Understanding transition probabilities

Michael Cysouw, LMU München

cysouw@lmu.de

Abstract

Dunn et al. (2011) propose an innovative approach to establishing typological correlations. Their fascinating method exemplifies the dynamicization of linguistic typology. Instead of estimating relative frequencies of types, they switch to estimating transition probabilities between types. This commentary will present some background to better understand this approach, because an in-depth explanation of the underlying assumptions is unfortunately lacking in their rather short paper. This additional explanation is intended to clarify why the Dunn et al. paper deserves careful examination in the field of linguistic typology.

Notwithstanding the important methodological innovation of this paper, there are problems with their interpretation of the empirical results. Basically, and contrary to their claim, it does not seem to be the case that "most observed functional dependencies between traits are lineage-specific rather than universal tendencies". Obviously, there is variation between lineages, but no previously proposed word-order universal is contradicted by the variation attested in the Dunn et al. paper. For the current criticism, I will only use the data and results from their paper, so the problems presented here really concern the interpretation of the results and not the method as such.

Keywords

dynamic typology, transition probabilities, universals, comparative methodology

1. Galton’s problem

A central empirical objective of the field of linguistic typology is to estimate the relative frequency of a particular language type (or cross-section of types) among human languages. In the traditional approach, the *empirical* frequency of a type among the known world's languages is interpreted as indicative of the *universal probability* for this type. However, it has been obvious to all practitioners ever since the dawn of linguistic typology that there are many unwanted coincidental factors influencing these empirical frequencies, like genealogical relationship or areal convergence. To correct for such influences, there is a long tradition in typology of various kinds of sampling of the world's languages. Typically, only a single language per known genealogical (sub)grouping is included in the sample and a diverse geographical spread of the sampled languages over the whole world is preferred. Many variants of such sampling have been discussed in the typological literature over the last few decades (cf. Cysouw 2005 for a survey).

This problem is more widely known in statistics as autocorrelation, and more specifically as ‘Tobler’s Law’ in geography (Tobler 1970) and ‘Galton’s problem’ in anthropology (Naroll 1961). Sir Francis Galton originally raised this problem in the discussion after a presentation by Sir Edward B. Taylor (1889) on establishing cross-cultural correlations. In his reply to Galton’s objection, Taylor immediately suggested to use some sort of stratified sampling, the favourite solution ever since: “The difficulty raised by Mr. Galton that some of the concurrences might result from transmission from a common source, so that a single character might be counted several times from its mere duplication, is a difficulty ever present in such investigations […]. The only way of meeting this objection is to make separate classifications depend on well marked differences, and to do this all over the world” (Taylor 1889: 272).

Although sampling is an easy solution for removing the most glaring effects of autocorrelation, it is far from an ideal solution. First, sampling reduces the already rather limited amount of data available about the world’s language, so any generalisation has to be made on the basis of less than possible data (cf. Perkins' 1989 proposal that it might not be possible to sample more than 50 independent languages). Second, there might be unrecognized genealogical or areal groupings, not acknowledged in the sampling, which leads to inflation of the frequency of a type, notwithstanding the sample. Even more problematic is the possibility that the actual world's languages are not representative of the possible human languages. It is possible that there are still founder effects available in the current distribution of the world's languages, i.e. that there are preferences in the current world's languages that go back to incidental events during the spread of languages over the world (Maslova 2000).

A much more suitable solution to Galton's problem has been proposed by Mark Pagel (1994, see also Mace & Pagel 1994), and the method used by Dunn et al. is a direct descendant of this proposal. The software used by Dunn et al, called *BayesTraits* (Pagel & Meade 2006) is developed by Mark Pagel and his collaborators as an implementation of the method proposed in the earlier papers. The basic underlying idea is to change perspective from trying to establish empirical frequencies of a particular linguistic type to estimating transition probabilities between linguistic types. To be able to obtain such estimates, we need a rather different kind of sample. Instead of sampling across known genealogical units, we now need to sample within such groups. By assessing the internal variation within a language family it becomes possible to estimate the probability of change. In the field of linguistic typology, this methodological approach has been independently proposed by Elena Maslova (2000: 328-329; 2002; 2004; Maslova & Nikitina 2008). Following her work, this approach might be called 'dynamic typology'.

In this commentary, I will first discuss the principle of using transition probabilities, and how these probabilities can be used to infer a stable state (Section 2). Then I will briefly sketch two different approaches to the empirical estimation of the transition probabilities (Section 3). Finally, on this basis I will discuss the claims made by Dunn et al. and argue that they seem to misinterpret their own results, both in relation to the typological correlations (Section 4) and with respect to the variability of the transition probabilities (Section 5).

Although I completely agree with their approach, I think that they too easily dismiss the consistency in their findings and focus too much on the variation attested. There is indeed variation between families (which can hardly be said to be surprising), but this variation mainly leads to absence of evidence for cross-linguistic generalization, not to to any evidence for their absence. The limited amount of convergent evidence available in their data actually substantiates widespread assumptions about the preferred word orders of human languages.

