Populations of Learners: the Case of Portuguese

Partha Niyogi (NIYOGI@RESEARCH.BELL-LABS.COM)
Robert C. Berwick (BERWICK@AI.MIT.EDU)
MIT and Bell Laboratories
MIT Center for Biological and Computational Learning
Room 765
545 Technology Sq.
Cambridge, MA 02139

Abstract

We present new results of a novel computational approach to the interaction of two important cognitive-linguistic phenomena: (1) language learning, long regarded as central to modern synchronic linguistics; and (2) language change over time, diachronic linguistics. We exploit the insight that while language learning takes place at the level of the individual, language change is more properly regarded as an ensemble property that takes place at the level of populations of language learners — while the former has been the subject of much explicit computer modeling, the latter been less extensively treated. We show by analytical and computer simulation methods that language learning can be regarded as the driving force behind a dynamical systems account of language change. We apply this model to the specific (and cognitively relevant) case of the historical change from Classical Portuguese (CP) to European Portuguese (EP), demonstrating how a particular language learning model (for instance, a maximum-likelihood model akin to many statistically-based language approaches), coupled with data on the differences between CP and EP, leads to specific predictions for possible language-change envelopes, as well as delimiting the class of possible language-learning mechanisms and linguistic theories compatible with a given class of changes. The main investigative message of this paper is to show how this methodology can be applied to a specific case, that of Portuguese. The main moral underscores the individual/population difference, and demonstrates the potential subtlety of language change: we show that simply because an individual child will, with high probability, choose a particular grammar (European Portuguese) does not mean that all other grammars (e.g., Classical Portuguese) will come to be eliminated; rather, contrary to surface intuition, that is property of the dynamical system and the population ensemble itself.

Language Change: the Population Approach

This paper presents a *computational* approach to the interaction of two important *cognitive* phenomena: language learning and language change. The first, language learning, occurs at the level of the individual — children acquire the language (grammar) of their caretakers, a cognitive ability that has been broadly investigated via a range of computational and experimental methodologies. We note that language change occurs at the level of a population: it is individual language learners whose collective, ensemble properties constitute a *distribution* of linguistic knowledge. This is, of course, the common biological view used to bridge between surficial properties of individual organisms (their phenotypes) and the distribution of those properties over time (phenotypic evolutionary change), as underlain by internal, individual constitutional change (genotypes and genotypic distributions over time). In our linguistic analogy, which is nearly exact, initial individual linguistic (or "grammatical") knowledge corresponds to genotypes, and final attained states of linguistic knowledge to phenotypes; the distribution over final states characterizes the linguistic properties of the population as a whole. Programmatically, this analogy has often been drawn in diachronic linguistic research, but in the main the corresponding computational analysis, which might then be drawn almost directly from results in mathematical and computational population biology, has been lacking.[1]

There are three logical distributional possibilities: distributions of languages over monolingual speakers; distributions of languages within non-monolingual speakers; and combinations of these two. For example, consider a population of monolingual speakers; each speaks only one language. What of the population? If all speakers speak the same language, then one might say that the community speaks that language. However, there can still be mixtures of monolingual speakers. The other possibility is that a single speaker could internalize more than one language (grammar), that is, the possibility of non-monolingual (bi-, tri-lingual, etc. speakers). In the remainder of this paper, we focus on just the first two cases, that is, distributions of languages over monolingual speakers; as it turns out, the mathematics for the remaining possibility (bilingualism) can be captured by our assumption, though space prohibits a complete demonstration of this property here. We proceed to outline the basic model, and then a particular cognitive case of language learning and language change, namely, Portuguese.

The Logical Basis of Language Change

In our model the logical basis of change is language learning: the possibility of *mis*learning a particular target grammar of one's caretakers. Note that if children *always* converged on

[1]Here there are two chief exceptions: the general work on cultural transmission and evolution by Cavalli-Sforza and Feldman (1981) and the more narrowly drawn work of Kroch (1990). These previous works have complementary strengths and weaknesses: Cavalli-Sforza and Feldman (1981), while providing a richer mathematical model explicitly connected to population biology, do not actually pursue specific cognitive predictions or linguistic implications. In contrast, Kroch (1990) focuses on specific language changes, like the well-known English Verb-second change, but does not consider a dynamical system model, and in fact posits only the possibility of different language distributions *within* a single speaker, rather than over a population of speakers, thus avoiding a true population approach to the problem.

the language of their parents, then their language would be the same as that of their parents, from each generation to the next. Consequently, for languages to change from one generation to the next it must be the case that children attain a language different from that of their parents. We next show how to model this computationally.

