To appear in *Language dispersal, diversification, and contact: a global perspective*,
ed. M. Crevels, J.-M. Hombert & P. Muysken, Oxford University Press

Large and ancient linguistic areas[*]

Balthasar Bickel

University of Zürich

1 Introduction

Areal diffusion is ubiquitous in human language. The maps of the *World Atlas of Language Structure* (Comrie et al. 2005) are full of geographical patterns. Many of the areas that we see, or think we see, are large and cover entire continents or even more (Dryer 1989, Nichols 1992, and many others since). This is no surprise since most languages are spoken in close neighborhood to each other (with 50% of languages having a neighbor as close as 30 km: Köhli 2013), and there is robust evidence for various scenarios of language contact (Muysken 2010) to result over time in the mutual assimilation of languages (as amply attested for example by the chapters in the present volume).

The geographical patterns that result from contact-induced processes have received increased attention during the past two decades. This is especially true for large-scale areas that are based on typological data because such areas can open up windows into the deep past (Nichols 1992, Dediu & Levinson 2012, among others) and define a key confound for discovering and demonstrating universal biases (Dryer 1989, Bickel in press, among others). But how can we demonstrate such areas? Plotting or analyzing maps is not enough, as correlations between language and space can of course just as well be entirely spurious (Roberts & Winters 2013).

The first challenge here is the formulation of a well-founded theory of the specific historical scenario that explains and thereby predicts an observable area. Following the principles of *Predictive Areality Theory* (Bickel & Nichols 2006), the foundation for such a theory must come from outside linguistics, to avoid circularity. For smaller-scale areas, such as Europe or Southeast Asia, such evidence can be derived from known population history (as e.g. in Haspelmath 2001, Heine & Kuteva 2006, Enfield 2005). For larger

---

[*] I thank the audiences at the KNAW Conference on Patterns of Diversification and Contact in Amsterdam (2012), the Conference on Diversity Linguistics in Leipzig (2015), and the Workshop on Capturing Phylogenetic Algorithms for Linguistics in Leiden (also 2015) for useful comments. Many thanks also go to the editors for helpful suggestions.

areas, such as Eurasia (Jakobson 1931) or the Trans-Pacific (also known as Pacific Rim or Circum-Pacific) areas (Nichols 1992), we need to resort to a combination of historical, archeological and population-genetic sources.

The predictions of such areality theories are probabilistic and diachronic: if a language is located within a historically formed area, it is more likely to develop in such a way that it is similar to the other languages in the area than to develop in the opposite direction. And if a language already fits the profile of the area, it is more likely to keep its structures than to develop away from the profile. How can we test such predictions? Simple counting languages or clustering them runs into the well-known Galton problem of hidden dependencies: languages enter many other dependencies apart from being in the same area; in particular, they bear genealogical relationships to each other, and any counts will be influenced by such relationships. These problems are particularly severe for hypotheses about large-scale and potentially ancient areas.

In this chapter, I first discuss traditional approaches to testing areal hypotheses in typology and note several shortcomings of these approaches when applied to large-scale areas (Section 2). I then offer an alternative approach in Section 3. Section 4 applies this alternative to a case study on the Trans-Pacific area, an area that was first proposed on the basis of traditional typological methods (Nichols 1992, Bickel & Nichols 2006). In Section 5, I summarize what current methods can do and suggest avenues for future research.

2 Traditional approaches

Most studies of geographical patterns seek to solve Galton's problem by directly comparing the predictions of areal and genealogical hypotheses. In the simplest case, a pattern can be declared as areal when the languages in the area are known to be unrelated, i.e. when they are all isolates. In other cases, a pattern can be shown to be areal if it clearly cross-cuts the branches of a family, cannot plausibly be reconstructed in a family, or covers only parts of a family, leaving divergent relatives outside the area of interest. Such approaches work well for relatively small regions like Europe, and considerable progress has been made in quantifying the evidence for geographical signals under such circumstances (van der Auwera 1998, McMahon & McMahon 2005, Janssen et al. 2006, Gray et al. 2010, Donohue & Whiting 2011, Muysken et al. 2014).

But this approach does not scale up to larger and older areas, e.g. potential areas like Eurasia or the Trans-Pacific. These are typically not limited to isolates, and they tend to comprise families exhaustively, e.g. Eurasia includes all of Indo-European or Sino-Tibetan. This means that we cannot evaluate areal signals by comparing the branches of a family or by checking relatives outside Eurasia.[1] Demonstrating that a feature does not reconstruct is an option in principle, but this enterprise is fraught with severe

---

[1] The present-day presence of such relatives in the Americas and elsewhere is too recent a phenomenon to reveal any signals from the time when large areas were formed.

problems when it comes to abstract typological features and when exploring families whose history is less well-understood than that of Indo-European or Austronesian.

Given this, we clearly need another way to solve Galton's problem when dealing with hypotheses of large areas. And solving the problem is important. Without addressing it, areal signals might be significantly inflated by genealogical factors, i.e. genealogical relatedness might be mistaken for geographical patterning. In the worst case, signals might be heavily distorted by Simpson's paradox, where an area-wide preference for some features hides systematic dis-preferences against the same features within each family in the area, or vice-versa (Jaeger et al. 2011).

The most common response to Galton's problem in areal typology is the same as in research on universals: genealogical relations are treated as a confounding factor. Once we weed out the impact of this factor, so the argument, the remaining distributions reflect the effects of the factor of interest, here an areal pattern. The assumption is that if languages are genealogically related, it is possible that they share structures not because of the relevant factor of interest, but because they descend from the same proto-language. Concretely: French and Italian have relative pronouns not because they are in the European Sprachbund but simply because Latin had relative pronouns. If we control for such cases, any areal (or universal) pattern can then be tested through standard statistical approaches, e.g. with clustering algorithms or regression models.

All that is needed from this point of view is a method for excluding the possibility of shared inheritance. This can be achieved by factoring out any effects from this possibility. This basic idea is implemented in the literature mostly through strategic sampling, e.g. by collapsing data-points that are identical within a family or within each subbranch of a family ('genealogically-balanced sampling'; Dryer 1989, Cysouw & Comrie 2009), or, less commonly, through re-sampling, e.g. by randomly picking one datapoint per family in thousands of samples (e.g. Everett et al. 2015). An alternative implementation is to build genealogical relatedness directly into statistical models as a confounding variable, either as a fixed (e.g. Dediu & Ladd 2007, Bickel et al. 2009) or as a random factor (e.g. Jaeger et al. 2011, Bentz & Winter 2013). This method adjusts the estimates of the areal effect of interest for possible confounding effects that arise from family-specific skewing patterns, possibly subject to interactions.

Three problems arise. They are all grounded in the fact that genealogical relatedness is not orthogonal to areal diffusion.

First, if a typological feature is shared within a family this does not necessarily reflect shared inheritance. In fact, if a feature is shared within a family, this can just as well be caused by areal diffusion. An example is the distribution of differential object marking (DOM) in Romance and in Indo-Iranian. Map 1 shows the Eurasian distribution of DOM, defined here as a formal distinction of P from {S,A} argument[2] marking that is conditioned by some category, e.g. a pronoun vs. noun condition or a definite vs. indefinite condition (Bickel et al. 2014). As the map suggests, all Romance languages

---

[2] S refers to the sole argument of one-argument verbs; A and P to the most and least agentive arguments of two-argument verbs, respectively.

