A model of population dynamics in phonetic change

James Kirby (University of Edinburgh), Morgan Sonderegger (McGill University)

Introduction The idea that phonetic change is driven by the gradual accumulation of error goes back to the Neogrammarians. However, this type of account problematically predicts change far more often than is actually observed [5]. Nonetheless, many computational models of phonetic change continue to assume, implicitly or explicitly, the Neogrammarian position (e.g. [3, 4]). We explore the effects of different assumptions regarding prior bias and population structure in computational models of the evolution of sound patterns. In a case study of the emergence of vowel harmony through phonologization of vowel-to-vowel coarticulation [2], 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/, /i/, /aCi/$\}$, assuming that the distributions of /a/ and /i/ are known to all learners and do not change. There is one dimension of variation: $F1$ for /a/ before /e/ differs from /a/ before /a/ by an offset to the mean p, indicating how much /a/ is affected by coarticulation in the /e/ context. Given a set of /aCi/ examples the learner's task is simply to infer p, which at the outset we assume is normally distributed in the population.

Models and results We tried models corresponding to four different assumptions about learners' prior bias under two different assumptions about population structure: one in which each learner's data is provided by a single teacher, and one in which the data is drawn from multiple teachers. Because which type of prior bias was assumed turned out to strongly affect the predicted population dynamics, while the number of teachers did not, we discuss only the former here.

Naive learning models. We first considered the dynamics of successive generations of maximum-likelihood (ML) learners, who replicate the amount of coarticulation observed in their training data. In both the single- and multiple-teacher scenarios, the mean of p is stable over generations. However, in the single-teacher case, the variance increases over time (Fig. 1), while in the multiple-teacher case, the variance moves towards a fixed point. These results suggest that ML learning directly from data is empirically inadequate (cf. [1, 3]).

Learners with priors. Next, we considered three scenarios in which learners have a prior distribution over p. The first assumes a Gaussian prior which increasingly disfavors values of p greater than 0. Under this regime, the population-level expectation and variance of p move towards 0 (Fig. 2), predicting that V-to-V coarticulation will never phonologize. We then considered a quadratic prior favoring values of p which are close to 0 or $\mu_a - \mu_i$, with a parameter a controlling the prior's 'flatness'. Using this prior, we ran simulations in which we considered a range of priors (varying a) for two cases: little coarticulation in the population ($p_0 \sim \mathcal{N}(10, 10)$), and extreme coarticulation ($p_0 \sim \mathcal{N}(100, 10)$). Population-level stability of p was possible when learners were equipped with a strong prior (a near 0) favoring values of p at one extreme or the other; however, change from one stable state to the other is never possible (Fig. 3). Intuitively, the reason is that change cannot occur is that there is no force biasing learners towards one outcome (little or full coarticulation) over the other. Thus, we considered a final model, where learners have both a quadratic prior and a *bias factor* λ: some fraction of tokens a learner receives have F1 moved towards μ_i by λ, due to an external force which increases the likelihood of coarticulated variants (cf. [3, 4]). Both stability and change are now possible: with a weak bias ($\lambda \leq 5$) and a strong prior (a near 0), the population stays near $p_0 = 10$. If λ is increased across a critical value $\lambda \approx 5$, the population rapidly changes to a stable state of full coarticulation (Fig. 4). Thus, we find that all empirically observed situations (stable minimal coarticulation, stable full coarticulation, change from minimal to full coarticulation) are possible, but only under strong assumptions about learning bias.

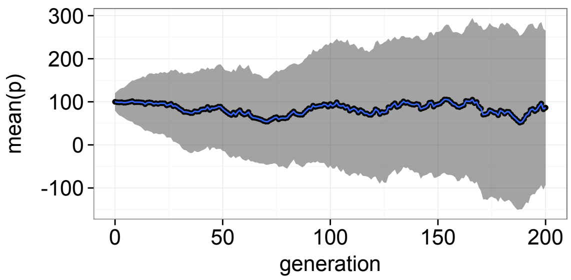

Figure 1: The evolution of p in a population of learners with no prior over $P(p)$.

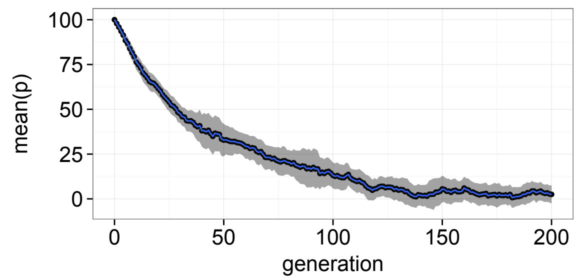

Figure 2: The evolution of p in a population of learners with a Gaussian prior over $P(p)$.

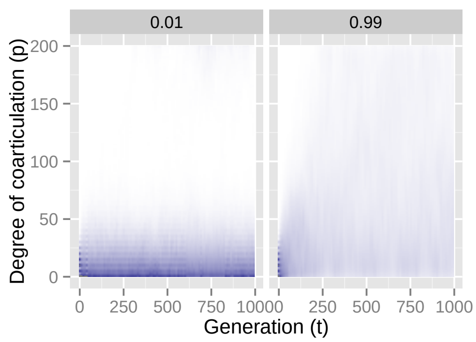

Figure 3: Evolution of $\pi(p)$ with strong ($a = 0.01$, left) and weak ($a = 0.99$, right) quadratic priors.

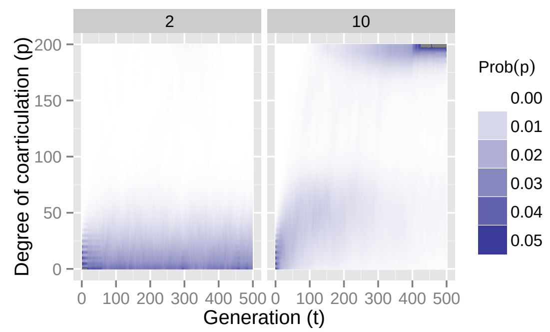

Figure 4: Evolution of $\pi(p)$ with a strong quadratic prior ($a = 0.01$), with 10% of tokens subject to bias factor $\lambda = 2$ (left) or $\lambda = 10$ (right).

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] 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.

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

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

[5] U. Weinreich, 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.