2. Introducing transition probabilities

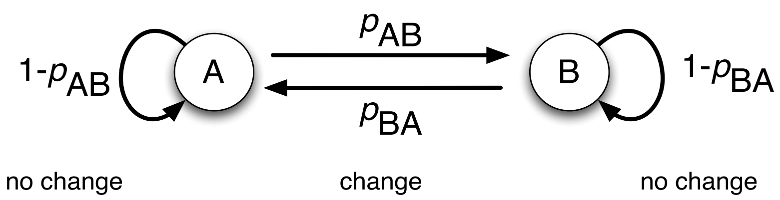
To explain the concept and implications of transition probabilities, consider first a simple typology T, classifying languages into just two different types, A and B. The traditional empirical objective of linguistic typology is to estimate the relative frequency $F(\mathrm{A})$ and $F(\mathrm{B})$ of these types, for example, that A occurs in 79% of the world's languages and B in 21% (i.e. $F(\mathrm{A}) = 0.79$ and $F(\mathrm{B}) = 0.21$). In dynamic typology, the objective is changed to estimating transition probabilities between A and B, for example, that A changes to B with a probability of $p_{\mathrm{AB}} = 0.13$, while B changes to A with a probability of $p_{\mathrm{BA}} = 0.48$.

There are three important aspects of transition probabilities that have to be clarified. First, such transition probabilities describe the probability that a change happens within a particular time frame. Of course, this time frame should ideally be concretely specified, for example, whether the transition probabilities describe a probability of change within 100 years, or 5000 years. However, in most practical applications of transition probabilities, to be explained shortly, the concrete size of the time frame turns out to be unimportant, as long at it is kept constant. Therefore, it is mostly unnecessary to specify the concrete time frame, and it is sufficient to simply speak about timeless transition probabilities. However, for any comparison of specific values for transition probabilities, it has to be ascertained that the time frames are comparable.

Second, transition probabilities are independent values, contrary to the relative frequencies. With reference to the numbers as used above as an example, the relative frequencies of A

(79%) and B (21%) necessarily add up to a 100%, i.e. $F(A) + F(B) = 1$.[1] In contrast, the transition probabilities from A to B (0.13) and from B to A (0.48) do not add up to one. This is the normal situation, because there is no definitional relation between the probability of a change from A to B (henceforth p_{AB}) and the probability of a change from B to A (henceforth p_{BA}). The complement of p_{AB} (i.e. the value $1\text{-}p_{AB} = 0.87$) represents the probability that there was no change from A to B, i.e. the probability that a language of type A remains to be of type A in the (unspecified) time frame considered. The probability of change (0.13) and the corresponding probability of no change (0.87) add up to one. But the probability of change in the one direction (0.13) is independent of the probability of change in the other direction (0.48).

(1)



Finally, the empirical establishment of transition probabilities is a process of estimation. There is no claim that it is necessarily possible to actually pinpoint a real value for any specific transition. It is possible that attempts to estimate such values turn out not to stabilize, so we might have to conclude that some assumed transition probabilities are situation dependent. Further, even when the estimates stabilize, any proposed value will have an error margin, which is ideally explicitly specified.

The most important auxiliary value that can be derived from a set of transition probabilities is the stable state. When transition probabilities remain the same over a longer period of time, then it is possible to predict the stable distributions of the types A and B, i.e. the situation in which the relative frequencies of languages of type A and B do not change anymore. This stable state

---

[1] This equation of course only holds under the condition that we only strictly allow languages to be of either type A or B and nothing else. Adding the possibility of intermediate types corresponds to allowing more possible types than just two.

represents the dynamic equivalent of the traditional estimate of relative frequencies. The basic idea of the stable state is that language change is still happening, but this language change does not result in different relative frequencies of A and B. Such a stable state arises when the number of languages that change from A to B is identical to the number of languages that change from B to A (within a specific time frame). Now, the number of languages that change from A to B is easily expressible as the product of the relative frequency of A and the probability that any of those languages changes to B, i.e $F(\text{A})\cdot p_{\text{AB}}$ and vice versa for the change from B to A, $F(\text{B})\cdot p_{\text{BA}}$. In the stable state, these two values have to be identical:

(2) $F_{\text{stable}}(\text{A})\cdot p_{\text{AB}} = F_{\text{stable}}(\text{B})\cdot p_{\text{BA}}$

Because in the stable state the frequencies of A and B still add up to 100%, we also have:

(3) $F_{\text{stable}}(\text{A}) + F_{\text{stable}}(\text{B}) = 1$

These two equations with two unknowns are easily solved to a formula for the stable state frequency as expressed by the transition probabilities:

(4) $F_{\text{stable}}(\text{A}) = p_{\text{BA}} / p_{\text{BA}} + p_{\text{AB}}$

$F_{\text{stable}}(\text{B}) = p_{\text{AB}} / p_{\text{AB}} + p_{\text{BA}}$

Using the exemplary values from above this would result in:

(5) $F_{\text{stable}}(\text{A}) = 0.48 / 0.48 + 0.13 = 0.79$

$F_{\text{stable}}(\text{B}) = 0.13 / 0.13 + 0.48 = 0.21$

As an aside, note that the same stable state frequencies can arise from rather different transition probabilities. Further, because of the division of transition probabilities, any specific time frame of the transition probabilities is removed from the stable state frequency, so the stable state is independent of the time frame in which the transition probabilities are established.

Another linguistically useful auxiliary notion that can be derived from the transition probabilities is the stability of a typological parameter T. The basic concept of typological stability is to estimate the probability that there is no change at all (Dediu 2011). So, in case of a typological parameter T with two possible types A and B, the probability of no change happening among the type-A-languages is $1 – p_{AB}$, while the probability of no change among the type-B-languages is $1 – p_{BA}$. As a measure of overall stability of the typological parameter T one could for example simply take the average of these two values:

(6) $S(\mathrm{T}) = (1 – p_{AB}) / 2 + (1 – p_{BA}) / 2$

This example of the usage of transition probabilities only represents the most basic situation with just two discrete types. When more types are allowed, or even continuous variation is included, the mathematics become quickly more complicated. In general, with n possible types, the number of transition probabilities to be estimated is n^2-n, which empirically becomes unmanageable for higher n.