The Computational Framework

The procedure for mapping language learning to language change has been developed in a series of previous works (Niyogi and Berwick, 1996) and is reviewed here for convenience.

The three main components of a language learning framework are:

1. $\mathcal{G}$: a class of grammars (languages) from which the child chooses one on the basis of example sentences.
2. $\mathcal{A}$: a learning algorithm used by the language learner to choose a grammar $g \in \mathcal{G}$. (In what follows we will systematically interchange the use of "grammar" and "language" when there is no possibility of confusion or difference in the resulting model results.)
3. $\mathcal{P}$: a probability distribution with which sentences are presented to the learner.

Once each of items (1), (2), and (3) are well-specified we have a complete description of language learning for a single generation. This has been dealt with in a variety of situations under a number of different assumptions about the class of languages, learning algorithms, and the like, from Gold (1967) to more recent work. As one can see, our framework is general enough to encompass even more recent learning methods such as Minimum Description Length criteria or other statistical methods; in fact, in our example below we use a Maximum Likelihood search method, but any well-defined procedure would do. A complete analysis of the behavior of the individual learner will allow us to analyze the behavior of the population as we see in the next section.

Individuals versus Populations of Learners

The language learning problem focuses on the individual child and attempts to characterize how it updates its hypothesis from example sentence to example sentence over its lifetime. Computational models of this phenomenon typically require the learner's hypothesis get closer and closer to the target grammar as more and more data becomes available — this is as true of the classic Gold "identification in the limit" model as it is of more recently statistically-based methods that work on corpuses. The Gold success criterion requires the learner to converge to the target as the data goes to infinity; a more psychologically plausible criterion requires the learner to be at the target with high probability after a psychologically realistic number of examples have been received. Let us assume that we are able to *completely* characterize the behavior of the individual leaner after receiving a finite number of examples (positive or negative), i.e., we are able to solve the following problem:

Suppose n examples are drawn according to P on Σ^* and presented to the child. Then, for every grammar g in $\mathcal{G}$, what is the probability that the child will have attained that grammar? Let:

$$p_n(g) = \text{Probability[Learner hypothesizes } g \text{ after } n \text{ examples]} \quad (1)$$

Eq. 1 characterizes the probability with which an individual child internalizes each of the possible grammars (languages) after n examples. Naturally, if the data were all drawn from some target language corresponding to grammar h (say) then realistic language learning requires that the corresponding probability ($p_n(h)$ to be high, i.e., the learner attains the target grammar with high probability.

Let us now consider the population as a whole. The population is composed of (i) a collection of individual adults that are the source of example sentences to the generation of children; and (ii) a collection of individual children who attempt to acquire the grammar of the parental generation on the basis of example sentences. If we make the population the object of our study, then we would characterize the linguistic composition of the population (what percentage speaks what language) and how it evolves from generation to generation. Since the population is an aggregate of individuals, if we take ensemble averages, we would arrive at the behavior of the population.

For the purposes of this paper we make the following idealizations for population modeling; however, all of these can be systematically dropped and their consequences explored: (1) non-overlapping generations, i.e., adults and children and the linguistic composition of a particular generation is comprised of its adult speakers, rather than a mix of children and adults; (2) no neighborhood effects, i.e., the mix of adults determines the source of sentences and this distributional source is identical for all children (clearly not the case for geographical boundary conditions, but as mentioned, easily modeled by a extension using conventional population biology methods); (3) adults do not change their grammar/language over their lifetime, i.e,. a monolingual maturation hypothesis[2] (4) children have a finite time to acquire the grammar, i.e., a learning maturation hypothesis.

