Bay Area Typology Workshop, 03/17/2001

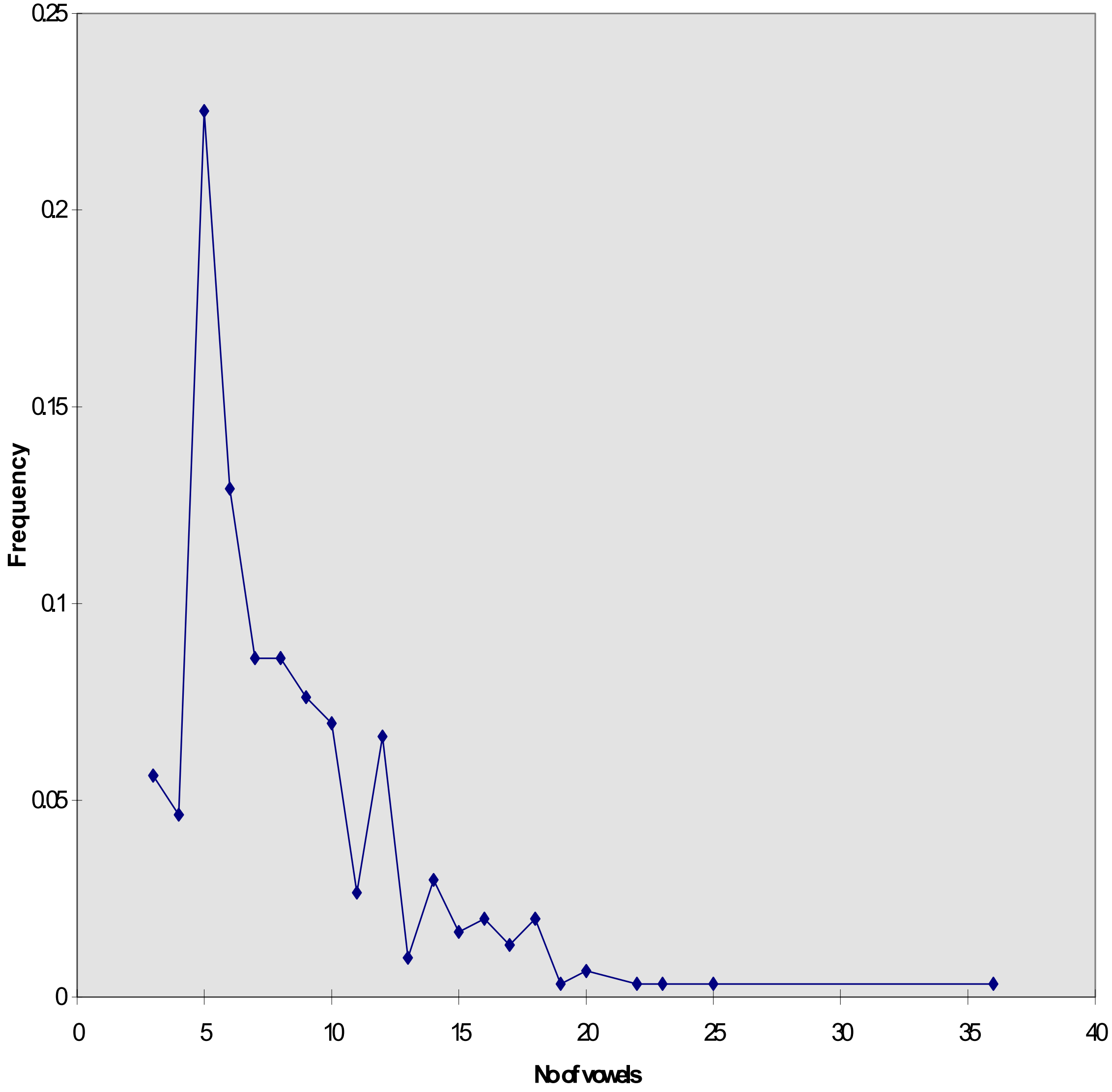
On linguistic interpretation of quantitative cross-linguistic data

Elena Maslova
maslova@jps.net

An example: Number of vowels per language (Maddieson 1984)

n	***f*(*n*)**
3	0.06
4	0.05
5	0.23
6	0.13
7	0.09
8	0.09
9	0.08
10	0.07
11	0.03
12	0.07
13	0.01
14	0.03
15	0.02
16	0.02
17	0.01
18++	0.04

f(n) - the frequency of languages with ***n*** vowel phonemes



How to interpret such a distribution? Is it a linguistically significant fact?

- Although the typological method has generated a variety of uneven distributions like this, we still have no reliable answers (cf. e.g. Rijkhoff & Bakker 1998:265), that is, we are unable to say, for each specific cross-linguistic distribution, whether (and to which extent) it is determined by linguistic factors (as opposed to random historical effects).

- It is widely believed, however, that the linguistic significance of a cross-linguistic distribution depends on the design of the sample employed to establish this distribution (e.g. this belief is integrated into the most recent typological textbook (Whaley 1997:36ff)). Roughly speaking, it is believed that a sample generates more 'linguistically accurate' results if only one language is selected from each genetic grouping (as opposed to proportional representation of genetic groupings, or just to random sample). The extreme version of this approach (one language per each major family) is advocated by Perkins (1989). Two correct observations are employed to justify this approach:
    - There is an enormous variation in size between genetic groupings, which has nothing to do with the hypothetical 'linguistic preferences'.
    - The type of a language is not stochastically independent of its genetic affiliation.