Map 1: Differential object marking by cases or adpositions in Eurasia. Filled symbols signal the presence of categorically conditioned P $\neq$ {S,A} marking somewhere in the language; empty symbols its absence. An upward triangle (△) represents Romance, a downward triangle (▽) Indo-Iranian. *Projection*: Gall. *Data source:* Bickel et al. (2014).

and almost all Indo-Iranian languages of the sample show some version of differential object marking. The concrete manifestations vary, however, and cannot be traced to the respective ancestors of these groups. Romance includes simple pronoun vs. noun splits (French) along with information-structure based distinctions as in Spanish or Romanian, which, moreover, have different etymologies (with an object marker *a* in Spanish and *pe* in Romanian). None of these DOM patterns are inherited. The same holds for Indo-Iranian where DOM is demonstrably younger than the last common ancestor and is derived from varied etyma (cf. e.g. Nepali *-lāi* vs. Hindi *-ko* vs. Persian *râ*) (Schikowski 2013:215-234). Under a traditional typological approach, all of these variants would be collapsed into a single datapoint, or they would be treated as a family-specific adjustment in a regression model. In either case, we could not even begin to ask whether the parallels within the families are triggered by language contact or not. From Map 1 one would expect large-scale contact patterns to have played a role (Bickel et al. 2014), but this hypothesis cannot be tested by relying on the traditional approach.

A second problem that arises from traditional approaches is that even if a feature reflects shared inheritance, this does not rule out the possibility that it was affected by contact and area formation (Masica 2001). Take the distribution of relative pronouns, defined as clause-initial pronouns specialized for the role of a head in a modifier clause (Lehmann 1984). The distribution of relative pronouns is a well-established areal feature of Europe (Haspelmath 2001, Comrie & Kuteva 2005), as shown for example by the fact that new relative pronouns have developed in some varieties of Basque through contact with Romance or in some varieties of Turkish in contact with Slavic and Albanian (Heine & Kuteva 2006: Chapter 6). In the case of Romance, relative pronouns reflect shared inheritance from Latin, but this does not imply that contact was unimportant because contact can also increase the probability of preservation. Indeed, to the extent that relative pronouns are dispreferred worldwide, contact is the most plausible explanation of why Romance daughter languages have strongly tended to keep the inherited structure. Similar observations can be made for other phenomena: for example, gender distinctions are not a feature that languages favor universally, but the feature tends to be better retained when it clusters geographically (Nichols 1992, 2003, Bickel 2013). This suggests that contact within or between language families can support retention. Such effects cannot in principle be detected when following the standard approach to solving Galton's problem.

Third, a large and ancient area is unlikely to be the result of a one-off, saltational effect. An area of this kind is more likely to slowly emerge through diverse contact events repeated over time (Heine & Kuteva 2006), i.e. through a series of diachronic processes that leave various signals in the long-term lineage of each language in the area. These signals are precisely what the traditional approach removes. Worse, by completely putting aside diachrony, traditional methods cannot distinguish between frequency signals that are recent local events in a sub-area (e.g. local effects inside Meso-America rather than larger-scale New World patterns) and signals that are much older because they affect not just the synchronic results but the history of the families in the areas, including effects on the proto-languages of these families and, by extrapolation, effects on the families that these proto-languages were daughters of.

A related problem arises with isolates. With its lack of a diachronic dimension, the traditional approach takes the typological features of isolates at face value. From a diachronic point of view, however, these features are particularly hard to estimate: a given feature may be what is inherited, or it may be a recent innovation. For example, imagine a world in which Hindi is the sole survivor of Indo-European: under a traditional approach we would infer that having nominal gender and displaying split ergativity were the preference in the family, perhaps caused by contact in its areal location. If English is the sole survivor, we would conclude the opposite — although of course the probability of surviving has nothing to do with the presence or absence of these features. Without any estimate on diachronic trends, we cannot even begin to address this problem.

3 An alternative: exploring family biases

The solution of these problems requires a radical break with the traditional approach in typology. Instead of taking genealogical relationships as a confounding factor, an alternative is t take these relationships as the very basis on which to test areal hypotheses (or hypotheses on universals, for that matter). Proven relationships among a set of languages means by definition that this set derives from a common ancestor (understood here as a single language or as a dialect continuum). Given this, we can estimate the most likely diachronic processes that characterize the derivation of each language from the ancestor within a given family (Greenberg 1978, Maslova 2000, 2004, Dunn et al. 2011). If we find that such family-specific diachronic processes are biased in different directions inside vs. outside an area, this is then likely to reflect the effects of diffusion events in the area at a time depth at least as old as the families examined (cf. Problem

3 in the preceding section). Also, the estimated biases are neutral as to whether they arise from retention or innovation (Problem 2), and as to whether they arise from joint or parallel innovations (Problem 1): they simply summarize the overall diachrony in the family.

Estimating such biases is what I call the *Family Bias* method (Bickel 2011, 2013). There are two key challenges for such a method. The first challenge is that we need a suitable approach to estimating diachronic biases on the basis of the synchronic distribution within each family. The second challenge is that we often lack sufficiently rich within-family samples. This is partly due to the traditional practice in typology to aim at genealogical balancing, i.e. to sample only one or at best a few data-points per family. But to some extent the problem is factual: even in the largest databases available, for almost half of the known families we only know a single language, an 'isolate'.[3] I will take up these challenges in turn.

Before proceeding, however, a caveat is in order: I limit my attention in the following to evidence from discrete (categorical) typological data. Continuous variables are relatively rare in current typological research, and in many cases, such variables are ultimately based on what are discrete processes from a diachronic perspective. Take for example counts of consonants or degrees of synthesis. From a diachronic point of view, gaining or loosing consonants (either as single segments or as classes) or gaining or loosing inflectional categories are discrete events, similar in nature to the change between an OV and a VO constituent order. The process here is different from the continuous process that applies when a biological species gains body mass over time. That said, for cases where a typological variable is truly continuous, e.g. when it is based on large-scale measurements across many paradigms or within corpora, alternatives exist. For an application using bio-informatic models of continuous change (known as Brownian motion and Ornstein-Uhlenbeck models), see Witzlack-Makarevich et al. (in press).

3.1 Diachronic biases

Two classes of approaches to estimating diachronic biases are currently available: set-based and tree-based methods. They both make specific assumptions about diachrony and put specific requirements on the kind of data that they need, summarized in Table 1 (and explained in the following). Some of these assumptions and requirements are problematic, as we will see. This makes it advisable to use both methods in tandem, cross-validating results.

The *set-based method* conceptualizes a family as a set of related languages but makes no further assumption about whether this set originated in a tree-like fashion or through a reticulate network. Each language, or even each sub-system (say, a specific inflectional paradigm or a clause type) or each variety (say, a sociolect or dialect) is interpreted as an independent diachronic trial: given a certain inherited value (say, VO order), each language (or subsystem thereof) can change this value (to OV) or keep it, like in a

---

[3] AUTOTYP (2927 languages, Nichols et al. 2013): 47% isolates;
GLOTTOLOG (7701 languages, Hammarström et al. 2014, accessed 2015-01-26): 44% isolates.

	Set-based	*Tree-based*
Family model	tree, wave, linkage, network	tree, or trees with uncertainty
Stochastic process	independent multinomial trials	Continuous-Time Markov Chain
Data requirement	none	non-constant
Family requirement	none	topology and branch lengths

Table 1: Two approaches to estimating diachronic biases in discrete variables within a given family: model assumptions and requirements (as explained in the main text)

standard multinomial (here, binomial) trial. A bias in a family can then be detected by using a multinomial (here, binomial) test. If such a test shows that a specific value (e.g. VO order) is significantly more frequent than the other values within a family, this can be interpreted as the result of a diachronic bias: either the family had this value from the beginning (e.g. the proto-language had VO order) and preferentially kept it, or the value was innovated early and then preferentially kept, or the value was innovated often in most branches and then preferentially kept, or there was any combination of these possibilities (Bickel 2013). It doesn't matter which option was true, or whether was a combination of these options, as every case reflects the same overall preference for innovation and/or retention. The method is equally applicable to families with constant (non-varying) values and families with non-constant (varying) value distributions, and it needs no information on the internal structure of the family (as a tree or a network). The disadvantage is of course that no such information can be used, even when it is available or can be estimated.

