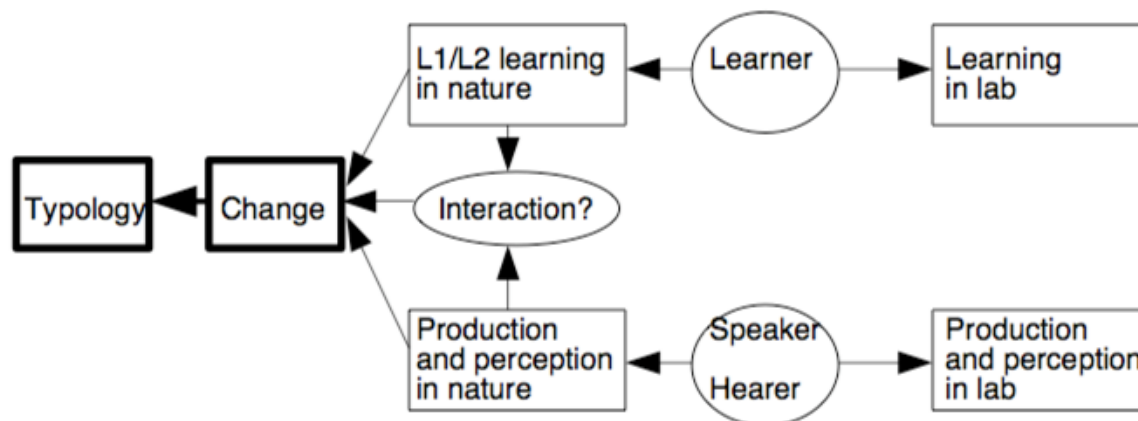


(1) Where we are:



(2) Preview

- a. Defining typological frequency.
- b. Relation of typological frequency to historical change.
- c. Application to concrete example (height-height and height-voice patterns).
- d. Problems of scale, feasibility, and explanatory value.

1 The Markov model (Bell, 1970)

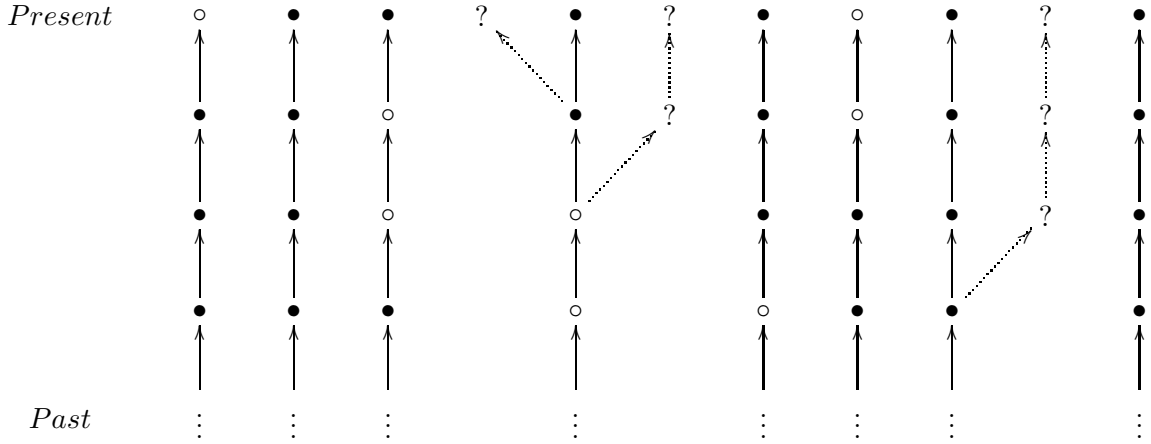
(3) Basic idea: Types are states in a Markov process, and their frequencies are its (unique) stationary distribution (Bell, 1970, 1971; Greenberg, 1978; Griffiths and Kalish, 2007; Griffiths et al., 2008).
⇒ Typological frequencies emerge from transition frequencies.

(4) Consider the simplest possible typological fact: S_1 occurs with probability π_1 , and $S_2 = \text{not-}S_1$ with probability $\pi_2 = 1 - \pi_1$. E.g., having vs. lacking /h/ in the inventory (Maddieson, 1984, §3.9):

S_1	Lacks /h/	$\pi_1 = 0.37$
S_2	Has /h/	$\pi_2 = 0.63$

What accounts for π_1 ?

(5) Take any present-day language and follow it back in time to get a lineage.