Outline

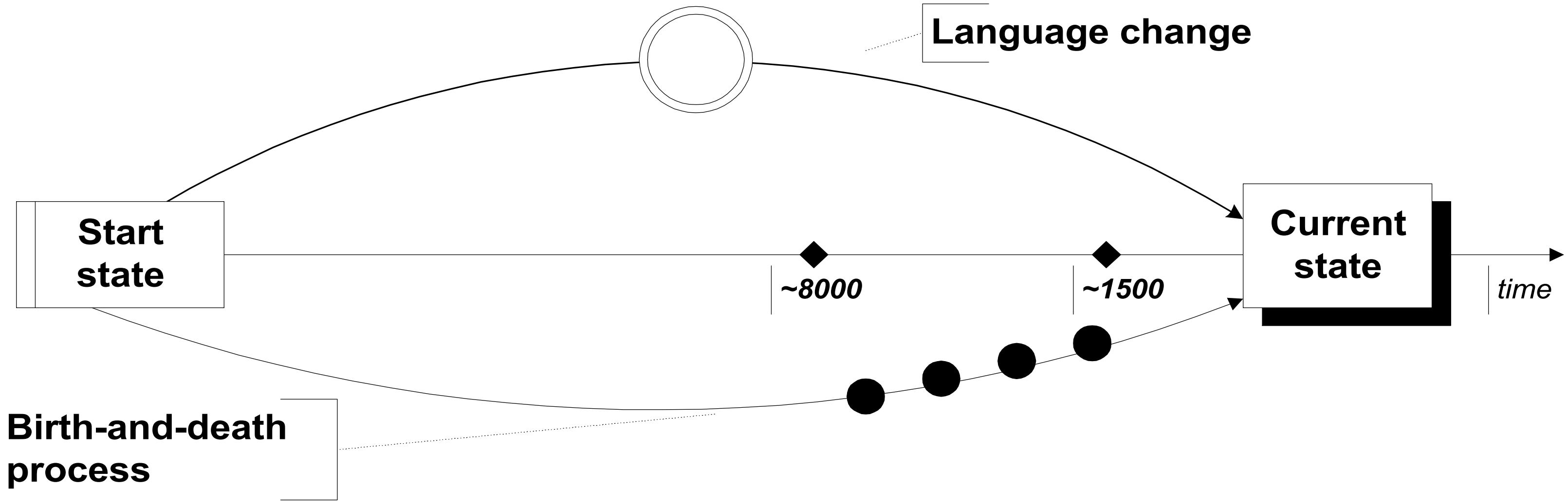
- The observations are correct, yet they do not justify the approach (nor the belief that this approach leads to (more) ‘linguistically accurate’ distributions):
  - Although the variation in family size is indeed striking, it has no significant impact on typological distributions (that is, whether we select only ‘independent’ languages, or design a proportional sample, the results would be roughly the same, granted that some basic requirements are met). This is relatively easy to show, and I’ll start with this issue.
  - The dependency of language type on its genetic affiliation reflects its stochastic dependency on the type of the ancestor language (the proto-language of the genetic grouping). If this underlying dependency exists, then it exists independently of the sample design; it does **not** disappear if we select only one language per grouping; it just cannot be revealed by statistical analysis of such a sample. In order to assess the role of such diachronic dependencies, we have to explore the type-shift (language change) processes. This is the only way to decide whether or not a given distribution is linguistically significant, and this is the central part of my talk.

Properties of the sample and the birth-and-death process

The sample in (Maddieson 1984) is designed in such a way as to include one language per ‘small’ genetic grouping, with “an intention to include no pair of languages which had not developed in their own independent speech communities for at least some 1000-1500 years, but to include one language from within each group of languages that shared a closer history than that” (ibid.:158-159). The result can be also considered a good proportional sample (the correlation between the representation of a group in the sample and its size according to *Ethnologue* is ca. 0.7).

The implications of this design are easy to see if we consider the following scheme of the history of the language population, which includes, apart from the current state of affairs (the distribution of linguistic traits among the existing languages), the ‘start state’ (be it the properties of the (single) Proto-language or of several ‘independent’ proto-languages), and two types of processes: the processes of language change (type shifts) in the history of each language and the ‘birth-and-death’ process in the population. An important point that there is nothing else that might have modified any given typological distribution.

The history of a typological distribution



The underlying idea of ‘quota’ sampling is to eliminate the effects of language break-ups (“birth” events) that happened after some moment *t*; in the extreme version, *t* corresponds to the time depth of the major families (~8000 on the Figure). Black circles indicate those birth events that would not affect a distribution attested in a Perkins-style sample, but have affected the distribution in Maddieson’s sample. My first claim is that this difference has virtually no impact on the resulting distribution.

Birth-and-death process in a large population

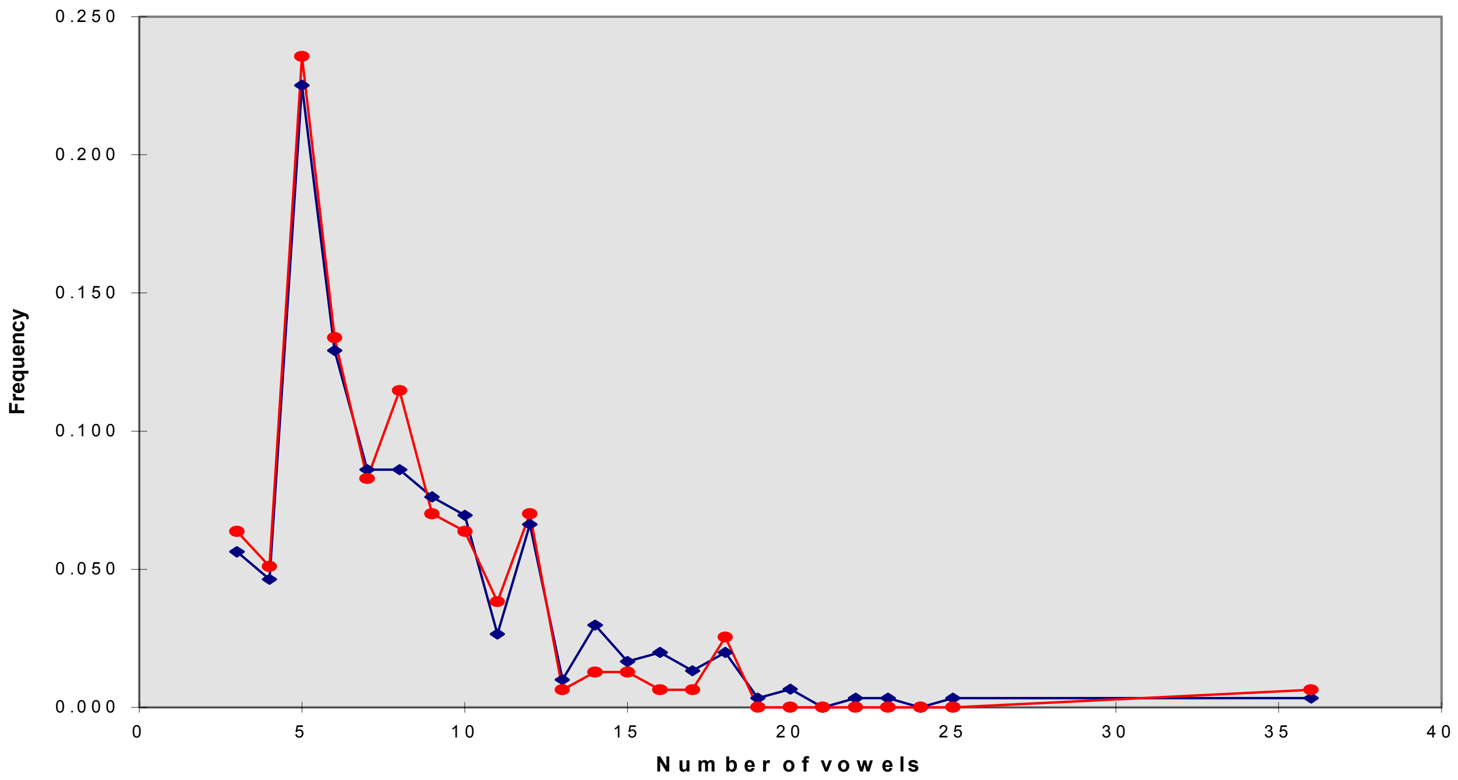
The mathematical analysis of the birth-and-death process is not a difficult task (in the introductory courses on probability and statistics, it most commonly figures as an exercise). It results in the following rule of thumb for potential impact of the birth-and-death process on a frequency of language type (see Maslova 2000 for mathematical details):