The *tree-based* method has its key strength in precisely using such information. The disadvantage is that the method is limited by definition to families that have evolved in a tree-like fashion (which may not be the case: François 2014), or that at least can be successfully modeled by tree-like evolution (e.g. by simultaneously assuming many different topologies compatible with the data). The tree-based approach is borrowed from biological phylogenetics and models linguistic change as a Continuous-Time Markov Chain along the branches of a tree. The method establishes diachronic biases by estimating the transition rate matrix of the Markov Chains from the data, either via a Maximum Likelihood (ML) or a Bayesian Markov Chain Monte Carlo (MCMC) approach. There is then evidence for a diachronic bias towards some value, say OV order, if a model fits better that assumes a higher rate for transitions towards that value (OV order) than a model that assumes equal rates. Model comparison is usually done via likelihood ratio tests or Bayes Factor assessments (Pagel 1999, Pagel et al. 2004, Dunn et al. 2011, Cysouw 2011).

Currently available transition rate estimators require data to be varying (non-constant), and this is sometimes problematic as typological data are often constant within smaller and mid-sized families (e.g. all of Dravidian has OV order, all of Tai-Kadai has VO order).[4] Another important requirement of the tree-based method is that we know, or can estimate, the length of branches between the nodes in a tree. These lengths define the number of opportunities for transitions in the Markov Chain. While such information can be gained for well-studied families from data on lexical and other idiosyncratic change, in most cases we often only have at best a vague idea about the raw tree topology. But the case is not lost. A reasonable way out is to simply set all lengths between two nodes to the same value (e.g. to 1). Doing so makes the assumption that each major type change (such as between VO and OV) requires the birth of a new language, i.e. a node, and that non-cladogenetic (anagenetic) change elsewhere in the tree — specifically, lexical change — have no impact on the number of opportunities for typological change. This assumption is consistent with the research tradition in historical linguistics, where one generally expects typological change (e.g. in word order) to be independent of whether a language is lexically conservative or underwent heavy cognate replacement. Indeed, especially in contact-induced processes, typological change is expected to be driven by different scenarios than lexical change (Thomason & Kaufman 1988). The assumption is also consistent with the observation that typological change tends to be characterized by punctuated evolution rather than gradual anagenetic change (Dediu & Levinson 2012). A possible problem, however, is that we might underestimate earlier branch lengths just because we don't know additional nodes that might have played a role in the past but cannot be reconstructed because all languages that split off died out without a trace.[5] A promising solution here is to experiment with alternative branch estimates and compare results (Dediu 2015).

Software for both set-based and tree-based methods is available. The set-based method can be accessed through the `R` package `familybias` (Zakharko & Bickel 2011ff) that also handles isolates (to which we will turn immediately). Tree-based methods are far better supported because they have a long tradition in biology. A Maximum Likelihood version is available in the `R` package `geiger` (Harmon et al. 2008), a Bayesian version in `BayesTraits` (Pagel & Meade 2013) and several other software packages.

3.2 Isolates and small families

Estimating diachronic biases in families requires that we have sufficient data-points within each family. As noted above, this is not the case for almost half of the families we know. A way out is to combine information from large and small families and arrive at an extrapolation estimate (Bickel 2013). From large families we can derive an estimate of the probability that a family shows a bias, in whatever direction. One way to arrive at this estimate is to apply Laplace's Rule of Succession, which avoids unrealistic $Pr(\text{bias}) = 1$: if we see 6 out 10 families in an area to be biased (some towards OV, some towards VO), we expect the next family to be biased with $Pr(\text{bias}) = \frac{6+1}{10+2} = .58$, even if we can-

---

[4] A survey of the data described in Section 4.1 for families with at least 5 members shows that in 55% of the variables, at least one quarter of the families have constant values, and in 22% of the variables, at least half of the families have constant values.

[5] I am grateful to Gerhard Jäger for bringing this up.

not directly examine the family because we know only one or two of its representatives. We can then draw random samples of small families from a binomial distribution with $Pr(\text{bias})$ and declare them as the sole survivors of larger unknown families with a bias. Some of these survivors will be representative of the bias in their unknown larger family, and so we can take their structural choice (e.g. OV order) to reflect the direction of the bias in this family.[6] However, some survivors might be deviates, e.g. showing VO order although the family as a whole was biased against this. We estimate the probability of deviation, $Pr(\text{deviation})$, from the strength of the bias in large families, so for example a mean 90% bias in large families would give $Pr(\text{deviation}) = .1$. Then, we again draw random samples of deviating members from a binomial distribution with $Pr(\text{deviation})$. The resulting samples of deviates are assigned the opposite value of what they actually have.

The two cases of random sampling (with $Pr(\text{bias})$ and $Pr(\text{deviation})$) during these extrapolations incur an error, but the error follows a normal distribution. By performing the extrapolations many times (say 1,000) we can take the mean of these extrapolations as our estimate of bias directions in small families.

This is one of several imaginable ways of extrapolating to small families and isolates, but so far the only one that is implemented (Bickel 2013, Zakharko & Bickel 2011ff). An important parameter is the threshold of what counts as a small vs. a large family, e.g. whether this is two or six languages. If this parameter is set at too low a number, the estimates of directions of biases in large families become unreliable; if the parameter is set at too high number, there are not enough large families per area and so the extrapolation parameters $Pr(\text{bias})$ and $Pr(\text{deviation})$ become unreliable. Experiments so far suggest that a reasonable threshold is that large families have $N \geq 5$ members.[7]

One important implication of the threshold problem is that for some hypotheses, the family bias method may not be applicable: if an area is small and there are only a couple of large families in it, extrapolation is very problematic. But as noted in Section 2, traditional approaches perform reasonably well already for such case, and many of the problems that the family bias approach seeks to solve do not arise to begin with.

Furthermore, family bias estimates require that we sample families as densely as possible, with at least half a dozen datapoints per family. When this is not possible, traditional approaches are the only viable option at present.

3.3 Inference

Family bias estimates suggest for each family whether its diachrony was shaped by a bias in a certain direction (say, towards OV or towards VO), or whether no bias is detectable. The latter case is uninformative, as it is compatible with conflicting scenarios: perhaps there was a bias, but the family is not old enough for a signal to emerge, and if so, the

---

[6] If a small family has beetween one and a few members, we can pick the majority value and make a random choice in the case of ties.

[7] The software implementation of the extrapolation algorithm reports credibility intervals for the extrapolation parameters, and these can be used to determine a reasonable threshold.

bias could be in any direction; or the family was old and diversified randomly. To the extent that these possibilities cannot be separated, inference on family biases needs to be based on families where biases are detected.

Where we have estimates on bias directions for many families, we can compare families inside with families outside an area of interest. In the simplest case, we can compare the counts of bias directions per area, e.g. the number of families biased towards OV vs. the number of families biased towards VO inside vs. outside Eurasia. A χ^2 or Fisher test can then be used to assess the evidence for a difference between areas. For assessing specific bias directions, Pearson residuals can be examined and subject to a permutation test (as implemented e.g. in the `R` package `vcd` by Meyer et al. 2009).[8] When the Pearson residual of certain bias direction (e.g. towards OV order) is significant in an area (e.g. in Eurasia), this means that this bias direction was driven by contact in the area. If a significant residual is negative, contact has lead to a decrease of the bias direction in the area; if the residual is positive, contact has lead to an increase in the area. If a residual in an area is not significant under a permutation test, this means that the bias directions in this area have been close to what is expected under the null hypothesis of equal distributions. If there is nevertheless an overall signal for the difference between two areas or between families e.g. inside vs. outside an area, one possibility is that the signal comes from significant residuals in the other area: for example, if there is an overall difference in biases towards case marking inside vs. outside Eurasia and the residual outside Eurasia is not significant, it may be significant inside Eurasia. The other logical possibility is that the overall difference stems from small differences between the areas that cannot be picked up when testing the residuals individually.