Dunn et al. use a second, more complex model with four states, representing a cross-section of two binary typological parameters (see Section S1.4 of the supplementary information of their paper for a detailed explanation). With four types, the number of transition probabilities to be estimated is $4^2-4 = 12$. However, to keep the estimation manageable, only eight of these twelve transition probabilities are estimated (following the proposal of Pagel 1994). The idea is to only consider transitions that change one type at a time and ignore (as a simplification) the transitions that change both parameters at once. For reasons of space, I will not spell out the complete derivation, but the equations to establish the stable state are the following in this situation:

(7)
$$F_s(\mathrm{A})\cdot(p_{\mathrm{AC}} + p_{\mathrm{AD}}) = F_s(\mathrm{C})\cdot p_{\mathrm{CA}} + F_s(\mathrm{D})\cdot p_{\mathrm{DA}}$$
$$F_s(\mathrm{B})\cdot(p_{\mathrm{BC}} + p_{\mathrm{BD}}) = F_s(\mathrm{C})\cdot p_{\mathrm{CB}} + F_s(\mathrm{D})\cdot p_{\mathrm{DB}}$$
$$F_s(\mathrm{C})\cdot(p_{\mathrm{CA}} + p_{\mathrm{CB}}) = F_s(\mathrm{A})\cdot p_{\mathrm{AC}} + F_s(\mathrm{B})\cdot p_{\mathrm{BC}}$$
$$F_s(\mathrm{D})\cdot(p_{\mathrm{DA}} + p_{\mathrm{DB}}) = F_s(\mathrm{A})\cdot p_{\mathrm{AD}} + F_s(\mathrm{B})\cdot p_{\mathrm{BD}}$$
$$F_s(\mathrm{A}) + F_s(\mathrm{B}) + F_s(\mathrm{C}) + F_s(\mathrm{D}) = 1$$

These equations are rather easily solvable, but the resulting formulas are too large and unwieldy to write down explicitly here. However, the basic point is that given a set of transition probabilities, it is relatively easy to derive the frequencies in the stable state, which is the equivalent in dynamic typology to the traditional estimation of frequencies through a sample of the world's languages.

3. Estimating transition probabilities

The remaining (big) problem of course is how to empirically estimate transition probabilities. The crucial insight is that variation within groups of related languages can be used to perform such estimation.[2] This insight to use variation within groups represents almost a complete reversal of sampling in typology. Traditionally, variation within groups of closely related languages has been considered a nuisance in typology (cf Dryer 1992 and Bickel 2008 for proposals how to include group-internal variation in the construction of a sample). In contrast, variation within groups is crucial for dynamic typology.

Maslova proposed a practical implementation of this approach on the basis of what she calls the *divergence rate*: "The divergence rate is measured for a sample of pairs of related languages with a relatively small time depth and corresponds to the frequency of pairs that exhibit different values of this variable" (Maslova & Nikitina 2008). So, instead of sampling one language per

---

[2] In most cases, such groups of languages will consist of genealogically related languages, but it is equally possible to include areal groupings. Due to lack of space I will restrict myself here to the discussion of groups of genealogically related languages.

genealogical group, she proposes to sample two (closely related) languages per group and to investigate how often these two languages differ typologically. In the case of a simple typology with only two types A and B and two transition probabilities p_{AB} and p_{BA}, Maslova (2002; 2004) derives the following astonishingly simple formula for the divergence rate $F(D)$:

(8) $F(D) = 2 \cdot F(A) \cdot (p_{AB} - p_{BA}) + 2 \cdot p_{BA} \cdot (1 - p_{AB})$

Because both $F(D)$ and $F(A)$ are known in any sample of language-pairs, the transition probabilities can be calculated when at least two such samples are available (because of the two unknowns, p_{AB} and p_{BA}), which can be obtained by e.g. by splitting the world's languages into two subsamples.

Dunn et al. (based on the proposals from Pagel 1994) use a different practical implementation of this same insight to use group-internal variability. They use highly detailed reconstructions of a single family tree to estimate the transition probabilities. For a large group of related languages they establish their typological types, and given the distribution of these types over the genealogical tree, the transition probabilities can be estimated. This method already works for just a single family, given that data for enough languages is available.

Both these practical implementations to estimate transition probabilities represent two nicely complementary approaches, and it would be highly interesting to investigate the agreement between their results. However, to be able to compare both these approaches, and to develop different methods to estimate transition probabilities, it is of central importance that more typological collections move away from the traditional method of typological sampling. Instead of sampling one language per genealogical unit, it is actually much more informative to sample various languages from the same unit. One practical approach could be the sampling of pairs of closely related languages, as proposed by Maslova. Another possibility is to select just a few lineages and sample them densely, as exemplified by Dunn et al. Practically, a typological sample of 100 languages will then not consist of 100 languages from different genealogical groups, but it will consist of, say, pairs from 50 different groups, or even just 10 groups with about 10 languages from each. Note that it is not necessary to collect equally sized groups.