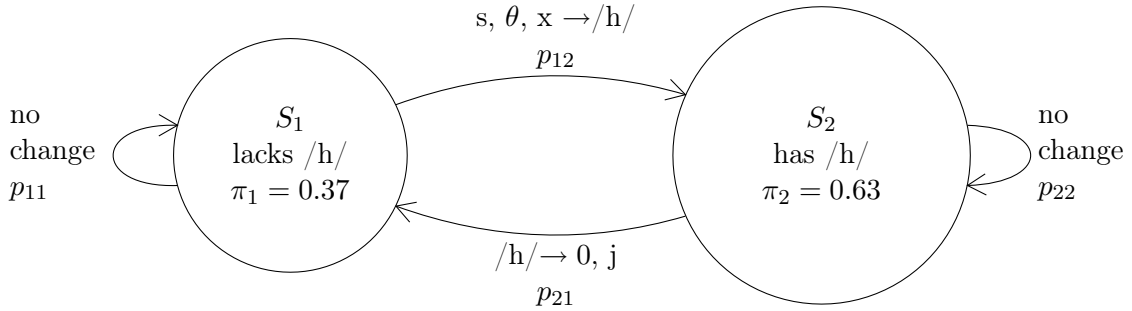
Languages gain and lose [h] over time, i.e., lineages wander from type to type.

(6) Assume typology is in equilibrium: Typological frequencies don't change as you go back in time. E.g., no long-term trend toward [h]-lessness.¹

(7) **Equilibrium** \Rightarrow **Innovations** = **Extinctions**. The number of languages going from “has-/h/” to “lacks-/h/” per unit time has to equal the number going the other way (otherwise, one type would grow and the other would shrink). If there are N lineages, then

$$N\pi_1 p_{12} = N\pi_2 p_{21}$$

where p_{ij} is the probability that a lineage will be in Type S_j on the next tick, given that it was in Type S_i on the current tick.



(8) The transition probabilities *completely determine* the typological frequencies:

$$\pi_1 = \frac{p_{21}}{p_{12} + p_{21}}, \quad \pi_2 = 1 - \pi_1$$

The typological frequencies *completely determine* the transition probabilities, up to a constant factor:

¹This might be wrong, if long-term social changes have systematic linguistic effects (Hay and Bauer, 2007).

$$\frac{p_{12}}{p_{21}} = \frac{\pi_2}{\pi_1} = \frac{1 - \pi_1}{\pi_1}$$

(9) I.e., constant typological frequencies \iff constant (ratio between) transition probabilities. Constant typology means that change may happen faster or slower at different times, but that the bias between different changes is always the same.

(10) When we scale up to typologies with more than two states,

- a. For a given transition matrix $\mathbf{P} = (p_{ij})$, there is only one equilibrium state $\bar{\pi}$, i.e., \mathbf{P} still completely determines $\bar{\pi}$. No matter how many languages are in each type “to begin with”, the frequencies converge to the equilibrium state as time passes.
- b. If the off-diagonal entries of \mathbf{P} differ from the corresponding entries \mathbf{Q} by a constant factor (i.e., if p_{ij}/q_{ij} is constant for $i \neq j$), then \mathbf{P} and \mathbf{Q} have the same equilibrium state. (However, it is no longer true that $\bar{\pi}$ uniquely determines the p_{ij} up to a constant factor.)
- c. The relationship between \mathbf{P} and $\bar{\pi}$ becomes very opaque, and we have to use numerical methods (i.e., we need specific numbers, not variables).

(11) The Markov model follows almost inevitably from assuming that typological frequency is constant through time, and that contact effects are small enough not to matter much.

Cogency does not depend on, e.g., beliefs about the existence of Universal Grammar, or of any grammar at all.

(12) Markov model accounts for $\bar{\pi}$ via \mathbf{P} , and for the stability of the former via the stability of the latter.

\Rightarrow Need to ask: What determines \mathbf{P} , and keeps it stable?