Given the general model and these assumptions, one can now characterize the evolution of a population of speakers from generation to generation as a dynamical system. Let the state of the population in generation i be defined by a probability distribution $s_p(i)$ on the set of grammars $\mathcal{G}$. Thus $s_p(i)[g]$ denotes the proportion of the population that speak a language corresponding to grammar $g \in \mathcal{G}$. Children are exposed to data that are a mix of the languages of the adult population — in this way one can model very simple geographical effects (not detailed areal distributions, however). Suppose we are able to characterize the behavior of the individual average child, as in eq. 1. Assuming that maturation occurs after n examples, when the current generation of children mature into adulthood, the composition of their population would be given by p_n. Thus the update rule for the entire system, determining the language mix of the next generation, is given by the equation:

$$s_p(i+1) = p_n$$

[2]In contradistinction to a bi-lingual assumption; however, as noted, the mathematics for mixtures of languages *within* an individual is approximately the same as that of mixtures of languages between individuals, at least for our initial purposes here.

where p_n clearly depends upon (1) the composition of the previous generation $s_p(i)$; (2) the learning algorithm that children use to learning languages (grammars); and (3) the probability distributions with which speakers of a particular language produce example sentences. In the remainder of this paper we show how to develop the form of the update rule for a concrete learning/language change situation, namely that of Portuguese, and how to use the resulting dynamical system to test whether the assumptions of the model are cognitive plausible, e.g., whether the observed time course of the dynamical system actually follows the observed linguistic change from classical to modern European Portuguese, while manipulating assumptions about the learning algorithm, and the like.

Portuguese: A Case Study

The main investigative message of this paper is to show how this methodology can be applied to a specific case, that of Portuguese. The main moral underscores the individual/population difference, and demonstrates the potential subtlely of language change: we show that simply because an individual child will, with high probability, choose a particular grammar (European Portuguese) does not mean that all other grammars (e.g., Classical Portuguese) will come to be eliminated; rather, contrary to surface intuition, that is property of the dynamical system and the population ensemble itself.

The Facts of Portuguese Language Change

In this paper, we focus on a particular change in phonological and syntactic Portuguese recently discussed by Galves and Galves (1995). Roughly, over a period of 200 years, starting from 1800, "classical" Portuguese (CP) underwent a change in clitic placement. From the 16th century or before until the beginning of the 19th century, both proclitics and enclitics were possible in root declarative sentences (nonquantified subjects), as given by G&G's examples (1) and (2), and in quantified subjects (3), which we will refer to henceforth via their reference numbers:

(1) Paulo a ama.
Paulo her loves
'Paulo loves her'
(proclitic)

(2) Paulo ama-a
Paulo loves-her
'Paulo loves her' (enclitic)

(3) Quem a ama?
Who her loves?
' Who loves her?'
(proclitic)

G&G summarize the relevant historical facts as follows: "During the 19^{th} century a change affecting the syntax of clitic-placement occurred in the language spoken in Portugal. . . As a result, sentences like (1) became agrammatical and (2) remained as the only option for root affirmative sentences with non-quantified subjects. This change, however, did not concern sentences like (3) with quantified or *Wh*-subjects in which proclisis was, and continues to be, the only option."

G&G offer an explanation of this change, proposing a link between phonology and syntax. Roughly speaking, Galves has argued that phonological changes in Portuguese altered the stress contours, and consequently the probabilities with which sentence types occurred; this difference is stress is what learning hinges on, and so the historical change. While this explanation is arguable, we will accept it to illustrate how different learning algorithms might have different evolutionary consequences for historical prediction, ignoring for the moment the linguistic implications of the various algorithms and concentrating only on their computational properties. To each sentence we will assign (a) a morphological word sequence; (b) a stress contour; and (c) a syntactic structure. For example, again following G&G's analysis, sentence type (2) will remain only in CP, while the two sentences (2)–(3) above will have *different* stress patterns for CP and EP. We omit a detailed description of the stress assignment and syntactic properties, as they are not necessary for our analysis. All we need to know is that G&G assume that the stress contours corresponding to sentence types (1), (2), and (3), which we denote simply as c_1, c_2, c_3, follow a Markov chain description and, more importantly, govern the probability with which sentences are produced.[3] Thus, if two sentences have the same stress contour, then they will be produced with the same probability (given by the probability of the stress sequence according to Markov production rules). In short, for the purpose of this paper, it is sufficient to assume that there are two simply two grammars (in accordance with Galves' assumptions): G_{CP}, denoting the grammar of Classical Portuguese (earlier) and G_{EP}, denoting the grammar of European Portuguese. Furthermore, the only data that is relevant (ignoring other aspects of the grammar) is as follows:

Classical Portuguese
(CP-1) c_1 : produced with probability p; (CP-2) c_2 produced with probability $1 - 2p$; and (CP-3) c_3 produced with probability p.