With a probability more than 0.95, the deviation *d(t)* of a typological frequency induced by the birth-and-death process during the last *t* years satisfies the following inequality:

$$d(t) < \frac{1}{\sqrt{N(t)}}$$

where *N(t)* is the estimated number of genetic groupings of time depth *t*, that is if *N(t)* is large, *d(t)* is very small. Furthermore, this maximum deviation can only be achieved for frequencies close to 0.50, since the formula above conceals a dependency of *d(t)* on the frequency itself: $d(t) \sim f(1-f)$, which means that *d(t)* achieves the maximum for $f = 0.5$. I will not go into mathematical details here; instead, I will show that the main prediction works for the test example. More specifically, I will show that if the sample were constructed as suggested by Perkins (that is, one language per family), the result would be roughly the same as for the essentially proportional sample under consideration.

The impact of the birth process for (~(-8000, -1500-1000))



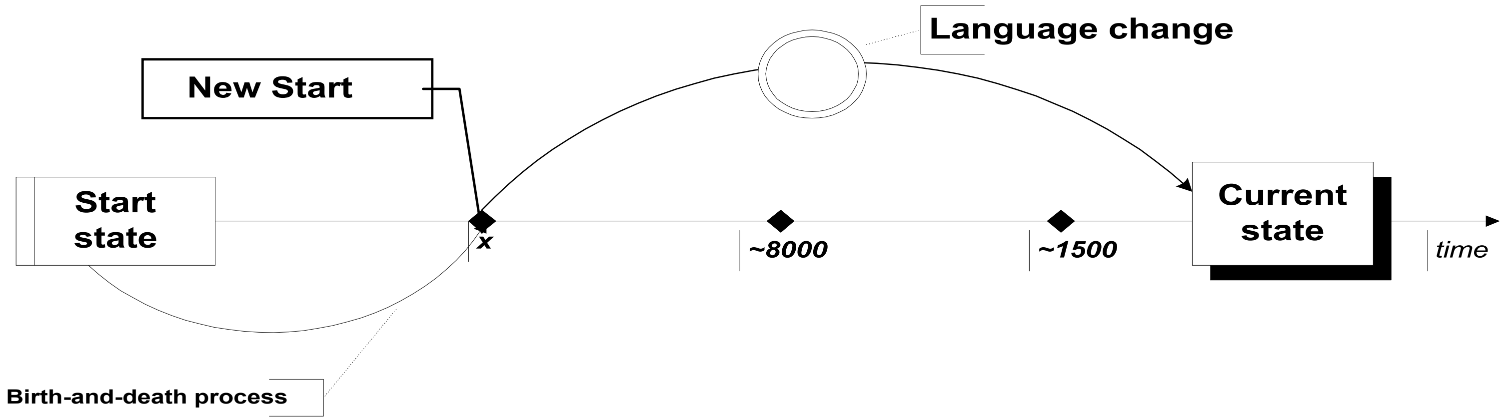
The red series with circles as data points is based on three different random sub-samples which include only language per family (according to *Ethnologue*). Thus, it serves as an estimate for the distribution that might have been observed if there had been no language splits within the given period of time. The blue series with squares shows the overall distribution in the sample. The maximum deviation is achieved for n = 8 and equals 0.03. According to the χ^2-test, each of the three sub-samples can be taken to display the same distribution as the sample as a whole.

Interim results

Thus, the overall impact of the language splits that occurred within the period of ca. 8000-1000 years ago on this particular typological distribution is shown to be negligible, exactly as predicted by the estimates based on general properties of the birth-and-death process. The same is most likely to be the case for the splits that occurred during the last 1000-1500 years (since there are even more genetic groupings of this time depth), which means that if the sample were just random, the attested distribution would be roughly the same.

Based on these estimates, we can, for the time being, just forget about the 'historical accidents' (birth-and-death process) that have occurred during the time span when the population has been large. Now let *x* be the (unknown) time moment when this historical period began. Although we do not know the exact value of *x*, it is apparently less than the time depth of the major families (Dryer 2000, Bill Croft, p.c.). It must be borne in mind, however, that before *x* (that is, when (if) the language population had been small), the effects of the birth-and-death process could be truly significant, e.g. it could easily induce a strong cross-linguistic 'preference' (Maslova 2000). With this in mind, the scheme for 'the history of a typological distribution' can be revised:

The history of a typological distribution revised



Before the ‘new start’, the effects of the birth-and-death process are too strong to license any significant impact of language change on the overall distribution, afterwards, the reverse is the case. The initial distribution (‘new start’) can be considered entirely “accidental”.