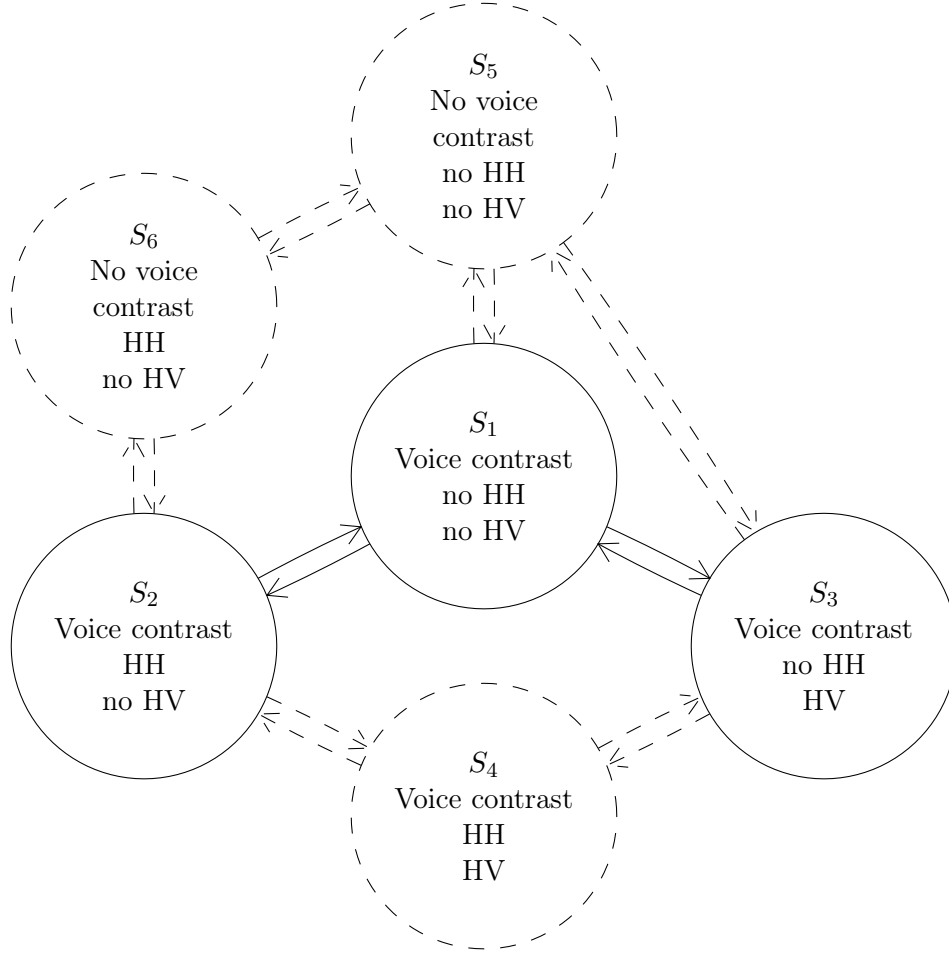
2 Example: height-height and height-voice interactions

(13) Comparing two analogous phonological patterns:

HH pattern Predictive dependency between vowel height in adjacent syllables (harmony or disharmony in height between V_1 and V_2 in $V_1C_0V_2$). Seems to be rather frequent.

HV pattern Predictive dependency between vowel height and “voicing” (= phonetic voicing, aspiration, or fortis/lenis contrast) of immediately-following obstruent (V_1C_1). Seems to be rare.

(14) Here’s a diagram of relevant typological states.



- a. The HV pattern can only occur in languages which have a voicing contrast, hence the “missing” state.
- b. S_4 is empirically rare; ignore it.
- c. Maddieson (1984, 26) suggests that about 84% of languages have a voicing contrast, so S_5 and S_6 are small relative to S_1 and S_2 .
- d. Moreover, we have no reason to think that innovating a voicing contrast interacts with innovating the HH pattern $\Rightarrow \pi_2/\pi_1$ is unaffected by deleting S_5 and S_6 .
- e. Also assume that it is negligibly likely for a language to acquire both a voice contrast and the HV pattern at the same time.
- f. \Rightarrow Delete S_4, S_5, S_6 to yield a three-state typology:
 - (i) One big state that most languages are in (S_1)
 - (ii) Two small states, perhaps of unequal sizes.

(15) The Markov model now tells us that

$$\begin{aligned}\frac{\pi_2}{\pi_3} &= \frac{\pi_2}{\pi_1} \cdot \frac{\pi_1}{\pi_3} \\ &= \frac{p_{12}}{p_{21}} \cdot \frac{p_{31}}{p_{13}}\end{aligned}$$

⇒ If we knew π_2/π_3 , we could make predictions about the transition probabilities (and hence about the factors that determine them).

(16) Since π_1 isn't involved in π_2/π_3 , we only need to count *positive* instances of S_2 and S_3 . (Don't need to take N languages and see how many are in S_1 , S_2 , and S_3 .)

Typological surveys: Brute-force search, aided by secondary literature and p.c.s. Restrictions:

- a. Limited to languages in which both patterns have opportunity to occur, i.e., languages with lexical contrasts in both relevant features.
- b. No phonetic confounds (glottalization, prenasalization, etc.).
- c. Pattern must neutralize contrast in some environment (excludes allophony, insures pattern isn't just phonetic).
- d. Alternations limited to single morphemes did not qualify.
- e. Language must have been described while still alive.
- f. Counted “families” (= top-level Ethnologue categories) rather than individual languages, so that each case represented a single independent lineage.

