Distributional universals and the rate of type shifts: towards a dynamic approach to "probability sampling"

Elena Maslova

1 Problem and outline

- A synchronic typological frequency f, however large, is not a "linguistic fact", insofar as precisely the same value of f may result from widely different combinations of circumstances.
- Hence, typological frequencies can only be interpreted linguistically if language sampling is "enhanced" by the diachronic dimension, but this seems empirically infeasible.

However, this can be done by means of a relatively simple and feasible technique that may significantly increase the linguistic value of statistical typological observations. The basic idea is very simple: what we need is a representative sample of **pairs** of closely related languages (with a roughly similar time of divergence). This gives us an additional statistic, the frequency h of **divergent pairs** – an estimate of **divergence rate**, which can significantly restrict the set of possible diachronic scenarios (hence, of plausible linguistic interpretations of typological tendencies). Methodologically, the proposed technique remains within the realm of (synchronic) typology, in the sense that only present feature values are sampled; no reconstructions are needed.

Plan:

1. The evolution of language population: a stochastic model
2. Stochastic versions of linguistic hypotheses
3. What can the divergence rate tell us?

2 Model

2.1 Temporal scale

For the sake of simplicity, I will talk in terms of discrete-time model, i.e., all processes going on in the language population will be modeled in terms of temporal "steps". I will assume that a reasonable time interval corresponding to such a "step" is one thousand years. I will further assume that there can be no plausible hypotheses concerning typological distributions 10000 years (10 "steps") ago, but

- The language population has been sufficiently large for typological statistics during these 10000 years.
- It is likely that languages spoken within this time interval instantiated essentially the same universal phenomenon as the modern languages.

2.2 Birth-and-death process

- The birth-and-death process in the language population (so-called "historical accidents") can significantly affect typological distributions if the population is relatively small (this is why there are no plausible hypotheses about typological distributions 10000 years ago), but
- Its impact cannot be significant in a large population; so, in the present context, these potential effects can be neglected.

The last result appears to be counterintuitive to many typologists (what may be true in a mathematical model, need not be valid in the "real life"). Therefore, I will present a "real-life" example that illustrates the same point. The example is based on Maddieson's (1984) phoneme inventory sample; there are multiple "types" corresponding to the number of vowel phonemes. Figure 2 shows the impact of birth-and-death process on this distribution over ca. 7000 years (from –8000 to –1000), that is, the difference between the distribution in Maddieson's sample and the distribution in a random sub-sample containing one language per major linguistic family.

2.3 Type-shift process

I will consider only two-type typologies $\mathbf{T} = \{A, B\}$; the present distribution is $\mathbf{F} = \{f(A) = f, f(B) = 1 - f\}$. For the sake of simplicity I will assume here and below that $f \geq 1 - f$. The type-shift process is described by the following matrix:

"target type":		A	B
"source type":	A	$1-q$	q
	B	p	$1-p$

If the frequency of type A at some step k is $f^{(k)}$, then its frequency at the next step is expected to be around $f^{(k+1)}$:

$$f^{(k+1)} = f^{(k)} \cdot (1 - q) + (1 - f^{(k)}) \cdot p \tag{1}$$

If p and q are constant during n steps, the resulting frequency $f^{(n)}$ of type A is given by the following equation:

$$f^{(n)} = f^{(0)}(\frac{p}{p+q} + \frac{q(1-p-q)^n}{p+q}) + (1 - f^{(0)})(\frac{p}{p+q} + \frac{p(1-p-q)^n}{p+q}) \tag{2}$$

The rate of the process (the number of "steps" which makes $f^{(n)}$ independent of $f(0)$) is determined by $\alpha = p + q$; the resulting limiting (stationary) distribution, by $\frac{p}{q}$. Thus, if the present distribution is independent of the initial distribution, then the following equation holds:

$$\frac{f}{1-f} \simeq \frac{p}{q} \tag{3}$$

This means, roughly speaking, that to say that a higher frequency f manifests a "linguistic preference" **means** to say that $p > q$, or in other words, that the would-be stationary frequency f^{∞} is higher than $\frac{1}{2}$.

3 Testing linguistic hypotheses

3.1 Stochastic versions of linguistic hypotheses

One can think of two "extreme" hypothesis about the regularities of language change (and their "relation" to the current synchronic distribution). First, we might want to test the **Hypothesis of Stationary Distribution**, that is, roughly speaking, that the synchronic distribution is determined solely by transition probabilities (plus random deviations δ due to the birth-and-death process). The opposite extreme is the hypothesis of absolutely random (linguistically unmotivated) language change; that is, $p = q$ (**null hypothesis**).