European Portuguese
(EP-1) c_1: not produced; (2) (EP-2) c_2 produced with probability $1 - q$; and (EP-3) c_3 produced with probability q.

Any (historically changing) population will now by assumption contain a mix of speakers of Classical and European Portuguese. The Classical Portuguese speakers produce the sentence types shown above with the probabilities (parameterized by p). The European Portuguese speakers produce the sentence types shown above with the probabilities (parameterized by q).

Thus we have defined (1) the class of grammars (1) $\mathcal{G} = \{G_{EP}, G_{CP}\}$; (2) Probabilities with which speakers of G_{EP} and G_{CP} produce sentences (parameterized by p and q. We therefore can derive the evolutionary consequences on the population for a variety of learning algorithms. We first consider a probabilistic, maximum likelihood method: to choose between CP and EP given some input sentences (conditioned on stress patterns), pick the language (grammar) that maximizes the probability of generating the available data (surface forms). This is probably the simplest probabilistic learning algorithm and leads to:

[3] We are of course aware that this assumption of G&G may also be questioned; one might substitute for it any other more plausible relation between stress and sentence types — if any; this assumption is simply designed as a bridge to get the child from a presumably observable surface fact to a sentence type.

Galves Batch Algorithm

The learning algorithm for grammatical acquisition proceeds as follows:

1. Draw N examples (sentences).
2. Compute likelihoods, i.e., $P(S_n|G_{CP})$ and $P(S_n|G_{EP})$.
3. Use the Maximum Likelihood principle to choose between the grammars.

Analysis of Individual Learning Algorithm As discussed earlier, to calculate the historical dynamics we must be able to analyze the behavior of the learning algorithm, i.e., characterize eq. 1. For the analysis of the algorithm, we assume that sentences are drawn in i.i.d. fashion according to a distribution dictated by their stress contours as indicated in the earlier section.

First, consider the form of the likelihoods. Let the example sentences be $S_n = \{s_1, s_2, \ldots, s_n\}$. Due to the i.i.d. assumption $P(S_n|G_{CP})$ is given by $\prod_{i=1}^{n} P(s_i|G_{CP})$. Suppose that the set of n examples consists of a draws of c_1, b draws of c_3 and $n-a-b$ draws of c_2. Then the following is immediately clear:

$$P(S_n|G_{CP}) = p^a(1-2p)^{(n-a-b)}p^b$$

and

$$P(S_n|G_{EP}) = (0)^a(1-q)^{(n-a-b)}q^b$$

Consequently, the individual child, following the likelihood principle will choose the grammar EP (G_{EP}) only if (1) no instances of c_1 occur in its sample; and (2) c_2 and c_3 occur in numbers so that $q^b(1-q)^{(n-b)} > (1-2p)^{(n-b)}p^b$. There are three cases to consider:

Case 1. $p < q < 2p$.
Decision Rule: For this case, it is possible to show that the child (following the maximum likelihood rule) always chooses G_{EP} if no instances of c_1 occur. This is simply because $1-q > 1-2p$ and $q > p$.
Equation 1 and Population Update: Suppose that the proportion of speakers of G_{CP} in the ith generation is α_i. Then the probability of drawing c_1 is given by $\alpha_i p$. Consequently, the probability of drawing a set of n examples without a single draw of c_1 is $(1-\alpha_1 p)^n$. This is of course the probability with which the individual child chooses the grammar of European Portuguese, G_{EP}. Thus the update rule has the following form:

$$\alpha_{i+1} = 1 - (1-\alpha_i p)^n$$

Case 2. $q < p < 2p$.
Decision Rule: In this case, the maximum likelihood decision rule reduces to the following. Choose G_{EP} *if and only if* (1) $a = 0$, i.e., no instances of c_1 occur; and (2) $b < n\gamma$ where $\gamma = \frac{\log(\frac{1-q}{1-2p})}{\log(\frac{1-q}{1-2p})+\log(\frac{p}{q})}$. For all other data sets, choose G_{CP}.
Equation 1 and Population Update: As usual, let there be α_i proportion of the previous generation speaking G_{CP}. It can be shown that events (1) and (2) above occur with probability $\sum_{k=0}^{n\gamma} \binom{n}{k} P^k Q^(n-k)$ where $P = \alpha_i p + (1-\alpha_i)q$ and $Q = \alpha_i(1-2p) + (1-\alpha_i)(1-q)$. Thus the update rule has the following form:

$$\alpha_{i+1} = 1 - \sum_{k=0}^{n\gamma} \binom{n}{k} P^k Q^(n-k)$$

Case 3. $p < 2p < q$.
Decision Rule: The maximum likelihood decision rule reduces to: choose G_{EP} *if and only if* (1) $a = 0$; and $b > n\gamma$ where $\gamma = \frac{\log(\frac{1-2p}{1-q})}{\log(\frac{q}{p})+\log(\frac{1-2p}{1-q})}$. Otherwise, choose G_{CP}.
Equation 1 and Population Update:
As usual, let α_i be the proportion of the previous generation speaking G_{CP}. It can be shown that the update rule has the following form:

$$\alpha_{i+1} = 1 - \sum_{k=n\gamma}^{0} \binom{n}{k} P^k Q^(n-k)$$

where P and Q are as in case 2.

System Evolution We have shown above how the behavior of the population can be characterized as a dynamical system and have derived the update rules for such a system for a maximum likelihood learning algorithm The dynamical system captures the evolutionary consequences of this particular learning algorithm. In this section we describe its evolutionary properties, and see how they mesh with observed cognitive (historical) trends.

Case 1.

1. $\alpha = 0$ is a fixed point, i.e., if the initial population consists entirely of European Portuguese speakers, it will always remain that way. Furthermore, if $np < 1$, then this is a *stable* fixed point. It is also the *only* fixed point between 0 and 1. Thus in this case a population speaking *entirely* Classical Portuguese would gradually be converted to one speaking *entirely* European Portuguese.
2. If $np > 1$, then $\alpha = 0$ remains a fixed point but now becomes *unstable*. For this case, an additional fixed point (*stable*) is now created between 0 and 1. All initial population compositions will tend to this particular *mix* of G_{CP} and G_{EP} speakers. Figure 1 shows the fixed (equilibrium) point as a function of n and p.

Case 2.

1. Unlike case 1, the dynamical evolution depends now depends upon both p and q in addition to n.
2. It is easily seen that $\alpha = 0$ is no longer a fixed (equilibrium) point (unless $p = q$). Consequently, populations, irrespective of their initial composition, will *always* contain some speakers of Classical Portuguese.
3. It is possible to show that there is exactly one fixed (*stable*) point and all initial populations will tend to this value. Shown in fig. 2 is plot of the fixed point as a function of q and p for a fixed value of n. Notice the multiple ridges in the profile suggesting sensitivity to the value of q around some critical points. Shown in fig. 3 is a plot of the fixed point as a function of p for various choices of n keeping q fixed at 0.1.

Case 3.

1. Like case 2, the dynamical evolution depends upon both p and q in addition to n.

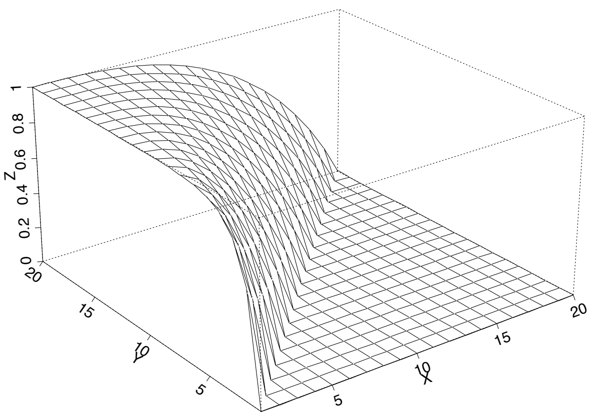

Figure 1: The fixed point of the dynamical system (on the Z axis) as a function of n (on the X axis) and $\frac{1}{p}$ (on the Y axis).