(17) *No* HV cases fit the criteria perfectly, so results are presented in two tiers, “strict” (clearly fits criteria) and “lax” (questionable on one or more criteria).

Height- Height (S_2)	<i>Strict</i> : Afro-Asiatic (Awngi), Altaic (Udihe), Basque (Basque), Indo-European (Buchan Scots), Niger-Congo (C'Lela), Oto-Manguean (Malinaltepec Tlapaneca), Sino-Tibetan (Lhasa Tibetan). <i>Lax</i> : Austronesian (Woleiaian), Chukotko-Kamchatkan (Chukchee), Dravidian (Tamil), Gulf (Tunica), Hokan (Washo), Korean (Korean)	7 +8
Height- Voice (S_3)	<i>Strict</i> : None. <i>Lax</i> : Indo-European (Polish, Canadian English), Sino-Tibetan (Lungtu Fujien Chinese)	0 +2
Both (S_4)	<i>Strict</i> : None. <i>Lax</i> : Nilo-Saharan (Murle)	0 +1

(18) 95% confidence intervals (using R's `binom.bayes` function in the `binom` package, with Jeffreys prior):

	$\frac{\pi_2}{\pi_2 + \pi_3}$			Equivalent $\frac{\pi_2}{\pi_3}$		
Sample	Lower	Mean	Upper	Lower	Mean	Upper
Strict (7:0)	0.768	0.938	1	3.31	15.1	∞
Lax (14:16)	0.656	0.853	0.973	1.91	5.80	36.0

- a. We can be pretty sure that $\pi_2 > \pi_3$.
- b. Estimates of how much bigger are fuzzy and depend on how you count, but probably π_2 is at least twice π_3 , maybe much larger.

(19) So we can conclude that the HH pattern is either more likely to arise, or less likely to be lost, than the HV pattern:

$$\frac{p_{12}}{p_{13}} \cdot \frac{p_{31}}{p_{21}} > 3.31$$

If $a \cdot b > c$, then at least one of a or b must be $> \sqrt{c}$, so

$$\Rightarrow \underbrace{\frac{p_{12}}{p_{13}} > 1.82}_{\text{more likely to arise}}, \quad \text{or} \quad \underbrace{\frac{p_{31}}{p_{21}} > 1.82}_{\text{less likely to be lost}}, \quad \text{or both.}$$

- a. Direct predictions about frequency of historical changes
- b. Indirect predictions about the factors that determine p_{ij} once we have a hypothesis about how they determine it. (See next handout.)

3 Comments on the Markov model

(20) Preview of comments:

- a. Can tolerate some contact effects.
- b. Impractical and unilluminating for more complex typologies \Rightarrow Need some way to reduce the number of explanatory variables. Possibilities:
 - (i) Ignore outgoing transitions.
 - (ii) Assume sound change probabilities are independent of typological states.
 - (iii) Find out what determines the transition probabilities.

3.1 Contact effects

(21) The model assumes that each language is descended from one and only one immediate ancestor. I.e., it ignores contact phenomena such as creolization, borrowing, and areal transfer (Bell (1970, 221); Greenberg (1978)).

However, the model *can* tolerate certain contact phenomena.

(22) Suppose languages encounter each other randomly, such that the probability at time t that L in state S_i encounters some L' in state S_j is proportional to $\pi_j(t)$, the typological frequency of S_j at time t . And suppose further that on each encounter, one of the following happens:

a. *Nothing*.

b. *Assimilation*: L moves to S_j . Models transfer of a feature from L' to L .

Since the probability of assimilation to S_j is proportional to $\pi_j(t)$, assimilation moves languages between types but doesn't change type frequencies.

c. *Saltation*: L moves to another type S_k with some fixed probability r_k . E.g., loss of tone in pidginization, even between two tone languages (Bakker, 1994).

Since the probability of saltation is the same for all i, j , and t , we can simply incorporate it into the p_{ij} s.

$\Rightarrow \bar{\pi}$ is still determined by \mathbf{P} , which is constant at all times.

(23) Some ways contact could still invalidate the model:

a. Outcome of contact between S_i and S_j depends on S_i and S_j in some more complex way. E.g., if creole vowel systems tend towards the intersection of the source-language vowel systems (Winford, 2003, 320)

b. Long-term changes in rate of contact introduces dependency on t . (Increasing mobility, decreasing diversity, etc.)

\Rightarrow Model is probably wrong; have to hope that it isn't wrong enough to matter very much. (We'll find out eventually if it is.)

3.2 Scale and feasibility

(24) The Markov model explains $\bar{\pi}$, a vector with m entries, using \mathbf{P} , a matrix with m^2 entries.

(Sounds like a bad bargain, but it does reduce by m the number of things we have to explain.)