4. Correlating transition probabilities

Dunn et al. do not restrict themselves to estimating transition probabilities for various typological parameters, which would already have been a highly innovative and useful addition to the field of linguistic typology. They even go one step further and propose a new method to establish typological correlations between parameters on this basis. The basic insight here is to investigate whether the typological characteristics co-evolve in the linguistic lineage as represented by a family tree. Unfortunately, they seem to misinterpret the results from their investigation. They claim that “contrary to the Greenbergian generalizations, we show that most observed functional dependencies between traits are lineage-specific rather than universal tendencies” (p.79). I will argue, in contrast to this claim, that their results show clear agreement across lineages. These agreements are indeed not very strong, so their results—to a large extend—only shows absence of evidence for word order correlations. However, absence of evidence should not be confused with evidence of absence. The absence of evidence might very well depend on the small number of just four lineages that are investigated. I strongly expect that the evidence will become clearer once more different families are investigated.

In practice, Dunn et al. investigate typological correlations by comparing two different hypotheses for each combination of two parameters.[3] For example, they consider the combination of the OV/VO parameter and the postpositions/prepositions parameter. First, the independent hypothesis represents the model in which these two parameters are not correlated. In this model, the transition probabilities are calculated for each parameter individually using the simple model from Section 2 with just two types and two transition probabilities. Second, the dependent hypothesis represents the model in which the two parameters are correlated. In this

---

[3] See Section S1.4 in the Supplementary Information from the Dunn et al. article for a more detailed explanation.

model, the transition probabilities are calculated for the intersection of the two parameters using the more complex model from Section 2 with four types and eight transition probabilities.

Dunn et al. use the "Bayes Factor" to compare the dependent hypothesis to the independent hypothesis. Ignoring the details, the Bayes Factor is a number that will be negative when the independent hypothesis is to be preferred, and it will be positive when the dependent hypothesis is to be preferred. How high (or low) the number is indicates how strong any such preference is; the closer to zero, the less telling. As always with such test-statistics, it is difficult to decide at which point the strength of the preference becomes significant. The problem is not a mathematical one, because the Bayes Factor is distributed according to a chi-square distribution. The problem arises because the significance on such a chi-square distribution has to be determined relative to the degrees of freedom, and it is not clear how many degrees of freedom have to be assumed in the current situation. The manual of *BayesTraits* (the software used by Dunn et al.) suggests to take "degrees of freedom equal to the difference in the number of parameters between the two models". That would imply four degrees of freedom for the current situation, because the independent model has four parameters (viz. two times two transition probabilities), while the dependent model has eight parameters (the eight transition probabilities). A chi-square distribution with four degrees of freedom would mean that a Bayes Factor greater than 9.5 occurs with a probability of less than 5%, while factors greater than 13.3 occur with a probability of less than 1%. In contrast, Dunn et al. assume for the interpretation of their Bayes Factors (Section S3 of the SI) that "values from 2-5 indicate weak support […]. Values from 5 are conventionally considered strong evidence." These values imply a much more lenient interpretation of the Bayes factors, approximately assuming a chi-square distribution with a single degree of freedom. Notwithstanding such details, the factors can surely be interpreted as saying that the stronger positive the value, the more evidence for the dependent hypothesis, and the stronger negative the value, the more evidence for the independent hypothesis. Values between 5 and -5 are only weakly indicative, with all values between 2 and -2 being not telling at all.

The Bayes Factors for all combinations of parameters are listed in Section S3 of the Supplementary Information in Dunn et al., separated for the four families investigated. The values are plotted in Figure 1, ordered by the significance as established by Fisher's combined probability test (more on that below). Basically, the more to the right, the more interesting is the correlation. Horizontal lines are added to indicate the non-indicative range between 2 and -2. The first important observation to make is that there are almost no negative values attested, and the few cases of 'weakly supportive' Bayes Factors smaller than -2 never go below -5. This is probably the case because the independent model is a special case of the dependent model, or, as the authors explain: "the independent model can be expressed as a special case of the dependent model, so the test cannot show that the independent model is preferred over the dependent; rather it can only show that the dependent model is not superior" (p. S6). This means that their approach inherently cannot strongly argue for the independent hypothesis, i.e. for the absence of a correlation.

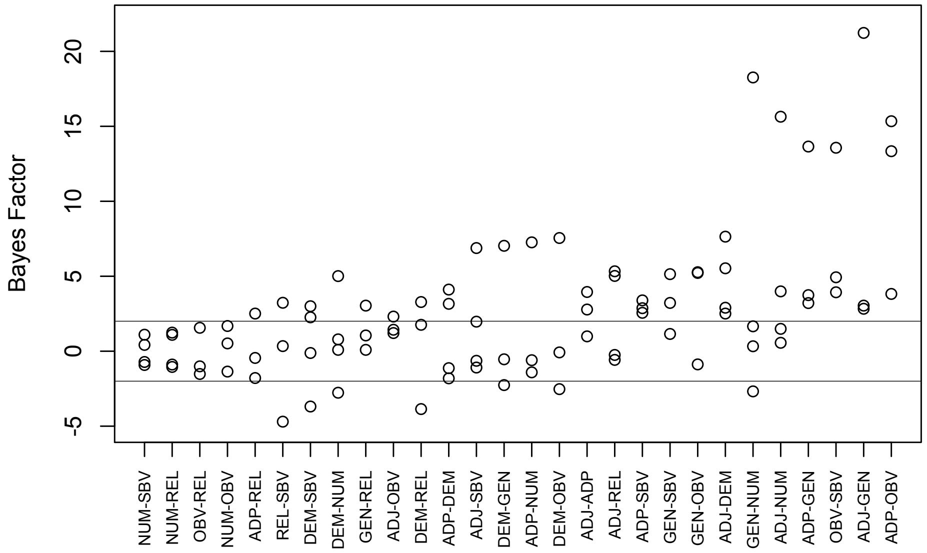


Figure 1. Bayes Factors for the different combinations of word order parameters, ordered by Fisher's meta-significance (stronger significant combinations to the right). Values between 2 and -2 (indicated by the horizontal lines) are not telling. Only values outside the range between 5 and -5 show strong indication of any preference. Note that there are no values below -5 attested, as the method inherently does not allow for strong evidence for the independent hypothesis (i.e. the absence of a correlation).