Now it is clear (if it was not before) that whenever we consider a typological distribution as a “linguistic fact”, we implicitly assume that it is determined by the corresponding type-shift process (= language change), that is, that it would be roughly the same for other initial distributions. This assumption is present (if not explicated) even if the linguistic explanation of a distribution is synchronic in essence.

An example: a possible type-shift process

	3	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18++**
3	0.64	0.06	0.17	0.05	0.02	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
4	0.04	0.52	0.17	0.06	0.06	0.05	0.05	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00
5	0.03	0.02	0.67	0.07	0.03	0.03	0.05	0.04	0.01	0.01	0.01	0.00	0.00	0.02	0.01	0.00
6	0.02	0.02	0.23	0.47	0.06	0.04	0.03	0.03	0.02	0.03	0.01	0.01	0.00	0.01	0.01	0.01
7	0.01	0.01	0.17	0.15	0.38	0.01	0.04	0.04	0.01	0.02	0.03	0.02	0.03	0.02	0.00	0.03
8	0.01	0.07	0.04	0.05	0.02	0.52	0.10	0.03	0.01	0.07	0.00	0.01	0.04	0.01	0.00	0.02
9	0.00	0.01	0.08	0.08	0.10	0.07	0.39	0.07	0.02	0.05	0.04	0.01	0.03	0.00	0.00	0.03
10	0.01	0.01	0.18	0.05	0.07	0.06	0.09	0.34	0.05	0.09	0.00	0.04	0.00	0.00	0.00	0.02
11	0.00	0.03	0.10	0.13	0.01	0.03	0.04	0.22	0.31	0.00	0.00	0.06	0.00	0.01	0.06	0.00
12	0.04	0.01	0.05	0.06	0.03	0.11	0.04	0.12	0.01	0.42	0.00	0.03	0.04	0.01	0.00	0.04
13	0.00	0.00	0.05	0.05	0.16	0.01	0.21	0.02	0.01	0.00	0.31	0.14	0.00	0.00	0.02	0.00
14	0.01	0.00	0.03	0.10	0.06	0.01	0.03	0.09	0.05	0.05	0.06	0.30	0.03	0.08	0.02	0.08
15	0.02	0.03	0.03	0.01	0.11	0.06	0.12	0.00	0.00	0.17	0.00	0.07	0.31	0.05	0.00	0.02
16	0.02	0.00	0.26	0.03	0.07	0.03	0.00	0.01	0.01	0.01	0.00	0.07	0.03	0.33	0.01	0.12
17	0.00	0.00	0.05	0.21	0.01	0.00	0.00	0.00	0.14	0.00	0.03	0.04	0.00	0.01	0.36	0.13
18++	0.00	0.00	0.01	0.02	0.10	0.02	0.04	0.05	0.01	0.10	0.00	0.06	0.02	0.06	0.04	0.47

Each row corresponds to a possible source state, and and each column, to a possible target state of the transition process. A cell shows the probability of a transition from the given source state to the given target state within a certain period of time (e.g. 1000 years). The diagonal cells show the probability to retain the given type (within the same period), cf. Greenberg’s “stability” parameter.

The steady-state distribution and linguistic significance

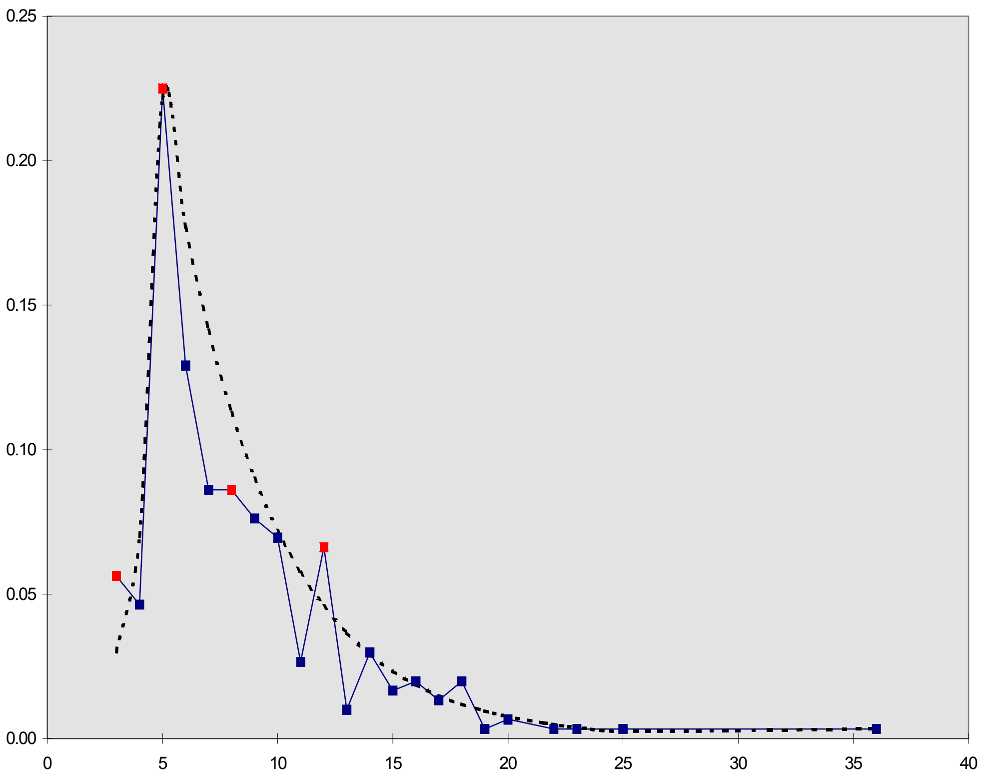
If the type-shift process for the typology under consideration is adequately described by this transition matrix, then it is bound to bring about a typological distribution very close to the attested one, independently of the initial distribution. In other words, the attested distribution is close to the **stationary** (or steady-state) distribution **determined** by this transition process.