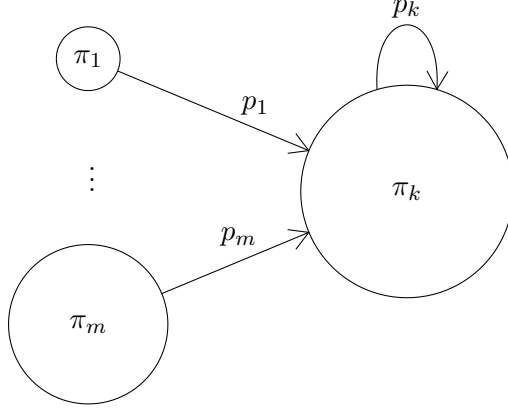
(25) How feasible is this in the real world?

a. If we want to predict $\bar{\pi}$ with a given precision, how many samples do we need of \mathbf{P} ? These are *historical transitions*, for which there is no UPSID. Is there enough data in the whole world?

b. How many states can our typology have and still be testable?

(26) Start with a more modest objective: We want to predict π_k , the frequency of S_k , from

- The frequencies of the other types, $(\pi_1, \dots, \pi_{k-1}, \pi_{k+1}, \dots, \pi_m)$, and
- The transition probabilities $(p_{1,m}, \dots, p_{m,m})$, hereinafter (p_1, \dots, p_m) for short.



- a. The predictive relationship is $\pi_k = (\sum_{i \neq k} \pi_i p_i) / (1 - p_k)$.
- b. We get all of the $\hat{\pi}_i$, our estimates of the π_i , from a single sample of N languages.
- c. We get each \hat{p}_i by sampling n_i historical languages in S_i to see how many of them change to S_m in the time unit.
- d. The rarer S_i , the harder it is to find cases, so say $n_i \approx n\pi_i$, where n is the total number of languages in our time-lapse sample. This means we get more info about changes out of more-frequent types (which have more influence on π_k).
- e. Applying the delta method (Agresti, 1990, 418–425) to Equation (a), we have, after doing some algebra and throwing away negative terms,

$$\text{Var}(\hat{\pi}_k) \leq \frac{\pi_k}{1 - p_k} \left(\frac{1}{N} + \frac{1}{n}(1 + p_k) \right)$$

(27) Example: Suppose we have a four-state typology with the transition matrix

$$\mathbf{P} = \begin{pmatrix} 0.6033 & 0.2054 & 0.1884 & 0.0029 \\ 0.7174 & 0.0004 & 0.0961 & 0.1862 \\ 0.2113 & 0.0074 & 0.5958 & 0.1855 \\ 0.6462 & 0.0515 & 0.0244 & 0.2779 \end{pmatrix}$$

We follow the procedure given in (26) 1000 times, with $N = n = 100$, each time calculating a predicted value for each π_i . The predicted and empirical sample standard deviations are

k	π_k	s.d. = $\sqrt{\text{Var}(\hat{\pi}_k)}$	
		Eqn (26e)	Actual
1	0.5137	0.1836	0.1507
2	0.1128	0.0475	0.0326
3	0.2724	0.1323	0.1288
4	0.1011	0.0565	0.0479

(28) The sampling variance relates precision to sample size:

- a. We can get $(1 - \alpha)$ probability that $|\hat{\pi}_k - \pi_k| \leq \delta \pi_k$ if we take N and n big enough that $z_{\alpha/2} \sqrt{\text{Var}(\hat{\pi}_k)} \leq \delta \pi_k$, or

$$\frac{1}{N} + \frac{1}{n}(1 + p_k) \leq (1 - p_k)\pi_k \left(\frac{\delta}{z_{\alpha/2}} \right)^2$$

- b. To satisfy that, we need

$$N \geq \left(\frac{z_{\alpha/2}}{\delta} \right)^2 \frac{1}{\pi_k} \frac{1}{1 - p_k}$$

and

$$n \geq \left(\frac{z_{\alpha/2}}{\delta} \right)^2 \frac{1}{\pi_k} \frac{1 + p_k}{1 - p_k}$$

N.B. These bounds do not depend (directly) on the number of states in the typology, or on any probabilities except π_k and p_k !