Still, looking at the cases with weak support for the independent hypothesis (which is the maximum possible given their method), the most strongly negative Bayes Factor are attested for the pairs like "relative clause position-subject/verb" (REL-SBV), "demonstrative/noun-relative clause position" (DEM-REL) and "demonstrative/noun-subject/verb" (DEM-SBV), which have never been claimed to be typologically related. So, in general, Dunn et al. do not find any evidence for absence of typological correlations, and the most extreme cases in favour of independence are attested for combinations of parameters that never have played any role in any typological investigation.

In contrast, there is a set of correlations in which all genealogical families have a Bayes Factor above 2, indicating consistent positive support for the dependent hypothesis (though not necessarily very strong support in all families). Such consistent positive support is attested most clearly for the following well-known correlations (cf. Dryer 1992):

- adpositions - object/verb (combined $p = 0.001$)
- adjective/noun - genitive/noun (combined $p = 0.005$)
- object/verb - subject/verb (combined $p = 0.03$)
- adpositions - genitive/noun (combined $p = 0.05$)

The above reported 'combined' *p*-values are the combined probabilities for each dependent hypothesis using Fisher's method of meta-analysis.[4] Basically, when there are various (statistically independent) test results trying to prove the same hypothesis, Fisher's method offers an assessment of the combined probability of all tests together. To calculate these combined probabilities, I first established the probabilities of the Bayes Factors using the strict interpretation with four degrees of freedom, and then combined these probabilities using Fisher's combined probability test. Negative Bayes Factors have been assigned a probability of zero. The above-noted set of four typological correlations all have a combined probability of 5%

---

[4] I thank Dan Dediu for this suggestion.

or lower, which strongly suggests that there is positive evidence in these cases for the dependent hypothesis, i.e. that the parameters are correlated in all lineages investigated.

Summarizing, there is of course variation between linguistic families, but the general tendencies across the families seem to be nicely consistent with common assumptions about word order regularities in linguistic typology. Contrary to the author's claim, there appears to be a significant correlation for at least four cross-sections of parameters across all lineages studied. Other than that, they mostly find only limited and inconclusive evidence, surely not sufficient to disprove widespread typological assumption. What would have been really ground-breaking is when Dunn et al. could show that any widely assumed universal correlation would be a chimera, but they do not have any evidence for such a case, and it even appears that the method in principle does not even allow for such conclusions.

5. Interpreting transition probabilities

Dunn et al. finish their paper with a second conclusion, namely that "even where we find dependencies shared across language families, the phylogenetic analyses show family-specific evolutionary processes at work" (p. 81). This claim specifically refers to the fact that the combination of the parameter "adpositions" (ADP) and "object-verb order" (OBV) show strong evidence for the dependent hypothesis for both Austronesian and Indo-European (both Bayes Factors are above 13.3, which amount to a probability of below 1% assuming four degrees of freedom). So in this case there is really strong evidence for a correlation in two out of three applicable families, while the third (Uto-Aztecan) shows weak evidence (and the combination of all three is highly significant, as noted above). Indeed, this combination of parameters seems to be the prime case of clear evidence for a typological correlation. However, Dunn et al. proceed to argue that even in this case there actually is no consensus between Austronesian and Indo-European, with reference to their Figure 3. However, this interpretation again seems to be misguided.

Their crucial Figure 3 shows a rough approximation of the estimated transition probabilities. The visual impression suggests rather different transition probabilities comparing Austronesian to Indo-European. The authors were kind enough to share some of the raw output of their analysis with me, so I was able to extract the real values of these transition probabilities.[5] The average transition probabilities as estimated by *BayesTraits* are shown in Figure 2, mimicking Figure 3 from Dunn et al. In their figure, they only distinguish between prominent transitions (thick arrows) and minor transitions (thin arrows).[6] This simplification hides the differentiation within the thin arrows. Even within the minor transitions there is mostly a strong asymmetry between the probabilities into the one direction compared to the reverse direction. In my Figure 2, I have drawn the arrows with the exact thickness corresponding to the transition probabilities. This illustration much clearer shows the parallelism between the two families, including the same asymmetries in the minor transition probabilities.

---

[5] The raw output that was made available to me consisted of 100.000 selected iterations of one run for each correlation pair in each family. I removed all iterations in which at least one of the transition probabilities is zero. A zero transition probability leads to unsolvable equations. I could have artificially added some random value to remove the zeros, but I decided to simply ignore the iterations with any zero probabilities, as there are still enough iterations remaining. In general, different solutions how to deal with the zeros did not seem to change any of the calculations here significantly.

[6] For some unknown reason, there is one arrow missing from Figure 3 in Dunn et al. I assume that this is simply a printing error.