In some cases, it will be interesting to control for possibly universal dependencies. Areal differences in family bias direction sometimes emerge only once universal dependencies are controlled for: for example, it turns out that left-branching NPs are more strongly associated with left-branching VPs (OV) in Eurasia than outside Eurasia. Thus, the association between NP and VP order is universal, but its strength shows an additional signal of areality (Bickel in press). A suitable framework for inference about such signals is log-linear modeling. In this approach one explores the fit of models that try to predict the observed frequencies of biases given a series of typological and areal predictors, e.g. models that try to predict the biases in NP order from predictors of VP order and Eurasian areality. Through systematic comparison of model fit one can then find the best-fitting model (Baayen 2008, Bickel 2013, 2015, in press).

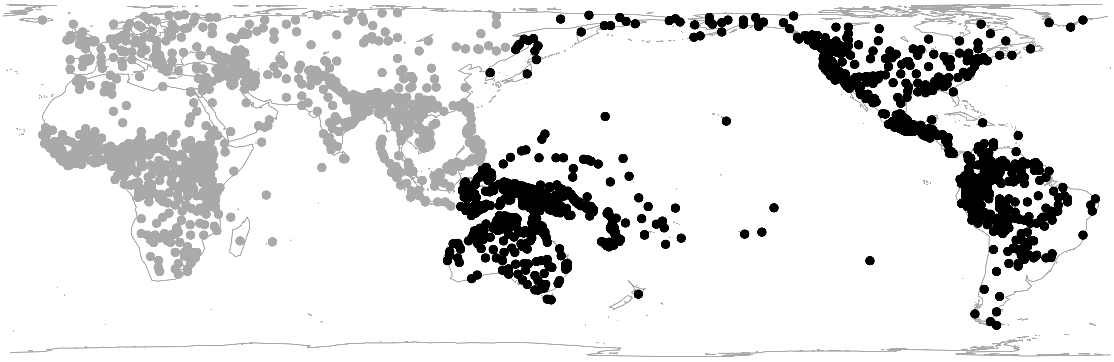
4 Case study: the Trans-Pacific area

The Trans-Pacific area was first proposed by Nichols (1992) and is shown in Map 2. Its motivation lies in the population history of the Americas out of the same East Asian

---

[8] Pearson residuals report the extent to which observed counts are higher (positive residuals) or lower (negative residuals) than what is expected under the null hypothesis of independence, here between areas and bias directions.

migration routes that were also at the source of the earlier settlement of New Guinea and Australia. While the details of these routes deserve a new review in the light of recent advances in population genetics (e.g. Rasmussen et al. 2011, 2014, Raghavan et al. 2014, 2015, Skoglund et al. 2015), I focus here on the methodological problem of detecting possible signals of shared linguistic ancestry in the Trans-Pacific area, leaving an in-depth discussion of the population history to another occasion.



Map 2: The Trans-Pacific area (Bickel & Nichols 2006) in the AUTOTYP database (Nichols et al. 2013). *Projection*: cylindrical equal area

Bickel & Nichols (2006) used a traditional method of genealogically balanced sampling from data on 100 variables from the *World Atlas of Language Structures* (WALS). After controlling for family-wise errors within re-codings of the same variable (e.g. various versions of coding verb-final vs. verb-initial structures), the study revealed about 40 variables with a significant difference in frequencies within vs. outside the Trans-Pacific area (as assessed by permutation tests). The false discovery rate (q) for these stays below .05 (as evaluated by the method in Dabney & Storey (2014)).

This is a surprisingly strong signal, but it is unclear to what extent this is an artefact of the method used. The results depend on which languages of a family happen to be sampled as representative, and they may miss more subtle trends in the history of the families in the area, either in support of or against the hypothesis. To evaluate these possibilities, I submitted an expanded typological database to a family bias estimate.

4.1 Data and Methods

I combined the AUTOTYP (Bickel & Nichols in prep.) and WALS (Comrie et al. 2005) databases and extracted variables with $N \geq 250$ languages and $k < 10$ levels per variable, ensuring that family bias estimates can be performed reasonably well. This resulted in 356 variables coded in between 250 and 1370 languages. Most variables have only few levels. Some are count variables, but based on the discussion above (cf. Section 3), I treated them here as discrete variables.

Some variables contain multiple values per language (e.g. ergative alignment in the past, accusative alignment in the nonpast). Multiple values of this kind are treated here as independent diachronic offshoots of a language. This is in line with the traditional assumption that splits arise because sub-systems can develop independently of each other.

Some variables come in multiple versions. For example, the locus of marking variable (Nichols & Bickel 2005) comes both in a version with a rough binary head-marking vs. dependent-marking coding and in a version with more fine-grained distinctions (distinguishing e.g. between cliticization on a head and genuine inflectional head-marking); and there are multiple ways in which one can define the argument role for which locus of marking is coded (e.g. identify 'objects' or 'P' arguments by syntactic or semantic criteria). We cannot (and should not) anticipate *a priori* which of these was most strongly affected by areal diffusion, and so I entered all such versions into the family bias estimations. But different versions are likely to correlate among each other. To some extent, such dependencies can be held in check by estimating the false discovery rate (Storey & Tibshirani 2003), but it is important to also control for dependencies qualitatively when interpreting the results. The same holds for correlations that are suspected on the basis of received theories, as is the case for the Greenbergian word order correlations. At the same time, however, we need to bear in mind that areal diffusion can also have effects on the strength of these correlations, independent of their universal validity (cf. Section 3.3).

For estimating family biases I used both set-based and tree-based methods as input to the extrapolation algorithm described above. However, as noted before, the tree-based method can only be used for families with non-constant value distributions. For extrapolation in these cases, I added the set-based results for families with constant value distributions, assuming that an equal-rate model is implausible for a family with an extreme bias towards a single value (e.g. with all daughter languages having OV order). In all cases, I set the threshold for what counts as a large family at $N \geq 5$.

For the set-based method, I relied on the algorithm in Zakharko & Bickel (2011ff), using a multinomial test with a rejection level of $\alpha = .1$ as the criterion for a bias. For the tree-based method, I used the `BayesTraits` package (Pagel & Meade 2013), declaring a family as biased if a different-rates model fits better than an equal-rates model. The difference in data fitting was assessed by a likelihood ratio test with $\alpha = .1$ for ML estimates and by a Bayes Factor comparison with a $BF > 2$ criterion for MCMC estimates.

These α and BF criteria are relatively liberal, but at least for the set-based method I found that a lower α has very weak power and incurs discreteness problems with families having between five and ten members, which is a frequent size range (Bickel 2013).

The combined dataset is available online (Bickel & Nichols in prep.)[9] and an `R` script performing all analyses is available in the Supporting Online Material.[10]

---

[9] The present study is based on a draft version of the database. The public release of the revised AUTOTYP data is planned during 2016.

[10] I thank Taras Zakharko for help with the script.

4.2 Results and Discussion

Figure 1 compares the overall results across the different methods used for family bias estimates. The results are captured by the p-value of Fisher Exact tests used for detecting frequency differences in bias directions between families in the Trans-Pacific area and families outside this area.[11] The results from the set-based and the MCMC (Markov-Chain Monte Carlo) version of the tree-based methods are very similar, while the ML (Maximum Likelihood) version of the tree-based method is more conservative.