(29) Historical samples are probably harder to get than contemporary ones, so let's focus on n as the limiting practical factor.

$$n \geq \underbrace{\left(\frac{z_{\alpha/2}}{\delta} \right)^2}_{\text{precision}} \cdot \underbrace{\frac{1}{\pi_k}}_{\text{frequency}} \cdot \underbrace{\frac{1 + p_k}{1 - p_k}}_{\text{stability}}$$

\Rightarrow Rarer and more-stable types need bigger samples because transitions involving that state are harder to observe. (And of course precision is costly.)

a. $\alpha = 0.05, \delta = 1$				b. $\alpha = 0.05, \delta = 1/2$				c. $\alpha = 0.025, \delta = 1/2$			
π_k	p_k			π_k	p_k			π_k	p_k		
	0.1	0.5	0.9		0.1	0.5	0.9		0.1	0.5	0.9
0.1	47	115	730	0.1	188	460	2920	0.1	246	603	3818
0.5	9	23	146	0.5	38	92	583	0.5	49	121	764
0.9	—	—	81	0.9	—	—	324	0.9	—	—	424

(N.B. Exports from a state have to go somewhere: $\pi_k(1 - p_k) \leq 1 - \pi_k$. Hence the empty cells. High-frequency types *must* be stable.)

(30) This is the cost of predicting the frequency of *one* type. What if we want to do them all?

- There are bounds on precision of $\hat{\pi}$ in terms of precision of \mathbf{P} (O'Connell, 1993; Xue, 1997), but they give huge overestimates through not being probabilistic.
- Empirically (simulations with randomly-generated \mathbf{P} s), the variance of each $\hat{\pi}_k$ obtained by estimating $\bar{\pi}$ as the stationary state of \mathbf{P} is almost always smaller than obtained via the piecemeal procedure of (26). E.g., the random matrix from (27):

s.d. = $\sqrt{\text{Var}(\hat{\pi}_k)}$				
k	π_k	Eqn (26a)	Actual	
			Piecemeal	Stationary-state
1	0.5137	0.1836	0.1507	0.0668
2	0.1128	0.0475	0.0326	0.0304
3	0.2724	0.1323	0.1288	0.0739
4	0.1011	0.0565	0.0479	0.0409

\Rightarrow (29) may apply equally well to the sample size needed to realize the Grand Vision. Suppose it does.

- c. To get $1 - \alpha$ probability that *all* $|\hat{p}_k - p_k| \leq \delta p_k$, we need $1 - \alpha/m$ probability that *each* is (“Bonferroni correction”), so

$$n \geq \max_k \left(\frac{z_{\alpha/2m}}{\delta} \right)^2 \cdot \frac{1}{\pi_k} \cdot \frac{1 + p_k}{1 - p_k}$$

(31) Example: For the example from (27) again, if we want a 95% chance that all of the $\hat{\pi}_k$ are within $\pm \pi_k/2$ of π_k , we should use $n \geq 353$ historical observations. Results from 1000 simulations with $n = 353$:

k	π_k	95% CI radius
1	0.5137	0.0762
2	0.1128	0.0374
3	0.2724	0.0843
4	0.1011	0.0433

(32) How does m , the number of types, affect n ? Say S_1 is the smallest type; then $\pi_1 \leq 1/m$, so pretend it is $1/m$ (the most-optimistic assumption). Say we want $\delta = 1/2$, $\alpha = 0.05$. Then n depends on m and p_1 as follows:

p_1	m							
	2	3	4	5	10	20	50	100
0.1	49	84	122	162	385	894	2647	5923
0.5	121	206	299	398	945	2194	6497	14539
0.9	764	1307	1897	2521	5988	13894	41145	92079

These are big numbers, and they will be proportionally bigger if $\pi_1 < 1/m$. For comparison, there are

- ~ 150 completely independent lineages in Ethnologue (Gordon, 2005) (100 families, 40 isolates, and 70 “unclassified”, exclusive of signed languages (121), mixed languages (21), creoles (86), pidgins (18), and artificial languages (3)).
- ~ 500 lineages separated for at least 3500 years Bell (1978, Table 2).
- ~ 7000 languages currently spoken on Earth (Gordon, 2005).

(33) We can get smaller (hence better) p_k ’s by using a bigger time unit, but we pay for that with increased probability that languages in our sample will undergo more than one change in the sample period, reducing the accuracy of the p_k estimate.

(34) Summary: Feasibility of explanation in the Markov model

- a. Not very practical: Predicting typological frequencies from type-transition probabilities can be costly.
 - (i) Perhaps feasible when applied to individual fairly-common, well-trafficked types in a typology of any complexity, and not too much precision is demanded.
 - (ii) When applied to an entire typology, it can require prohibitively large samples when there are more than a few states.
- b. *Not very illuminating*: Uses an $m \times m$ matrix to explain an $m \times 1$ vector.
- c. *Not very explanatory*: What made the transition probabilities be that way? What keeps them constant over time? Something must!

$$\boxed{\text{something}} \longrightarrow \boxed{\mathbf{P}} \longrightarrow \boxed{\bar{\pi}}$$