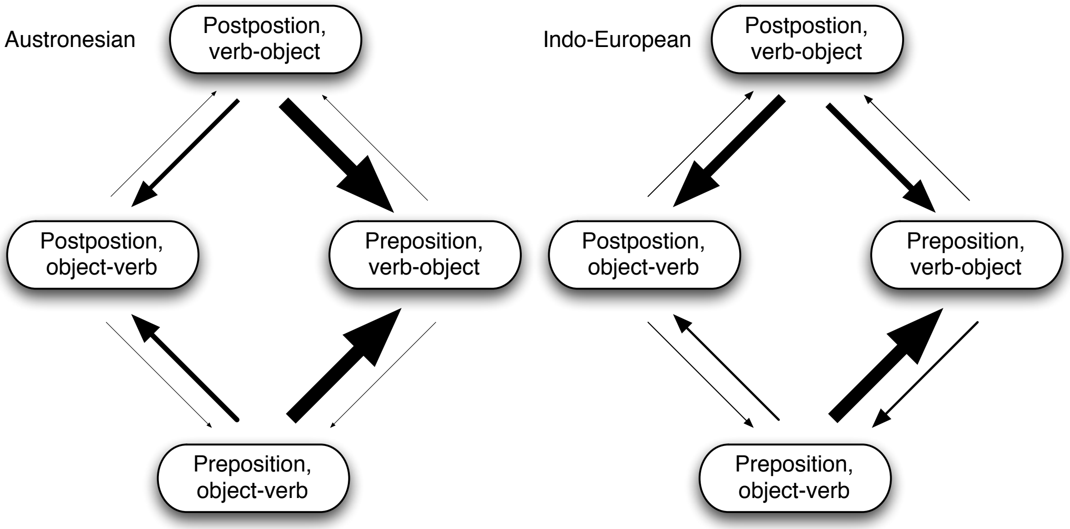


Figure 2. Transition probabilities mimicking Figure 3 from Dunn et al. In this graph, the thickness of the arrows corresponds to the actual average transition probabilities. In contrast, Figure 3 from Dunn et al. only distinguishes between the four major arrows (drawn thick) and all others (drawn thin). This rough approximation hides the detailed parallelism between the families visible in this more appropriate figure.

The actual values are plotted in Figure 3 here, comparing Austronesian (x-axis) with Indo-European (y-axis). The grey diagonal line indicates identical probabilities. As can be seen, the only substantial difference in the estimated transition probabilities is a preference for prepositions+OV to change to prepositions+VO in Austronesian, while there is a contrary preference for a change to postpositions+OV in Indo-European. Although this might seem like a major difference, the other transition probabilities do not change very much between these two families. The limited impact of this single difference becomes even clearer when the stable distribution is calculated on the basis of the complete set of transition probabilities. This comparison is shown in Figure 4. This figure shows estimates for the stable distribution separately for the four different combinations of parameters.[7] There is a striking agreement

---

[7] On the basis of the raw data made available to me by the authors, I removed all iterations with some zeros in the estimated transition probabilities. For each remaining iteration, I calculated

between the estimated stable distributions of Austronesian and Indo-European. Basically the combinations postpositions+OV and prepositions+VO are frequent in both stable state, while the other two combinations are predicted to be extremely rare. As expected from the transition probabilities, the stable distribution for postpositions+OV is slighly lower for Austronesian, but the hinges in the boxplot still overlap. Likewise, the stable distribution of prepositions+OV is slightly lower for Indo-European, but here also the difference is still within the range of the hinges of the boxes.