Any claim of linguistic significance of a cross-linguistic distribution (or, for that matter, any linguistic interpretation of this distribution) implies that the language population achieved the steady-state distribution for the given typology (or at least came close to this distribution).

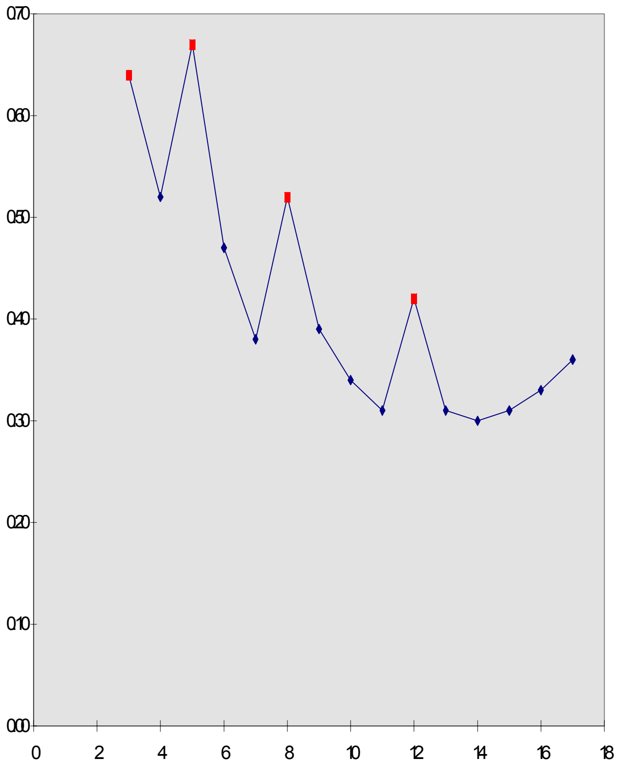
The other way round, in order to decide whether a distribution is linguistically significant, we must be able to test whether it approximates the steady-state distribution of the underlying type-shift process. Apart from one special case to be described shortly, such a test will necessarily involve some estimates for transition probabilities. Note that if we are able to assess these probabilities, this will give a lot of additional information for linguistic explanations and inferences.

An example: “Stability” as an explanatory factor

Synchronic Frequency



Stability



Apart from the mode (5), the curve for synchronic frequency shows several points of non-monotone behavior (3, 8, 12). As shown by the right-hand curve, these points correspond to the peaks of ‘stability’ (~ the probability to retain the same number of vowels).

Is this distribution stationary?

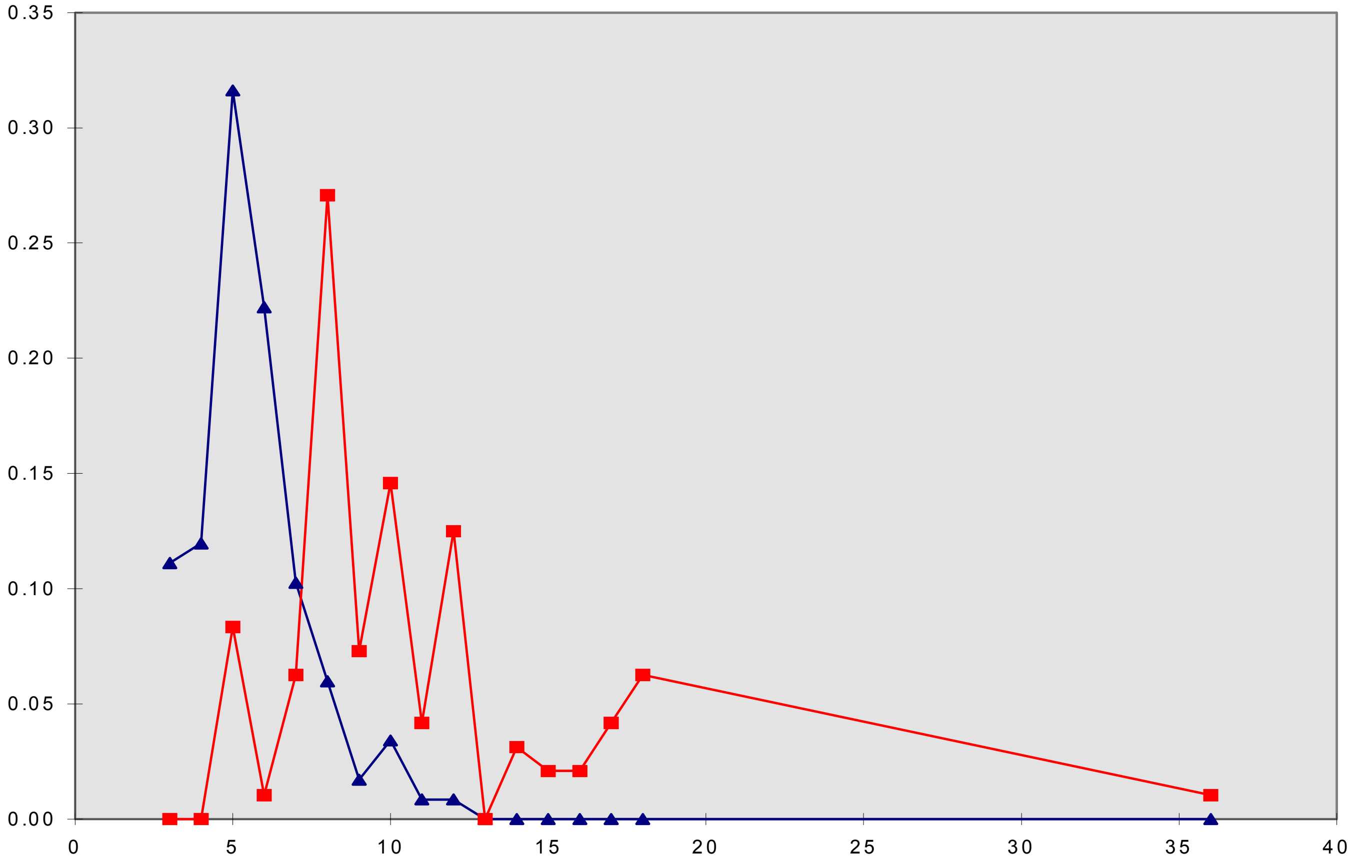
In principle, there are two ways to decide whether an attested distribution is close to the steady-state distribution of the type-shift process:

- One possibility is that the time period of a few thousand years (say, 8000 or 10000), which separates the current language population from the earlier population that included ancestors of the major families (below, A-population), can be shown to be sufficiently long for the type-shift processes to bring about the steady-state distribution for ***T***. This hypothesis can be verified (or falsified) without invoking the transition matrix directly (unfortunately, we have no evidence to test this hypothesis (directly) for a longer time span).
- If this time period is not sufficiently long (i.e. a dependency on the properties of the A-population is retained by the current population), then a test for stationary distribution must be based on some hypotheses on transition probabilities. The major challenge is to derive plausible hypotheses of this sort on the basis of typological data.