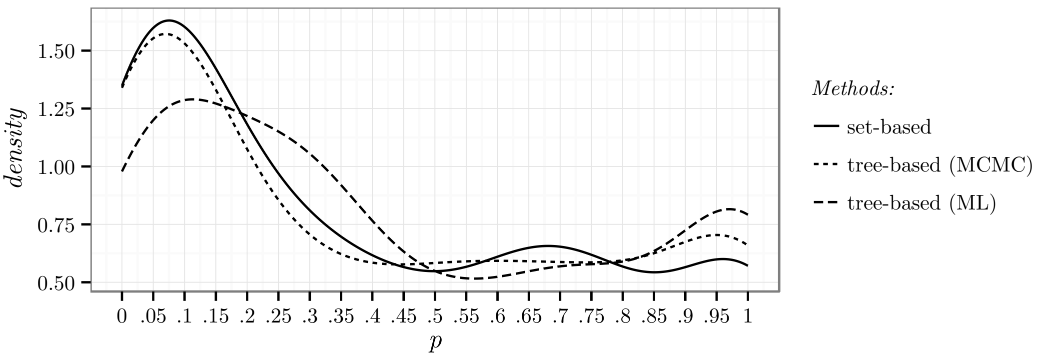


Figure 1: Comparison of results across methods: density distribution of p-values from a Fisher Exact test on the difference between the Trans-Pacific area and the rest of the world.

The convergence between the set-based and the MCMC approach inspires confidence in the results from these two methods, especially since they make completely different theoretical assumptions (cf. Table 1) and rely on a radically different estimation algorithm. The divergence of the ML approach needs interpretation in a separate study. For now, I proceed chiefly with the results from the set-based and MCMC approaches.

The table in the Appendix lists all variables for which the set-based method has a p-value below .05. The risk of false rejections of the null hypothesis that arises from multiple testing can be assessed by false discovery rate (q) estimates. For this, I used the bootstrap method implemented by Dabney & Storey (2014). The results are summarized in Table 2. What this summary suggests is that for set-based and MCMC-based results about 16 – 17% variables with $p < .05$ are false discoveries — possibly, but not necessarily the ones at the end of the table in the Appendix, i.e. those with a high p-value. Given the expected weakness of the signal, such a false discovery rate is bearable. Lower rates risk a loss in statistical power. At any rate, pushing the rate below 10% (last column in Table 2) leaves about half of the variables. The results for the ML-based estimates are much weaker.

The table in the Appendix suggests that about 28 (set-based results) and 21 (MCMC) of the significant variables are variants of each other (annotated as such in the last

---

[11] An important question (raised by Damián Blasi) is whether the p-values here simply depend on sample size. Local (LOESS) regression plots of p-values against samples size reveal no linear trend whatsoever.

	Significant at $\alpha < .05$	q at that level	Significant at $q < .1$
Set-based	73	0.16	28
Tree-based (MCMC)	69	0.17	40
Tree-based (ML)	47	0.26	19

Table 2: Counts of variables where family bias directions differ between the Trans-Pacific area and the rest of the world. q represents the false discovery rate as estimated by the bootstrap method implemented by Dabney & Storey (2014).

column). Subtracting these from the figures in Table 2 leaves us with about 46 significant effects (45 in the set-based and 48 in the MCMC results). Of these, about 39 (38 in the set-based and 40 in the MCMC results) are true signals under the false discovery rates of 16 – 17% (Table 2). These results confirm the findings from Bickel & Nichols (2006), but on a firmer methodological basis and with much more systematic controls for false discovery rates and the effects of recoded variables.

In addition, the present study allows an analysis of Pearson residuals (as explained in Section 3.3 which gives more detailed insights into the relevant areal processes. The bias directions with significant residuals are listed in the table in the Appendix. There are substantially more significant characteristics of the families outside (86% in the set-based and 77% in the MCMC results) than of the families inside the Transpacific area (27% in the set-based and 28% in the MCMC results).

This is consistent with the suggestion that the Trans-Pacific area is not so much characterized by the spread of biases within the area, but rather by the retention of earlier bias distributions that were more diverse (Bickel & Nichols 2003). Some family bias directions do seem to show a signal from areal diffusion, e.g. biases against voicing distinctions or towards possessive prefixes, but in most cases the Trans-Pacific area shows a roughly uniform internal distribution of bias directions, with no privileged patterns. This is different from the world outside the Trans-Pacific area where the residual analysis reveals a large number of heavily skewed bias directions, i.e. the area is characterized by many family biases that are significantly more or significantly less common than expected under the null hypothesis of no areality. If a bias is more common than expected (‘+’ in the table in the Appendix), the families in the area systematically evolved towards a given feature; if the bias is less common than expected (‘–’ in the table), the families systematically evolved against a given feature.

Wherever we find significant residuals, the patterns seem to have been driven by contact during the history and pre-history of the families, e.g. biases towards tone or against head-marking. From inspecting maps, however, it appears likely that the patterns represent various different contact histories in large sub-regions, such as within Eurasia or within Africa. This invites further research on the detailed historical scenarios that are at the root of the patterns we find.

5 Conclusions and prospects

This chapter proposed a systematic method for testing hypotheses on large ancient areas. By estimating likely diachronic biases in families, the method can capture effects that were operative during the time depth of known families and their immediate ancestors: if there are systematic trends in family bias directions, and they differ between areas, these different directions are likely to have been triggered by events that affected the immediate ancestors of the families and their subsequent development. This gives us a time frame of over at least 10ky. In the case study above, the method picked up many signals from contact events in Eurasia and Africa during such a time frame.

However, if family bias signals statistically characterize an area as wide as the Trans-Pacific, i.e. an area that has not been been continuously connected through area-wide contact events during the past 10ky, the signals are likely to stem from much older contact events. In the case of the Trans-Pacific area, for example, it is possible that the few family bias directions that we found to encompass the entire area, from Australia to South America, arose in Asia 50-60ky BP.

While these possibilities need to be examined in separate work, I would like to conclude by pointing to two methodological desiderata. First, the tree-based and the set-based approaches are both radical in using all or no information from tree-like diachronies. But neither method uses information from innovation waves, as understood in dialectology but also relevant for entire families (François 2014). What is needed, then, is a third, more general approach that can make use of any kind of diachronic model. Second, I have concentrated in this paper on hypothesis testing, but we clearly also need exploratory clustering methods. None of the available methods here incorporates the diachronic dynamics of families. What we need here is approaches that directly cluster diachronic trends, not languages. This is another area where much methodological progress can and should be made.

Appendix: Variables with a significant difference in family biases, with $p < .05$ for at least the set-based or the MCMC tree-based estimates (ordered by the the set-based p-values). Plus and minus signs indicate Pearson residuals that are significant under a permutation test with a .05 rejection level. Missing p-values mean that family bias estimates failed for statistical or computational reasons. A = AUTOTYP, W = WALS