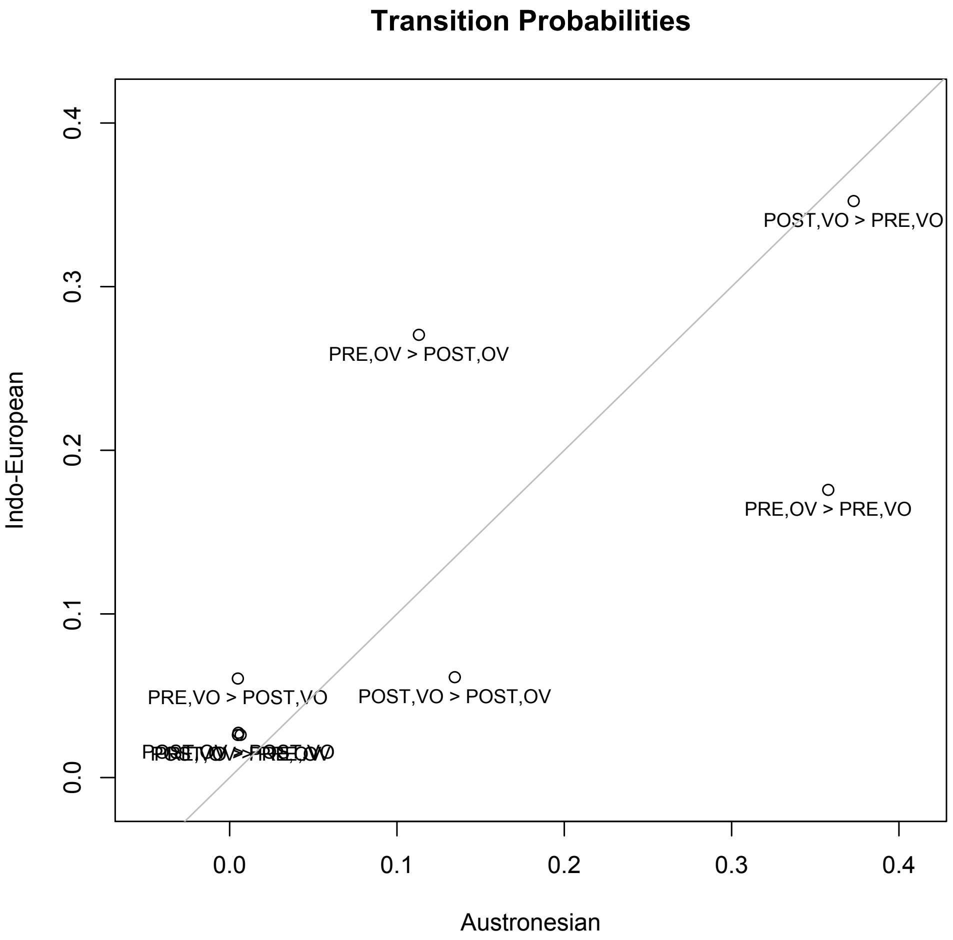


Figure 3. Comparison of the estimated transition probabilities between Austronesian and Indo-European for the combination postpositions/prepositions and OV/VO word order.

---

the stable state, using the equations as discussed in Section 2. The resulting variation in stable state proportions between the iterations is shown using boxplots in Figure 3.

The most extreme effect of calculating the stable state can be observed for the combination postpostions+OV in Austronesian. This combination is extremely rare in Austronesian. In the sample as used by Dunn et al, only 7 languages (6%) are of this combination of types. Yet, because of their specific distribution in the genealogical tree, the estimate for the stable state is much higher (viz. a mean of about 24%). This striking difference between the actual frequency and the predicted stable state frequency highlights the strength of the dynamic approach to typology. Even in extremely biased distributions as in Austronesian, the stable state still can reconstruct more general preferences.

This difference between the actual frequencies and the stable state frequencies in Austronesian can be explained as the residue of a founder effect in the rather young Austronesian family. Proto-Austronesian most probably had prepositions+VO and almost all Austronesian languages still are of the type. In the relatively recent and extremely rapid spread of the Austronesian languages throughout the Pacific there has not been enough time for many changes to accumulate. This results in a rather skewed distribution of the actual frequencies. Given the geographical and genealogical distribution of the few Austronesian languages with postpositions+OV (they are all on the eastern coast of New Guinea) it even seems likely that there has just been one or two events of this change (possibly contact induced). So the probability to lose the original state of prepositions and VO and change to postpositions+OV is small. However, the few languages with postpositions+OV are all very close in the family tree, which suggests a high stability of this combination. Once the combination arises, it does not easily change anymore. So, it seems probable that, given more time, the Austronesian languages might develop more instances of postpositions+OV, and this is exactly what the predicted stable state describes. It is precisely these kind of dynamics of change that can be reconstructed by switching to transition probabilities.