I will start with exploring the first option, in order to show that a transition-based test for stationary distribution is an inescapable task. This will mean that we cannot assume that the ‘recent’ history (~10000 years) has been enough to achieve steady-state distributions of typological parameters, and we can only hope that the real history of the language population has been longer than that.

- There is a straightforward way to verify the first hypothesis. If the time period separating the existing language population from the A-population is sufficient to achieve the steady-state distribution for ***T***, this means, by definition of the steady-state distribution, that the current state of a language (with respect to ***T***) is stochastically independent from the state of its ancestor in the A-population. This entails that, for any subset ***R*** of the A-population, the current distribution for ***T*** among its descendants must be roughly **the same** as the overall distribution. Indeed, by choosing a subset ***R*** of the A-population, we choose certain initial states. If the current type of a language is indeed independent of its initial state, then the distribution must be the same for all initial states, hence, for any ***R***. Now let ***P(R)*** be the distribution displayed by the descendants of ***R***. If we find two subsets (R_1 and R_2) such that $P(R_1)$ significantly differs from $P(R_2)$, then hypothesis is falsified, that is, the relevant time period is not sufficiently long to ensure that the steady-state distribution is achieved by the population. Note that a single pair of subsets with this property will be enough to falsify the hypothesis. It was not a difficult task to split the genetic groupings represented in the sample into such subsets. The resulting distributions $P(R_1)$ and $P(R_2)$ look as shown by the following chart.

P(R$_1$*)* and *P(R*$_2$*)*



- The curves are so obviously different that it seems superfluous to apply any statistical tests. However, the value of χ^2 is 118.3 (the maximum acceptable value for 15 degrees of freedom is 37.7), which means that the hypothesis of independence of the current type from the initial type is rejected.

Two subsets of major families

R_1-families (ancestors) [‘blue distribution’ (triangles)]

- Afro-Asiatic
- Araucanian
- Arawakan
- Australian
- Austronesian
- Aymaran
- Barbacoan
- Basque
- Carib
- Chimakuan
- Chukotko-Kamchatkan
- East Papuan
- Eskimo-Aleut
- Gulf
- Hokan
- Japanese
- Mataco-Guaicuru
- Mayan
- Muskogean
- Panoan
- Penutian
- Quechuan
- Sino-Tibetan
- South Caucasian
- Tacanan
- Tarascan
- Totonacan
- Trans-New Guinea
- Uto-Aztecan
- Witotoan
- Yenisei Ostyak
- Yukaghir
- Yuki

R_2-families (ancestors) [‘red distribution’ (squares)]

- Algic
- Altaic
- Andamanese
- Austro-Asiatic
- Caddoan
- Chibchan
- Coahuiltecan
- Dravidian
- Guahiban
- Indo-European
- Iroquoian
- Jivaroan
- Keres
- Khoisan
- Kiowa Tanoan
- Macro-Ge
- Mixe-Zoque
- Na-Dene
- Nambiquaran
- Niger-Congo
- Nilo-Saharan
- North Caucasian
- Oto-Manguean
- Paezan
- Siouan
- Tai-Kadai
- Tucanoan
- Tupi
- Uralic
- Zaparoan

NB: what I have just demonstrated is the well-known **stochastic dependency** of the language type on its genetic affiliation. Note that the dependency is demonstrated by comparing two samples of isolated (“independent”) languages, i.e. it cannot be surmounted by any sampling procedure.

Reconstruction of transition probabilities and tests for stationary distribution

Thus, we have to turn to the second option, that is, we need some reconstructions of transition matrices, even if only to test the hypothesis of linguistic significance of a synchronic distribution.

Generally, it is theoretically impossible to infer transition probabilities from a synchronic distribution (hence, to decide whether this distribution is the stationary one) without some addition information on the transition process itself. Yet we do have some additional information, namely, the information on the genetic relationships between languages. More specifically, for any group of related languages, we know that they have originated from the same type-state (although, in the general case, not what this original type-state was). My basic idea was to find a way to generate hypotheses on transition probabilities on the basis of this additional information. In other words, the question is, whether we can infer any information on transition probabilities from **synchronic** typological data, in particular, whether it is possible to test the hypothesis of stationary distribution.

The answer is that, in principle, this is a soluble problem. Furthermore, I will claim that the attested distribution of the number of vowels per language is close to the steady-state distribution determined by the corresponding type-shift process.

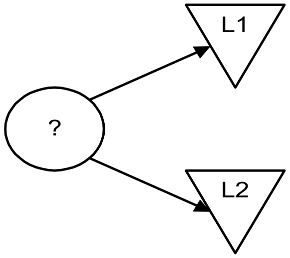
Reconstructing transition matrices: random search

As far as I was able to figure out, this type of problem is not solved by the mathematical statistics (probably not because it is so difficult, but just because the statistics solves problems posed by other disciplines). The challenge was essentially mathematical, and it is only to a limited extent that I was able to meet this challenge (so far). At the present time, I have two algorithms that "reconstruct" transition processes on the basis of the synchronic typological data and the corresponding sub-tree of the genetic classification and calculate the stationary distributions. The *Ethnologue* is used as the model of genetic classification (despite its limited applicability for this task).

