A model of population dynamics in phonetic change

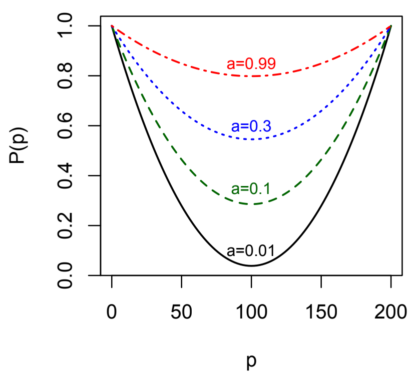
Introduction The idea that phonetic change is driven by the gradual accumulation of error goes back to at least the Neogrammarians [5]. However, as famously noted by Weinrich et al. [9], this type of account is problematic in that it predicts change far more often than is actually observed. Nonetheless, many computational models of phonetic change continue to assume, implicitly or explicitly, the Neogrammarian position [3, 6, 8], cf. [1, 2]. Here, we explore the effects of different assumptions regarding prior bias and population structure on the actuation and evolution of sound patterns. In particular, we consider population-level models that allow for considerable variability in the social network: which teachers individuals learn from. In a case study of the emergence of vowel harmony through phonologization of vowel-to-vowel coarticulation [4], we demonstrate that both stability and phonologization can emerge as attractor states, but only under certain assumptions about the prior bias of the learners.

Models We consider a language with a simple lexicon $\Sigma = \{$/a/, /e/, /aCe/$\}$, assuming that the distributions of /a/ and /e/ are known to all learners, are the same for all learners, and do not change over time. The relevant dimension of variation here is $F1$: the distribution of $F1$ for /a/ in the context of /e/ (i.e. /aCe/) differs from that of /a/ only by an offset to the mean p, indicating how much /a/ is affected by coarticulation in the /e/ context. Given a list of training examples $\vec{y}$, learner's task is then simply to infer this parameter p, which at the outset we assume is normally distributed in the population.

Results We first considered the dynamics of a maximum-likelihood (ML) learner under two *social learning* scenarios: one in which each learner's data is provided by a single teacher (Fig. 2b), and one in which the data may be drawn from multiple teachers (Fig. 2c). Both scenarios allow for more complex and (arguably) realistic dynamics than the *iterated learning* structures (Fig. 2a) commonly assumed in computational studies of sound change (e.g. [3]). In both instances, the population mean of the coarticulation parameter p is stable over generations, but the standard deviation increases by an amount inversely proportional to the number of training examples received. This suggests that ML learning directly from data is indeed empirically inadequate, as argued by [1, 6].

Next, we considered two scenarios in which learners have a prior distribution over p. In the first scenario, we assumed a Gaussian prior $\alpha \sim \mathcal{N}_x(0, \tau^2)$ which increasingly disfavors values of p greater than 0. For both single and multiple teachers, the expectation and variance of p move towards 0 and a fixed attractor α_* respectively, with $\alpha_* \to 0$ as $n \to \infty$. Thus, the Gaussian prior predicts learners should never phonologize the effects of coarticulation, regardless of the how strongly this is indicated by the input. In the second scenario, we considered a quadratic prior centered at $(\mu_a - \mu_a)/2$ which is concave between 0 and $\mu_a - \mu_e$:

Figure 1: Quadratic prior over $P(p)$.



(1) $\quad P(p) \propto \left\{a(\mu_a - \mu_e)^2 + (p - (\mu_e - \mu_a)/2)^2\right\}$

where a is a scale parameter controlling the 'flatness' of the prior (see Fig. 1). The value of p which maximizes the posterior can thus be found by evaluating the log posterior $\log(P(\vec{y}|p)) + \log(P(p))$ for various values of p and a. Here, we report the results of simulations where we assumed $F1$ values of 730 for /a/ and 530 for /e/, and considered a range of priors (by varying a) for both a case of little coarticulation in the population ($p_0 \sim \mathcal{N}(10, 10)$) as well as a case of an extreme degree of coarticulation ($p_0 \sim \mathcal{N}(100, 10)$). The results (Fig. 3) illustrate how both stability and change are possible under this regime, depending on the flatness of the prior, but that crucially, 'intermediate' states of coarticulation do not emerge. With a steep prior ($a = 0.01$), the population tends to eschew coarticulation; with a shallower prior ($a = 0.5$), a population where $p_0 = 10$ stays stable (with little coarticulation), while a population where $p_0 = 100$ oscillates between full and no coarticulation; with a very shallow prior ($a = 0.99$), we observe a roughly sigmoidal shift to full coarticulation (i.e. vowel harmony). Note that the population-level variance of p stays stable over time.

Conclusions Our results suggest that simple models of sound change based on accumulation of error are inadequate at the population level, but that including biases towards attractor states—in this case, either full coarticulation or no coarticulation—allows us to to model both stability and change. That we observe bifurcations in the population dynamics suggests that this approach may be an appropriate strategy for modeling changes to continuous as well as the discrete parameters considered in previous work [7]. We suggest that while misapprehension may still play a role in the actuation of sound change, its effects are attenuated by the distribution of priors in the population.