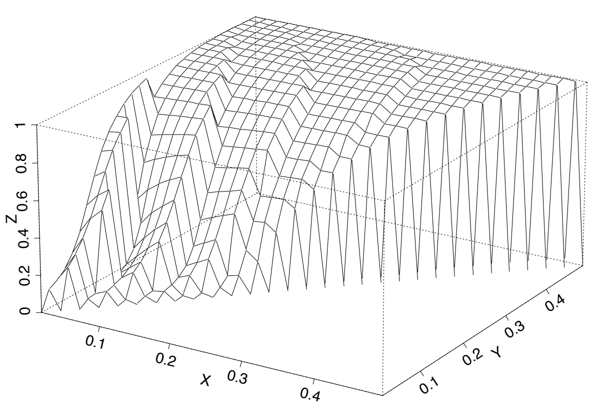

Figure 2: The fixed point of the dynamical system (on the Z axis) as a function of q (on the X axis) and p (on the Y axis). The value of n was held fixed at 5.

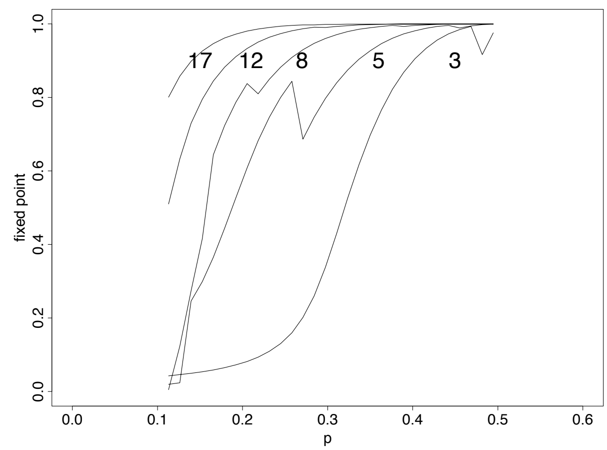

Figure 3: The fixed point of the dynamical system as a function of p (on the X axis) for various values of n. Here q was held fixed at 0.1 and p was allowed to vary from 0.1 to 0.5.

2. Again, it is easily seen that $\alpha = 0$ is no longer a fixed point. Therefore, the speakers of Classical Portuguese can never be eliminated altogether for p and q in this range.

We can again plot the fixed points of the resulting dynamical system as a function of the q and p where n is held fixed at 5 or for various values of n, keeping p fixed. We omit the figures for reasons of space. The results are: again the ridges in the landscape suggest a great sensitivity of the final equilibrium point to slight changes in the values of p and q. Classical Portuguese speakers are never completely eliminated, although their frequency can get quite low in certain regions.

What are the important conclusions from this analysis? In short, children using the maximum likelihood rule will choose G_{EP} over G_{CP}. However, a dynamical systems analysis must be carried out to see if that will suffice to "wipe out" Classical Portuguese. Only in case 1 will Classical Portuguese be lost completely (provided $p < 1/n$). *In all other cases*, there will always remain some speakers of Classical Portuguese within the community. In fact, the evolutionary properties can be quite subtle.[4] Consider the following three example cases.

Example 1 *Let* $p = 0.05, q = 0.02$ *and* $n = 4$. *In this case, if the parental generation were all speaking Classical Portuguese (*$\alpha = 1$*) then a simple computation shows that the probability with which the child would pick* G_{EP} *(European Portuguese) is* 0.66, *i.e., it is greater than one-half. Thus,* **in spite of the fact** *that the* **majority** *of children choose the grammar of European Portuguese, the speakers of Classical Portuguese will never die out completely. In fact, the fixed point is 0.11. Roughly 11 percent of the population will continue to speak Classical Portuguese.*

[4] The same dynamical properties would hold of a "bi-lingual" model where the distribution of languages was over *single* individuals — like that advocated by Kroch (1990). Here too the dynamical properties and resulting cognitive changes are simply too subtle to pick out by intuition alone; that is the key result of this paper.

Example 2 *Let $p = 0.05, q = 0.06$ and $n = 8$. If this were the case, and the parental generation all spoke Classical Portuguese, it turns out that the probability with which the individual child would pick G_{EP} would again be 0.66. However, now the speakers of classical Portuguese would all be lost and the population would move to its stable, fixed point containing only speakers of European Portuguese.*

Example 3 *If $p = 0.05, q = 0.06$ and $n = 21$ however, it is easily seen that Classical Portuguese speakers can never be completely lost.*

Batch Subset Algorithm

Most importantly, we see that the above learning algorithm makes specific predictions about the change of the linguistic composition of the population as a whole. For purposes of exploration, let us turn our attention to a simple modification of the previous learning algorithm that we call the Batch Subset Algorithm because all the data is processed "at once." Our aim is to demonstrate how readily one may carry out changes in the learning algorithm and investigate their model consequences.