Variable	*Source*	N(lang.)	p(sets)	p(MCMC)	p(ML)	*Trans-Pacific*	*Other*	*Variant of*
DRYPOS0	W	591	0.0000	0.0000	0.0000	+poss pf;-poss sf	-poss pf;+poss sf;-both	DRYPOS
MADVOI2	W	565	0.0000	0.0000	0.0001	-voicing;+no voicing	+voicing;-no voicing	MADVOI
DRYPOS	W	794	0.0000	0.0009	0.0069		-poss pf;+poss sf	DRYPOS
MADVOI	W	565	0.0000	0.0099	0.0128	-voi in plos/fric;+no voi	+voi in plos/fric;-no voi	MADVOI
DRYGEN	W	1102	0.0001	0.0042	0.0014	-NGen	+NGen	DRYGEN
MADLAT2	W	565	0.0002	0.0001	0.0002	+no lat	-no lat	MADLAT
BAKADP2	W	377	0.0002	0.0002	0.0009	+no adp	-no adp	BAKADP
DRYGEN0	W	1020	0.0002	0.0002	0.0001	-Nnp	-npN;+Nnp	DRYGEN
MADLAT	W	565	0.0002	0.0227	0.0041	-/l/no obstr;+no lat	+/l/no obstr;-no lat	MADLAT
POLYAGR.STRICT	A	331	0.0004	0.0004	0.0018	-absent;+present	+absent;-present	
DRYDEM0	W	1011	0.0004	0.0006	0.0017	+DemN;-NDem	-DemN;+NDem	DRYDEM
MADPRS	W	565	0.0006	0.0000	0.0019		+lab.velar	
LOCUS.POSS.BROAD	A	278	0.0006	0.0129	0.2974	+H	-H	LOCUS.POSS
MADTON02	W	525	0.0008	0.0007	0.0029	+atonal;-tonal	-atonal;+tonal	MADTON
HASWAN03	W	269	0.0011	0.0006	0.0055	+some desid	-some desid	HASWAN
LOCUS.POSS.AGGR	A	271	0.0014	0.0031	0.2974		-H	LOCUS.POSS
SIMPLE.LOCUS.S.BROAD	A	282	0.0017	0.0049	0.1770		+D;-H	
DRYSOV3	W	1226	0.0020	0.0015	0.0509	-other	-final/free;+other	DRYSOV
LocPOSSdm	A	282	0.0024	0.0024	0.0480		+present;-absent	LOCUS.POSS
ANDANG2	W	467	0.0027	0.0020	0.0038	-ŋ;+no ŋ	+ŋ;-no ŋ	ANDANG
DRYDEM	W	1084	0.0027	0.1625	0.0873	+Dem.wordN	-Dem.wordN;+NDem.word	DRYDEM
HASNPL22	W	290	0.0049	0.0049	0.0189	-pl anim;+no/opt pl	+pl anim;-no/opt pl	HASNPL
LocPOSShm	A	282	0.0059	0.0059	0.0084	-absent;+present	+absent;-present	LOCUS.POSS
DRYXOV	W	448	0.0081	0.0737	0.1270		+VOX	
DRYSOV2	W	1226	0.0083	0.0151	0.0064		-mixed;+medial	DRYSOV
BROFIN	W	590	0.0092	0.0092	0.0048		-hand=finger	
VP	W	1276	0.0101	0.0077	0.0335		-OV;+VO	DRYOBV
DRYOBV	W	1367	0.0106	0.0106	0.0357		+VO	DRYOBV
ANDANG	W	467	0.0128	0.1525	0.1943		-ŋ	ANDANG
ANY.V.AGR	A	349	0.0136	0.0136	0.0411		+absent	
Nr.POSS.CLASS	A	262	0.0139	0.0418	0.2509		-1	POSS.CLASSES
Nr.INALIEN.CLASS	A	255	0.0143	0.0801	0.3738			
CORASS	W	256	0.0145	0.0168	0.0982		+sem and formal	
LocPhm	A	294	0.0148	0.0105	0.0067		+absent;-present	LOCUS.P
CASE.ALIGN.PER.LANG.N	A	345	0.0152	0.0104	0.0102			
SIMPLE.LOCUS.P.NARROW	A	286	0.0153	0.0438	0.2594		-H	LOCUS.P
SIMPLE.LOCUS.P.BROAD	A	290	0.0157	0.0236	0.2594		-H	LOCUS.P

(continued)

Variable	*Source*	N(lang.)	p(sets)	p(MCMC)	p(ML)	*Trans-Pacific*	*Other*	*Variant of*
BAKADP	W	377	0.0162	0.0138	0.4332	+no adp	-no adp	BAKADP
INAL.CLASS	A	255	0.0188	0.0188	0.1841		+absent	INAL.CLASSES
Nr.ALIEN.CLASS.BINNED	A	251	0.0188	0.0188	0.0937		+none	
Nr.INAL.CLASS.BINNED	A	255	0.0188	0.0395	0.0985		+none	INAL.CLASSES
HASDIT0	W	339	0.0192	0.0117	0.0424	+SO	-SO	
DRYSOV4	W	1054	0.0193	0.0091	0.0161		+non-final	DRYSOV
LOCUS.ARG	A	294	0.0194	0.0265	0.0194		+absent	
LOCUS.P.BROAD	A	290	0.0195	0.1023	0.1023		-H	LOCUS.P
ANY.V.AGR.STRICT	A	281	0.0204	0.0297	0.0308		+absent	
HASNPL0	W	262	0.0224	0.3822	0.1851		+pl oblig	HASNPL
MADVOW	W	562	0.0228	0.0198	0.0999		+3	
MADVOWC	W	562	0.0228	0.0674	0.1263		+(7-14)	
AL.CLASS.SEM.TYPE	A	295	0.0229	0.0366	0.0366		-formal;+neutral	
MADGLO	W	565	0.0229	0.0723	0.0319		+impl only	
MADTON	W	525	0.0248	0.0064	0.0386			MADTON
HASWAN02	W	269	0.0256	0.0072	0.0361		-desid strategy	HASWAN
HAAEVD	W	416	0.0259	0.0686	0.1017			HAAEVD
LocAorPhm	A	294	0.0265	0.0265	0.0265		+absent	
N.PLURAL	A	258	0.0286	0.0286	0.0286		-absent	
CASE.ALIGN.PER.LANG.P	A	668	0.0311	0.0145	0.0873		+acc	
SIEPAS	W	372	0.0319	0.0891	0.0275		-no pass;+pass	
SIEGEN	W	377	0.0332	0.0214	0.6751			
DRYPRO2	W	644	0.0345	0.1177	0.0877		+oblig.Spro	DRYPRO
DRYCOQ0	W	782	0.0361	0.0320	0.0347		-Q oblig initial	DRYCOQ
DRYCOQ	W	802	0.0368	0.0592	0.0422			DRYCOQ
DRYSOV1	W	1226	0.0372	0.0094	0.0217			DRYSOV
HASWAN	W	283	0.0381	0.0330	0.1917			HASWAN
LOCUS.P.NARROW	A	286	0.0396	0.1116	0.1023			LOCUS.P
POSS.CL.AGGR.PER.LANG	A	268	0.0408	0.0379	0.1121		+absent;-present	
POSS.CL.ANY.PER.LANG	A	275	0.0420	0.0420	0.1121		+absent;-present	
DRYADP0	W	992	0.0425	0.0665	0.1832			DRYADP
DRYPRO3	W	583	0.0432	0.0488	0.0423		+oblig.Spro (bin)	DRYPRO
PP	W	985	0.0446	0.0495	0.1264		+Pnp	DRYADP
HASNPL	W	290	0.0458				+pl oblig	HASNPL
POSS.CLASS.AGGR	A	274	0.0469	0.0698	0.1121		-present;+absent	POSS.CLASSES
POSS.CL.PER.LANG	A	257	0.0480	0.0650	0.2693		-1	
DUAL	A	476	0.0512	0.0002	0.0002			
HAAEVD2	W	416	0.0631	0.0338	0.0268		-some evid	HAAEVD
A.P.ORDER.AGGR	A	1370	0.0749	0.0036	0.0480		-mixed	

(continued)

Variable	*Source*	N(lang.)	p(sets)	p(MCMC)	p(ML)	*Trans-Pacific*	*Other*	*Variant of*
POSS.CLASS.ALL	A	282	0.0749	0.0432	0.1961			POSS.CLASSES
CASE.ALIGN.WALS.STYLE.N	A	310	0.1006	0.0265	0.1140			
CASE.ALIGN.WALS.STYLE.P	A	379	0.1220	0.0390	0.0390			
S.ALIGN.PRO.AGGR	A	411	0.1224	0.0120	0.0557			
SIMPLE.LOCUS.A.BROAD	A	291	0.1354	0.0343	0.2856			LOCUS.A
SIMPLE.LOCUS.A.NARROW	A	289	0.1354	0.0464	0.2856			LOCUS.A
LOCUS.A.AGGR	A	282	0.3433	0.0184	0.0542			
BAKADP02	W	314	0.6079	0.0475	0.7098			BAKADP
SONNON	W	310	0.6616	0.0289	0.8875			
AUWHOR	W	375	0.7126	0.0112	0.3260			

References

van der Auwera, J. 1998. Revisiting the Balkan and Meso-American linguistic areas. *Language Sciences* 20. 259–270.