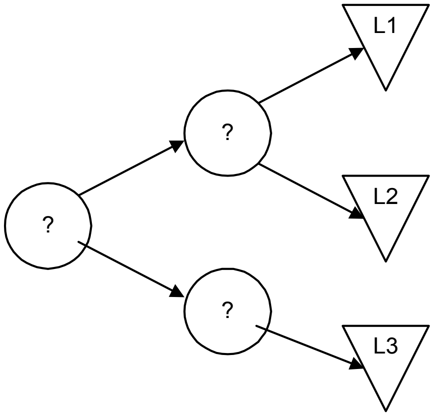
The first algorithm is essentially a random search for most probable transition matrices, based on the whole genetic (sub)-tree. The transition matrix shown before is generated by this algorithm. The main drawback of this algorithm is that I cannot offer any precise estimate of its robustness (it generates some 'answers' even for those cells of the transition matrix where evidence is not sufficient to make it reliable.) Some of the answers, however, do represent plausible (and testable!) hypotheses. Furthermore, this algorithm consistently **confirms** that the attested distribution is close to the steady-state distribution of the underlying type-shift process, i.e. **any** process generated by the algorithm (ca. 10000 trials) has a steady-state distribution close to the attested distribution.

Two-types test for stationary distribution

The major advantage of the second algorithm is that its results license an exact mathematical interpretation: it is more like a usual parametric statistical test, whereby the transition probabilities are taken to be unknown parameters. Its basic idea is to look not only at the frequencies of **single** languages of each type (as is usual), but also at the frequencies of **type sets** attested in the groups of (most) closely **related** languages (the only thing that is assumed is that genetically related languages originate from the same type, no assumptions on what this type was). The only version of this algorithm that could be tested on the basis of this sample makes use of derived samples of **pairs** and **triples** of related languages and works **only** for two-way typologies:



Three types of pairs:
AA, BB, AB



Six types of triples:

(AA)A, (AA)B,

(BB)A, BB(B),

(AB)A, (AB)B

Interpretation and limitations

Although the algorithm can be applied only to two-way typologies, it can serve as a test for stationary distribution for the entire typology: since the parameter is quantitative, we can easily “scan” the typology, that is, apply the algorithm to all possible type boundaries. This procedure confirms that the distribution is close to the stationary one.

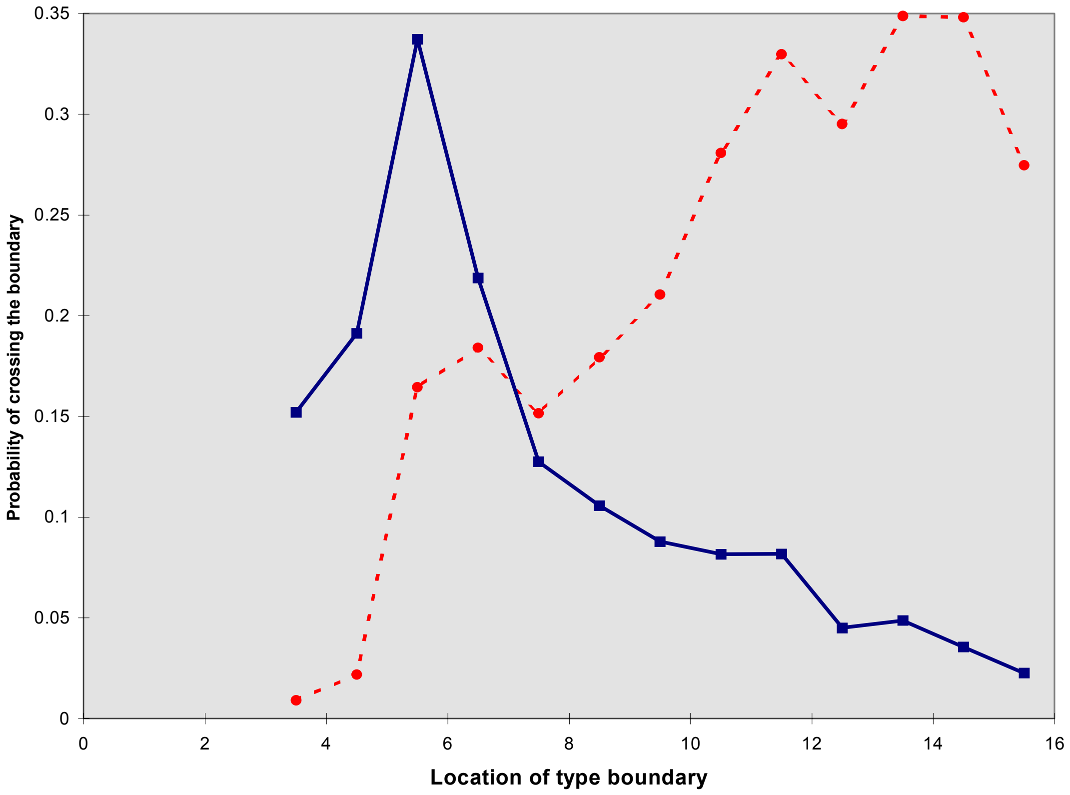
Unfortunately, the estimates for transition probabilities obtained by means of this procedure are of limited linguistic significance: for each *n*, we know the probability for a language that has *n* or less vowels to shift towards a state with more than *n* vowels, and vice versa. The linguistic interpretation of such data is hindered by the fact that many very different shifts are ‘merged’ together (cf. these data with the ‘complete’ transition matrix generated by the first algorithm). In other words, we have reliable information on the probabilities to cross various ‘boundaries’ (in a given direction), but not on the actual source states and target states of languages crossing the boundary. However, some interesting observations appear to emerge even on the basis of this limited information.

