & Hall/CRC. doi:10.1201/b10905-5
- Ross, Malcolm. 2013. Diagnosing contact processes from their outcomes: The importance of life stages. *Journal of Language Contact* 6(1). 5–47. doi:10.1163/19552629-006001002
- Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace and Company.

- Schaller, Helmut Wilhelm. 1975. *Die Balkansprachen: eine Einführung in die Balkanphilologie*. Heidelberg: Carl Winter.
- Swadesh, Morris. 1952. Lexicostatistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96. 452–463.
- Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21. 121–137. doi:10.1086/464321
- Trudgill, Peter. 1974. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society* 3(2). 215–246. doi:10.1017/S0047404500004358
- Trudgill, Peter. 2001. Contact and simplification: Historical baggage and directionality in linguistic change. *Linguistic Typology* 5. 371–374.
- Vennemann, Theo. 1994. Linguistic reconstruction in the context of European prehistory. *Transactions of the Philological Society* 92. 215–284. doi:10.1111/j.1467-968X.1994.tb00432.x
- Viti, Carlotta. 2010. The information structure of OVS in Vedic. In Gisella Ferraresi & Rosemarie Lühr (eds.), *Diachronic studies on information structure: Language acquisition and change*, 37–62. Berlin: Walter de Gruyter.
- Viti, Carlotta. 2014. Reconstructing syntactic variation in Proto-Indo-European. *Indo-European Linguistics* 2. 73–111. doi:10.1163/22125892-00201004
- Wichmann, Søren. 2015. Diachronic stability and typology. In Claire Bower & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 212–224. London: Routledge.
- Willis, David. 1998. *Syntactic change in Welsh: A study of the loss of the verb-second*. Oxford: Oxford University Press.
- Winford, Donald. 2005. Contact-induced changes: Classification and processes. *Diachronica* 22(2). 373–427. doi:10.1075/dia.22.2.05win
- Yang, Ziheng. 2014. *Molecular evolution: A statistical approach*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199602605.001.0001
- Yanovich, Igor. 2016. Genetic drift explains Sapir's "drift" in semantic change. In Seán G. Roberts, Christine Cuskley, Luke McCrohon, Lluís Barceló-Coblijn, Olga Fehér & Tessa Verhoef (eds.), *The evolution of language: Proceedings of the 11th international conference (EVOLANGX11)*, Online at <http://evolang.org/neworleans/papers/24.html>.

Résumé

Cet article traite de la dynamique évolutionnaire et spatiale des caractéristiques typologiques dans 117 langues indo-européennes. Trois catégories de développements linguistiques ont été établies: (1) simplification, (2) neutralisation, et (3) complexification. Nous constatons que les développements qui créent de la complexité démontrent un niveau de concentration spatiotemporelle inférieur, de façon significative, par rapport aux deux autres types de changement. Ce résultat est compatible avec une situation dans laquelle certains changements proviendraient de contacts bilatéraux et de contacts entre des communautés linguistiques proches géographiquement, tandis que les autres seraient purement aléatoires, et n'auraient aucun lien avec le degré de contact entre différentes communautés linguistiques.

Zusammenfassung

Dieser Beitrag behandelt die evolutionäre und räumliche Entwicklung einiger typologischer Charakteristika in 117 indoeuropäischen Sprachen. Wir untergliedern die Studie hinsichtlich der Kategorien des Sprachwandels, welcher bei jeder der untersuchten Varianten auftritt: (1) Wandel, der zu einer Simplifizierung erlernter morphosyntaktischer Strukturen führt (2) neutraler Wandel (also weder Simplifizierung noch Komplexifizierung), und (3) Wandel, der eine komplexere Struktur mit sich bringt. Dabei können wir beobachten, dass Wandelphänomene, die zu einem Anstieg an Komplexität führen, gemäß der von uns entwickelten Metrik signifikant weniger räumliche Ausprägungen zeigen als Simplifizierungen und neutraler Wandel; allerdings finden wir keine signifikanten Unterschiede innerhalb der letzten beiden Gruppen. Dieses Ergebnis lässt sich verbinden mit einem Szenario, bei dem bestimmte Typen des Sprachwandels eher durch Advergenz und Kontakt zwischen benachbarten Sprachgemeinschaften vermittelt wird, während andere Entwicklungen lediglich auf Drift zurückzuführen und großenteils unabhängig von interkulturellem Kontakt sind.

Authors' addresses

Chundra Cathcart (corresponding author)
Department of Comparative Linguistics
University of Zurich
Plattenstrasse 54
8032 ZÜRICH
Switzerland
chundra.cathcart@uzh.ch

Gerd Carling
Department of Linguistics and Phonetics
Lund University

Filip Larsson
Department of Linguistics and Phonetics
Lund University

Niklas Johansson
Department of Linguistics and Phonetics
Lund University

Erich Round
Ancient Language Lab and Centre of Excellence for the Dynamics of Language
University of Queensland (St. Lucia, Australia)
Surrey Morphology Group
University of Surrey (Guildford, UK)