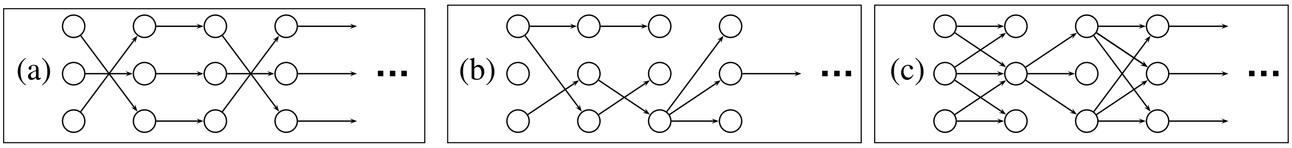


Figure 2: (a) Iterated learning. Each agent receives input from a single teacher, and each agent acts as both teacher and learner. (b) Social learning from a single teacher. The input for each learner comes from a single teacher, but the same teacher may supply input to multiple learners. (c) Social learning from multiple teachers. Here, input may come from multiple teachers, although some teachers may not provide data to any learners in the following generation.

References

[1] A. Baker. Addressing the actuation problem with quantitative models of sound change. In *Proc. 31st Annual Penn Ling. Colloq.*, 14(1), pp. 29–41, 2008.

[2] A. Baker, D. Archangeli, and J. Mielke. Variability in American English s-retraction suggests a solution to the actuation problem. *Lang. Variation and Change*, 23:347–374, 2011.

[3] J. Kirby. The role of probabilistic enhancement in phonologization. In A.C.L. Yu, ed., *Origins of Sound Patterns: Approaches to Phonologization*. Oxford University Press, Oxford, 2013.

[4] J. Ohala. Towards a universal, phonetically-based, theory of vowel harmony. In *Proc. 3rd Int. Conf. on Spoken Lang. Process.*, pp. 491–494, Yokohama, Japan, 1994.

[5] H. Paul. *Prinzipien der Sprachgeschichte* Macmillan & Co., New York, 1880 [1889]. Tr. H. A. Strong.

[6] J. Pierrehumbert. Word-specific phonetics. In C. Gussenhoven and N. Warner, eds., *Papers in Laboratory Phonology VII*, pp. 101–139. de Gruyter, Berlin, 2002.

[7] M. Sonderegger and P. Niyogi. Combining data and mathematical models of language change. In *Proc. 48th Ann. Meeting of the Assoc. for Comp. Ling.*, pp. 1019–1029, Uppsala, Sweden, 2010.

[8] A. Wedel. Exemplar models, evolution and language change. *Ling. Rev.*, 23:247–274, 2006.

[9] U. Weinrich, W. Labov, and M. Herzog. Empirical foundations for a theory of language change. In W. Lehmann and Y. Malkiel, eds., *Directions for historical linguistics*, pp. 95–195. University of Texas Press, Austin, 1968.

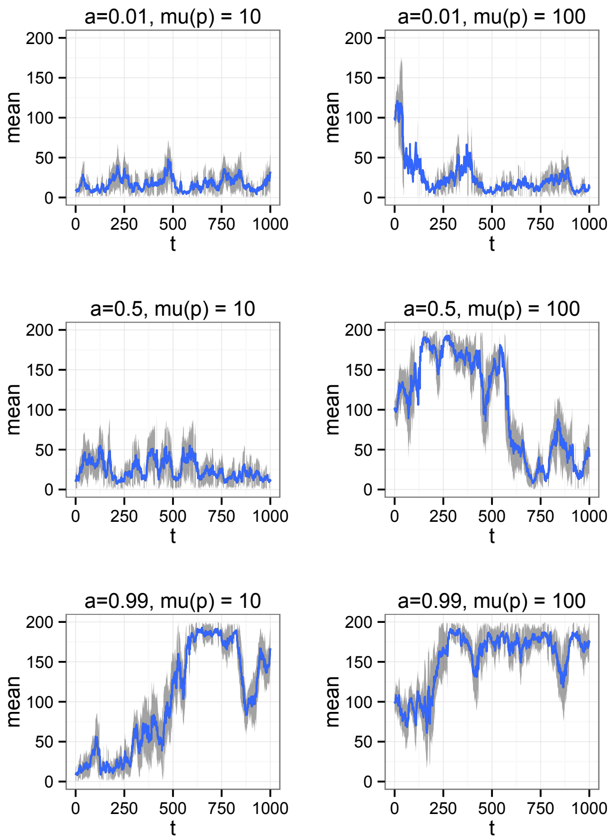


Figure 3: Stability and change in social learning with multiple teachers over 1000 generations. Blue line gives mean value of p for that generation, with variance shown in grey. Left column: $p_0 \sim \mathcal{N}(10, 10)$; right column: $p_0 \sim \mathcal{N}(100, 10)$.