Baayen, R. H. 2008. *Analyzing linguistic data: a practical introduction to statistics*. Cambridge: Cambridge University Press.

Bentz, C. & B. Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3. 1–27.

Bickel, B. 2011. Statistical modeling of language universals. *Linguistic Typology* 15. 401–414.

Bickel, B. 2013. Distributional biases in language families. In B. Bickel, L. A. Grenoble, D. A. Peterson & A. Timberlake (eds.), *Language typology and historical contingency*, 415–444. Amsterdam: Benjamins.

Bickel, B. 2015. Distributional typology: statistical inquiries into the dynamics of linguistic diversity. In B. Heine & H. Narrog (eds.), *The Oxford handbook of linguistic analysis, 2nd edition*, 901 – 923. Oxford: Oxford University Press.

Bickel, B. in press. Areas and universals. In R. Hickey (ed.), *The Cambridge Handbook of Areal Linguistics*, Cambridge: Cambridge University Press [pre-print available at `http://www.ivs.uzh.ch/bickel-files/papers/Bickel2015Areas_and_universals-preprint.pdf`].

Bickel, B., K. Hildebrandt & R. Schiering. 2009. The distribution of phonological word domains: a probabilistic typology. In J. Grijzenhout & B. Kabak (eds.), *Phonological domains: universals and deviations*, 47–75. Berlin: Mouton de Gruyter.

Bickel, B. & J. Nichols. 2003. Typological enclaves. Paper presented at the 5th Biannual Conference of the Association for Linguistic Typology, Cagliari, September 18; available at `http://www.autotyp.uzh.ch/download`.

Bickel, B. & J. Nichols. 2006. Oceania, the Pacific Rim, and the theory of linguistic areas. *Proc. Berkeley Linguistics Society* 32. 3 – 15.

Bickel, B. & J. Nichols. in prep. The AUTOTYP database. Electronic database; `http://www.autotyp.ch`.

Bickel, B., A. Witzlack-Makarevich & T. Zakharko. 2014. Typological evidence against universal effects of referential scales on case alignment. In I. Bornkessel-Schlesewsky, A. Malchukov & M. Richards (eds.), *Scales: a cross-disciplinary perspective on referential hierarchies*, 7–43. Berlin: De Gruyter Mouton.

Comrie, B., M. S. Dryer, D. Gil & M. Haspelmath (eds.). 2005. *The World Atlas of Language Structures Online*. Oxford: Oxford University Press.

Comrie, B. & T. Kuteva. 2005. Relativization strategies. In B. Comrie, M. S. Dryer, D. Gil & M. Haspelmath (eds.), *World atlas of language structures*, 494–501. Oxford: Oxford University Press.

Cysouw, M. 2011. Understanding transition probabilities. *Linguistic Typology* 15. 415–431.

Cysouw, M. & B. Comrie. 2009. How varied typologically are the languages of Africa? In R. Botha & C. Knight (eds.), *The cradle of language, vol 2: African perspectives*, 189–203. Oxford: Oxford University Press.

Dabney, A. & J. D. Storey. 2014. `qvalue`: Q-value estimation for false discovery rate control. `R` package, version 1.40.0.

Dediu, D. 2015. `lgfam-newick`: Language family classifications as Newick trees. Repository at [`https://github.com/ddediu/lgfam-newick`].

Dediu, D. & R. D. Ladd. 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, *ASPM* and *Microcephalin*. *Proceedings of the*

National Academy of Sciences of the USA 104. 10944–10949.

Dediu, D. & S. C. Levinson. 2012. Abstract profiles of structural stability point to universal tendencies, family-specific factors, and ancient connections between languages. *PloS One* 7. e45198.

Donohue, M. & B. Whiting. 2011. Quantifying Areality: A Study of Prenasalisation in Southeast Asia and New Guinea. *Linguistic Typology* 15(1). 101–121.

Dryer, M. S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13. 257–292.

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

Enfield, N. J. 2005. Areal linguistics and Mainland Southeat Asia. *Annual Review of Anthropology* 34. 181–206.

Everett, C., D. E. Blasi & S. G. Roberts. 2015. Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences* doi:10.1073/pnas.1417413112.

François, A. 2014. Trees, waves and linkages: models of language diversification. In C. Bowern & B. Evans (eds.), *The Routledge handbook of historical linguistics*, 161–189. London: Routledge.

Gray, R. D., D. Bryant & S. J. Greenhill. 2010. On the shape and fabric of human history. *Philosophical Transactions of the Royal Society B* 365. 3923–3933.

Greenberg, J. H. 1978. Diachrony, synchrony and language universals. In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (eds.), *Universals of human language I: method and theory*, 61–92. Stanford: Stanford University Press.

Hammarström, H., R. Forkel, M. Haspelmath & S. Nordhoff. 2014. *Glottolog 2.3.* Leipzig: Max-Planck-Institute for Evolutionary Anthropology [`http://glottolog.org`].

Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor & W. Challenger. 2008. `geiger`: investigating evolutionary radiations. *Bioinformatics* 24. 129–131.

Haspelmath, M. 2001. The European linguistic area: standard average European. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds.), *Language typology and language universals: an international handbook*, 1492–1510. Berlin: Mouton de Gruyter.

Heine, B. & T. Kuteva. 2006. *The changing languages of Europe.* Oxford: Oxford University Press.

Jaeger, T. F., P. Graff, W. Croft & D. Pontillo. 2011. Mixed effect models for genetic and aral dependencies in linguistic typology. *Linguistic Typology* 15. 281–320.

Jakobson, R. 1931. K xarakteristike ėvrazijskogo jazykovogo sojuza. In *Selected Writings 1*, 144–201. The Hague: Mouton 1970.

Janssen, D., B. Bickel & F. Zúñiga. 2006. Randomization tests in language typology. *Linguistic Typology* 10. 419–440.

Köhli, M. 2013. *Quantitative Analyse des Zusammenhangs zwischen der globalen Sprachendiversität und geographischen Faktoren.* Department of Geography, University of Zürich MA thesis.

Lehmann, C. 1984. *Der Relativsatz.* Tübingen: Narr.

Masica, C. 2001. The definition and significance of linguistic areas: methods, pitfalls, and possibilities (with special reference to the validity of South Asia as a linguistic area). In P. Bhaskararao & K. V. Subbarao (eds.), *Tokyo Symposium on South Asian languages: contact, convergence, and typology [= The Yearbook of South Asian Languages and Linguistics 2001]*, 205–267. New Delhi: Sage Publications.

Maslova, E. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4. 307–333.

Maslova, E. 2004. Dinamika tipologičeskix raspredelenij i stabil'nost' jazykovyx tipov. *Vorprosy Jazykoznanija* 2004. 3–16.

McMahon, A. M. S. & R. McMahon. 2005. *Language classification by numbers.* Oxford: Oxford University Press.

Meyer, D., A. Zeileis & K. Hornik. 2009. `vcd`: visualizing categorical data. `R` package, `http://www.R-project.org/`.

Muysken, P. 2010. Scenarios for language contact. In R. Hickey (ed.), *The handbook of language contact*, 265 – 281. London: Wiley-Blackwell.

Muysken, P., H. Hammarström, J. Birchall, R. V. Gijn, O. Krasnoukhova & N. Müller. 2014. Linguistic areas: bottom-up or top-down? The case of the Guaporé-Mamoré. In B. Comrie & L. Golluscio (eds.), *Language Contact and Documentation / Contacto lingüístico y documentación*, 205–238. Berlin: Walter de Gruyter.