1. Draw n examples.
2. If c_1 occurs even once, choose G_{CP}, otherwise choose G_{EP}.

Since European Portuguese is a subset of Classical Portuguese for the data at hand, such a learner would choose the grammar of European Portuguese G_{EP} as its default grammar unless it received contradictory data (in this case c_1: which informs it that the target is not G_{EP} but G_{CP}. Of course, such a learning algorithm is guaranteed to converge to the correct target as the data goes to infinity. A natural question to ask is whether it makes a different prediction about how the population would evolve. As it turns out, it is possible to prove:

Theorem 1 *Let α_i denote the proportion of the community speaking Classical Portuguese (G_{CP}) in the ith generation. Then this evolves as*

$$\alpha_{i+1} = 1 - (1 - \alpha_i p)^n$$

where n is the number of examples drawn and p is as usual.

One can already see that the evolutionary properties for this learning algorithm are different from the previous one. The dynamics is always given by the same update rule *irrespective* of the values of p and q. In fact, the evolution, which is totally independent of q, is identical to the Case 1 of the previous learning algorithm dynamics. Naturally, it has the same equilibrium behavior as Figure 1.

Since a batch algorithm is presumably psychological unreal (due to memory limitations), one could substitute, as we have done, a memoryless algorithms, such as local gradient ascent (Gibson and Wexler's "Triggering Learning Algorithm" or TLA, 1994). Due to reasons of space, we leave a detailed presentation of the results of this modification to one side, and simply note that one obtains yet a different historical dynamic. It is possible to prove that in this case, the population evolves according to the update rule:

$$\alpha_{i+1} = 1 - \frac{1}{2}(1 - \alpha_1 p)^n$$

where α_i and α_{i+1} are the proportion of the population speaking Classical Portuguese in generation i and $i+1$ respectively. As usual, n is the number of examples drawn. Here, note that CP speakers can *never* be eliminated to less than $\frac{1}{2}$ of the population. Consequently, one is able to see immediately that the TLA does *not* have the right evolutionary properties to explain the change from Classical to European Portuguese. Second, it is possible to show that there is exactly one (stable) fixed point (between $1/2$ and 1) to which such a system evolves, for various values of n and p.

Conclusions: Letters from the Portuguese

As the case of Portuguese language change shows, *individual* language change need not be the same as *ensemble* language change. Differences in language learning strategies can lead to differences in language change over time. The dynamical systems mathematics is essential because intuitions can lead one astray. To probe more deeply requires a new, more sophisticated approach like the one adopted by evolutionary population biologists. Mathematical modeling may be required to tease out subtle differences between language-learning driven language change — so much so that, to our minds, diachronic language analysis now demands the same armamentarium applied to "cognitive" genotypes and phenotypes that population biologists have brought to bear on the study of genotypic and phenotypic change over time.

Acknowledgements

Thanks to Antonio Galves, Morris Halle, Tony Kroch, and David Lightfoot for useful discussions. This research was performed at the Center for Biological and Computational Learning at MIT. Support for the Center is provided in part by a grant from the National Science Foundation under contract ASC–9217041.

References

Cavalli-Sforza, L. &Feldman, M. (1981). *A Mathematical Model of Cultural Transmission*. Princeton, NJ: Princeton University Press.

Clark, R. & Roberts, I. (1991). A Computational Model of Language Learnability and Language Change. *Linguistic Inquiry*, *24(2)*, 299-345.

Galves, A., & Galves, C. (1995). A case study of prosody-driven language change: from classical to modern European Portuguese. Unpublished M.S., University of Sao Paolo, Sao Paolo, Brasil.

Gibson, E., & Wexler, K. (1994). Triggers. *Linguistic Inquiry*, *25(4)*, 407–454.

Gold, E., (1967). Language identification in the limit. *Information and Control*, *10*, 447–474.

Kroch, A. (1990) Reflexes of grammar in patterns of language change. *Language Variation and Change*, 199–243.

Niyogi, P., & Berwick, R., (1997, to appear). A dynamical systems model for language change. *Linguistics and Philosophy*.