Since HSD is impossible to confirm or reject in most cases, we would like to be able to test weaker hypotheses: for example, if the current frequency of some type is 0.95 and we are unable to confirm that this distribution is stationary (that is, that $r = p/q \approx 19$), we may still be able to confirm something more significant than $p > q$, e.g., that $r > 2$ or $r > 3$. Similarly, if the current distribution is even ($f \approx 1 - f$), it would be nice to be able to say something like $0.5 < r < 2$ – that is, even if we cannot claim that $p = q$, we might be still able to claim that these parameters do not differ too much. I will refer to this family of hypotheses as R-hypotheses.

One can also think of two **Unidirectionality Hypotheses** ($p = 0$ and $q = 0$): what we observe now is just an intermediate stage of the evolution of language population from one "universal" type to another (either from A to B, or vice versa).

3.2 What can the divergence rate tell us?

The vaule of **divergence rate** in the present language population is determined by the following equation:

$$DV = P(AB) = 2f^{(-1)}q(1-q) + 2(1-f(-1))p(1-p) \quad (4)$$

where $f^{(-1)}$ is the frequency of A one "step" ago. The frequency h of divergent pairs in a random sample of pairs can serve as an estimate for DV. Equation (1) can be used to exclude the unknown parameter $f^{(-1)}$:

$$h \simeq P(AB) = 2(f-p)(q-p) + 2p(1-p) \quad (5)$$

Now we have two equations, (5) and 2, for three unknown parameters ($f^{(0)}$, p and q), which may seem to make the situation hopeless. Indeed, we cannot give precise estimates for p and q. What we need, however, is to be able to confirm R-hypotheses. To do so, we will need to consider the "worst" possible f_0.

Let the R-hypothesis we wish to confirm be $p \geq rq$: in other words, we want to confirm that the **large** number of A-languages now is determined by large p and not by a large number of A-languages in the initial distribution. Hence, the worst f_0 will be $f_0 = 1$ (that is, roughly speaking: 10000 years ago there were only A-languages, and the population slowly drifts towards the B-only distribution). Similarly, if our R-hypothesis is $p \leq q$, then the worst case is $f_0 = 0$.

Consider a constructed example illustrated in Figure 3. The observed frequency is $f = 0.75$ (like in Figure 1). However, now that we have an estimate for $p(AB)$ and equation 5, some of the theoretically possible diachronic paths can be rejected. In particular, if $p(AB) > 0.10$, then we can say that

$$1.1q < p < 17.9q, \tag{6}$$

which means that:

$$0.52 < P^{stat}(A) < 0.95. \tag{7}$$

Thus, although we can neither reject nor confirm the hypothesis of stationary distribution, we do get some essential information: p must be at least slightly higher than q (otherwise, the current frequency would be **lower** than $0.75-\delta$ even if the initial frequency was 1. On the other hand, it must be less than $18q$ – otherwise, the current frequency would be **higher** than $0.75 + \delta$ even if the initial frequency was zero. Note that this means that we have also rejected both unidirectionality hypotheses.

With some reservations to be discussed below, the higher the divergence rate, the better our estimates: roughly speaking, a higher divergence rate means that the process is faster, so the current frequency must be closer to the stationary probability. For example, if $p(AB) > 0.20$, then we can say that

$$1.9q < p < 5.2q, \tag{8}$$

which means that:

$$0.66 < P^{stat}(A) < 0.84. \tag{9}$$

On the other hand, if $p(AB)$ can be less than 0.05, than we cannot confirm/reject any hypotheses (with this method, of course).

3.3 A "real-life" example

According to Maddieson, the present frequency of vowel systems with five or more vowels is $f = 0.90$. The minimum of divergence rate can be estimated as 0.066 (5 pairs out of 74 found in Maddieson's sample). This gives us a (relatively) strong lower bound for p/q ($p > 2q$), which corresponds to the minimum of stationary probability $P^{stat}(n(v) > 4) > 0.67$. Yet the unidirectionality hypothesis cannot be rejected (i.e., it may be the case that $q = 0$, that is, 5-vowels languages do not lose vowels.)

On the other hand, the frequency of vowel systems with six or more vowels is $f = 0.67$, the minimum of DV can be estimated as 0.256 (21 pairs), which gives us the following estimates:

$$1.33q < p < 3.35q, \tag{10}$$

which means that:

$$0.57 < P^{stat}(n(v) > 5) < 0.77. \tag{11}$$

This means that we can safely assume that (1) the statistical predominance of languages with more than five vowels is motivated by stochastic properties of language change, but (2) the existence of smaller systems is also not a trace of the previous stage of evolution of the language population.