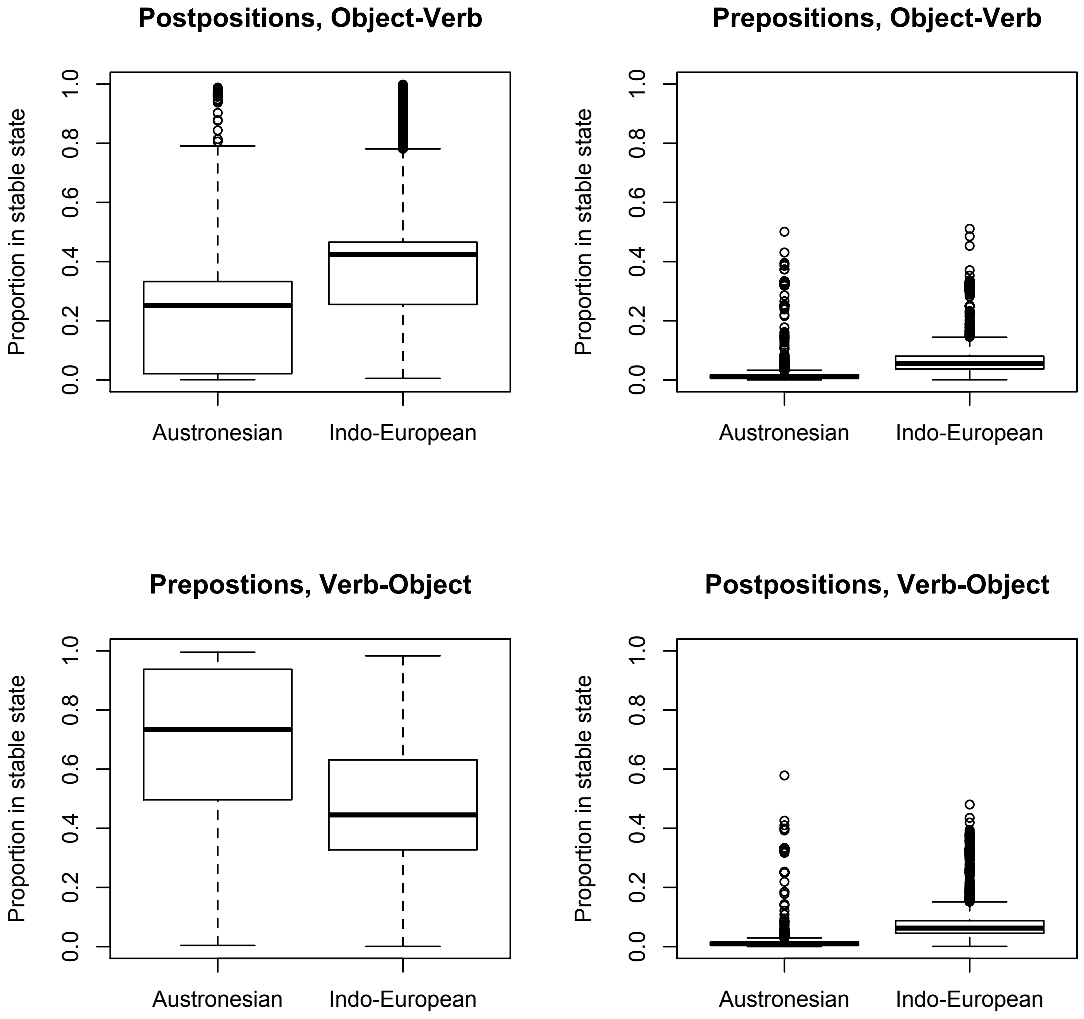


Figure 4. Comparison of predicted stable states between Austronesian and Indo-European. Although there are lineage-specific details discernible, the general pattern is strikingly similar. Two combinations are common, while two others are extremely rare.

6. Conclusion

The paper by Dunn et al. (2011) presents a fascinating and important case for a more dynamic approach to linguistic typology. Instead of trying to quantify the current state of the world's languages, their approach to language variation attempts to quantify the dynamics of linguistic change. If that can be achieved, and I think that Dunn et al. have shown that it is indeed possible, then we can finally surpass the old debate of how to integrate historical factors into synchronic linguistic typology. Linguistic typology will turn into a diachronic discipline, in which synchronic tendencies and universals will be derived from the dynamics of language change.

Unfortunately, the main selling point of the paper by Dunn et al. seems to be a misinterpretation of their own results. They conclude their paper saying that “what the current analyses unexpectedly reveal is that systematic linkages of traits are likely to be the rare exception rather than the rule” (p.82). This does not seem to be accurate. In at least four cases their results show strong evidence for the same tendency across all lineages studied, and most others cases simply do not show any clear evidence, neither in favour of systematic linkage, nor against it. The strongest case of cross-lineage agreement even shows, again in contrast to their explicit renouncement, that the same word order dynamics are at work in both Austronesian and Indo-European, even though the synchronic profiles of these families are strikingly different.

Regrettably, the rather provocative rhetoric of the paper does not serve its goal. By unfoundedly stressing the variability of languages and denying off-hand the results of decades of typological research, the authors alienate exactly the main group of scientists that would in principle most welcome the revolutionary methods and insights from this paper. I hope that this commentary can be of assistance in changing the perception of the main results of the Dunn et al. paper. It is not the “lineage specific trends” from the title of their paper that are new and noteworthy, it is their breakthrough renovation of decades old methodology in linguistic typology.

Acknowledgements

I thank the authors of the Dunn et al. paper, and in particular Michael Dunn, for making part of the raw output of their analyses available for this post-hoc analysis. This collegial approach to sharing results is deeply appreciated, and I think that this commentary once again shows that scientific understanding is as much a communal effort as it is a competitive game for personal fame. Real progress can only be achieved through open-minded sharing of ideas, data, analyses and results (of course with reciprocating acknowledgements) . Further, I thank Dan

Dediu and Steven Moran for many helpful suggestions and correction on earlier versions of this paper.

References

Bickel, Balthasar. 2008. A refined sampling method for genealogical control. *Sprachtypologie und Universalienforschung* 61(3). 221-233.

Cysouw, Michael. 2005. Quantitative methods in typology. In: Gabriel Altmann, Reinhard Köhler & R Piotrowski (eds.), *Quantitative Linguistics: An International Handbook,* 554-578. Mouton de Gruyter: Berlin. 554-578.

Dediu, Dan. 2011. A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. *Proceedings of the Royal Society B: Biological Sciences* 278(1704). 474-479.

Dryer, Matthew S. 1992. The Greenbergian Word Order Correlations. *Language* 68(1). 80-138.

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. 79-82.

Mace, Ruth & Mark Pagel. 1994. The comparative method in anthropology. *Current Anthropology* 35(5). 549-564.

Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4(3). 307-333.

Maslova, Elena. 2002. Distributional universals and the rate of type shifts: towards a dynamic approach to "probability sampling". Lecture given at the *3rd Winter Typological School*, Moscow. Available online at http://anothersumma.net/Publications/Sampling.pdf

Maslova, Elena. 2004. Dinamika tipologičeskix raspredelenij i stabil'nost' jazykovyx tipov. [Dynamics of typological distributions and stability of language types]. *Voprosy jazykoznanija*, 5. 3-16.

Maslova, Elena and Tatiana Nikitina. 2008. Stochastic universals and dynamics of cross-linguistic distributions: the case of alignment types. Unpublished Manuscript, available online at http://www.anothersumma.net/Publications/ProbabilityPubl.html

Naroll, Raoul. 1961. "Two solutions to Galton’s Problem". *Philosophy of Science* 28. 15-29.

Pagel, Mark. 1994. Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. *Proceedings of the Royal Society B: Biological Sciences* 255(1342). 37-45.

Pagel, Mark & Andrew Meade. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *American Naturalist* 167(6). 808-825.

Perkins, Revere D. 1989. Statistical techniques for determining language sample size. *Studies in Language* 13(2). 293-315.

Taylor, Edward B. 1889. On a Method of Investigating the Development of Institutions; Applied to Laws of Marriage and Descent. *The Journal of the Anthropological Institute of Great Britain and Ireland.* 18. 245-272.

Tobler, W R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46. 234-240.