Nichols, J. 1992. *Linguistic diversity in space and time.* Chicago: The University of Chicago Press.

Nichols, J. 2003. Diversity and stability in language. In R. D. Janda & B. D. Joseph (eds.), *Handbook of Historical Linguistics*, 283–310. London: Blackwell.

Nichols, J. & B. Bickel. 2005. Locus of marking (in the clause; in possessive noun phrases; and whole-language typology). In M. Haspelmath, M. S. Dryer, D. Gil & B. Comrie (eds.), *The world atlas of language structures*, 98–109. Oxford: Oxford University Press.

Nichols, J., A. Witzlack-Makarevich & B. Bickel. 2013. The AUTOTYP genealogy and geography database: 2013 release. Electronic database, `http://www.uzh.ch/autotyp`.

Pagel, M. 1999. The maximum-likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology* 48. 612–622.

Pagel, M. & A. Meade. 2013. `BayesTraits V2`. Software and manual, `http://www.evolution.rdg.ac.uk/BayesTraitsV2Beta.html`.

Pagel, M., A. Meade & D. Barker. 2004. Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* 53. 673–684.

Raghavan, M., P. Skoglund, K. E. Graf, M. Metspalu, A. Albrechtsen, I. Moltke, S. Rasmussen, T. W. Stafford Jr, L. Orlando, E. Metspalu, M. Karmin, K. Tambets, S. Rootsi, R. Magi, P. F. Campos, E. Balanovska, O. Balanovsky, E. Khusnutdinova, S. Litvinov, L. P. Osipova, S. A. Fedorova, M. I. Voevoda, M. DeGiorgio, T. Sicheritz-Ponten, S. Brunak, S. Demeshchenko, T. Kivisild, R. Villems, R. Nielsen, M. Jakobsson & E. Willerslev. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505. 87–91. `http://dx.doi.org/10.1038/nature12736`.

Raghavan, M., M. Steinrücken, K. Harris, S. Schiffels, S. Rasmussen, M. DeGiorgio, A. Albrechtsen, C. Valdiosera, M. C. Ávila-Arcos, A.-S. Malaspinas, A. Eriksson, I. Moltke, M. Metspalu, J. R. Homburger, J. Wall, O. E. Cornejo, J. V. Moreno-Mayar, T. S. Korneliussen, T. Pierre, M. Rasmussen, P. F. Campos, P. d. B. Damgaard, M. E. Allentoft, J. Lindo, E. Metspalu, R. Rodríguez-Varela, J. Mansilla, C. Henrickson, A. Seguin-Orlando, H. Malmström, T. Stafford, S. S. Shringarpure, A. Moreno-Estrada, M. Karmin, K. Tambets, A. Bergström, Y. Xue, V. Warmuth, A. D. Friend, J. Singarayer, P. Valdes, F. Balloux, I. Leboreiro, J. L. Vera, H. Rangel-Villalobos, D. Pettener, D. Luiselli, L. G. Davis, E. Heyer, C. P. E. Zollikofer, M. S. Ponce de León, C. I. Smith, V. Grimes, K.-A. Pike, M. Deal, B. T. Fuller, B. Arriaza, V. Standen, M. F. Luz, F. Ricaut, N. Guidon, L. Osipova, M. I. Voevoda, O. L. Posukh, O. Balanovsky, M. Lavryashina, Y. Bogunov, E. Khusnutdinova, M. Gubina, E. Balanovska, S. Fedorova, S. Litvinov, B. Malyarchuk,

M. Derenko, M. J. Mosher, D. Archer, J. Cybulski, B. Petzelt, J. Mitchell, R. Worl, P. J. Norman, P. Parham, B. M. Kemp, T. Kivisild, C. Tyler-Smith, M. S. Sandhu, M. Crawford, R. Villems, D. G. Smith, M. R. Waters, T. Goebel, J. R. Johnson, R. S. Malhi, M. Jakobsson, D. J. Meltzer, A. Manica, R. Durbin, C. D. Bustamante, Y. S. Song, R. Nielsen & E. Willerslev. 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349(6250). doi:10.1126/science.aab3884. `http://www.sciencemag.org/content/349/6250/aab3884.abstract`.

Rasmussen, M., S. L. Anzick, M. R. Waters, P. Skoglund, M. DeGiorgio, T. W. Stafford Jr, S. Rasmussen, I. Moltke, A. Albrechtsen, S. M. Doyle, G. D. Poznik, V. Gudmundsdottir, R. Yadav, A.-S. Malaspinas, S. S. W. V, M. E. Allentoft, O. E. Cornejo, K. Tambets, A. Eriksson, P. D. Heintzman, M. Karmin, T. S. Korneliussen, D. J. Meltzer, T. L. Pierre, J. Stenderup, L. Saag, V. M. Warmuth, M. C. Lopes, R. S. Malhi, S. Brunak, T. Sicheritz-Ponten, I. Barnes, M. Collins, L. Orlando, F. Balloux, A. Manica, R. Gupta, M. Metspalu, C. D. Bustamante, M. Jakobsson, R. Nielsen & E. Willerslev. 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506. 225–229. `http://dx.doi.org/10.1038/nature13025`.

Rasmussen, M., X. Guo, Y. Wang, K. E. Lohmueller, S. Rasmussen, A. Albrechtsen, L. Skotte, S. Lindgreen, M. Metspalu, T. Jombart, T. Kivisild, W. Zhai, A. Eriksson, A. Manica, L. Orlando, F. M. De La Vega, S. Tridico, E. Metspalu, K. Nielsen, M. C. Ávila-Arcos, J. V. Moreno-Mayar, C. Muller, J. Dortch, M. T. P. Gilbert, O. Lund, A. Wesolowska, M. Karmin, L. A. Weinert, B. Wang, J. Li, S. Tai, F. Xiao, T. Hanihara, G. van Driem, A. R. Jha, F.-X. Ricaut, P. de Knijff, A. B. Migliano, I. Gallego Romero, K. Kristiansen, D. M. Lambert, S. Brunak, P. Forster, B. Brinkmann, O. Nehlich, M. Bunce, M. Richards, R. Gupta, C. D. Bustamante, A. Krogh, R. A. Foley, M. M. Lahr, F. Balloux, T. Sicheritz-Pontén, R. Villems, R. Nielsen, J. Wang & E. Willerslev. 2011. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* 334(6052). 94–98. doi:10.1126/science.1211177.

Roberts, S. & J. Winters. 2013. Linguistic Diversity and Traffic Accidents: Lessons from Statistical Studies of Cultural Traits. *PloS One* 8. e70902.

Schikowski, R. 2013. *Object-conditioned differential marking in Chintang and Nepali.* Zürich: University of Zürich PhD dissertation.

Skoglund, P., S. Mallick, M. C. Bortolini, N. Chennagiri, T. Hunemeier, M. L. Petzl-Erler, F. M. Salzano, N. Patterson & D. Reich. 2015. Genetic evidence for two founding populations of the Americas. *Nature* 525. 104–108. `http://dx.doi.org/10.1038/nature14895`.

Storey, J. D. & R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* 100(16). 9440–9445. doi:10.1073/pnas.1530509100. `http://www.pnas.org/content/100/16/9440.abstract`.

Thomason, S. G. & T. Kaufman. 1988. *Language contact, creolization, and genetic linguistics.* Berkeley: University of California Press.

Witzlack-Makarevich, A., T. Zakharko, L. Bierkandt, F. Zúñiga & B. Bickel. in press. Decomposing hierarchical alignment: co-arguments as conditions on alignment and the limits of referential hierarchies as explanations in verb agreement. *Linguistics* .

Zakharko, T. & B. Bickel. 2011ff. `familybias`: Family bias estimation. R package, `https://github.com/IVS-UZH`.