3.3 Reducing the number of explanatory variables

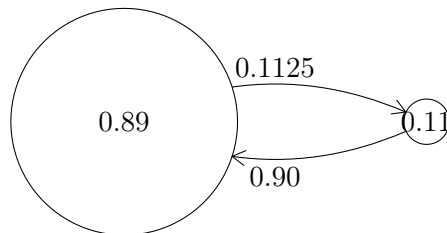
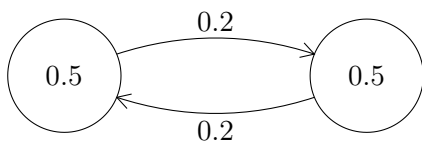
(35) Current implementations of the original state-and-process model emphasize the overall frequency of incoming transitions as the chief determinant of typological frequency, e.g., Evolutionary Phonology:

The working hypothesis supported throughout this volume is that *recurrent synchronic sound patterns have their origins in recurrent phonetically motivated sound change...* Common instances of sound change give rise to commonly occurring sound patterns. Certain sound patterns are rare or unattested, because there is no common pathway of change which will result in their evolution.

Blevins (2004, 8–9)

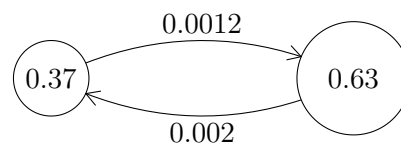
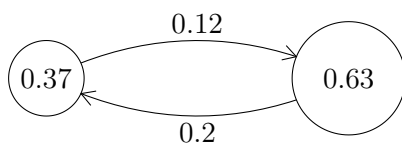
(36) However, some caveats apply:

- a. Typological frequency depends equally on innovations and extinctions.
- b. The sheer *rate of innovation per unit time* does not predict typological frequency. The same rate of innovation is consistent with a high or low typological frequency.



$$0.5 \times 0.2 = 0.1 = 0.89 \times 0.1125.$$

- c. *Relative frequency* of changes does not predict typological frequency. There are always just as many transitions from S_1 to S_2 as the reverse, yet S_1 is bigger than S_2 .
- d. *Absolute conditional probability* of a change does not predict typological frequency. The same typological frequency can come from high or low transition probabilities, e.g.,



(37) Patterns have whatever typological frequency is needed to balance innovations and extinctions. Greenberg (1978) has constructed a meta-typology on this basis:

If a particular phenomenon can arise very frequently and is highly stable once it occurs, it should be universal or near universal. This could be true of front unrounded vowels.

If it tends to come into existence often and in various ways, but its stability is low, it should be found fairly often but distributed relatively evenly among genetic linguistic stocks. A possible example is vowel nasalization.

If a particular property rarely arises but is highly stable when it occurs, it should be fairly frequent on a global basis but largely confined to a few linguistic stocks, e.g., vowel harmony.

If it occurs only rarely and is unstable when it occurs, it should be highly infrequent or non-existent and sporadic in its geographical and genetic distribution, e.g., velar implosives.

(Greenberg, 1978, 76)

⇒ Can't do without the outgoing transitions.

(38) Bell assumed that processes (of change) were independent of typological state:

To construct the model, many specific hypotheses about syllable structures and the processes that change them must be made. They are described in the following sections in detail. Overshadowing them are three general assumptions. First is the Markovian property. For the states postulated, the processes are independent of all past states that a language has traversed. The second, another assumption of independence, is that process likelihoods are independent of the states per se. By process I mean here a structural change plus a structural description. A state affects only whether a process's structural description is satisfied. If it is satisfied, then the process occurs with an intrinsic likelihood. This implies, for example, that processes affecting final clusters will have the same effect on languages with and without initial clusters. The third assumption is that only process internal to a language are considered; an ideal situation in which loans do not affect syllable structure is assumed.

Bell (1970, 221)

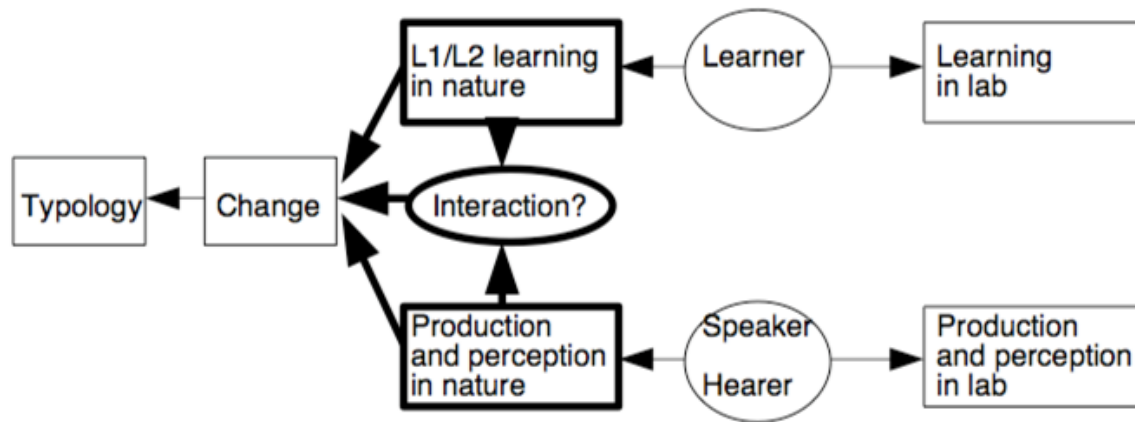
This allowed him to estimate several transition probabilities at once, from a larger pool of data.