The results of the two-types test for various type boundaries

n	p_1	p_{21}	p_{12}	χ^2 for triples	
3	0.06	0.01	0.15	7.90	5%
4	0.10	0.02	0.19	9.99	5%
5	0.33	0.16	0.34	5.46	5%
6	0.46	0.18	0.22	3.26	5%
7	0.54	0.15	0.13	3.29	5%
8	0.63	0.18	0.11	8.47	5%
9	0.71	0.21	0.09	15.28	0.1%
10	0.77	0.28	0.08	16.48	0.1%
11	0.80	0.33	0.08	13.14	1%
12	0.87	0.30	0.05	17.33	0.1%
13	0.88	0.35	0.05	19.12	0.1%
14	0.91	0.35	0.04	17.50	0.1%
15	0.92	0.27	0.02	14.27	1%

This table shows the results of the two-types test for stationary distribution for various boundaries between types (*n*). For each *n*, $\boldsymbol{p_1}$ is the hypothetical steady-state probability for a language to have *n* or less vowels, $\boldsymbol{p_{12}}$ and $\boldsymbol{p_{21}}$ are the transition probabilities estimated on the basis of the sample of pairs of most closely related languages: $\boldsymbol{p_{21}}$ is the probability for a language that has more than *n* vowels to have or less vowels after ~1000-1500 years, $\boldsymbol{p_{12}}$ - the probability of the reverse transition. The last two columns give the results of χ^2 test based on the sample of “triples”. It can be easily seen that the triples are distributed as predicted by the hypothesis of stationary distribution (that is, the test does not reject this hypothesis).

‘Traffic’ through various boundaries



Blue series - crossing from ‘small’ to ‘large’

Red (dashed) series - crossing from ‘large’ to ‘small’

One interesting fact is that the probability for a language of the ‘small’ type to shift towards the ‘large’ type does not decrease monotonically as the boundary shifts to the right (i.e. from frequent ‘large’ types to rare ‘large types), but has a clear mode at the boundary between 5 and 6-vowel systems, which means that the 5-vowel systems are most likely to be ‘enhanced’ by additional vowels. An implication is that the mode of the steady-state distribution is ensured not only by the stability of this system, but also by a strong drift in the opposite direction.

Historical interpretation and its linguistic implications

According to Comrie, the typological method assumes that “at least within time span of several thousand years in either direction from the present, there has been no significant sense in which human language has evolved, i.e. no sense in which human language as a whole today is different in essence from that of ten thousand years ago” (1989:9). An assumption of this sort is in fact also associated with the (unknown) temporal point referred to as “new start” in the scheme above (S.12): *x* can be viewed as the start of the population’s drift toward the steady-state distribution only if the transition probabilities have remained the same since this moment of time. Now what the test example has shown is this:

(a) ten thousand years is **not enough** to achieve the steady-state distribution,
but (b) the language population **has achieved** this distribution.

The historical implication of this result is quite trivial: at least as far as phonology is concerned, the same ‘level of evolution’ can be assumed for a significantly greater time span than ten thousand years from the present. In this sense, the result has more far-reaching linguistic implications than just demonstration of linguistic significance of one specific cross-linguistic distribution: if the population has had enough time to achieve at least one steady-state distribution, it is (more) likely that it has achieved such distributions for other typological parameters as well.

Implications for typological sampling

- The reason why the cross-linguistic data which served as a test example license testing the hypothesis of stationary distribution and some (albeit limited) inferences on transition probabilities is that the sample does include a relatively large number of groups of (relatively) closely related languages, which made it possible to derive reasonably large sub-samples of “pairs” and “triples”, something that a Perkins-like sample would not allow ‘by definition’.
- The limitations on linguistically interesting inferences on transition probabilities are partly determined by the fact that the most closely related languages in the sample are too distant, i.e. the time period for which the transition probabilities (1000 - 1500 years) can be calculated is longer than desirable for an accurate exploration of the transition process: as can be inferred from this sample, a vowel system can easily undergo more than one modification within this period. Otherwise, the inferences on transition probabilities could be extended (at least) to three- or four-way typologies.
- Whether we want just to test if a distribution is linguistically motivated, or also to explore the diachronic implications of cross-linguistic data, we need to sample groups (at least ‘pairs’ and ‘triples’) of closely related languages (or even dialects), not (or not only) ‘independent’ languages.

References

- Bell, Alan (1978). Language sampling. In: Joseph H. Greenberg et al. (eds.) *Universals of human languages*. Vol. 1. *Method & Theory*. Stanford: Stanford University Press. 125-156.
- Comrie, Bernard (1989). *Language universals and linguistic typology*. (2 ed.) Oxford: Blackwell Publishers.
- Dryer, Matthew S. (2000) A response to (Maslova 2000). *Linguistic Typology 4-3*.
- Greenberg, Joseph H. (1978) Diachrony, synchrony and language universals. In: Joseph H. Greenberg et al. (eds.) *Universals of human languages*. Vol. 1. *Method & Theory*. Stanford: Stanford University Press, 61-91
- Greenberg, Joseph H. (1995) The diachronic typological approach to language. In: Masayoshi Shibatani and Theodora Bynon (eds.) *Approaches to language typology*. Oxford: Clarendon Press. 143-166.
- Grimes, Barbara F. (ed.) (1997) *Ethnologue: Languages of the world (plus Supplement: Ethnologue Index).* 13th edition. Dallas: Summer Institute of Linguistics.
- Hawkins, John A. (1990) Seeking motives for change in typological variation. In: Croft, William, Keith Denning and Suzanne Kemmer (eds.) *Studies in typology and diachrony. Papers presented to Joseph H. Greenberg on his 75th birthday. [Typological Studies in Language 20].* Amsterdam, Philadelphia: John Benjamins. 95-128.

- Maddieson, Ian (1984) *Patterns of Sounds.* Cambridge Studies in Speech Science and Communication. Cambridge University Press.
- Maddieson, Ian (1991) Investigating Linguistic Universals. *Proceedings of the XIIth International Congress of Phonetic Sciences, vol. I.* Aix-en-Provence: Service des Publications, Université de Provence, 346-354.
- Maslova, Elena (2000) A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4-3.
- Perkins, Revere D. (1989) Statistical techniques for determining language sample size. *Studies in Language, 13*: 293-315.
- Rijkhoff, Jan & Dik Bakker (1998) Language sampling. In: *Linguistic Typology 2-3*, 263-314.
- Tomlin, Russel S. (1986) *Basic word order: functional principles*. London et al.: Croom Helm. *[Croom Helm Linguistic Series].*
- Whaley, Lindsay J. (1997) Introduction to Typology. The Unity and Diversity of Language. SAGE Publications.