(39) Useful idealization, but:

- a. L1 has large effects on channel bias, in both production (e.g., Keating (1985); Kingston and Diehl (1994); Wells-Jensen (2007)) and perception (e.g., Hallé et al. (1998); Berent et al. (2007)).

- b. Typological state can affect sound-change probability. Ex.: Compensatory lengthening is much more common in languages with a pre-existing length contrast than in those without (de Chene and Anderson (1979), Kavitskaya (2002)). Blevins (2004, §6.4), gives the ratio as eighty to eight, and attributes the asymmetry to an analytic bias (Structural Analogy).

(40) \Rightarrow We can't get around having to empirically study the factors determining the transition probabilities, which means we can't get around having to study analytic bias and its interaction with channel bias.



Next topic: How?

References

- Agresti, A. (1990). *Categorical data analysis*. New York: Wiley Interscience.
- Bakker, P. (1994). Pidgins. In P. Muysken and N. Smith (Eds.), *Pidgins and creoles*, Chapter 3, pp. 25–40. Amsterdam: John Benjamins.
- Bell, A. (1970). *A state-process approach to syllabicity and syllabic structure*. Ph. D. thesis, Stanford University.
- Bell, A. (1971). Some patterns of the occurrence and formation of syllabic structure. *Working Papers on Language Universals* 6, 23–138.
- Bell, A. (1978). Language samples. In J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Eds.), *Universals of human language, volume 1, method and theory*, pp. 123–156. Stanford, California: Stanford University Press.
- Berent, I., D. Steriade, T. Lennertz, and V. Vaknin (2007). What we know about what we have never heard: evidence from perceptual illusions. *Cognition* 104, 591–630.
- Blevins, J. (2004). *Evolutionary phonology*. Cambridge: Cambridge University Press.
- de Chene, B. and S. R. Anderson (1979). Compensatory lengthening. *Language* 55(3), 505–535.
- Gordon, R. G. (2005). *Ethnologue: languages of the world*. 15th ed. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Greenberg, J. H. (1978). Diachrony, synchrony, and language universals. In J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Eds.), *Universals of human language, volume 1, method and theory*, pp. 61–91. Stanford, California: Stanford University Press.
- Griffiths, T. L. and M. L. Kalish (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science* 31(3), 441–480.
- Griffiths, T. L., M. L. Kalish, and S. Lewandowsky (2008). Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society B (Biological Sciences)* 363(1509), 3503–3514.
- Hallé, P. A., J. Segui, U. Frauenfelder, and C. Meuiet (1998). Processing of illegal consonant clusters: a case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception & Performance* 24(2), 592–608.
- Hay, J. and L. Bauer (2007, June). Phoneme inventory size and population size. *Language* 83(2), 388–400.
- Kavitskaya, D. (2002). *Compensatory lengthening: phonetics, phonology, diachrony*. New York: Routledge.
- Keating, P. A. (1985). Universal phonetics and the organization of grammars. In V. A. Fromkin (Ed.), *Phonetic linguistics: essays in honor of Peter Ladefoged*, pp. 115–130. Orlando: Academic Press.
- Kingston, J. and R. L. Diehl (1994). Phonetic knowledge. *Language* 70(3), 419–454.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge studies in speech science and communication. Cambridge University Press.
- O’Cinneide, C. A. (1993). Entrywise perturbation theory and error analysis for Markov chains. *Numerische Mathematik* 65, 109–120.
- Wells-Jensen, S. (2007). A cross-linguistic speech error investigation of functional complexity. *Journal of Psycholinguistic Research* 36(2), 107–157.
- Winford, D. (2003). *An introduction to contact linguistics*. Malden, Mass.: Blackwell.
- Xue, J. (1997). A note on entrywise perturbation theory for Markov chains. *Linear Algebra and its Applications* 260, 209–213.